Lecture Notes in Artificial Intelligence 6780

Edited by R. Goebel, J. Siekmann, and W. Wahlster

Subseries of Lecture Notes in Computer Science

Dylan D. Schmorrow Cali M. Fidopiastis (Eds.)

Foundations of Augmented Cognition

Directing the Future of Adaptive Systems

6th International Conference, FAC 2011
Held as Part of HCI International 2011
Orlando, FL, USA, July 9-14, 2011
Proceedings

Springer

Series Editors

Randy Goebel, University of Alberta, Edmonton, Canada
Jörg Siekmann, University of Saarland, Saarbrücken, Germany
Wolfgang Wahlster, DFKI and University of Saarland, Saarbrücken, Germany

Volume Editors

Dylan D. Schmorrow
United States Navy
1777 N Kent Street, Arlington, VA 22209, USA
E-mail: schmorrow@yahoo.com

Cali M. Fidopiastis
University of Alabama at Birmingham
336 SHPB, 1530 3rd Avenue South, Birmingham, AL 35294, USA
E-mail: cfidopia@uab.edu

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-21851-4 e-ISBN 978-3-642-21852-1
DOI 10.1007/978-3-642-21852-1
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011929349

CR Subject Classification (1998): I.2, I.4, J.3, H.2.8, H.3-5, C.2

LNCS Sublibrary: SL 7 – Artificial Intelligence

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

The 14th International Conference on Human–Computer Interaction, HCI International 2011, was held in Orlando, Florida, USA, July 9–14, 2011, jointly with the Symposium on Human Interface (Japan) 2011, the 9th International Conference on Engineering Psychology and Cognitive Ergonomics, the 6th International Conference on Universal Access in Human–Computer Interaction, the 4th International Conference on Virtual and Mixed Reality, the 4th International Conference on Internationalization, Design and Global Development, the 4th International Conference on Online Communities and Social Computing, the 6th International Conference on Augmented Cognition, the Third International Conference on Digital Human Modeling, the Second International Conference on Human-Centered Design, and the First International Conference on Design, User Experience, and Usability.

A total of 4,039 individuals from academia, research institutes, industry and governmental agencies from 67 countries submitted contributions, and 1,318 papers that were judged to be of high scientific quality were included in the program. These papers address the latest research and development efforts and highlight the human aspects of design and use of computing systems. The papers accepted for presentation thoroughly cover the entire field of human–computer interaction, addressing major advances in knowledge and effective use of computers in a variety of application areas.

This volume, edited by Dylan D. Schmorrow and Cali M. Fidopiastis, contains papers in the thematic area of augmented cognition (AC), addressing the following major topics:

- Theories, models, and technologies for augmented cognition
- Applied neuroscience and brain monitoring
- Augmented cognition, social computing, and collaboration
- Augmented cognition for learning
- Augmented cognition and interaction
- Augmented cognition in complex operational environments

The remaining volumes of the HCI International 2011 Proceedings are:

- Volume 1, LNCS 6761, Human–Computer Interaction—Design and Development Approaches (Part I), edited by Julie A. Jacko
- Volume 2, LNCS 6762, Human–Computer Interaction—Interaction Techniques and Environments (Part II), edited by Julie A. Jacko
- Volume 3, LNCS 6763, Human–Computer Interaction—Towards Mobile and Intelligent Interaction Environments (Part III), edited by Julie A. Jacko
- Volume 4, LNCS 6764, Human–Computer Interaction—Users and Applications (Part IV), edited by Julie A. Jacko
- Volume 5, LNCS 6765, Universal Access in Human–Computer Interaction—Design for All and eInclusion (Part I), edited by Constantine Stephanidis

- Volume 6, LNCS 6766, Universal Access in Human–Computer Interaction—Users Diversity (Part II), edited by Constantine Stephanidis
- Volume 7, LNCS 6767, Universal Access in Human–Computer Interaction—Context Diversity (Part III), edited by Constantine Stephanidis
- Volume 8, LNCS 6768, Universal Access in Human–Computer Interaction—Applications and Services (Part IV), edited by Constantine Stephanidis
- Volume 9, LNCS 6769, Design, User Experience, and Usability—Theory, Methods, Tools and Practice (Part I), edited by Aaron Marcus
- Volume 10, LNCS 6770, Design, User Experience, and Usability—Understanding the User Experience (Part II), edited by Aaron Marcus
- Volume 11, LNCS 6771, Human Interface and the Management of Information—Design and Interaction (Part I), edited by Michael J. Smith and Gavriel Salvendy
- Volume 12, LNCS 6772, Human Interface and the Management of Information—Interacting with Information (Part II), edited by Gavriel Salvendy and Michael J. Smith
- Volume 13, LNCS 6773, Virtual and Mixed Reality—New Trends (Part I), edited by Randall Shumaker
- Volume 14, LNCS 6774, Virtual and Mixed Reality—Systems and Applications (Part II), edited by Randall Shumaker
- Volume 15, LNCS 6775, Internationalization, Design and Global Development, edited by P.L. Patrick Rau
- Volume 16, LNCS 6776, Human-Centered Design, edited by Masaaki Kurosu
- Volume 17, LNCS 6777, Digital Human Modeling, edited by Vincent G. Duffy
- Volume 18, LNCS 6778, Online Communities and Social Computing, edited by A. Ant Ozok and Panayiotis Zaphiris
- Volume 19, LNCS 6779, Ergonomics and Health Aspects of Work with Computers, edited by Michelle M. Robertson
- Volume 21, LNAI 6781, Engineering Psychology and Cognitive Ergonomics, edited by Don Harris
- Volume 22, CCIS 173, HCI International 2011 Posters Proceedings (Part I), edited by Constantine Stephanidis
- Volume 23, CCIS 174, HCI International 2011 Posters Proceedings (Part II), edited by Constantine Stephanidis

I would like to thank the Program Chairs and the members of the Program Boards of all Thematic Areas, listed herein, for their contribution to the highest scientific quality and the overall success of the HCI International 2011 Conference.

In addition to the members of the Program Boards, I also wish to thank the following volunteer external reviewers: Roman Vilimek from Germany, Ramalingam Ponnusamy from India, Si Jung "Jun" Kim from the USA, and Ilia Adami, Iosif Klironomos, Vassilis Kouroumalis, George Margetis, and Stavroula Ntoa from Greece.

This conference would not have been possible without the continuous support and advice of the Conference Scientific Advisor, Gavriel Salvendy, as well as the dedicated work and outstanding efforts of the Communications and Exhibition Chair and Editor of HCI International News, Abbas Moallem.

I would also like to thank for their contribution toward the organization of the HCI International 2011 Conference the members of the Human–Computer Interaction Laboratory of ICS-FORTH, and in particular Margherita Antona, George Paparoulis, Maria Pitsoulaki, Stavroula Ntoa, Maria Bouhli and George Kapnas.

July 2011 Constantine Stephanidis

Organization

Ergonomics and Health Aspects of Work with Computers

Program Chair: Michelle M. Robertson

Arne Aarås, Norway
Pascale Carayon, USA
Jason Devereux, UK
Wolfgang Friesdorf, Germany
Martin Helander, Singapore
Ed Israelski, USA
Ben-Tzion Karsh, USA
Waldemar Karwowski, USA
Peter Kern, Germany
Danuta Koradecka, Poland
Nancy Larson, USA
Kari Lindström, Finland
Brenda Lobb, New Zealand
Holger Luczak, Germany
William S. Marras, USA
Aura C. Matias, Philippines
Matthias Rötting, Germany
Michelle L. Rogers, USA
Dominique L. Scapin, France
Lawrence M. Schleifer, USA
Michael J. Smith, USA
Naomi Swanson, USA
Peter Vink, The Netherlands
John Wilson, UK

Human Interface and the Management of Information

Program Chair: Michael J. Smith

Hans-Jörg Bullinger, Germany
Alan Chan, Hong Kong
Shin'ichi Fukuzumi, Japan
Jon R. Gunderson, USA
Michitaka Hirose, Japan
Jhilmil Jain, USA
Yasufumi Kume, Japan
Mark Lehto, USA
Hirohiko Mori, Japan
Fiona Fui-Hoon Nah, USA
Shogo Nishida, Japan
Robert Proctor, USA
Youngho Rhee, Korea
Anxo Cereijo Roibás, UK
Katsunori Shimohara, Japan
Dieter Spath, Germany
Tsutomu Tabe, Japan
Alvaro D. Taveira, USA
Kim-Phuong L. Vu, USA
Tomio Watanabe, Japan
Sakae Yamamoto, Japan
Hidekazu Yoshikawa, Japan
Li Zheng, P. R. China

Human–Computer Interaction

Program Chair: Julie A. Jacko

Sebastiano Bagnara, Italy
Sherry Y. Chen, UK
Marvin J. Dainoff, USA
Jianming Dong, USA
John Eklund, Australia
Xiaowen Fang, USA
Ayse Gurses, USA
Vicki L. Hanson, UK
Sheue-Ling Hwang, Taiwan
Wonil Hwang, Korea
Yong Gu Ji, Korea
Steven A. Landry, USA
Gitte Lindgaard, Canada
Chen Ling, USA
Yan Liu, USA
Chang S. Nam, USA
Celestine A. Ntuen, USA
Philippe Palanque, France
P.L. Patrick Rau, P.R. China
Ling Rothrock, USA
Guangfeng Song, USA
Steffen Staab, Germany
Wan Chul Yoon, Korea
Wenli Zhu, P.R. China

Engineering Psychology and Cognitive Ergonomics

Program Chair: Don Harris

Guy A. Boy, USA
Pietro Carlo Cacciabue, Italy
John Huddlestone, UK
Kenji Itoh, Japan
Hung-Sying Jing, Taiwan
Wen-Chin Li, Taiwan
James T. Luxhøj, USA
Nicolas Marmaras, Greece
Sundaram Narayanan, USA
Mark A. Neerincx, The Netherlands
Jan M. Noyes, UK
Kjell Ohlsson, Sweden
Axel Schulte, Germany
Sarah C. Sharples, UK
Neville A. Stanton, UK
Xianghong Sun, P.R. China
Andrew Thatcher, South Africa
Matthew J.W. Thomas, Australia
Mark Young, UK
Rolf Zon, The Netherlands

Universal Access in Human–Computer Interaction

Program Chair: Constantine Stephanidis

Julio Abascal, Spain
Ray Adams, UK
Elisabeth André, Germany
Margherita Antona, Greece
Chieko Asakawa, Japan
Christian Bühler, Germany
Jerzy Charytonowicz, Poland
Pier Luigi Emiliani, Italy
Michael Fairhurst, UK
Dimitris Grammenos, Greece
Andreas Holzinger, Austria
Simeon Keates, Denmark
Georgios Kouroupetroglou, Greece
Sri Kurniawan, USA
Patrick M. Langdon, UK
Seongil Lee, Korea

Zhengjie Liu, P.R. China
Klaus Miesenberger, Austria
Helen Petrie, UK
Michael Pieper, Germany
Anthony Savidis, Greece
Andrew Sears, USA
Christian Stary, Austria
Hirotada Ueda, Japan
Jean Vanderdonckt, Belgium
Gregg C. Vanderheiden, USA
Gerhard Weber, Germany
Harald Weber, Germany
Panayiotis Zaphiris, Cyprus

Virtual and Mixed Reality

Program Chair: Randall Shumaker

Pat Banerjee, USA
Mark Billinghurst, New Zealand
Charles E. Hughes, USA
Simon Julier, UK
David Kaber, USA
Hirokazu Kato, Japan
Robert S. Kennedy, USA
Young J. Kim, Korea
Ben Lawson, USA
Gordon McK Mair, UK
David Pratt, UK
Albert "Skip" Rizzo, USA
Lawrence Rosenblum, USA
Jose San Martin, Spain
Dieter Schmalstieg, Austria
Dylan Schmorrow, USA
Kay Stanney, USA
Janet Weisenford, USA
Mark Wiederhold, USA

Internationalization, Design and Global Development

Program Chair: P.L. Patrick Rau

Michael L. Best, USA
Alan Chan, Hong Kong
Lin-Lin Chen, Taiwan
Andy M. Dearden, UK
Susan M. Dray, USA
Henry Been-Lirn Duh, Singapore
Vanessa Evers, The Netherlands
Paul Fu, USA
Emilie Gould, USA
Sung H. Han, Korea
Veikko Ikonen, Finland
Toshikazu Kato, Japan
Esin Kiris, USA
Apala Lahiri Chavan, India
James R. Lewis, USA
James J.W. Lin, USA
Rungtai Lin, Taiwan
Zhengjie Liu, P.R. China
Aaron Marcus, USA
Allen E. Milewski, USA
Katsuhiko Ogawa, Japan
Oguzhan Ozcan, Turkey
Girish Prabhu, India
Kerstin Röse, Germany
Supriya Singh, Australia
Alvin W. Yeo, Malaysia
Hsiu-Ping Yueh, Taiwan

Online Communities and Social Computing

Program Chairs: A. Ant Ozok, Panayiotis Zaphiris

Chadia N. Abras, USA
Chee Siang Ang, UK
Peter Day, UK
Fiorella De Cindio, Italy
Heidi Feng, USA
Anita Komlodi, USA
Piet A.M. Kommers, The Netherlands
Andrew Laghos, Cyprus
Stefanie Lindstaedt, Austria
Gabriele Meiselwitz, USA
Hideyuki Nakanishi, Japan
Anthony F. Norcio, USA
Ulrike Pfeil, UK
Elaine M. Raybourn, USA
Douglas Schuler, USA
Gilson Schwartz, Brazil
Laura Slaughter, Norway
Sergei Stafeev, Russia
Asimina Vasalou, UK
June Wei, USA
Haibin Zhu, Canada

Augmented Cognition

Program Chairs: Dylan D. Schmorrow, Cali M. Fidopiastis

Monique Beaudoin, USA
Chris Berka, USA
Joseph Cohn, USA
Martha E. Crosby, USA
Julie Drexler, USA
Ivy Estabrooke, USA
Chris Forsythe, USA
Wai Tat Fu, USA
Marc Grootjen, The Netherlands
Jefferson Grubb, USA
Santosh Mathan, USA
Rob Matthews, Australia
Dennis McBride, USA
Eric Muth, USA
Mark A. Neerincx, The Netherlands
Denise Nicholson, USA
Banu Onaral, USA
Kay Stanney, USA
Roy Stripling, USA
Rob Taylor, UK
Karl van Orden, USA

Digital Human Modeling

Program Chair: Vincent G. Duffy

Karim Abdel-Malek, USA
Giuseppe Andreoni, Italy
Thomas J. Armstrong, USA
Norman I. Badler, USA
Fethi Calisir, Turkey
Daniel Carruth, USA
Keith Case, UK
Julie Charland, Canada
Yaobin Chen, USA
Kathryn Cormican, Ireland
Daniel A. DeLaurentis, USA
Yingzi Du, USA
Okan Ersoy, USA
Enda Fallon, Ireland
Yan Fu, P.R. China
Afzal Godil, USA

Ravindra Goonetilleke, Hong Kong
Anand Gramopadhye, USA
Lars Hanson, Sweden
Pheng Ann Heng, Hong Kong
Bo Hoege, Germany
Hongwei Hsiao, USA
Tianzi Jiang, P.R. China
Nan Kong, USA
Steven A. Landry, USA
Kang Li, USA
Zhizhong Li, P.R. China
Tim Marler, USA
Ahmet F. Ozok, Turkey
Srinivas Peeta, USA
Sudhakar Rajulu, USA
Matthias Rötting, Germany
Matthew Reed, USA
Johan Stahre, Sweden
Mao-Jiun Wang, Taiwan
Xuguang Wang, France
Jingzhou (James) Yang, USA
Gulcin Yucel, Turkey
Tingshao Zhu, P.R. China

Human-Centered Design

Program Chair: Masaaki Kurosu

Julio Abascal, Spain
Simone Barbosa, Brazil
Tomas Berns, Sweden
Nigel Bevan, UK
Torkil Clemmensen, Denmark
Susan M. Dray, USA
Vanessa Evers, The Netherlands
Xiaolan Fu, P.R. China
Yasuhiro Horibe, Japan
Jason Huang, P.R. China
Minna Isomursu, Finland
Timo Jokela, Finland
Mitsuhiko Karashima, Japan
Tadashi Kobayashi, Japan
Seongil Lee, Korea
Kee Yong Lim, Singapore
Zhengjie Liu, P.R. China
Loïc Martínez-Normand, Spain
Monique Noirhomme-Fraiture, Belgium
Philippe Palanque, France
Annelise Mark Pejtersen, Denmark
Kerstin Röse, Germany
Dominique L. Scapin, France
Haruhiko Urokohara, Japan
Gerrit C. van der Veer, The Netherlands
Janet Wesson, South Africa
Toshiki Yamaoka, Japan
Kazuhiko Yamazaki, Japan
Silvia Zimmermann, Switzerland

Design, User Experience, and Usability

Program Chair: Aaron Marcus

Ronald Baecker, Canada
Barbara Ballard, USA
Konrad Baumann, Austria
Arne Berger, Germany
Randolph Bias, USA
Jamie Blustein, Canada
Ana Boa-Ventura, USA
Lorenzo Cantoni, Switzerland
Sameer Chavan, Korea
Wei Ding, USA
Maximilian Eibl, Germany
Zelda Harrison, USA

Rüdiger Heimgärtner, Germany
Brigitte Herrmann, Germany
Sabine Kabel-Eckes, USA
Kaleem Khan, Canada
Jonathan Kies, USA
Jon Kolko, USA
Helga Letowt-Vorbek, South Africa
James Lin, USA
Frazer McKimm, Ireland
Michael Renner, Switzerland
Christine Ronnewinkel, Germany
Elizabeth Rosenzweig, USA
Paul Sherman, USA
Ben Shneiderman, USA
Christian Sturm, Germany
Brian Sullivan, USA
Jaakko Villa, Finland
Michele Visciola, Italy
Susan Weinschenk, USA

HCI International 2013

The 15th International Conference on Human–Computer Interaction, HCI International 2013, will be held jointly with the affiliated conferences in the summer of 2013. It will cover a broad spectrum of themes related to human–computer interaction (HCI), including theoretical issues, methods, tools, processes and case studies in HCI design, as well as novel interaction techniques, interfaces and applications. The proceedings will be published by Springer. More information about the topics, as well as the venue and dates of the conference, will be announced through the HCI International Conference series website: http://www.hci-international.org/

General Chair
Professor Constantine Stephanidis
University of Crete and ICS-FORTH
Heraklion, Crete, Greece
Email: cs@ics.forth.gr

Table of Contents

Part I: Theories, Models and Technologies for Augmented Cognition

Part III: Augmented Cognition, Social Computing and Collaboration

Part IV: Augmented Cognition for Learning

Part V: Augmented Cognition and Interaction

Part VI: Augmented Cognition in Complex Environments

Part I

Theories, Models and Technologies for Augmented Cognition

The Brain as Target Image Detector: The Role of Image Category and Presentation Time

Anne-Marie Brouwer[1], Jan B.F. van Erp[1], Bart Kappé[1], and Anne E. Urai[1,2]

[1] TNO Human Factors, Kampweg 5, 3769 ZG Soesterberg, The Netherlands
[2] University College Utrecht, P/O. Box 80145, 3508 TC Utrecht, The Netherlands
{anne-marie.brouwer,jan.vanerp}@tno.nl, bart.kappe@xs4all.nl, anne.urai@gmail.com

Abstract. The brain can be very proficient in classifying images that are hard for computer algorithms to deal with. Previous studies show that EEG can contribute to sorting shortly presented images in targets and non-targets. We examine how EEG and classification performance are affected by image presentation time and the kind of target: humans (a familiar category) or kangaroos (unfamiliar). Humans are much easier detected as indicated by behavioral data, EEG and classifier performance. Presentation of humans is reflected in the EEG even if observers were attending to kangaroos. In general, 50ms presentation time decreased markers of detection compared to 100ms.

1 Introduction

Recent technological developments have lowered the costs of gathering and storing high volumes of images. Enormous amounts of images are digitally available in fields ranging from internet search engines to security cameras and satellite streams. Finding an image of interest requires a system of image triage through which only a subset of images is selected for further visual inspection. However, in some cases, automatic analysis of image contents is difficult because computer vision systems lack the sensitivity, specificity and generalization skills needed for efficient image triage. The human brain, on the other hand, can be extremely apt at image classification and can recognize target images quickly and precisely. Participants in a study by Thorpe et al. [1] had to indicate whether a previously unseen photograph, flashed for just 20 ms, contained an animal or not by releasing or holding a button. Already 150 ms after stimulus onset EEG (electroencephalography) signals for target and non-targets started to differ reliably– a frontal negativity developed for non-target images. Similar results were found by Goffaux et al. [2] where observers had to categorize types of landscape. An image classification BCI (Brain Computer Interface) may provide us access to these very powerful brain mechanisms to interpret images and enable observers to reliably classify images at very high speeds.

Several groups have already implemented image classification BCIs, usually based on a particular event related potential (ERP) present in the EEG, called the P3. The P3 is a positive peak in EEG that occurs approximately 300 ms after a target stimulus (a stimulus that the observer is attending to) is presented [3]. Sajda, Parra, Gerson and colleagues [4-7] presented their observers with sequences of 50 to 100 images of

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 3–12, 2011.

natural scenes, where each image was presented for 100 ms. Observers had to press or release a button right after detecting a natural scene containing people, or after the sequence had ended. Each sequence contained 1 or 2 of these targets. They found that both EEG and button presses contributed to correctly ordering images from more to less likely to be a target. Similarly, Huang, Pavel, and colleagues [8-10] presented sequences of 50 satellite images, where each image was presented for 60 to 200 ms. Targets were man-made objects such as ships, oil storage depots or golf courses. Half of the sequences contained 1 target, the other half none. Observers pressed a button directly after detecting the target or after the sequence had ended. They also found that both EEG and button presses contributed to correct classification.

The previous studies show the feasibility of image classification BCIs. In our research we want to build a BCI to classify shortly presented images, but in line with virtually all real-life image classification situations and (partly) in contrast to the previous studies, observers are unaware of the number of targets. This may be an important factor. If observers know that one target will be present, they may quit paying attention after target detection, or, if they did not see the target yet, anticipate it towards the end of the sequence. Also note that few compared to many targets may enhance P3 size [3]. We here focus on the role of the image category of the target, or target type, within a fixed collection of context images. It may not be possible to generalize results of the studies mentioned before when other types of targets (within other types of contexts) are searched. When e.g. looking for a human in a natural environment, the observers' expertise of human appearances can support performance. In this study we compare brain responses to attended or unattended images of humans to those of kangaroos. Thus, we compare between groups of images that are always the same, the only difference being which group attention is focused on. Since our European observers are more familiar with recognizing humans than kangaroos, detecting humans amongst other animals may be easier than detecting kangaroos and correspondingly, produce stronger P3s. In addition, specific ERP components that are associated with faces or other highly familiar stimuli such as the N170 may be present [11-13]. If so, and if in a particular image classification case the target of interest corresponds to such a familiar stimulus, these could be used in classifiers. Together with the effect of target type, we examine the effect of presentation time (100 or 50 ms). Interactions between target type and presentation time may occur, such as kangaroo images eliciting P3s when they are presented long, but not when they are presented for a short time. Besides examining the ERPs directly, we also look for effects of target type and presentation time on classifier performance.

2 Methods

2.1 Participants

Twenty observers (10 men and 10 women) participated in the experiment. Their mean age was 38.9 years (SD= 16.6). As verified by a questionnaire, all participants were neurologically healthy and had normal or corrected to normal vision. Participants gave their informed consent before the start of the experiment and were given a monetary reward for their time.

2.2 Stimuli

All images were obtained from the Caltech-256 Object Category Dataset [14]. Images that were not clearly recognizable or had written text on them were excluded from the experiment. Only images in portrait format were used. In total 952 images were selected for use in the experiment, including 55 images of humans and 40 images of kangaroos. Images were normalized in size and in luminance using Matlab. Their size was reduced to 280 x 420 pixels. They were then transformed to the CIELAB Lab color space, where the average and standard deviation of luminance (L-component) were set to 30 and 25.2 respectively. Then, the images were transformed back to sRGB. Custom built software presented sequence of 60 images on a Dell 1907 LCD flat panel display (19 inch, 60 Hz, 1280 x 1024 pixels) at a viewing distance of about 70 cm. Each image was presented for 50 or 100 ms. In between image sequences, a white screen was shown for 1s followed by a white screen with a black fixation cross that was presented for a randomly chosen interval between 0.8 and 1.2 s.

2.3 Design

For each presentation time (50 and 100 ms), each participant completed 10 runs with target type human and 10 runs with target type kangaroo. Each run consisted of 5 image sequences of 60 images each. Sequences could contain between 0 and 4 targets as well as 0 to 4 non-targets. Non-targets were images of kangaroos for the human target type and images of humans for the kangaroo target type. The resulting 25 combinations of target and non-target numbers were randomly distributed across runs and occurred twice within each of the four conditions (four combinations of target type and presentation time). Image sequences were generated taking into account the following constraints. There were always at least six fillers (images of animals that were neither humans nor kangaroos) between targets and non-targets. Targets and non-targets were never among the first or last 4 images. Within one run, images were never shown more than once. Half of the participants first performed the task at 100 ms/image and then at 50 ms/image, the other half the other way around. The order of target types was counterbalanced across participants. The 10 runs were order balanced using a latin square.

2.4 Task and Procedure

Participants were seated comfortably in front of a monitor in a shielded room. Before the experiment started, the complete procedure was explained to the participants. The participants were asked to blink as little as possible and to limit any other movements during image sequences. The task was to concentrate on target images and count the number of times they were presented in an image sequence. The participants were informed of the nature of the targets (either humans or kangaroos) before every new target type. Every time that the target type and/or presentation speed changed, participants were given a training run to get used to the target type they had to detect and the speed of presentation. Participants entered the number of targets they had counted on a keypad during a time window of about 2s in between image sequences.

2.5 EEG Recording

EEG activity was recorded at the Fpz, Fz, Fp1, Fp2, Cz, Pz, P3, P4, Oz, POz, PO7 and PO8 electrode sites of the 10-20 system [15] using Au electrodes mounted in an EEG cap (g.tec medical engineering GmbH). A ground electrode was attached to the scalp at the AFz electrode site. The EEG electrodes were referenced to linked mastoid electrodes. The impedance of each electrode was below 5 kΩ. Data were sampled with a frequency of 256 Hz and filtered before storage by a 0.1 Hz high pass-, a 60 Hz low pass- and a 50 Hz notch filter (USB Biosignal Amplifier, g.tec medical engineering GmbH). Additional electrodes (Kendall Neonatal ECG electrodes from Tyco Healthcare Deutschland GmbH) were positioned above and below the left eye, and close to the outer canthi of the eyes to monitor EOG (electro-oculography - blinks and eye movements). EOG electrodes were referenced to each other. Data recording was controlled by a combination of custom-built software and Matlab/Simulink tools.

2.6 Analysis

EEG Signal Analysis. The EEG/EOG data were processed using Brain Vision Analyzer 2.0 (BrainProducts). We started out with data from the interval between the first fixation cross to 2s after the last image of the run. All EEG channels were automatically inspected for bad episodes, using standard settings of Brain Vision Analyzer. This identified most of the eye blinks, which mainly occurred in between image sequences in accordance with experimental instructions. Bad episodes were excluded from the analysis. The remaining data were manually inspected for further irregularities to remove all eye blinks and other artifacts from the data. EOG data were not further used. Segments were then selected starting at 200 ms before image onset and 600 ms after image onset. Since there were many more filler segments than target and non-target segments, only every fourth of the filler segments was used. Segments were baseline corrected using an interval of 200 ms to 0 ms before stimulus onset. Averages were calculated for targets, non-targets and fillers per participant and condition. Visual inspection of these averages revealed that the N170 component appeared between 100 and 350 ms after stimulus onset. The P3 component appeared between 300 and 550 ms after stimulus onset. The area in μV*ms within these timeframes was taken as a measure for the magnitude of the respective components.

For further P3 analysis, we selected data recorded at Pz because Pz is known to be a good location for measuring P3 [16] and these indeed distinguished well between targets and non-targets. More specifically, all electrodes distinguish well between targets and non-targets except for Fpz, Fp1 and Fp2. At these locations, paired t-tests on P3s per participant, electrode, targets and non-targets do not indicate significant differences between targets and non-targets (p-values > 0.11). For all other electrodes, p-values are <0.01.

For further N170 analysis, we selected data from P4 since in our study, these appeared to distinguish best between images of humans and kangaroos. When target type is human, the N170 is larger for targets (human) than non-targets (kangaroo) at all electrodes (paired t-tests on N170 per participant, electrode, target and non-target resulted in p-values <0.03). When target type is kangaroo, the N170 tends to be larger for non-targets (human) than targets (kangaroo) at all electrodes; significantly so for Pz (t_{19}=2.60, p=0.02), P3 (t_{19}=2.75, p=0.01), P4 (t_{19}=2.88, p<0.01) and PO7 (t_{19}=2.21, p=0.04).

We computed dP3 as the difference between the P3 elicited by targets and the P3 elicited by non-targets, separately for participants and conditions. A positive dP3 reflects a larger P3 for targets than for non-targets. This is the part of the P3 that we are interested in and that we want to check for sensitivity to target type and presentation time. Similarly, we computed dN170 as the difference between the N170 elicited by human images and the N170 elicited by kangaroo images, separately for participants and conditions. A positive dN170 reflects a larger N170 for human images than for kangaroo images.

Classification. We performed a cross-validation using a type of discriminant analysis as described in [17] which, in short, works as follows. EEG data was segmented in 1 sec epochs, starting at stimulus onset. EEG segments were normalized such that their average value was zero. For each participant and each run, segments following target stimuli were averaged across all electrodes. Segments following filler and non-target stimuli were averaged likewise. This was done using all sequences minus one. The difference between the two average responses (target minus filler/non-target) or 'template' was multiplied by single segments from the sequence left out when constructing the template. If for a certain segment, the mean result exceeded a certain threshold, it was interpreted as a response to a target, and otherwise as a response to a filler/non-target. The threshold was chosen such that we had equal percentages of target responses that were wrongly classified as filler/non-target responses and filler/non-target responses wrongly classified as target responses. This prevented the problem of judging a classifier that labels all segments as filler/non-targets as a good classifier since it only makes an error about twice in a sequence. The percentage of wrongly identified responses that we get when applying the threshold as described above is termed Equal Error Rate (EER). It is a measure of classification performance, where an EER of 50 means that a target or a non-target/filler has a 50% chance of not being identified as such, thus reflecting chance performance, and an EER of 0 means perfect performance.

Statistical analysis. Repeated measures ANOVAs were used to test for effects of target type (human or kangaroo) and presentation time (50 or 100 ms) on dP3, dN170, EER and counting error. Counting error was the absolute difference between the number of targets present in an image sequence and the number as reported by the participant. When appropriate, Tukey HSD post-hoc tests were performed to establish the nature of an effect. One sample t-tests were used to examine whether dP3 and dN170 were different from zero in all four conditions.

3 Results

3.1 ERPs

General. Figure 1 gives a general impression of the ERPs. It shows Pz grand averages of EEG separately for targets and non-targets, and for human and kangaroo images. Human targets appear to elicit both a negative component between 100 and 350 ms after stimulus onset (N170) and a positive component between 300 and 550ms (P3). Non-target images of humans seem to elicit an N170 as well, but no P3. Images of kangaroos do not seem to generate a N170 or a P3, although the EEG seems slightly more positive for kangaroo targets than for kangaroo non-targets towards the end.

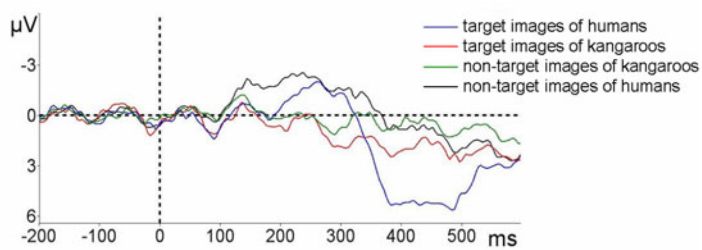


Fig. 1. EEG grand averages recorded at Pz, separately for targets and non-targets, and for human and kangaroo images. Time 0 corresponds to stimulus onset.

P3. Figure 2A shows dP3 (P3 at Pz for the target minus that for the non-target) for each target type and presentation time. The mean values are larger than zero, except for the kangaroo target presented at 50 ms, suggesting that generally an attention driven P3 is present. However, only for the human target presented at 100 ms the dP3 is significantly larger than 0 (one sample t-test t_{19}=3.78, p<0.01, other p-values >0.32). A repeated measures ANOVA indicates that the dP3 is affected by target type ($F_{(1,19)}$=5.61, p=0.03) with the human target producing larger dP3s than the kangaroo, and presentation time ($F_{(1,19)}$=7.04, p=0.02) with 100 ms producing larger dP3s than 50 ms. There is no interaction between target type and presentation time ($F_{(1,19)}$=0.04, p=0.84).

N170. Figure 2B shows dN170 (N170 at P4 for the human image minus that for the kangaroo) for each target type and presentation time. An N170 is present as indicated by mean dN170 values being significantly larger than zero for the human target types (t_{19}=3.08 for 100 ms presentation time and t_{19}=3.12 for 50 ms, both p-values<0.01) and the kangaroo target type at 100 ms presentation time (t_{19}=3.37, p<0.01). For the kangaroo type - 50 ms condition dN170 is not significantly larger than zero (t_{19}=0.34, p=0.98). A repeated measures ANOVA indicates that there is no main effect of target type ($F_{(1,19)}$=0.01, p=0.93) on dN170. There is an effect of presentation time ($F_{(1,19)}$=5.59, p=0.03) and an interaction between target type and presentation time ($F_{(1,19)}$=6.36, p=0.02). Tukey HSD post-hoc tests shows that for human target type conditions, presentation time does not affect the dN170 while for kangaroo target type conditions, dN170 is larger when presentation time is 100 ms compared to 50 ms (difference between the different presentation times for kangaroo target type conditions: p=0.02; all other comparisons p-values>0.28).

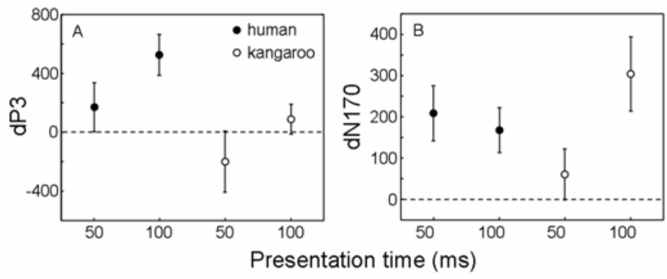


Fig. 2. dP3: P3 at Pz for the target minus that for the non-target (A) and dN170: N170 at P4 for the human image minus that for the kangaroo (B) for each target type and presentation time. Error bars represent standard errors of the mean.

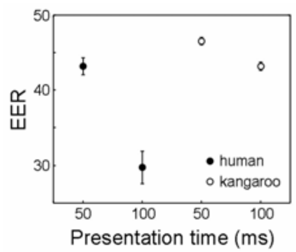


Fig. 3. Equal error rate for each target type and presentation time. Perfect classification performance would be at 0 and chance performance at 50. Error bars represent standard errors of the mean.

3.2 Classification

Figure 3 shows mean equal error rate (EER) for each target type and presentation time where 0 indicates perfect classification performance and 50 chance performance. A repeated measures ANOVA showed an effect of presentation time ($F_{(1,19)}$=94.43, p<0.01), target type ($F_{(1,19)}$=36.69, p<0.01) and an interaction ($F_{(1,19)}$=42.34, p<0.01). Tukey HSD post-hoc tests indicated that the EER for human target at 100 ms was lower than the EER in all other conditions (p-values <0.01 for all comparisons) while there were no other significant differences between the EERs (p-values >0.11).

3.3 Counting Error

Figure 4A plots the number of targets reported against the number of targets presented, separately for each target type and presentation time. The figure shows a clear pattern of better performance for human compared to kangaroo targets, and for long compared to short presentation times. Better performance is reflected by stronger positive relations between counted and presented number of targets. Figure 4B shows the mean counting error for each target type and presentation time. Target type ($F_{(1,19)}$=405.00, p<.01) and presentation time ($F_{(1,19)}$=75.39, p<.01) both have an effect on counting error. There is an interaction between presentation time and target type ($F_{(1,19)}$=8.80, p<.01) that indicates that counting human targets benefits more from a longer presentation time than kangaroo targets does.

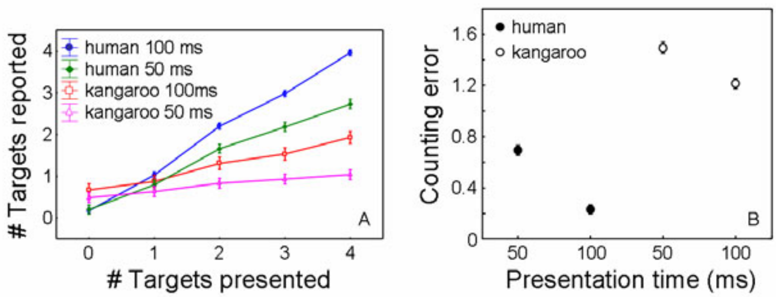


Fig. 4. Counting performance for each target type and presentation time as indicated by the number of targets reported against the number of targets presented (A) and counting error (B). Error bars represent 95% confidence intervals.

4 Conclusions

Our results clearly indicate that target type (in combination with the type of fillers) can affect performance of image classification BCIs. Images of humans as targets elicit stronger P3s than images of a less familiar species, kangaroos. In line with this, classification performance was better for human targets (when images were presented for 100 ms) as well as counting performance. Detecting kangaroos in between animals appeared close to impossible with the short presentation times used and this was reflected in all dependent measures.

Human images also elicited an N170 whereas kangaroo images did not. For the presentation time of 100 ms, this even held when observers focused on detecting images of kangaroos (in line with [18]). This means that besides the P3, the N170 can be used in classification algorithms when the goal of the image classification BCI is to find images of humans. Our finding that an electrode at the right hemisphere tended to display the N170 best is in line with previous findings (e.g. [19]). However, better N170 results are expected when EEG electrodes are placed more temporally than in the current study [20, 21].

Besides target type, presentation time also affected the results. As demonstrated before, shorter presentation times (below 500 ms – [22]) result in smaller or harder to detect P3s and lower classification rates [4, 8, 23]. This is probably caused both by the fact that short presentation times limit processing time and because they increase overlap of ERPs resulting in noisier signals. Presentation time also interacted with target type. Presentation time did not affect the N170 with a human target, but when the target was a kangaroo, 100 ms presentation time produced a clearer N170 than 50 ms (in which case dN170 was not significant). For classification, longer presentation time increases performance, though not significantly so for the kangaroo, perhaps because this target type was just too difficult. Similarly, an interactive effect of presentation time and target type occurred for counting performance with longer presentation time increasing performance, but especially so for the human target.

In the present study, we focused on targets and non-targets, leaving ERPs elicited by the fillers largely aside. In general we can say that for the P3, filler and non-target data are virtually indistinguishable. The fact that non-targets do not elicit larger P3s than fillers, indicates that observers succeeded in switching attention between target types. Fillers elicit N170s that are in between those elicited by human images and kangaroo images. This is probably caused by filler images that resemble humans, like images of other primates. Note that the N170 would probably have been stronger when only images of human faces had been used instead of the broader category of images of humans that also included whole body pictures.

In conclusion, performance of image classification BCIs cannot simply be generalized to situations in which other types of targets are being searched. Image classification BCIs will probably be most successful when typical human expertise can be used, such as detecting human faces. In that case even unintentional detection could be tapped. In addition, future studies should take care not to design image classification BCIs for cases that can be effectively dealt with by computer algorithms.

Acknowledgements. The authors gratefully acknowledge the support of the BrainGain Smart Mix Programme of the Netherlands Ministry of Economic Affairs and the Netherlands Ministry of Education, Culture and Science, and David van Leeuwen for the classification analysis.

References

1. Thorpe, S., Fize, D., Marlot, C.: Speed of processing in the human visual system. Nature 381, 520–522 (1996)
2. Goffaux, V., Jacques, C., Mouraux, A., Oliva, A., Schyns, P.G., Rossion, B.: Diagnostic colours contribute to the early stages of scene categorization: Behavioural and neurophysiological evidence. Vis. Cogn. 12, 878–892 (2005)
3. Farwell, L.A., Donchin, E.: Talking off the top of your head: A mental prosthesis utilizing event-related brain potentials. Electroencephalography and Clinical Neurophysiology 70, 510–523 (1988)
4. Sajda, P., Gerson, A., Parra, L.: High-throughput image search via single-trial event classification in a rapid serial visual presentation task. In: Proc. First International IEEE EMBS Conference on Neural Engineering, pp. 7–10 (2003)
5. Gerson, A.D., Parra, L.C., Sajda, P.: Cortically-coupled Computer Vision for Rapid Image Search. IEEE Trans. on Neural Systems & Rehabilitation Engineering 14, 174–179 (2006)
6. Sajda, P., Gerson, A.D., Philiastides, M.G., Parra, L.C.: Single-trial analysis of EEG during rapid visual discrimination: Enabling cortically-coupled computer vision. In: Dornhege, G., Mueller, K.-R. (eds.) Brain-Computer Interface. MIT Press, Cambridge (2007)
7. Parra, L.C., Christoforou, C., Gerson, A.D., Dyrholm, M., Luo, A., Wagner, M., Philiastides, M.G., Sajda, P.: Spatio-temporal linear decoding of brain state: Application to performance augmentation in high-throughput tasks. IEEE Signal Processing Magazine 25, 95–115 (2008)

8. Huang, Y., Erdogmus, D., Mathan, S., Pavel, M.: Comparison of Linear and Nonlinear Approaches on Single Trial ERP Detection in Rapid Serial Visual Presentation Tasks. In: International Joint Conference on Neural Networks, pp. 1136–1142 (2006)
9. Huang, Y., Erdogmus, D., Mathan, S., Pavel, M.: A Fusion Approach for Image Triage using Single Trial ERP Detection. In: 3rd International IEEE/EMBS Conference on Neural Engineering, pp. 473–476 (2007)
10. Huang, Y., Erdogmus, D., Mathan, S., Pavel, M.: Large-scale image database triage via EEG evoked responses. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 429–432 (2008)
11. Bentin, S., McCarthy, G., Perez, E., Puce, A., Allison, T.: Electrophysiological studies of face perception in humans. J. Cogn. Neurosci. 8, 551–565 (1996)
12. Rossion, B., Jacques, C.: Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. NeuroImage 39, 1959–1979 (2008)
13. Busey, T.A., Vanderkolk, J.R.: Behavioral and electrophysiological evidence for configural processing in fingerprint experts. Vis. Res. 45, 431–448 (2005)
14. Griffin, G., Holub, A., Perona, P.: Caltech-256 Object Category Dataset. California Institute of Technology (2007), `http://resolver.caltech.edu/CaltechAUTHORS:CNS-TR-2007-001`
15. Jasper, H.: Report of the committee on methods of clinical examination in electroencephalography. Electroencephalography and Clinical Neurophysiology 10, 370–375 (1958)
16. Ravden, D., Polich, J.: On P300 measurement stability: habituation, intra-trial block variation, and ultradian rhythms. Biol. Psych. 51, 59–76 (1999)
17. Bandt, C., Weymar, M., Samaga, D., Hamm, A.O.: A simple classification tool for single-trial analysis of ERP components. Psychophysiol. 46, 747–757 (2009)
18. Rousselet, G.A., Macé, M.J., Fabre-Thorpe, M.: Animal and human faces in natural scenes: How specific to human faces is the N170 ERP component? JOV 4, 13–21 (2004)
19. Rossion, B., Joyce, C.J., Cottrell, G.W., Tarr, M.J.: Early lateralization and orientation tuning for face, word and object processing in the visual cortex. Neuroimage 20, 1609–1624 (2003)
20. Kanwisher, N., McDermott, J., Chun, M.M.: The fusiform face area: A module in human extrastriate cortex specialized for face perception. J. Neurosci. 17(11), 4302–4311 (1997)
21. Luck, S.J.: An Introduction to the Event-Related Potential Technique. MIT Press, Cambridge (2005)
22. Shenoy, P., Tan, D.S.: Human-Aided Computing: Utilizing Implicit Human Processing to Classify Images. In: CHI 2008 Proceedings Cognition, Perception, and Memory (2008)
23. Yazdani, A., Vesin, J.-M., Izzo, D., Ampatzis, C., Ebrahimi, T.: Implicit retrieval of salient images using brain computer interface. In: Proceedings of International Conference on Image Processing, ICIP (2010)

Implementation of fNIRS for Monitoring Levels of Expertise and Mental Workload

Scott C. Bunce[1,2], Kurtulus Izzetoglu[2], Hasan Ayaz[2], Patricia Shewokis[2,3], Meltem Izzetoglu[2], Kambiz Pourrezaei[2], and Banu Onaral[2]

[1] Penn State Hershey Medical Center and Penn State College of Medicine
[2] Drexel University School of Biomedical Engineering, Sciences, and Health Systems
[3] College of Nursing and Health Professions, Drexel University

Abstract. An accurate measure of mental workload would help improve operational safety and efficacy in many environments that involve multitasking or sustained vigilance. The current study utilized functional near-infrared spectroscopy (fNIRs) to examine the relationship of the hemodynamic response in dorsolateral prefrontal cortex (DLPFC) as it related to mental workload, level of expertise, and task performance. DLPFC responses were monitored with fNIRs while 8 participants (4 with high practice, 4 novices) completed a quasi-realistic computerized Warship Commander Task with various levels of difficulty. The results show that greater expertise was associated with relatively lower oxygenation (less neural activity) at low to moderate levels of taskload, but higher oxygenation and better performance at high levels of taskload. For novices, oxygenation was higher at moderate levels of taskload, but dropped precipitously at higher levels of taskload, along with performance, consistent with disengaging from the task. Results are interpreted within a “scaffolding-storage" framework.

Keywords: Optical Brain Imaging; functional near infrared spectroscopy; mental workload; expertise; practice, fNIR.

1 Introduction

Mental workload plays a critical role in many complex command and control systems. It is particularly important to understand operator workload in situations where performance failures could result in catastrophic losses (e.g., warship command, Air Traffic Control (ATC)). Accurate assessment of mental workload could help prevent operator error and allow timely intervention by predicting performance decline that can arise from either work overload or from understimulation [1-3]. For some time, investigators have been working on adaptive aiding schemes to facilitate optimal performance in critical mission systems by dynamically matching the momentary mental capabilities of the operator to the task demands [4-5]. To be viable, such a system must improve performance above the levels possible with unaided and fully automated systems [3, 6]. Further, adaptive aiding systems should provide aid only when required [7], as providing an unnecessary intervention can lead to performance errors as readily as not providing aid when it is required [1-3]. An accurate and

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 13–22, 2011.

reliable measure of the operator's mental workload is a critical component of any such adaptive aiding system [8].

Within the basic methods of measuring mental workload, physiological measures are the most promising for field applications. Neurophysiological and psychophysiological measures are commonly used to index the level of mental demand associated with a given task because they can provide continuous and unobtrusive monitoring that does not contribute to the operator's workload [9-16].

There is considerable evidence that neurophysiological and psychophysiological variables respond to cognitive demand in a predictable manner [17]. Direct measures of central nervous system function such as electroencephalography (EEG) and event-related potentials (ERPs) have been particularly strong candidates for accurate, objective measures of operator workload. Increasing task difficulty, for instance, is known to be associated with EEG changes such as increased power in the beta bandwidth, increased theta activity at frontal sites and the suppression of alpha activity [12, 18-19]. Although EEG has many excellent qualities for monitoring mental workload, including excellent temporal resolution, it is limited in its capacity for spatial resolution.

In this study, we utilized an optical measure of neural activity, functional near infrared spectroscopy (fNIRs), a good candidate for measuring mental workload under field conditions. fNIRS is safe, highly portable, user-friendly and relatively inexpensive, with rapid application times and near-zero run-time costs [20]. The most commonly used form of fNIRs uses light, introduced at the scalp, to measure changes in blood oxygenation as oxy-hemoglobin converts to deoxyhemoglobin during neural activity, ie., the hemodynamic response. Although its temporal resolution is limited to that of the hemodynamic response, fNIRs provides good spatial localization relative to EEG, on the order of $1cm^2$, and has the capacity to be integrated with EEG/event-related potentials (ERPs).

The capacity for spatial resolution has important implications for the use of fNIRs as a measure of workload. Current models of automaticity related to the development of expertise in certain tasks suggest that there are shifts in the functional neuroanatomy of task performance that support ongoing cognitive efforts. Operator skill and mental work load are generally inversely related [21-22]. This inverse relationship between expertise and the cognitive demand of a given task impacts the accuracy and interpretation of psychophysiological variables as measures of mental workload [23-24]. But as automaticity develops in various tasks, shifts in the functional neuroanatomy of task performance free up attentional resources, largely associated with prefrontal cortices, for other efforts. The development of expertise, or automaticity, can be characterized as freeing up of the limitations on those attentional resources [25-26].

The literature regarding the effect of practice and expertise on the functional anatomy of task performance is extensive and complex. Practice and the development of expertise have been studied across a range of motor, visuomotor, perceptual and cognitive tasks, and from disparate research perspectives. In summary, four main patterns of practice-related activation change can be distinguished [27]. Practice can lead to 1) an increase in activation in the brain areas involved in task performance, 2) a decrease in those areas, or, 3) a functional redistribution of brain activity, in which some initial areas of activation decrease, whereas other initial areas of activation increase, and 4) a functional reorganization of brain activity, i.e., the pattern of activation increases and decreases occurs in distinct brain areas as well as the initial areas.

There are significant differences in the brain areas involved across these various tasks. However, when practice is associated with the attainment of automatic or asymptotic performance, as is often the case with the development of expertise, the research consistently finds decreased demand on control or attentional processes and a shift to increased demand on storage and processing in task-specific areas, often referred to as the development of "automaticity." Petersen et al. [28] have referred to this phenomenon as a 'scaffolding-storage' framework.

According to Petersen et al. [28], a set of brain areas (the scaffolding) is used to support or cope with novel demands during unskilled, effortful performance. Practice allows processes or associations that are more efficiently stored and accessed elsewhere to be offloaded to those areas, and the scaffolding network is pruned away. The decreased reliance on the 'scaffolding' attentional and control areas is demonstrated by decreased activation in those areas, while an increase in activation is observed in those areas underlying the task-specific processes. Activations seen earlier in practice therefore involve generic attentional and control areas. The prefrontal cortex (PFC), anterior cingulate cortex (ACC) and posterior parietal cortex (PPC) are the main areas considered to perform the 'scaffolding' role, consistent with theories of PFC function and the involvement of these areas in the distributed working memory system. On the other hand, increases associated with highly practiced performance are primarily seen in task-specific areas such as representational cortex — primary and secondary sensory or motor cortex, or in areas related to the storage of those representations, such as parietal or temporal cortex.

The majority of studies examining task practice have found decreases in the extent or intensity of activations, particularly in the attentional and control areas [27]. This finding is true whether the task is primarily motor, as in a golf swing [29] or primarily cognitive in nature, as in the Tower of London problem [30]. Decreases in activation represent a contraction of the neural representation of the stimulus [31] or a more precise functional circuit [32]. This finding provides an important overlap with the literature on expertise. There is considerable evidence that overall, experts show lower brain activity relative to novices, particularly in prefrontal areas [29]. Both practice and the development of expertise (the latter of which includes individual differences in ability) typically involve decreased activation across attentional and control areas, freeing these neural resources to attend to other incoming stimuli or task demands. As such, measuring activation in these attentional and control areas relative to task performance can provide an index of level of expertise. One way to conceptualize this approach is that a relative quantification of the attentional and control resources necessary to perform at a given level can serve as an index of the trainee's neural "reserves," a capacity that can be used to perform effectively under greater situational demands.

A neuronal measure of expertise must be defined in relationship to behavioral performance. However, at a given level of performance, a neuronal measure of expertise that monitors the attentional and control resources the individual must utilize to maintain that level of performance could be expected to differentiate between relatively lesser and greater expertise. That is, even at 98-100% performance levels, where performance measures cannot differentiate between trainee capacities, some individuals will be performing at close to their peak performance, whereas others will be well below their performance capacity. An assessment of the cortical activity necessary to perform at a given level would indicate the cognitive resources available for more situational demands, consistent with greater expertise.

Below we provide preliminary evidence that fNIR can be used to 1) differentiate among individuals with more or less practice at a given task, and 2) differentiate among individuals who perform better or worse on a task that requires little practice, i.e., an example of neural efficiency. Previously, we have reported that the hemodynamic response over dorsolateral and ventrolateral prefrontal cortex, assessed using fNIRs, was responsive to workload in a realistic command and control task [33]. In this report, we examined the role of expertise (relative levels of practice) on the hemodynamic response over dorsolateral and ventrolateral prefrontal cortex using the same data set.

2 Method

Methods were identical to those reported in Izzetoglu et al. [33]. In brief, participants were 8 healthy adults (3 females), ages 18 – 50 years. Participants had varied experience with the Warship Commander Task (WCT), ranging from 3-4 hours to 300 hours. For the purposes of this study, a median split was used to divide participants into those with high levels of practice versus those with relatively few hours of practice. All participants signed informed consent statements approved by the Human Subjects Institutional Review Board at Drexel University.

Experimental Task. The WCT (*Pacific Science & Engineering Group, Inc.*) is a quasi-realistic, ship-based navy command and control environment task that requires spatial and verbal working memory and decision-making processes [34]. The version of WCT employed in the current experiment was comprised of two component tasks: Air Warfare Management and the Ship Status (SS) task. Air Warfare Management required the user to monitor “waves” of incoming airplanes, to identify them as friendly or hostile, and to warn and then destroy hostile airplanes using specified rules of engagement. Each wave lasted 75 seconds. The level of cognitive effort in the Air Warfare Management component was manipulated by 1) increasing the number of planes per wave (6, 12, 18 or 24 planes), and 2) changing the proportion of planes whose identity was unknown (hostile or friendly). Airplanes with unknown status require more decision and processing time, and therefore make the task more difficult. The WCT had two levels of difficulty relative to the proportion of unknown planes, *low* in which 33% of the planes were unknown, and *high*, in which 67% of the planes were unknown.

The presence or absence of the Ship Status (SS) Task, a secondary verbal task, was also used to manipulate cognitive workload. The SS task was comprised of auditory messages containing information about the ship and its operation (encoding), and periodic queries, or “recall” about earlier auditory messages. In this auditory task, the participants were required to listen to sporadic auditory messages and memorize various ship status data while answering periodic queries that appeared on the computer screen.

The Air Warfare Management task involves spatial and verbal working memory and decision-making processes, whereas the Ship Status Task is primarily a verbal memory task. When both tasks are operational, the WCT becomes a divided attention task. Although there is no universally accepted definition of cognitive workload, please see St. John et al. [34] for the rationale behind this definition of workload.

Each participant completed four sets of WCT. Each set was comprised of 12 waves, 3 repetitions of each wave size in the order of 6, 18, 12, and 24 planes. The factors of Wavesize (6, 12, 18, or 24 planes), Complexity (*high* versus *low* percentage of unknown airplanes), and *full* versus *divided* attention (secondary SS Task On or Off) were crossed to create a 4 x 2 x 2 repeated-measures design. Performance was assessed using the Percent Game Score, or PctGS, calculated as the percentage of game points that a participant accumulated during any given wave relative to the total game points that were possible for that wave [34].

Analyses. The average rate of change in oxygenation was calculated separately for each 75s wave. Oxygenation values were averaged across the 8 optodes monitoring left and right hemispheres respectively, yielding two values for each wave. Initial analyses showed that hemispheric differences were not significant, so subsequent analyses were collapsed across hemispheres. In these analyses, we were primarily interested in the effects of practice (or level of expertise) on the fNIRs signal as it related to both workload and performance. Analyses were conducted using univariate analyses of variance (MANOVA) in SPSS v.19.

3 Results

Warship Commander Task Performance. Performance data were subjected to a 4 x 2 x 2 factor mixed model ANOVA (with Wavesize and SS Task ON-OFF as within subject factors, and Practice (High, Low) as the between subjects factor). A main effect for Practice ($F(1,6) = 10.25$, $p=.02$, $h=.63$), revealed that more expertise was, as expected, associated with greater game score. The interaction between Practice and Wavesize was also significant ($F(3,18) = 3.15$, $p=.05$, $h=.34$). As displayed in Figure 1, total game score decreased with increasing task load (Wavesize). However, post-hoc analyses revealed that participants with higher levels of practice maintained relatively greater levels of performance as task demand increased, whereas total game scores dropped off for participants with less practice. A main effect for the secondary SS Task was also found ($F(1,6) = 11.45$, $p=.015$, $h=.89$), with the decreased performance when the SS Task was ON (M=87.34, (sd=9.6)) vs. OFF (91.35 (11.9)). There were no other significant effects involving the SS Task.

fNIRs data. fNIRs data were also subjected to a 4 x 2 x 2 mixed model ANOVA (Wavesize, SSTask ON/OFF as within-subjects factors, with Practice (High, Low) as a between-subjects factor). A main effect for Wavesize ($F(3,18) = 7.27$, $p = .002$, $h = .55$) was qualified by an interaction between Practice and Wavesize ($F(3,18) = 4.02$, $p = .02$, $h = .40$). As depicted in Fig. 2a, low to moderate levels of workload, increasing workload (greater wavesize) resulted in increasing oxygenation. Post hoc analyses revealed that at a moderate level of difficulty (12 planes), participants with greater expertise had lower levels of oxygenation. At 18 planes, the difference in oxygenation between high and low levels of was not significant. However, at the highest level of task difficulty (24 planes), oxygenation was greater among participants with more practice.

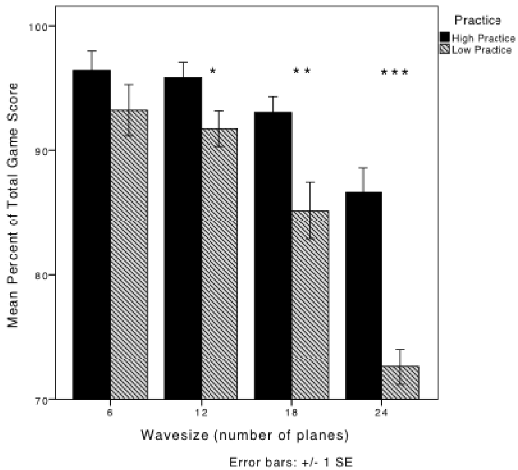


Fig. 1. Percentage of Warship Commander total game score as a function of wavesize (number of planes) and level of expertise. *p < .05; **p < .01; ***p < .001.

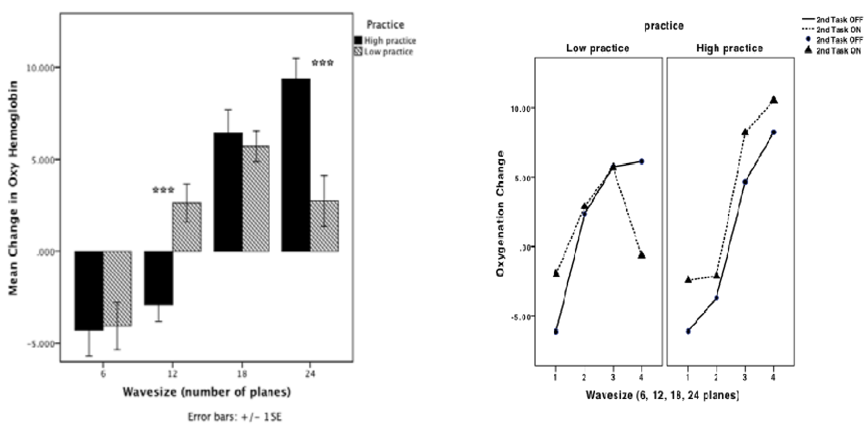


Fig. 2a. fNIRs measures of oxygenation change as a function of wavesize (number of planes) and level of expertise; ***p < .001. **2b.** fNIRs measures of oxygenation change plotted as a function of practice, wavesize, and secondary SStask.

Although there was a main effect for the SSTask (F(1,6) = 6.02, p =.05, *h* = .50), with greater oxygenation when the secondary task was on, the interaction with Practice did not obtain significance (F(3,18) = 2.67, ns. Although this interaction did not obtain significance, a graph of the means helps clarify the nature of the pattern of fNIRs oxygenation measures (Fig 2b). As can be seen in Fig. 3, greater task difficulty was generally associated with greater oxygenation. However, in the highest level of task difficulty, oxygenation in operators with less practice dropped precipitously, as did their game scores.

4 Discussion

The purpose of this study was to examine the impact of practice, or relative levels of expertise, on neurophysiological measures of the hemodynamic response in prefrontal cortex during a complex cognitive task. The complex, but interpretable, pattern of prefrontal cortical response in the WCT is consistent with a 'scaffolding-storage' framework of functional reorganization during the development of expertise [28].

At the lowest level of difficulty (6 planes), neither brain activity nor performance differed much between groups. Moderate levels of objective task difficulty elicited relatively less oxygenation among more experienced operators even as they maintained a high level of task performance relative to novice operators. At the highest levels of objective task difficulty, the oxygenation levels of experienced operators passed that of the novices. Both oxygenation and performance dropped off precipitously for novices at these difficult levels, suggesting a relative disengagement from the task, which was reflected in their performance scores. The more experienced operators maintained high performance scores, suggesting that relatively high levels of neural activity were supporting the more effortful processing necessary to manage the workload. This pattern is consistent with Petersen et al. 's theory that one function of practice and the development of automaticity is to free up attentional processes, i.e., a "cognitive reserve" to be used when task demands increase further, or when unexpected events arise.

Given the spatiotemporal pattern of results, it is likely the the fNIRs was measuring the hemodynamic correlates of sustained mental effort, or vigilance [35-36]. As effort dropped off, so did oxygenation. This suggests an important observation relative to the interpretation of the fNIRs signal. Lower oxygenation values may accurately reflect lower mental effort; however, the current results did not differentiate between an engaged but lower level of workload and disengagement from the task. As such, accurate interpretation of the fNIR signal within an adaptive system will likely need to provide some measure of either task difficulty or operator engagement.

One result that is somewhat difficult to interpret is the lack of differentiation across prefrontal cortex. Based on previous fMRI as well as fNIR studies, we expected more localized findings, particularly as activation during vigilance tends to be stronger in the right hemisphere [37-38]. It may be that the nature and complexity of this particular task elicited activation across both hemispheres; this is an area for future research. One limitation of the current study is the small sample size. With a sample size of four participants per cell, some effects were clearly underpowered. On the other hand, the goal of using such measures in adaptive automation systems requires the measure be sensitive for a single individual. It is likely that some factors that increased workload were operating in areas of the brain that were not being monitored in the current study. More information could be gained by monitoring parietal areas as well as frontal cortex, or combining fNIRs with EEG. Alternatively, signal processing algorithms for fNIRs may not be optimized at this time, or the hemodynamic response itself may not be sufficiently sensitive to pick up subtle changes in workload. This is an area for future research.

The current findings are consistent with numerous studies that have examined the neural correlates of the development of expertise in several cognitive domains [27-29]. These studies provide evidence that cortical activation in DLPFC increases with increasing attention/vigilance, as long as performance is held constant. These DLPFC activations reflect the effort of the participant, not the "objective" level of difficulty perse.

This study provides important albeit preliminary information about fNIR measures of DLPFC hemodynamic response and its relationship to mental workload, expertise, and performance, in a complex multitasking environment. Level of expertise does appear to influence the hemodynamic response in dorsolateral/ventrolateral prefrontal cortices, at least for some complex tasks. Since fNIRs technology allows the development of mobile, non-intrusive and miniaturized devices, it has the potential to be deployed in future learning & training environments to personalize the training regimen and/or assess the effort of human operators in critical multitasking environments.

References

1. Hancock, P.A., Parasuraman, R.: Human-Factors and Safety In The Design Of Intelligent Vehicle-Highway Systems (IVHS). J. Saf. Res. 23, 181–198 (1992)
2. Desmond, P.A., Hoyes, T.W.: Workload variation, intrinsic risk and utility in a simulated air traffic control task: evidence for compensatory effects. Safety Science 22, 87–101 (1996)
3. Hancock, P., Verwey, W.: Fatigue, workload and adaptive driver systems. Accident Analysis & Prevention 29, 495–506 (1997)
4. Young, M.S., Stanton, N.A.: Attention and automation: new perspectives on mental underload and performance. Theoretical Issues in Ergonomics Science 3, 178–194 (2002)
5. Young, M.S., Stanton, N.A.: Malleable attentional resources theory: A new explanation for the effects of mental underload on performance. Human Factors: The Journal of the Human Factors and Ergonomics Society 44, 365 (2002)
6. Parasuraman, R., Riley, V.: Humans and automation: Use, misuse, disuse, abuse. Human Factors 39 (1997)
7. Scerbo, M.W.: Theoretical perspectives on adaptive automation. In: Automation and Human Performance: Theory and Applications, pp. 37–63 (1996)
8. Parasuraman, R.: Neuroergonomics: Research and practice. Theoretical Issues in Ergonomics Science 4, 5–20 (2003)
9. Backs, R.W.: Engineering psychophysiology as a discipline: Historical and theoretical aspects. Engineering Psychophysiology: Issues and Applications (2000)
10. Hancock, P.A., Desmond, P.A.: Stress, workload, and fatigue. CRC, Boca Raton (2001)
11. Byrne, E.A., Parasuraman, R.: Psychophysiology and adaptive automation. Biological Psychology 42, 249–268 (1996)
12. Gevins, A., Smith, M.E., Leong, H., McEvoy, L., Whitfield, S., Du, R., Rush, G.: Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. Human Factors 40 (1998)

13. Gevins, A., Smith, M.E., McEvoy, L., Yu, D.: High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. Cerebral Cortex 7, 374 (1997)
14. Scerbo, M.W., Freeman, F.G., Mikulka, P.J.: A brain-based system for adaptive automation. Theoretical Issues in Ergonomics Science 4, 200–219 (2003)
15. Wilson, G.F., Russell, C.A.: Operator functional state classification using multiple psychophysiological features in an air traffic control task. Human Factors: The Journal of the Human Factors and Ergonomics Society 45, 381 (2003)
16. Wilson, G.F., Russell, C.A.: Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. Human Factors: The Journal of the Human Factors and Ergonomics Society 45, 635 (2003)
17. Fairclough, S.H., Venables, L., Tattersall, A.: The influence of task demand and learning on the psychophysiological response. International Journal of Psychophysiology 56, 171–184 (2005)
18. Gevins, A., Smith, M.E.: Neurophysiological measures of cognitive workload during human-computer interaction. Theoretical Issues in Ergonomics Science 4, 113–131 (2003)
19. Klimesch, W.: EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Research Reviews 29, 169–195 (1999)
20. Bunce, S.C., Izzetoglu, M., Izzetoglu, K., Onaral, B., Pourrezaei, K.: Functional near-infrared spectroscopy. Engineering in Medicine and Biology Magazine 25, 54–62 (2006)
21. Gopher, D., Kimchi, R.: Engineering psychology. Annual Review of Psychology 40, 431–455 (1989)
22. Liu, Y., Wickens, C.D.: Mental workload and cognitive task automaticity: an evaluation of subjective and time estimation metrics. Ergonomics 37, 1843–1854 (1994)
23. O'Donnell, R.D., Workload, F.T.E.: assessment methodology. In: Boff, K.R., T., L.K.A.J.P. (eds.) Handbook of Perception and Human Performance, vol. 2, pp. 42–49. Wiley Interscience, New York (1986)
24. Wierwille, W., Eggemeier, F.T.: Recommendations for mental workload measurement in a test and evaluation environment. Human Factors: The Journal of the Human Factors and Ergonomics Society 35, 263–281 (1993)
25. Schneider, W., Shiffrin, R.M.: Controlled and automatic human information processing: I. Detection, search, and attention. Psychological Review 84, 1 (1977)
26. Shiffrin, R.M., Schneider, W.: Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. Psychological Review 84, 127 (1977)
27. Kelly, A., Garavan, H.: Human functional neuroimaging of brain changes associated with practice. Cerebral Cortex 15, 1089 (2005)
28. Petersen, S.E., Van Mier, H., Fiez, J.A., Raichle, M.E.: The effects of practice on the functional anatomy of task performance. Proceedings of the National Academy of Sciences of the United States of America 95, 853 (1998)
29. Milton, J.G., Small, S.S., Solodkin, A.: On the road to automatic: dynamic aspects in the development of expertise. Journal of Clinical Neurophysiology 21, 134 (2004)
30. Beauchamp, M., Dagher, A., Aston, J., Doyon, J.: Dynamic functional changes associated with cognitive skill learning of an adapted version of the Tower of London task. Neuroimage 20, 1649–1660 (2003)
31. Poldrack, R.A.: Imaging brain plasticity: conceptual and methodological issues–a theoretical review. Neuroimage 12, 1–13 (2000)
32. Garavan, H., Kelley, D., Rosen, A., Rao, S.M., Stein, E.A.: Practice-related functional activation changes in a working memory task. Microscopy Research and Technique 51, 54–63 (2000)

33. Izzetoglu, K., Bunce, S., Onaral, B., Pourrezaei, K., Chance, B.: Functional optical brain imaging using near-infrared during cognitive tasks. International Journal of Human-Computer Interaction 17, 211–227 (2004)
34. St John, M., Kobus, D., Morrison, J.: A multi-tasking environment for manipulating and measuring neural correlates of cognitive workload, p. 7. IEEE, Los Alamitos (2002)
35. Parasuraman, R., Caggiano, D.: Neural and genetic assays of human mental workload. Quantifying human information processing (2005)
36. Warm, J.S., Parasuraman, R., Matthews, G.: Vigilance requires hard mental work and is stressful. Human Factors: The Journal of the Human Factors and Ergonomics Society 50, 433 (2008)
37. Parasuraman, R., Warm, J.S., See, J.W.: Brain systems of vigilance. In: Parasuraman, R. (ed.) The Attentive Brain, pp. 221–256. MIT Press, Cambridge (1998)
38. Ayaz, H., Willems, B., Bunce, B., Shewokis, P.A., Izzetoglu, K., Hah, S., Deshmukh, A., Onaral, B.: Cognitive Workload Assessment of Air Traffic Controllers Using Optical Brain Imaging Sensors. In: Marek, T., Karwowski, W., Rice, V. (eds.) Advances in Understanding Human Performance: Neuroergonomics, Human Factors Design, and Special Populations, pp. 21–31. CRC Press Taylor & Francis Group (2010)

Challenges and Solutions with Augmented Cognition Technologies: Precursor Issues to Successful Integration

Joseph Cohn

Human and Bioengineered Systems Division, Office of Naval;
Research, Arlington VA 22203
joseph.cohn@navy.mil

Abstract. Today's combat environment requires increasingly complex interactions between human operators and their systems. Whereas in the past, the roles of human and system were clearly delineated, with the integration of advanced technologies into the C4ISR toolkit, the distinct parsing of tasks has given way to paradigms in which the human operator's roles and responsibilities must dynamically change according to task and context. Yet, current methodologies for integrating the human into the system have not kept pace with this shift. An important consequence of this mismatch between human operator and system is that failures often lead to catastrophic and unrecoverable accidents (O'Connor & Cohn, 2010). In order to reintegrate the human element back into the system, new approaches for representing operator performance, in terms of their individual cognitive and behavioral capacities, limitations and changing needs are required.

Keywords: Neuroscience, Cognition, Autonomy, Human Systems, Information Processing, Adaptive, Cognitive Architecture.

1 Introduction

As early as the 1940s, researchers were concerned with representing the human element in human – systems in order to develop effective control schemes for allocating tasks between the two (Bates, 1947, Craik (1947/1948; 1948; Birmingham & Taylor, 1954). Initial solutions to this challenge involved automating task allocation, using predefined heuristics to parse tasks into those at which a machine excelled and those at which a human excelled (Licklider, 1960). Yet, over the course of even a basic task or mission, situations change, information changes and people's capabilities change. Early automation methods were unable to adapt to these changes, proving too brittle for the range of complex tasks humans and their systems were called upon to accomplish (Woods, 1996).

As modeling and simulation techniques became more pervasive, the focus of automation shifted towards creating more dynamic and adaptive techniques to allocate tasks in response to user performance (Scerbo, 1996). An important consideration with this approach, known as adaptive automation, is how the adaptation is triggered, which is tightly linked to how human performance is represented. Early attempts at creating

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 23–29, 2011.

adaptive techniques focused on linking expert systems – representing the ideal way of performing a task - with observation based representations of human performance, to guide task allocation (Rouse, 1977;). While representations of performance may be gleaned from observed behavior, they do not account well for an individual's mental and physiological states, their unique and evolving experiences and their distinct inclinations and preferences. As well, observed human performance typically evolves over a timescale measured in seconds or even minutes, while machine action – and external events - may occur over a millisecond timescale. Adaptive automation is therefore limited by how human performance is represented.

1.1 Augmented Cognition Control Schemes

One approach to overcoming the limitations of adaptive automation by using a control scheme that includes representations of human performance that are based on the neural action underlying behavior in addition to observations about the actual behavior. These augmented cognition control approaches are based on the idea that a more flexible, adaptive and individualized representation of human performance can be realized by including a neural component. In turn, these representations will lead to a control scheme that will bring the human and the machine into closer alignment so that they can work as a symbiotic team on the much wider range of dynamic tasks required of today's multi-tasking Warfighters (Cohn & Wheeler, 2010).

1.2 Why Now?

Emerging advances in neuroscience provide the basis for building these representations and for using them to establish and maintain robust human - system interactions. In recent years, the notion that the brain is composed of discrete centers, each responsible for a portion of the information processing chain (Wickens & Hollands, 2000), has given way to a more dynamic and integrative theory of brain in which different regions each contribute, in different ways depending on the task environment and user state, to the processing of information leading to behavior (Singer, 1999; Philiastides & Sajda, 2007). These regions form 'ad hoc' networks across the physical substrate of the brain via synchronization signals that briefly bind them together, leading to cognitive and motor behavior, like decision making, or physical action (Kahana, 2006; Osipova, et al, 2006; Palva, Palva, & Kaila, 2005; Singer, 1999). Guided by these promising advances in understanding the brain's operating principles, and harnessing and extending brain imaging technologies, advanced signal processing techniques and data modeling approaches to build neuroadaptive representations, it will be possible to firmly place the human into the loop in human - systems interactions..

2 Understanding the Challenges

There are two different ways of illustrating this challenge (Newell, 1994; Schmorrow, Cohn & Nicholson, 2009). First, this challenge can be represented in terms of the amount of information or the channels of information available to populate a model. Neural data clearly provides a significant amount of information, over a very rapid time course, leaving open the possibility of creating high fidelity, real time models of user performance that can intelligently drive adaptive automation. In contrast, behavior data provides comparatively little information, leading to the development of basic models of performance that are highly constrained. Second, this challenge can be represented in terms of the time required to capture the required information to populate a model. Neural activity takes place on the millisecond time scale, opening up the possibility for near-real time development and updating of models based on that type of information. In contrast, behavior takes place on the second or seconds time scale, even though the actual intent to act may have been formed within 100s of milliseconds of sensing relevant information (Luu, et al., 2010). Models based on behavior data simply cannot keep pace with the actions of the adaptive systems which they are meant to inform.

2.1 Breaking the Challenge Down

Recently, though, there has been a shift in theories of brain, made possible by advances in neuroimaging technologies and data analysis techniques. The notion that the brain is made of discrete centers, each responsible for a portion of the information processing chain, has given way to a more dynamic and integrative conception. Under this framework, different areas of the brain each contribute, in different ways depending on the task environment and user state, to ensemble brain processes leading ultimately to behavior (Singer, 1999; Philiastides & Sajda, 2007). Cognition, in this view, is the interaction and integration of 'building blocks' like perception, attention & memory, which, in turn results from activity across multiple brain regions.

These regions form 'ad hoc' networks across the physical substrate of the brain via synchronization signals that bind different neural centers together, for brief periods of time (Kahana, 2006; Osipova, et al.., 2006; Palva, Palva, & Kaila2005; Singer, 1999). These advances have been made possible as direct result of developments in three core domains:

- Detection Technologies
- Decoding Methodologies
- Modeling Frameworks

2.2 Detection

Access to the brain has been one of the key limiting steps in demonstrating coordinated activity across the brain as behavior develops. Technologies, like functional Magnetic Resonance Imaging (fMRI); (Logothetis, 2001), dense array Electroencephalography (dEEG); (Junghöfer, Elbert, Tucker, & Rockstroh, 2000), and others are at the point where they may now be applied to the challenge of capturing integrated neural action as it occurs simultaneously across multiple brain regions.

2.3 Decoding

Access to integrated neural data is necessary but not sufficient for representing performance. New methods for analyzing these multivariate data sets must also be established, refined and applied to data captured as users perform a range of cognitive and motor tasks. One promising emerging technique, multivariate decoding (Mitchell et al 2004), has the ability to take into account the full spatial pattern of brain activity, measured simultaneously across many regions, enabling the decoding and translation of measured brain activity. Other techniques (e.g. Kay et al, 2008) show similar promise.

2.4 Modeling

The final challenge is developing representation frameworks into which the detected and decoded information may be organized, to simulate and predict human performance as part of a neuroadaptive human - system control scheme. Possible solutions include extending current modeling architectures to handle neural data (Anderson, et al 2008) or developing software or silicon-based representations of the detected and decoded data that produce behaviors similar to those that result from actual neural activity (Fleisher & Krichmar, 2007).

3 Applications

Three areas of particular relevance to today's military could benefit from advances in augmented cognition control schemes– Information Processing, Discovery and Presentation; Intelligent and Autonomous Systems; and, Training, Education and Human Performance as each relies on representations of human performance.

3.1 Information Processing

Advanced information technologies require the rapid processing and analysis of massive amounts of data. Computers can process super-ordinate amounts of data in a short time but programming them to discover new patterns is extremely difficult. Humans are adept at discovering new patterns in large amounts of data, yet are limited by how quickly they process and analyze very large data sets (Hodgkinson, Langan-Fox, & Sadler-Smith, 2008; Volz and von Cramon, 2006). Currently, there is no simple way to combine the strengths of humans with the strengths of computers to enable better information analysis. Neuroadaptive representations could be developed that detect, decode and depict human pattern recognition, which could then be combined with adaptive data analysis techniques like genetic algorithms. The result would be a human - system that decreases information processing time, enables richer and deeper information discovery and tailors information presentation to individual users' needs

3.2 Intelligent and Autonomous Systems

Human team members, practicing together over time, develop 'mental models' of their team members' performance (Lim & Klein, 2006), enabling them to anticipate each others' responses and compensate as needed (Wilson, Salas, Priest, & Andrews, 2007). Using current approaches to representing human performance, intelligent and autonomous systems are not able to develop such anticipatory and predictive models. This creates an artificial barrier between humans and their systems, imposes additional personnel requirements, introduces significant performance lags in time critical scenarios, and reduces overall human – system effectiveness. Neuroadaptive representations would provide these systems with the basic abilities to interact with people using human-like representations, strategies, and knowledge.

3.3 Training and Education

Human behavior is the result of underlying neural processes. These processes may be modified over the long term, using *Education* to enduringly encode information into long term memory (Wickens & Hollands, 2000); they may be reinforced over a shorter term, using *Training* to enable the extension and application of this information to new situations (Anderson, 2000); or, they may be augmented in real time, using *Performance Enhancing* technologies to optimize their output in response to specific task requirements (Cohn and Forsythe, 2008). Technologies supporting these interventions form a human - system and rely heavily on adaptive and individualized representations of human performance (Gluck and Pew, 2005). Neuroadaptive representations will provide this flexibility, enabling significantly more effective training, education and human performance technologies.

4 Conclusion

The field of augmented cognition lies at the intersection of neuroscience and engineering. Advances in these areas continue to enable the development of a comprehensive theory, and associated technologies, that will ultimately allow humans and their machines to cooperatively achieve a shared set of goals. While past attempts at developing augmented cognition systems have not completely achieved the vision of closed-loop symbiosis, they have paved the way for increasingly more sophisticated theories and technologies that will enable the attainment of this vision. Most recently, advances on several fronts have significantly advanced the state of the art in human-machine interactions. It remains to be seen to what extent researchers will be able to integrate these advances in new and exciting ways to support human performance.

Acknowledgments. The author wishes to thank Ms. D. Brumer for her critical review and comments.

References

1. Anderson, J.R.: ACT: A simple theory of complex cognition. American Psychologist 51, 355–365 (1996)
2. Anderson, J.R., Carter, C.S., Fincham, J.M., Qin, Y., Ravizza, S.M., Rosenberg-Lee, M.: Using fMRI to Test Models of Complex Cognition. Cognitive Science 32, 1323–1348 (2008)
3. Bates, J.A.V.: Some characteristics of a human operator. Journal of the Institute of Electrical Engineering 94, 298–304 (1947)
4. Birmingham, H.P., Taylor, F.V.: A design philosophy for man-machine control systems. Proceedings of th e I.R.E. 42(12), 1748–1758 (1954)

5. Cohn, J.V., Wheeler, T.: Neuroadaptive systems: challenges and opportunities with creating symbiotic relationships between humans and the machines they use. In: Fafrowicz, M., Marek, T., Karwowski, W., Schmorrow, D. (eds.) Neuroadaptive Systems: Research, Theory, and Applications, CRC Press, Boca Raton (2010)
6. Cohn, J.V., Forsythe, C.J.: The Effective Use of Performance Enhancing Technologies: Mechanisms, Applications and Policies. Technology 11, 107–126 (2008); Craik, K.J.W.: Theory of the human operator in control systems I: The operation of the human operator in control systems. British Journal of Psychology 38, 56–61 (1947/1948)
7. Craik, K.J.W.: Theory of the human operator in control systems II: Man as an element in a control system. British Journal of Psychology 38, 142–148 (1948)
8. Fleischer, J.G., Krichmar, J.L.: Sensory integration and remapping in a medial temporal lobe model during maze navigation by a brain-based device. Journal of Integrative Neuroscience 6(3), 403–431 (2007)
9. Hammer, J.M., Small, R.L.: An intelligent interface in an associate system. In: Rouse, W.B. (ed.) Human/Technology Interaction in Complex Systems, vol. 7, pp. 1–44. JAI Press, Greenwich (1995)
10. Hodgkinson, G.P., Langan-Fox, J., Sadler-Smith, E.: Intuition: A fundamental bridging construct in the behavioral sciences. British Journal of Psychology 99(1), 1–27 (2008)
11. Junghöfer, M., Elbert, T., Tucker, D.M., Rockstroh, B.: Statistical control of artifacts in dense array EEG/MEG studies. Psychophysiology 37(4), 523–532 (2000)
12. Kahana, M.J.: The cognitive correlates of human brain oscillations. The Journal of Neuroscience 26(6), 1669–1672 (2006)
13. Kay, K.N., Naselaris, T., Prenger, R.J., Gallant, J.L.: Identifying natural images from human brain activity. Nature 452, 352–355 (2008)
14. Kelso, J.A.S.: Dynamic patterns: The self-organization of brain and behavior. MIT Press, Cambridge (1995)
15. Lim, B.C., Klein, K.J.: Team mental models and team performance: A field study of the effects of team mental model similarity and accuracy. Journal of Organizational Behaviour 27, 403–418 (2006)
16. Licklider, J.C.R.: Man-computer symbiosis. IEEE Transactions on Human Factors in Electronics HFE-1, 4–11 (1960)
17. Logothetis, N.K.: Neurophysiological investigation of the basis of the fMRI signal. Nature 412, 150 (2001)
18. Luu, P., Geyer, A., Wheeler, T., Campbell, G., Tucker, D., Cohn, J.: The Neural Dynamics and Temporal Course of Intuitive Decisions. Public Library Of Science (2010)
19. Mitchell, T., Hutchinson, R., Niculescu, R.S., Pereira, F., Wang, X., Just, M.A., Newman, S.D.: Learning to decode cognitive states from brain images. Machine Learning 57, 145–175 (2004)
20. Mitchell, T.M., Shinkareva, S.V., Carlson, A., Chang, K.-M., Malave, V.L., Mason, R.A., Just, M.A.: Predicting human brain activity associated with the meanings of nouns. Science 320, 1191–1195 (2008)
21. O'Connor, P.E., Cohn, J.V. (eds.): Human Performance Enhancement in High-Risk Environments. Praeger Security International, Santa Barbara (2010)
22. Osipova, D., Takashima, A., Oostenveld, R., Fernandez, G., Maris, E., Jensen, O.: Theta and gamma oscillations predict encoding and retrieval of declarative memory. The Journal of Neuroscience 26(28), 7523–7531 (2006)
23. Palva, J.M., Palva, S., Kaila, K.: Phase Synchrony among Neuronal Oscillations in the Human Cortex. Journal of Neuroscience 25(15), 3962–3972 (2005)

24. Philiastides, M.G., Sajda, P.: EEG-informed fMRI reveals spatiotemporal characteristics of perceptual decision making. Journal of Neuroscience 27(48), 13082–13091 (2007)
25. Rensink, R.A.: Visual sensing without seeing. Psychological Science 15, 27–32 (2004)
26. Rouse, W.B.: Human-computer interaction in multitask situations. IEEE Transactions Systems, Man, and Cybernetics SMC-7, 293–300 (1977)
27. Rouse, W.B.: Human-computer interaction in multitask situations. IEEE Transactions Systems, Man, and Cybernetics SMC-7, 293–300 (1977)
28. Scerbo, M.W.: Theoretical perspectives on adaptive automation. In: Parasuraman, R., Mouloua, M. (eds.) Automation and Human Performance: Theory and Applications, pp. 37–63. Lawrence Erlbaum Associates, Mahwah (1996)
29. Shinkareva, S.V., Mason, R.A., Malave, V.L., Wang, W., Mitchell, T.M., Just, M.A.: Using fMRI brain activation to identify cognitive states associated with perception of tools and dwellings. PLoS ONE 3, e1394 (2008)
30. Singer, W.: Neuronal synchrony: a versatile code for the definition of relations? Neuron 24, 49–65 (1999)
31. Volz, K.G., von Cramon, D.Y.: What Neuroscience Can Tell about Intuitive Processes in the Context of Perceptual Discovery. Journal of Cognitive Neuroscience 18(12), 2077–2087 (2006)
32. Wickens, C.D., Hollands, J.G.: Engineering Psychology and Human Performance, 3rd edn. Prentice Hall, Upper Saddle River (2000)
33. Wilson, K.A., Salas, E., Priest, H.A., Andrews, D.: Errors in the heat of battle: Taking a closer look at shared cognition breakdowns through teamwork. Human Factors 49, 243–256 (2007)
34. Woods, D.D.: Decomposing automation: Apparent simplicity, real complexity. In: Parasuraman, R., Mouloua, M. (eds.) Automation and Human Performance: Theory and Applications, pp. 3–18. Lawrence Erlbaum Associates, Mahwah (1996)

Augmenting Brain and Cognition by Aerobic Exercise

Kirk I. Erickson

Department of Psychology
Center for the Neural Basis of Cognition
University of Pittsburgh, USA
kiericks@pitt.edu

Abstract. Cognitive function declines in late adulthood and this is preceded by atrophy of the prefrontal cortex, hippocampal formation, and parietal cortex. Despite significant loss of brain tissue in late adulthood, decline is not ubiquitous across all older adults. In fact, some adults age quite successfully with minimal decline. This suggests that brain deterioration might not be an inevitable consequence of aging. In fact, mounting evidence suggests that participation in regular aerobic exercise is effective at enhancing cognitive and brain health in late adulthood. In this paper we discuss the evidence that cardiorespiratory fitness and aerobic exercise augments cognition by increasing gray matter volume in prefrontal and hippocampal brain regions.

Keywords: Aging, brain, atrophy, exercise.

1 Introduction

The United States Census Bureau expects that the percentage of people over the age of 65 will increase from approximately 11% of the population in 2000 to nearly 23% of the population in 2050 [1]. With an increase in the percentage of the population over the age of 65 comes an increase in the expected prevalence of age-related diseases. For example, a recent report from the Alzheimer's Association suggests that the number of persons with Alzheimer's Disease will increase from approximately 5.1 million in 2010 to nearly 13.5 million by 2050 [2]. Such an increase in the prevalence of Alzheimer's Disease is paralleled by elevated costs to treat and care for people with the disease. For example, it is estimated that health care costs associated with Alzheimer's Disease will be 172 billion dollars in 2010 and will increase to nearly 1.078 trillion dollars by 2050.

Not everyone will develop Alzheimer's disease. Yet, even in those people who do not develop Alzheimer's Disease, there is still evidence of cognitive decline. For example, longitudinal studies have found that there is relative stability or even growth in cognitive function, across several different cognitive domains until the age of about 55 or 60. After the age of 60 there is gradual decline in most measures of cognitive function including inductive reasoning, perceptual speed, spatial orientation, episodic memory and verbal memory [3]. Therefore, even in adults not experiencing Alzheimer's disease, cognitive decline is still prevalent.

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 30–38, 2011.

Preceding, and leading to, cognitive dysfunction in late life is a decline in the volume of cortical and subcortical brain tissue. For example, in one cross-sectional study in two-hundred adults between 18 and 80 years of age, it was found that there is a linear decline in the volume of gray matter in the dorsolateral prefrontal cortex starting as early as about 30 years of age [4]. Similarly, the hippocampus, a small structure located in the medial temporal lobe, also deteriorates in late adulthood, but instead of a liner decline like the dorsolateral prefrontal cortex, the hippocampus shows a non-linear rate of decline. That is, both cross-sectional and longitudinal studies of the hippocampus have found that there is rather little decline in the volume of this structure before the age of 55. After the age of 55 there is a steady decline in volume at about 1-2% per year in individuals without dementia [4] and about 3-5% per year in individuals with dementia [5].

Despite the average decline in volume of the hippocampus and dorsolateral prefrontal cortex in late life, there is a significant amount of individual variability, with some people showing more rapid decay and others showing minimal decay. This variability begs the question: what factors contribute to the individual variability in the rate and extent of brain atrophy? And if we can identify factors that explain this variation could we develop interventions that protect against brain decay or even reverse atrophy that is already manifest. As will be described below, there is now convincing evidence that a modest amount of aerobic exercise not only explains individual variability in the rate of brain deterioration but is also effective at augmenting brain health when an exercise regimen is initiated.

2 Why Aerobic Exercise?

When we think of methods to exercise our brains, we generally think of intellectual activities such as crossword puzzles, Sudoku, or reading. However, it turns out that when we work our muscles we also work our brains. In fact, we should no longer think of aerobic exercise as simply affecting our bodies from the neck down. The early seminal research on how exercise influences the brain was discovered by animal studies with rats and mice in which the intensity and duration of exercise could be easily monitored and manipulated. From these studies, it has been found that exercising increases the number of new neurons produced in the dentate gyrus of the hippocampus, even in aged animals. Although the rate of cell proliferation resulting from exercise in aged animals is less than the rate in young animals [6], the finding that older animals are still capable of neurogenesis and that aerobic exercise can take advantage of this plasticity is quite promising. The possibility of neurogenesis in old animals suggests that atrophy may not be inevitable and might even be reversible.

With the growth of new neurons comes an increased need for nutrients. Increased nutrients are supplied to the brain by increasing the vascularization of brain tissue. Exercise has been found to increase blood flow and vascularization in rodents. Angiogenesis, or the proliferation of new vasculature, has been found in several brain regions including the cerebellum, hippocampus, motor cortex, frontal cortex, and basal ganglia [6, 7].

Exercise has also been shown to increase synaptic connections between neurons and to enhance learning and memory [8]. For example, in a hippocampal-dependent maze task, rodents that had voluntary access to a running wheel demonstrated faster learning rates and enhanced retention compared to their non-exercising counterparts [9]. Both cell proliferation and learning and memory are thought to be dependent on similar cellular cascades. One molecule that is secreted by neurons, is considered to be involved in neurogenesis, and is critical in cellular analogs of learning and memory is brain-derived neurotrophic factor (BDNF). Levels of BDNF generally decline in aging and Alzheimer's disease, but exercise increases the production and secretion of this molecule [9]. Furthermore, blocking the binding of BDNF to its receptor essentially eliminates the exercise-induced enhancement of cognitive function [10]. This is strong evidence to suggest that BDNF plays an important mediating role in determining how aerobic exercise improves brain health.

In sum, studies in rodents have revealed the underlying molecular and cellular mechanisms by which exercise exerts its effects on the brain. These findings provide a low-level biological justification for examining the effects of exercise on brain integrity in humans.

3 Aerobic Exercise, Cognition and Brain Morphology in Humans

Research in the 1970's found that older adult athletes outperformed their more sedentary peers on several different cognitive and motor tasks [11]. Indeed, many cross-sectional studies have now found similar associations between higher aerobic fitness levels and better cognitive function in late adulthood [7]. Cross-sectional studies, although informative about associations between exercise and cognition, are inherently limited in determining causal relationships. In order to determine if aerobic exercise is effective at improving cognitive and brain function, randomized clinical trials are needed in which participants are randomly assigned to either receive exercise or a control condition. Importantly, exercise interventions in which older adults are randomly assigned to receive monitored and structured exercise for a period of 3-6 months have found that starting an exercise regimen can enhance cognitive function. In fact, a recent meta-analysis of 18 different randomized exercise interventions found that exercise improves cognition in both a general and specific fashion [12]. That is, the effects of aerobic exercise are general in the sense that nearly all cognitive domains are enhanced with exercise, but specific in the sense that executive functions are enhanced more than other cognitive domains. The term executive function is an umbrella term that broadly refers to several different higher-level cognitive functions such as selective attention, task-coordination, planning, sequencing, and maintaining items in working memory. Executive functions are largely supported by prefrontal and parietal brain circuits and are often found to show the most significant deficits in late adulthood compared to other cognitive domains. The fact that exercise appears to have its greatest effect on executive functions suggests that although executive functions show the greatest decline with advancing age, they remain tractable.

The results from the meta-analysis of exercise interventions [12] suggested that the brain regions supporting executive function, such as the prefrontal and parietal brain regions, would be the ones most affected by exercise. To test this prediction, Colcombe and colleagues [13] randomly assigned a group of older sedentary adults to either a moderate intensity activity group that walked for about 40 minutes three days per week or to a non-aerobic stretching and toning control group that came into the lab for the same amount of time as the exercising group. Both groups participated for a period of six months. High-resolution brain scans using magnetic resonance imaging (MRI) were obtained both before and after the intervention. Using a voxel-based morphometry technique to examine brain volume on a point-by-point basis throughout gray matter and white matter tissues, Colcombe and colleagues reported that exercise was effective at increasing gray matter volume in the prefrontal, parietal, and lateral temporal regions and at increasing white matter volume in the genu of the corpus callosum. This study was important as it suggested for the first time that brain tissue of older adults remains modifiable and that only six months of exercise is sufficient for taking advantage of the brains natural capacity for plasticity.

As described earlier, the hippocampus has been a region of great focus in aging research because of its role in memory formation, and specifically in the formation of declarative memory. The hippocampus is also important because it shows considerable atrophy in late adulthood and leads to Alzheimer's disease and memory loss. Research in rodents, however, have unequivocally found that exercise can influence the morphology and function of the hippocampus and that BDNF is highly concentrated in the hippocampus and increases with bouts of exercise [8-10]. This evidence leads directly to the speculation that higher fitness levels may be associated with less hippocampal atrophy and spared memory function. To test this hypothesis, Erickson and colleagues examined cardiorespiratory fitness levels in a sample of 165 older adults without dementia and used MRI techniques to identify the volume of the hippocampus [14]. They found that after controlling for potentially confounding factors like age, sex, and education, older adults that were more aerobically fit had larger hippocampal volumes than their less fit peers. In addition, a spatial memory task was used to test memory function in this study. It was found that higher fit older adults performed better on the task, and greater hippocampal volume partially mediated the fitness-cognition association. These results directly linked for the first time, cardiorespiratory fitness, age-related hippocampal atrophy, and memory function. There have now been several other studies in both older adults and children that have replicated this effect showing that higher fitness levels are associated with greater hippocampal volumes [15-16].

The effects of cross-sectional studies demonstrating associations between cardiorespiratory fitness levels and hippocampal volumes are provocative, but fail to demonstrate direct causal links between increasing exercise and hippocampal volume. In order to determine causality an intervention must be conducted. To address this concern, Erickson and colleagues [17] conducted a randomized one-year intervention in which 120 sedentary older adults without dementia were assigned to either a moderate intensity exercise intervention or to a stretching and toning control group. Similar to previous interventions, both groups received the same amount of social interaction and health instruction from trained health professionals. The main difference between the groups was that the walking group participated in aerobic exercises for one year while the control group participated in non-aerobic activity for one year. Using MRI technology again to examine the volume of the hippocampus and a spatial memory task to measure cognition, Erickson and colleagues [17] reported that one year of exercise was sufficient for increasing the size of the hippocampus. Furthermore, they reported that increased hippocampal volume was correlated with improvements in spatial memory function suggesting a direct link between hippocampal size and behavioral outcomes. Finally, they also reported that although the exercise intervention did not increase circulating levels of BDNF, the change in BDNF from baseline to post-intervention was associated with increases in hippocampal volume. These findings help to support the claim that modest amounts of exercise can increase the size of brain regions that normally undergo deterioration as we age and may help to prevent memory loss. Furthermore, these results claim that the brain remains modifiable well into late adulthood and that starting an exercise regimen in late adulthood is not futile; even those adults that have been sedentary can still benefit from starting to exercise.

The study by Erickson and colleagues [17] help to support the claim that exercise can augment brain and cognition in late adulthood. However, there are several important questions that remain unanswered from this. For example, how much exercise is necessary to observe its effects on brain and cognition? Epidemiological studies suggest that more strenuous activities are associated with a reduced risk of cognitive impairment [18] and meta-analyses of interventions suggest that about three to six months is sufficient for observing improvements in cognition. However, these studies are not dose-response studies in which the duration and intensity of exercise is manipulated. In fact, dose-response studies for the effects of exercise on cognitive and brain function have not yet been conducted.

To help address the question of the dosage of physical activity and the retention of the benefits of physical activity on brain morphology, Erickson and colleagues [19] conducted a thirteen-year longitudinal study of 299 adults over the age of 65. In this study, physical activity was assessed at baseline by asking participants how many blocks they walked on average over a one-week period. Nine years after this assessment, high-resolution brain images were collected and VBM was used to examine whether physical activity measured nine-years earlier was predictive of brain volume later in life. Consistent with the predictions, Erickson and colleagues [19] found that a greater amount of physical activity was associated with greater gray matter volume in prefrontal, hippocampal, and occipital regions. However, they also found that this occurred in a dose-dependent fashion. That is, sparing of gray matter volume with physical activity was only apparent in those individuals reporting more than 72 blocks of walking per week, or roughly one mile of walking per day. Those walking less than one mile per day showed less brain volume than their more active peers. This finding suggested that not only might there be a lower-bound threshold for the amount of activity needed to observe the benefits of exercise on brain morphology, but also that the effects of physical activity might be long-lasting. Furthermore, a four-year follow-up after the MRI assessment found that those individuals with greater gray matter volume in the inferior frontal gyrus, hippocampus, and supplementary motor area had a two-fold reduced risk of developing cognitive impairment.

In sum, the research described in this section now convincingly demonstrates that aerobic exercise is effective at augmenting brain and cognitive health in late adulthood and that even modest amounts of exercise is sufficient for increasing brain size and improving memory. At a time of life when memory impairment is a salient fear and brain atrophy is progressing at a faster rate, aerobic exercise could be an important low-cost and low-tech prevention and treatment that is accessible to most people. Although exercise will not be a magic bullet cure for Alzheimer's disease, even if it delays the onset or reduces the risk for developing cognitive impairment, it may save millions of dollars in health care costs and reduce the emotional toll on caregivers and those afflicted with impairment.

4 Aerobic Exercise and Brain Function

The research described above has focused on the role of aerobic exercise and fitness in relation to brain morphology. However, other neuroimaging techniques, such as functional MRI (fMRI) have also provided some important insight regarding how the functioning of the brain is affected by cardiorespiratory fitness and aerobic exercise. In one seminal study by Colcombe and colleagues [20], older adults were randomized either to an exercise walking group or to a stretching and toning control group similar to those interventions described above. In the first part of the study they used fMRI to examine brain function during a selective attention task in a cross-sectional sample of older adults. They found that more highly fit older adults had greater brain activation in prefrontal and parietal brain areas and less activity in the anterior cingulate cortex. These activation differences were accompanied by elevated performance on the selective attention task. This cross-sectional investigation was followed by results from a randomized intervention that demonstrated that these same regions also showed increased brain activity after the intervention. That is, the exercise intervention resulted in elevated performance on the task and increased brain activity in prefrontal and parietal areas. This study was important because it demonstrated that the effect of exercise extends beyond brain morphology by influencing the function of the supporting brain circuitry. These effects have recently been replicated in several other fMRI studies [21-23].

One way in which the brain networks are improved with exercise might be by an augmentation of the connectivity between brain regions. Enhanced brain connectivity resulting from exercise might help to explain how improvements in cognition are elicited. To test this hypothesis, Voss and colleagues [24] first demonstrated in a cross-sectional sample of older adults that individuals who were more aerobically fit had greater resting state connectivity than those adults who were less aerobically fit. Furthermore, enhanced functional connectivity between frontal and hippocampal nodes reliably mediating the fitness-cognition association suggesting that enhanced connectivity plays an important causal role in the augmentation of cognitive function. This cross-sectional study was followed-up by a randomized trial in which older adults were assigned to either a walking exercise group or to a stretching and toning control group and resting state fMRI was collected both before and after the one-year intervention. In this study, Voss and colleagues [25] found that exercise increased the connectivity between fronto-hippocampal regions and that this was paralleled by improvements in memory function. These results are important because they highlight the need to look beyond just the simple descriptive patterns of brain activity towards a more unified conceptualization of how the brain systems and communication network is being influenced by exercise.

In sum, aerobic exercise influences brain and cognitive function and evidence from fMRI demonstrate that enhancements in cognition are associated with enhancements in brain function in specific regions of cortex. Further, improvements in the communication and connectivity between regions probably underlies several of the enhancements commonly observed in studies that find better cognitive function in more fit individuals.

5 Conclusion

We have outlined the evidence in favor of the argument that modest amounts of aerobic exercise are sufficient for enhancing cognition and brain function. By using MRI technology, several studies have found that higher fit older adults have greater amounts of gray matter volume in several regions including the prefrontal cortex and hippocampus – areas of the brain that often show the most consistent patterns of deterioration and atrophy in late adulthood. Furthermore, exercise also increases brain activity in these same regions, which appears to parallel improvements in cognition. Overall, this evidence suggests that aerobic exercise can be envisioned as an effective method to prevent brain deterioration, maintain cognitive and brain function, and reverse atrophy that is already present. Despite these consistent and convincing findings, there remain many unanswered and unexplored questions.

One remaining set of questions involves the dose-response of exercise on brain and cognition. That is, as described above, we have relatively little information that tells us how much exercise is necessary, what intensity should be achieved, and what types of exercises are best to enhance cognition. The answers to these questions are critically important if aerobic exercise is to be used in clinical contexts and be prescribed to patients as a prevention or treatment for loss in cognitive function.

We also have very little information about the underlying mechanisms of aerobic exercise in humans. Is exercise working by predominantly influencing the creation of new vasculature? In fact, both morphological and fMRI results could be interpreted in relation to new vasculature resulting from exercise. Some studies have suggested that the effects of exercise extend beyond vascularization of brain tissue and that exercise works directly on the brain. In short, more research is needed to understand how aerobic exercise is exerting its effects on brain and cognition.

The results from research described in this review have been largely limited to older adult populations that are not experiencing signs of dementia. We do not yet know the extent to which aerobic exercise could prevent decay of brain tissue in those already experiencing cognitive impairment nor do we fully understand whether similar exercise interventions could influence cognition in other populations. In short, more research is needed to understand the generalizability of these effects.

Finally, not everyone benefits equally from exercise. Some people show very little benefits from the intervention while others show significant cognitive enhancement. What are the factors that contribute to this individual variability? Could there be genetic factors that moderate the extent to which any single person would benefit from exercise? Are there other factors such as intellectual stimulation or dietary habits that either accentuate or attenuate the effects of exercise.

In conclusion, we can argue that (1) the brains of older adults remain modifiable and that exercise can take advantage of this plasticity to prevent or even reverse brain decay, (2) it is never to late to start exercising, even adults that have been sedentary most of their lives can still reap the benefits of an exercise regimen, (3) the effects of exercise are not global throughout the entire brain, but have some specificity to hippocampal and prefrontal brain areas. Overall, this research suggests that brain atrophy and cognitive decline might not be as inevitable a consequence of aging as previously thought.

References

1. US Census Bureau, `http://www.census.gov/`
2. : Alzheimer's disease facts and figures: Alzheimers Dement., vol. 6, pp. 158–194 (2010)
3. Hertzog, C., Kramer, A.F., Wilson, R.S., Lindenberger, U.: Enrichment effects on adult cognitive development: can the functional capacity of older adults be preserved and enhanced? Psychological Science in the Public Interest 9, 1–65 (2009)
4. Kennedy, K.M., Erickson, K.I., Rodrigue, K.M., Voss, M.W., Colcombe, S.J., Kramer, A.F., et al.: Age-related differences in regional brain volumes: A comparison of optimized voxel-based morphometry to manual volumetry. Neurobiology of Aging 30, 1657–1676 (2009)
5. Jack Jr., C.R., Peterson, R.C., Xu, Y., et al.: Rate of medial temporal lobe atrophy in typical aging and Alzheimer's disease. Neurology 51, 993–999 (1998)
6. Kronenberg, G., Bick-Sander, A., Bunk, E., Wolf, C., Ehninger, D., Kempermann, G.: Physical exercise prevents age-related decline in precursor cell activity in the mouse dentate gyrus. Neurobiol Aging 27, 1505–1513 (2006)
7. Kramer, A.F., Erickson, K.I., Colcombe, S.J.: Exercise, cognition and the aging brain. Journal of Applied Physiology 101(4), 1237–1242 (2006)
8. Cotman, C.W., Berchtold, N.C.: Exercise: a behavioral intervention to enhance brain health and plasticity. Trends Neurosci. 25, 295–301 (2002)
9. van Praag, H., Shubert, T., Zhao, C., Gage, F.H.: Exercise enhances learning and hippocampal neurogenesis in aged mice. J. Neurosci. 25, 8680–8685 (2005)
10. Vaynman, S., Ying, Z., Gomez-Pinilla, F.: Hippocampal BDNF mediates the efficacy of exercise on synaptic plasticity and cognition. Eur. J. Neurosci. 20, 2580–2590 (2004)
11. Spirduso, W.W.: Reaction and movement time as a function of age and physical activity level. Journal of Gerontology 30, 435–440 (1975)
12. Colcombe, S.J., Kramer, A.F.: Fitness effects on the cognitive function of older adults: a meta-analytic study. Psychological Science 14, 125–130 (2003)
13. Colcombe, S.J., Erickson, K.I., Scalf, P.E., Kim, J.S., Wadhwa, R., McAuley, E., Kramer, A.F.: Aerobic exercise training increases brain volume in aging humans: evidence from a randomized clinical trial. Journal of Gerontology: Biological and Medical Sciences 61, 1166–1170 (2006)
14. Erickson, K.I., Prakash, R.S., Voss, M.W., Chaddock, L., Hu, L., Morris, K.S., et al.: Aerobic fitness is associated with hippocampal volume in elderly humans. Hippocampus 19, 1030–1039 (2009)
15. Chaddock, L., Erickson, K.I., Prakash, R.S., Kim, J.S., Voss, M.W., VanPatter, M., et al.: A neuroimaging investigation of the association between aerobic fitness, hippocampal volume and memory performance in preadolescent children. Brain Research 1358, 172–183 (2010)

16. Honea, R.A., Thomas, G.P., Harsha, A., Anderson, H.S., Donnelly, J.E., Brooks, W.M., Burns, J.M.: Cardiorespiratory fitness and preserved medial temporal lobe volume in Alzheimer's disease. Alzheimer Dis. Assoc. Disord. 23, 188–197 (2009)
17. Erickson, K.I., Voss, M.W., Prakash, R.S., Basak, C., Szabo, A., Chaddock, L., et al.: Exercise training increases size of hippocampus and improves memory. Proceedings of the National Academy of Sciences (in press)
18. Kramer, A.F., Erickson, K.I.: Capitalizing on cortical plasticity: influence of physical activity on cognition and brain function. Trends in Cognitive Sciences 11, 342–348 (2007)
19. Erickson, K.I., Raji, C.A., Lopez, O.L., Becker, J.T., Rosano, C., Newman, A.B., et al.: Physical activity predicts gray matter volume in late adulthood: The cardiovascular health study. Neurology 75, 1415–1422 (2010)
20. Colcombe, S.J., Kramer, A.F., Erickson, K.I., Scalf, P., McAuley, E., Cohen, N.J., et al.: Cardiovascular fitness, cortical plasticity, and aging. Proceedings of the National Academy of Sciences of the United States of America 101, 3316–3321 (2004)
21. Rosano, C., Venkatraman, V.K., Guralnik, J., Newman, A.B., Glynn, N.W., Launer, L., et al.: Psychomotor speed and functional brain MRI 2 years after completing a physical activity treatment. The Journals of Gerontology, Series A: Biological Sciences and Medical Sciences 65, 639–647 (2010)
22. Smith, J.C., Nielson, K.A., Woodard, J.L., Seidenberg, M., Durgerian, S., Antuono, P., et al.: Interactive effects of physical activity and APOE-e4 on BOLD semantic memory activation in healthy elders. Neuroimage 54, 635–644 (2011)
23. Prakash, R.S., Voss, M.W., Erickson, K.I., Lewis, J.M., Chaddock, L., Malkowski, E., et al.: Capitalizing on PASA: Cardiorespiratory fitness predicts neural flexibility of anterior processing regions in older adults. Frontiers in Human Neuroscience (2011)
24. Voss, M.W., Erickson, K.I., Prakash, R.S., Chaddock, L., Malkowski, E., Alves, H., et al.: Functional connectivity: a source of variance in the association between cardiorespiratory fitness and cognition? Neuropsychologia 48, 1394–1406 (2010)
25. Voss, M.W., Prakash, R.S., Erickson, K.I., Basak, C., Chaddock, L., Kim, J.S., et al.: Plasticity of brain networks in a randomized intervention trial of exercise training in older adults. Frontiers in Aging Neuroscience 2, 1–17 (2010)

Neurological Advances and Ethical/Legal Conundrums: Lessons from History

Cheryl Erwin

Assistant Professor and Director, Medical Humanities Certificate Program
University of Texas Medical School at Houston
6431 Fannin St., JJL 410, Houston, Texas 77030, USA
Cheryl.erwin@uth.tmc.ed, Cheryl.erwin@ttu.edu

Abstract. The scientific advances in the neurosciences are exciting and promise to advance our understanding of the human mind. The ethical and legal issues raised by neuroscience are distinctive but they are not unique to the twenty-first century. The ethical issues raised by these technologies deserve attention even while the science is in development. History teaches us to reflect on our humanity using insights from many disciplines and many times.

Keywords: Neuroethics, neurolaw, neuroprivacy, neuropolicy, research ethics, regulation of emerging technologies.

1 Introduction

The neurosciences are quickly outstripping our ability to fully assign human meaning to the very subject of our inquiry. What does it mean to say that we are mapping the mind? What does the map direct to our attention? How do we understand the meaning of what a mind is, or what it is capable of becoming? When we then image the mind, do we really understand fully how these images may be utilized for good or ill in society? How might we promote the good and dissuade the ill uses of our knowledge?

While the questions I pose are intriguing, they are not unique to the field of twenty first century cognitive neuroscience. A review of the issues we have confronted with prior neurotechnologies brings out important features that we may wish to consider and continue in conversation with one another. As an academic who works at the intersection of law-science-medicine, I am often asked why my law students see the world so differently from my medical students, and again differently from the science students in my ethics courses. While there are no simple answers, there are answers that are simply wrong. These wrong, and simple, answers often involve misattribution of divisions between “us” and “them”. It is more complex, but more accurate to see the enterprise of science and society as one that necessarily involves all of us: inclusive, disparate, but interconnected and interdependent at the same time.

1.1 Historical Trends in Scientific Ethics and Law

It is often possible to see from a distance what seems confusing at close range. The distance of time allows history to offer such a perspective.

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 39–45, 2011.

The history of scientific research is filled with examples of scientific enthusiasm outstripping moral reflection. In 1966, Henry Beecher wrote a classic article detailing the ethical lapses in a survey of the experiments then underway at Harvard University's teaching hospital [1]. Beecher documented twenty-two cases of unethical research which went unnoticed prior to his report. Seven years later the world discovered that the United States government had sponsored a study of syphilis in black men in Macon County Alabama, and had actively denied these men medical care [2]. Public outrage followed, and included calls for ethical conduct of human subjects research as well as a civil rights lawsuit which was settled for ten million dollars. In quick succession the field of bioethics was born and raised questions of the propriety of research techniques, clinical uses of new technologies, and respect for human self-determination. The law followed with regulations governing research with human subjects [3], privacy regulations, and best practices guidelines.

Neuroethics and neurolaw, as topic areas of intellectual inquiry, focus on the relationship of established principles of bioethics and legal doctrine to emerging cognitive science. Much more than previous encounters between the research community and law, the fields of neuroscience, ethics, and law are in dynamic tension and shaping one another as we progress in our discoveries. Neuroscience is challenging many of our legal traditions by exposing fallacies in our legal thinking and provoking new questions of how best to incorporate understandings of the mind into legal doctrine [4]. Neuroethics is in turn challenging both law and neuroscience to rethink notions of justice, access, individual dignity, privacy, and authenticity. It is my goal to raise these issues, to show how they are interconnected, and to build on the lessons that history offers in finding ways forward into a future that respects our shared humanity as much as our shared knowledge.

2 Neuroscience and Neuro-Knowledge

Neuroscience began over 100 years ago with scientists who studied the most remarkable human organ – the brain. Franz Josef Gall, an anatomist and physiologist in Austria, observed in the nineteenth century that students who had prominent foreheads also had good memories. Gall set out his hypotheses that "it is possible to ascertain different dispositions and inclinations by the elevations and depressions upon the head" [5]. Gall was excommunicated from the Church and forbidden from lecturing in Austria in 1802 [6]. Although first condemned on moral and theological grounds, Gall's theories soon found practical applications in business, criminology, and educational contexts [7]. Judges gave weight to phrenology in decisions about the size of a brain and the capacity to make a will [8] and whether the fact that a man was "remarkably ugly" should be factored into decisions about whether he was guilty of murder [9]. The rise of phrenology occurred in a scientific world before randomized double-blind controlled trials and the inductive method. Its impact on society can be seen as arising out of that less scientific time, but the law is called upon to make use of the best science available. Today phrenology is called junk science, pseudoscience, and non-admissible [10]. During its heyday, however, it was used to exclude individuals from rights otherwise guaranteed under the law.

Functional magnetic resonance imaging (fMRI) is now in its second decade. fMRI localizes changes in blood oxygenation when an individual performs a mental task [11]. Far from the pseudoscience of phrenology, fMRI is utilized to identify prodromal Huntington's Disease, and has numerous uses in established scientific investigation. It is also being used to correlate brain structure to dozens of physical and mental conditions, behaviors, characteristics and predispositions. These include major depression, schizophrenia, bipolar disorder, ADHD, social and racial evaluation, social cooperation, altruism, sexual arousal, ethical decision making, pedophilia, intelligence, humanity, empathy (and its lack), trust, humor and even the difference between the way men's and women's brains process information [6]. Ordinary individuals who lack specialized training in neuroscience interpret brain images as "proof" of the existence or non-existence of these traits [cite]. Businesses have found clever ways to market fMRI to screen potential employees for desirable mental traits [12].

Brain mapping has the potential to prospectively identify biological functions amenable to treatment, new drugs to treat disease, and dozens of other uses [13]. Contemporary neuroscience offers new tools for understanding human thought and will be incorporated into society. As Michael Gazzinaga noted almost twenty years ago: "The modular organization of the human brain is now fairly well accepted. The functioning modules do have some physical instantiation" [14]. These important accomplishments point to an equally impressive future, but they only tell part of the story of neuroscience.

Even the best neuroscience techniques cannot ascertain an individual's moral beliefs or the content of their religious convictions or hopes for the future [15]. As Martha Farah put it: "Although brain waves do not lie, neither do they tell the truth; they are simply measures of brain activity" [16]. It is important to keep in mind as scientists, while the science is important, it is only a part of the puzzle that informs how neuroscience will be used in a complex and interdependent society.

3 Neuroethics: Intriguing But Not Necessarily New

The issues raised by fMRI, PET scans, and other neurosciences are foreshadowed by the history of phrenology, and to some extent by the broader history of research with human subjects. Whether we are talking about neurodeterminism, neuroexceptionalism, neurorealism, neurosurrealism, neuroage, neuroeconomics, neuromarketing or any of the other vast array of social topics impacted by neuroscience, it is important to keep in mind that neuroscience is a part of the larger world that we all live in. History cautions us to reflect upon the consequences of the meanings that we ascribe to scientific knowledge. Even if the new neurosciences are as powerful as some would hope they might be, they are limited. Some of the issues raised in ethics and legal scholarship include:

- If the mind is a biological construct, who is the real me? [17]
- Where is volition located in the brain, and if not identifiable, does this imply a more holistic notion of the brain/mind question? [18]
- Should we engage in cognitive enhancement through neurotechnology? [19]
- If so, how should we assure conditions of access to all who might benefit? [19]

- How should we assign responsibility in criminal law? How can we protect neuroprivacy? Does the law have the right to force an accused to undergo neurological examination? [20] Is this a violation of the Fifth Amendment? [21]
- These are not necessarily new issues, although they may be packaged in neurolanguage.

4 Neuropolicy, Neurorealism, Neurolaw: Garnering Attention as the Science is Developed

The fields of study affected by neurosciences has been estimated to produce over 1000 articles a month [18]. Neuroethics is a large and expanding sub-discipline within the field of bioethics, and is the subject of numerous journals devoted to the ethical issues that arise from these technologies [22].

Science is expanding our understanding of the brain, but it cannot fully answer the questions that are raised about the contents or sources of the mind. Many thoughtful commentators have attempted to offer a conceptualization of the mind. Stephen Pinker has written extensively, and exhaustively on the mind. He describes the project thus:

> The opening chapter presents the big picture: that the mind is a system of organs of computation designed by natural selection to solve the problems faced by our evolutionary ancestors in their foraging way of life. Each of the two big ideas - computation and evolution - then gets a chapter. I dissect the major faculties of the mind in chapters on perception, reasoning, emotion, and social relations (family, lovers, rivals, friends, acquaintances, allies, enemies). A final chapter discusses our higher callings: art, music, literature, humor, religion, and philosophy. There is no chapter on language; my previous book *The Language Instinct* covers the topic. [23]

Pinker observes the complexity of the mind at close range and finds a fascinating range of potential human endeavors enabled by the mind. However, the book ends with the observation that some things are simply not explainable in scientific, humanistic, or psychological terms. In response to the question of whether the mind is located in the brain or in the soul, I would answer "neither." A full explanation of the mind and how we should regard the mind in society entails much more than science or philosophy. As scientists, ethicists, and lawyers, we should all be concerned with how neuroscience is appropriated by society. None of us has the full answer, but perhaps we each understand a part of the whole.

5 Proposal for Connection: Engaging the Broader Conversation

If history teaches us anything, it is that scientists must become involved in the conversation about the ethical, social, and legal uses of the knowledge created. The law, judges, policy makers and the attorneys who represent clients are all concerned with the appropriate use of neuroscience in the courtroom. As scientists, the courtroom may be the most common point of intersection with the law as an expert witness, but it is not the only contribution available. Ethical issues have an enduring quality, and there is always something to contribute from one's own experience. New technologies have always garnered attention, and neuroscience is similarly situated. Neuroscientists will be increasingly called upon to explain the significance of their work in terms of the ethical meaning they hold, as well as in scientific terms. There are several contributions that scientists can make following a review of the lessons from history.

Lessons from Tuskegee. The Tuskegee Syphillis Study is an important part of American history, and still resonates with the many groups in American society [24]. Researchers engaged with human subjects have an ethical and legal obligation to inform subjects of their right not to participate in research, and explain the risks and benefits of participation. The broader message from Tuskegee may well be the obligation of science to stop and reflect on the ethical implications of the way that knowledge is produced. Many studies today have an ethics core as an integral part of the design and methodology of the study. The trend towards integration of ethics and science, started in the aftermath of Tuskegee, should be a required connection of all neuroscience with larger social issues.

Lessons from Phrenology. New scientific advances are often met with great expectations. The optimism of a greater knowledge should be tempered with the wisdom of those who came before us. Neuroscience that becomes a part of the popular consciousness can become more influential than the science is actually capable of delivering. Businesses that value truthfulness and cooperation as virtues in their employees may want to find these qualities of character in a neurological image. We should make limited claims about our ability to make predictions about anything as difficult to define as the mind. Similarly, the law is premised on notions of the free will of individuals and responsibility for decisions. The courts have struggled with assigning criminal responsibility based on scientific information. Yet even if the best images could locate decision making in action, it does not follow that there is no free will. Neuroscience has the ability to inform the thinking of our best philosophers and jurists, but it does not supplant human experience and reflection.

Lessons from fMRI. Like phrenology, fMRI techniques raise important questions about the mind and personal identity. Researchers have demonstrated that individuals in the prodromal stage just prior to onset of Huntington's disease have reduced striatal volume and changes in the caudate putamen compared to earlier images of the same patient. Some clinicians argue that disclosure of these changes should be disclosed to patients, arguing that it will facilitate long-term planning. Others urge that disclosure should not occur until clinical manifestations of the disease are evident, arguing that the burdens of the disease are best postponed. We will not know how best to act on this knowledge until some consensus has been achieved, with the viewpoints of patients and scientists leading the way towards an understanding of "who" the patient is, and "when" they become a patient.

Lessons from Prior Ethical Deliberation. Ethicists have raised additional issues that will benefit from a broad conversation about the impact of neurotechnologies. Should we use neuroscience to enhance human capabilities? The answer to the question will depend upon our ability to deliver the science reliably and in a cost-effective manner. The current state of these mind-altering drugs is likely not sufficiently developed to cause a large societal upheaval. Many mind-altering substances are used by millions of people without undue societal disruption. However, drugs designed to alter memory or to change moral emotions such as trust could easily become problematic should they become truly effective and accessible.

An especially important and related topic for the brain-machine interface is the potential for the manipulation and abuse of others through what may be coined "mind wars". The United States government utilized secret experiments with LSD in the military [25] and provided evidence that the administration of intensely mind-altering drugs can be accomplished at will. In the only documented case in the legal literature, the United States government suffered no sanctions for this activity. Memory enhancement through machine interface could be enormously helpful in a time of combat, and could be required of troops, and perhaps without their consent if prior case law is extended to these new circumstances. Scientists and researchers should be aware of these possibilities for the misuse of neuroscience, and they should consider the agenda of those who would fund their work.

Like the atomic scientists and geneticists before them, neuroscientists and those working at the brain-machine interface will likely bear the burden of the possibility for great harm from the scientific knowledge produced. Such is the burden of success.

References

1. Beecher, H.: Ethics and Clinical Research. New Eng. J. Med. 274, 367 (1966)
2. Curran, W.J.: The Tuskegee Syphillis Study. New Eng. J. Med. 289, 730–731 (1973)
3. Regulation of Research with Human Subjects 45 CFR 46 (2011)
4. Goodenough, O.R.: Mapping Cortical Areas Associated with Legal Reasoning and Moral Intuition. Jurimetrics Journal 41, 429 (2001)
5. Gall, Franz Josef. Letter from Dr. F. J. Gall to Joseph Fr[ether] von Retzer, upon the Functions of the Brain, in Man and Animals. In: David G. Goyder, My Battle for Life: The Autobiography of a Phrenologist, 143-52 (1857), http://books.google.com/books?hl=en&lr=&id=6GsBAAAAQAAJ&oi=fnd&pg=PA1&dq=david+g+goyder&ots=zIT9k2XNPA&sig=ALn2WKRO3os_geJAbGXJpH27f4E#v=onepage&q&f=false (site accessed March 1, 2011)
6. Tovino, S.A.: Imaging Body Structure and Mapping Brain Function: A Historical Approach. Am. J. L. & Med. 33, 193 (2007)
7. Schlag, P.: Law and Phrenology. Harvard Law Review 110, 877 (1997)
8. Brock v Luckett's Executors 5 Miss. 459, WL 2421 (Miss. Err. & App. (1840)
9. Farrar v State 2 Ohio St. 54 (1853)
10. General Electric Co. v Joiner 522 U.S. 136, 153 & n.6 (1997) (Stevens, J., concurring)
11. Norris, D.G.: Principles of Magnetic Resonance Assessment of Brain Function. J. Magnetic Resonance Imaging 23, 794 (2006)
12. No Lie MRI, http://www.noliemri.com/default.htm (accessed March 1, 2011); Cephos Corporation, http://www.cephoscorp.com/ (accessed March 1, 2011)

13. Soneson, C., Fontes, M., Zhou, Y., Denisov, V., Paulsen, J.S., Kirik, D., Petersen, A.: Huntington Study Group PREDICT-HD Investigators. Early changes in the hypothalamic region in prodromal Huntington disease revealed by MRI analysis. Neurobiology of Disease 40, 531–543 (2010)
14. Gazzinaga, M.S.: Nature's Mind: The Biological Roots of Thinking, Emotions, Sexuality, Language, and Intelligence, 124 (1992)
15. Kennedy, D.: Neuroimaging: Revolutionary Research or a Post-Modern Phrenology? Am. J. Bioethics 5, 19 (2005)
16. Farah, M., Wolpe, P.R.: Monitoring and Manipulating Brain Function: New Neuroscience Technologies and Their Ethical Implications. Hastings Center Report, May-June 2004, at 35, 40 (2004)
17. Moreno, J.D.: Neuroethics: an Agenda for Neuroscience and Society. Nature Rev. Neuroscience 4, 149 (2003)
18. Pardo, M.S.: Philosophical Foundations of Law and Neuroscience. 2010 U. Ill. L. Rev. 1211 (2010)
19. Hyman, S.E.: Cognitive Enhancement: Promises and Perils. Neuron 69, 595 (2011)
20. Fox, D.: The Right to Silence as Protecting Mental Control: Forensic Neuroscience and "the spirit and history of the Fifth Amendment". Akron L. Rev. 42, 763 (2009)
21. The Self-Incrimination Clause of the Fifth Amendment provides that "No person ... shall be compelled in any criminal case to be a witness against himself." U.S. Const., Amend. V
22. Levy, N.: Introducing Neuroethics. Neuroethics 1, 1 (2008)
23. Pinker, S.: How the Mind Works (1997), `http://www.macroevolution.narod.ru/mindworks/pinker.htm` (Site accessed March 1, 2011)
24. Reverby, S.: More Fact than Fiction: Cultural Memory and the Tuskegee Syphilis Study. September-October Hastings Center Report 22-28 (2001)
25. U.S. v Stanley 483 US 669 (1987)

Individual Differences and the Science of Human Performance

Michael Trumbo, Susan Stevens-Adams, Stacey M.L. Hendrickson, Robert Abbott, Michael Haass, and Chris Forsythe

Sandia National Laboratories, MS 1188, Albuquerque NM, 87185-1188
{mctrumb,smsteve,smhendr,rgabbot,mjhaass,jcforsy}@sandia.gov

Abstract. This study comprises the third year of the Robust Automated Knowledge Capture (RAKC) project. In the previous two years, preliminary research was conducted by collaborators at the University of Notre Dame and the University of Memphis. The focus of this preliminary research was to identify relationships between cognitive performance aptitudes (e.g., short-term memory capacity, mental rotation) and strategy selection for laboratory tasks, as well as tendencies to maintain or abandon these strategies. The current study extends initial research by assessing electrophysiological correlates with individual tendencies in strategy selection. This study identifies regularities within individual differences and uses this information to develop a model to predict and understand the relationship between these regularities and cognitive performance.

Keywords: Individual Differences, EEG, Memory Span, RAT, Attentional Beam, Mental Rotation, Ruff Attention Task, Raven's Matrices, Box Folding, Dual Task, Barton's, Binary, Stroop, N-back, Mismatch Negativity, P300, Oddball, Semantic Memory, Episodic Memory, Go/No-Go, Flanker, Line Drawing, MAT-B.

1 Introduction

Generally, within the realm of experimental psychology, individual differences are treated as uncontrollable statistical variability (i.e., noise). However, much like one person's junk is another person's treasure, one study's noise is another study's signal. In that spirit, the current project will seek to identify regularities within individual differences and use this information to develop a model to predict and understand the relationship between these regularities and cognitive performance. This has implications for the effectiveness of training as well as various performance enhancement technologies, as both are contingent upon the ability to adapt to the knowledge, skills, and tendencies of specific individuals.

The current research will comprise the third year of the Robust Automated Knowledge Capture (RAKC) project. In the previous two years, preliminary research was conducted by collaborators at the University of Notre Dame and the University of Memphis. The focus of this preliminary research was to identify relationships

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 46–54, 2011.

between cognitive performance aptitudes (e.g., short-term memory capacity, mental rotation) and strategy selection for laboratory tasks, as well as tendencies to maintain or abandon these strategies.

In initial studies, research focused on a simple drawing task in which participants reproduced a figure 8 as well as other drawings under varying conditions to assess when the participant might switch strategies. Prior to this drawing task, participants were administered a battery of tests to assess aptitude for different facets of cognitive performance. It was shown that participants who performed well on the Remote Associates Task (RAT), which measures the ability to see relationships between concepts and has been advanced as a measure of creativity [1], tended to explore more alternative strategies [2]. In contrast, participants who scored high on working memory capacity attempted fewer strategies, with these attributes highly correlated with academic performance (i.e., SAT score). Perhaps most notable, based on data obtained from this study a model that incorporates both task-based (e.g., changes in task demands, time on task) and experience-based (e.g., frequency of strategy shifting, performance dips) factors has been developed that predicts on a trial-by-trial basis whether participants will shift strategies with 84% accuracy [2].

Further research assessed whether the findings from the initial studies would generalize to a more complex task environment. For these studies, the NASA Multi-Attribute Task Battery (MAT-B) [3] was employed. This task requires participants to divide their attention between performance of four simultaneous tasks and monitoring of two information systems. In addition to behavioral performance, these studies also recorded eye movements as an indication of ongoing strategy. A detailed description of this task can be found in subsequent sections. At this time, data collection has been completed, yet data analysis is still in progress.

2 Current Research

The current study extends the previous research by assessing electrophysiological correlates with individual tendencies in strategy selection. This study involves a replication of the previous research, but with the addition of a battery of tests to identify individual differences in brain electrophysiological parameters. Specifically, a collection of tests has been assembled consisting of tests frequently reported in the literature for which individual differences occur for certain electrophysiological parameters (e.g., amplitude of selected electroencephalography (EEG) signals, predominant EEG frequency). It is hypothesized that individual electrophysiological parameters (e.g., ratio of variability in theta (4-7 Hz) and beta (13-20 Hz) bandwidths) will correlate with propensity to shift strategies, as well as with the responsiveness to changing stimulus conditions.

Participants are asked to complete a series of experimental tests, including the Cognitive Aptitude Battery, the Brain Electrophysiology Battery, the Figure-8 Drawing Task, and the Multi-Attribute Test Battery (see subsequent section for task details). EEG data is recorded during the Brain Electrophysiology Battery, the Figure-8 Drawing Task, the Multi-Attribute Test Battery, and three of the cognitive tasks (mental rotation, Raven's matrices, and the Stroop task) using a 128-channel EEG system manufactured by Advanced Neuro Technologies.

Participants consist of 40 employees recruited from the Sandia National Laboratories population. Participants must be 18 years or older, speak English as a first language, possess normal or corrected-to-normal vision, and must not have a history of neurological diagnosis or of head trauma (including skull fracture, brain injury, or brain surgery).

2.1 Test Batteries

As mentioned above, participants undergo a series of test batteries. Testing requires a total of approximately seven hours. What follows is a list of the test batteries and what specific tasks they entail. Test batteries are listed in the order in which they are administered. Prior to commencing testing, basic demographic information is collected.

Cognitive Aptitude Battery, 1st Phase. During the first phase of the cognitive aptitude battery, participants complete a report of SAT and/or ACT score, the adult attention deficit hyperactivity disorder (ADHD) self-report scale, a memory span task, the Attentional Beam Task, the Remote Associates Task, the Pittsburgh Sleep Quality Index (PSQI), a video game survey, a box-folding task, the Barton/visual pattern task, the Ruff 2 & 7 Attention Task, a binary decision task, and a dual task. All of the tasks for this phase are completed on a Dell Precision M60 laptop with a screen measuring 40cm diagonal.

SAT Score. Participants are asked if they took the SAT. If they did, they are asked to report their score. If they did not, they are asked to report their ACT score. This information isused to correlate their standardized score with their performance on the other tasks.

Adult Attention Deficit Hyperactivity Disorder (ADHD) Self-Report Scale. For this task, participants are asked to fill out a questionnaire asking about attention performance and problems. This questionnaire is meant to identify those individuals who have problems concentrating their attention.

Memory Span. For the working memory span test, participants are given two sentences to read. Their task is to indicate whether each sentence is sensible or not by pressing a button labeled “yes” or “no” on a computer mouse. After both sentences have been presented, participants are asked to recall the last word from each of the sentences they just read via typing them into the computer.

Attentional Beam Task. The Attentional Beam task is illustrated in Figure 1. Participants first gaze at a fixation point for 600ms. Then they view a stimulus, which is an array of dots, one of which is colored in, for 30ms. Afterward, a visual mask of random lines is presented for 600ms. Next, a response display is presented. The task is to indicate in which of the eight directions the darkened circle was on that trial. This is a measure of the breadth of one’s visual attention focus.

Remote Associates Task (RAT). For the RAT, participants are presented with a set of three words on a computer (e.g., law, birthday, and swim). The task is to indicate whether they are related or not. This is a measure of creative thinking [1].

Pittsburgh Sleep Quality Index (PSQI). For this task, participants are asked to fill out a questionnaire assessing their sleep satisfaction and quality [4].

Video Game Survey. For this task, participants are asked to fill out a questionnaire assessing their video game experience. This information is important as it may dictate how well participants perform on some of the tasks [5].

Box-folding. For this task, participants are shown a diagram of a box that has been unfolded into a flat surface. Two edges are marked and the participants must indicate whether those two edges would meet if the object were folded into a box again. This is a measure of mental rotation.

Barton/Visual Pattern Task. This task is a measure of spatial memory. On each trial, participants see a grid of squares, some of which are colored. Then, on a blank grid, they select which squares were colored. The more positions a person can correctly remember, the better their spatial memory.

Ruff 2 & 7Attention Task. For the Ruff 2 & 7Attention Task, participants are presented with a series of strings of characters and must find all of the 2s and 7s in the string. They do this by clicking on the 2s and 7s using a computer mouse. Timing and accuracy are recorded. This is a measure of attention [6].

Binary Decision Task. For this task, participants are shown two lights and are asked to pick which light they think will light up over multiple trials. Unbeknownst to the participants, there is a fixed probability with which the lights will light up. Over time, the participants learn this probability and learn to pick which light has the greater chance of being lit. This is a measure of decision making and learning.

Dual Task. For this task, participants are asked to follow a dot with their cursor while also listening to and remembering a list of words. This is a measure of attention.

Cognitive Aptitude Battery, 2nd Phase. At this point, the 128-channel EEG cap is applied. Following application of the cap and attaining satisfactory electrode conductance the participant is seated in a sound attenuating booth. The booth creates a controlled environment that minimizes distractions during the study. At that point, the participant is presented with the remainder of the cognitive tasks: the mental rotation task, Raven's matrices, and the Stroop task. Tasks for this phase, and all subsequent tasks completed for the remaining duration of the study, are performed on a Wacom drawing tablet monitor with a screen measuring 55cm diagonally.

Mental Rotation. For this task, participants are presented with a series of 20 pairs of figures, such as that shown in Figure 2. The task is to indicate whether the two figures correspond to the same object or not. The number correct that can be classified in 60 seconds is taken as a measure of mental rotation ability. This is a general measure of visual-spatial processing.

Raven's Matrices. For this task, participants are asked to solve a matrix by identifying the missing item that completes the pattern. This is a measure of abstract reasoning.

Stroop Task. In this task, participants see a series of words written in a certain color of ink and are told to respond with the color of ink. The crucial part of the experiment is that the words are color names written with an incongruent ink color (e.g., the word "BLUE" written in red ink). This is a measure of mental attention and flexibility.

Multi-Attribute Test Battery (MAT-B). As shown in Figure 3, this task involves a computer program in which participants must keep track of events happening in 6 different windows of a display screen. The participant must simultaneously perform the roles of:

1. System Monitoring – watching for when the values go out of normal parameters
2. Tracking – keeping a randomly moving reticule in the center using a joystick
3. Scheduling – watching for important events that come up
4. Communicating – setting radio frequencies in response to messages presented over headphones
5. Resource Managing – keeping tanks filled by opening and closing valves in a system
6. Monitoring Pump status – on or off

This task requires participants to develop strategies with respect to how they divide their attention between the different displays.

Following general instructions, participants undergo a practice round for the communication task, followed by practice rounds for the system monitoring, resource management, and tracking tasks. During the practice rounds, participants only perform the specified task. Following completion of practice, participants are asked to perform all of the tasks simultaneously. Each task has a fluctuating difficulty level. EEG is recorded for the duration of the task (including practice). In addition, a Smart Eye tracking system is used along with Smart Eye Pro 5.5 software to monitor the direction of gaze of the participant relative to the display.

Brain Electrophysiology Battery. This battery includes a measure of resting EEG, an N-back task, a measure of mismatch negativity, the P300 auditory oddball, a semantic memory task, an episodic memory task, a go/no-go task, and a flanker task. EEG data is recorded for each of the measures within this battery.

Resting EEG. Resting EEG parameters are recorded for conditions with eyes open and eyes closed. In the eyes open conditions, participants are asked to fixate on a cross presented on the computer screen to minimize eye movements. Eyes open and eyes closed conditions extend for 1 minute intervals with a counterbalanced order (e.g. O–C–C–O–C–O–O–C).

N-Back. The spatial position of an uppercase letter is presented pseudo randomly in one of 8 locations on a computer screen. Participants are asked to compare the position of the stimulus to the position of the stimulus presented one or more trials previously. Generally, participants do well when comparing to the immediately preceding stimulus and find it very difficult when comparing to stimuli two or more trials back [7].

Mismatch Negativity. As participants watch a silent movie with subtitles on the computer screen, tones are presented at regular intervals through headphones. They are told to not pay attention to the tones. A standard tone (900 Hz) is presented in 85% of the instances and a deviant tone (1000 Hz) in the remaining 15%, using a semi-randomized order.

P300 Auditory Oddball. As with the mismatch negativity test, participants watch a silent movie with subtitles on the computer screen as tones are presented at regular intervals through headphones. However, for this test, participants are asked to press the "1" key every time they hear the deviant tone, as quickly and accurately as possible.

Semantic Memory. Participants are presented a series of words denoting objects (e.g. tree, car) and for each word, they are asked to press the "1" key if "living" and the "2" key if "nonliving."

Episodic Memory. Without prior warning, after a 5 minute interval following completion of the semantic memory task, participants are asked to recall as many of the words presented during the Semantic Memory task as possible. Participants are allowed to continue until they report being unable to recall any more words.

Go/No Go. Participants are presented a series of letters for 200ms each (e.g. A,B,C,D,E,F,G,H,J,L,O,X) and asked to press the "1" key every time the letter X (go trial) is presented after the letter "O" (primer trial) had been presented on the previous trial. For half the trials, the primer trial "O" is followed by a different letter (A,B,C,D,E,F,G,H,J,L) and participants have to suppress the key press (No go trial).

Flanker. Participants are presented a string of 5 letters consisting of either S's or H's (e.g., HHHHH, or SSSSS). The middle letter is the target and there are compatible strings (e.g., HHHHH or SSSSS) and incompatible strings (e.g., HHSHH or SSHSS). Participants are asked to press the "1" key if the target letter is an "S", and the "2" key if target letter is an "H," as quickly and accurately as possible.

Figure-8 Drawing Task. Following completion of the Brain Electrophysiology Batter, participants are given a break, after which they will be presented with the Figure-8 Drawing Task. For this task, participants are presented with a series of outlines of predominantly figure 8s (various other figures are included) on a computerized drawing tablet. They are asked to use a stylus to draw outlines of the figure 8 on the tablet. Participants receive immediate feedback following each trial, and are scored based on the accuracy and speed with which they complete each trial.

Over the course of 50 trials, the figure 8 changes (such as circles moving apart) as a means to induce participants to shift drawing strategy to accommodate these changes. EEG data is collected for the duration of this task. This is a measure of strategy shifting and replicates the measures used in previous studies conducted. Figures 4-6 depict a screenshot of this task, three predominant drawing strategies, and the various morphs that tend to lead to these strategies.

3 Discussion

Since the time of Darwin, scientists have been aware that members of a single species may show marked differences from one another [8]. Though the concept that individual variation in behavior may create particular categories of individuals within a species has been recognized by certain disciplines (e.g., personality psychology), and for particular species (e.g., rats, pigs) [9], the prevailing assumption in human cognitive theory and related modeling has been that cognitive processes are largely invariant across individuals and across different conditions for an individual [10]. Seeming discrepancies across studies are often attributed to within or between-subject variability that is treated as measurement error. Alternatively, we propose that this variance may be attributable to *cognitive adaptability*, a trait necessary to explain the inherently dynamic nature of cognitive processes as individuals adapt their available resources to ongoing circumstances. This does not imply a “blank slate”; humans are predisposed to process information in particular ways. Instead, it is asserted that given variation in the structure and functioning of the brain, there exists inherent flexibility that may be quantified and used to predict differences in cognitive performance between individuals and for a given individual over time [10].

These assertions are based on the theoretical proposition that the brain functions as a dynamic system, involving the continuous interaction of genetic, organismic, and environmental factors [11]. Given a particular problem to solve, an individual will have access to a variety of resources. However, individuals will differentially apply the resources available to them based on aptitude, past experience, recent success/failure, etc. Thus, the study of individual differences is important in that it allows individuals to be characterized with regard to their various aptitudes (i.e., the resources they have and may differentially relay upon) and variability in strategy selection for a given task may be explained on the basis of individual differences.

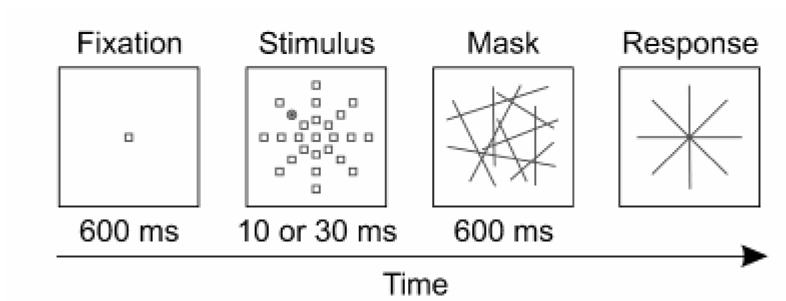


Fig. 1. The display sequence for a single trial of the attentional beam task

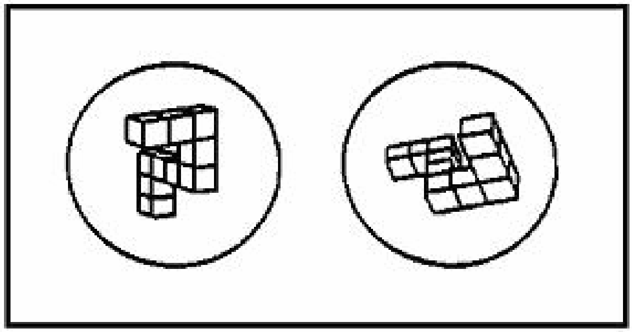


Fig. 2. An example of the stimuli for the mental rotation task

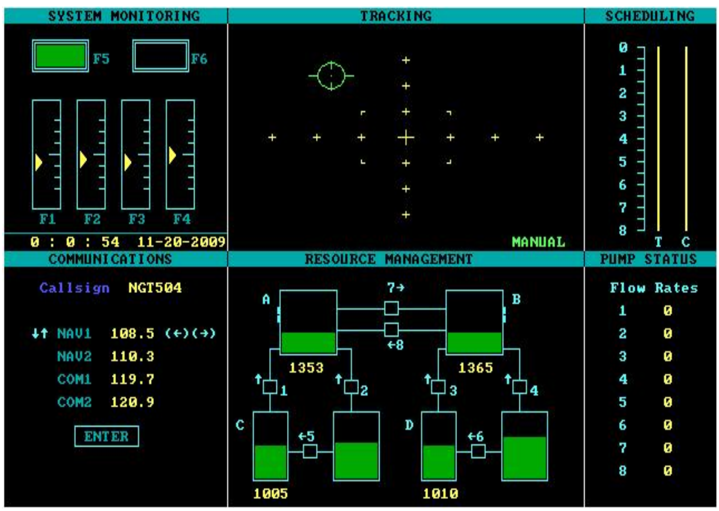


Fig. 3. A depiction of the MAT-B task. Each of the six boxes is labeled with respect to the task it contains.

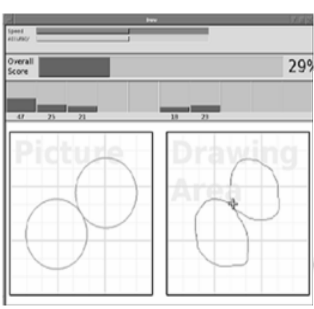


Fig. 4. A screenshot of the line drawing task

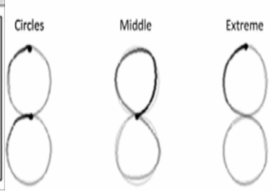


Fig. 5. Examples of three common strategies

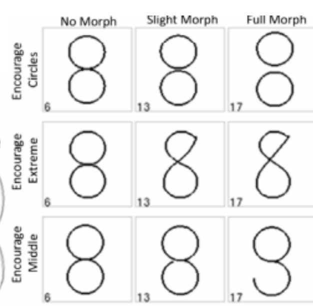


Fig. 6. Morph conditions, and the strategies they induced

Acknowledgement. Sandia is a multi-program laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United Sates Department of Energy's national Nuclear Security Administration under contract DE-AC04-94AL85000. The authors would like to thank the lab directed research and development (LDRD) for funding this work.

References

1. Razumnikova, O.M.: Creativity Related Cortex Activity in the Remote Associates Task. Brain Res. Bull. 73, 96–102 (2007)
2. Radvansky, G.A., D'Mello, S., Abbott, R.G., Moregan, B., Fike, K., Tamplin, A.K.: Rum Runner: A Model of Strategy Switching. J. Cogn. Neurosci. Submitted (2011)
3. Comstock, J. J.R., Arnegard, R.J.: The Multi-Attribute Task Battery for Human Operator Workload and Strategic Behavior Research. NASA Technical Memorandum 104174 (1992)
4. Buysse, D.J., Reynolds, C.F., Monk, T.H., Berman, S.R., Kupfer, D.J.: The Pittsburgh Sleep Quality Index (PSQI): A New Instrument for Psychiatric Research and Practice. Psychiatry Research 28(2), 193–213 (1989)
5. Feng, J., Spence, I., Pratt, J.: Playing an Action Video Game Reduces Gender Differences in Spatial Cognition. Psychol. Sci. 18(10), 850–855 (2007)
6. Knight, R.G., McMahon, J., Skeaff, C.M., Green, T.J.: Reliable Change Indices for the Ruff 2 and 7 Selective Attention Test in Older Adults. Appl. Neuropsychol. 17(4), 239–245 (2010)
7. McEvoy, L., Pellouchoud, E., Smith, M., Gevins, A.: Neurophysiological Signals of Working Memory in Normal Aging. Cog. Brain Res. 11(3), 363–376 (2001)
8. Darwin, C.: On the Origins of the Species by means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life, London (1859)
9. Jensen, P.: Individual Variation in the Behaviour of Pigs – Noise or Functional Coping Strategies? App. Animal Beh. Sci. 44, 245–255 (1995)
10. Hsu, N.S., Kraemer, D.J., Oliver, R.T., Schlichting, M.L., Thompson-Schill, S.L.: Color, Context, and Cognitive Style: Variations in Color Knowledge Retrieval as a Function of Task and Subject Variables. J. Cogn. Neurosci. Submitted [Epub ahead of print] (2011)
11. Stiles, J.: On Genes, Brains, and Behavior: Why Should Developmental Psychologists Care About Brain Development? Child Devel. Perspec. 3(3), 196–202 (2009)

Cognition: What Does It Have to Do with the Brain?

Alexandra Geyer

Aptima, Inc., 12 Gill Street, Suite 1400, Woburn, MA 01801
ageyer@aptima.com

Abstract. The emergence of new non-invasive technologies for assessing the structure and the function of the human brain has provided us with means to investigate the neural substrates underlying cognitive processes with the goal of achieving a better understanding of cognition. This paper is focused on discussing the contributions that assessment of neural processes underlying cognition brings to our understanding of cognition as well as the impact of this understanding of cognition on military operations.

Keywords: cognition, neural substrates, intent, warfighters.

1 Introduction

Over the past three decades, the emergence of new non-invasive technologies for assessing the structure and, more importantly, the function of the human brain has provided us with the means to investigate the neural substrates underlying cognitive processes with the goal of achieving a better understanding of cognition. Prior to that, cognition was explored by cognitive psychologists primarily through the examination of human behavior. This approach was successful in gaining a fair amount of insight into human cognition. However, given that virtually all human cognition relies on activity in various areas of the brain, it order to really understand it, we need to consider not just the behavior, but also the activity in the brain and how this activity influences the behavior. Since the late 1970s, when George Miller and Michael Gazzaniga coined the term "cognitive neuroscience" in the back of a taxi cab [1], which refers to the study of how the brain enables the mind [2], cognitive neuroscientists have been working hard at integrating the theories that though the work of cognitive psychologists with approaches in experimental psychology, neuropsychology and neuroscience in order to link the behavioral outcomes of cognition with brain activity.

Since the emergence of cognitive neuroscience, questions have been raised whether its newer, less tested, methodologies which still require time and effort to be fully validated bring any benefit to assessing and understanding cognition over what the proven and largely validated traditional cognitive psychology approaches to assessing cognition have brought. On the one hand, as Michael W. Eysenck and Mark T. Keane correctly point out in their textbook of Cognitive Psychology: "Experimental cognitive psychologists possess many well-worked-out empirical methods built up over 100 years of experimentation. In the neurosciences, in contrast, the methodologies for studying phenomena are still being developed… In brain imaging, there are still many

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 55–59, 2011.

issues about the methods used to rile out "noise" and exclude activity in brain regions that are merely byproducts of the focus of the study" (p. 4) [3]. On the other hand, the cognitive psychologists' approach of assessing cognition by measuring behavior does not provide a full picture of cognition. For example, there is sometimes an assumption that there is a direct mapping from specific cognitive process to specific behavior. This is not always the case. In fact, often, there are multiple cognitive processes that underlie one behavior. Also, even when it is recognized that multiple cognitive processes are involved, it is very difficult to tell, without employing methodologies from cognitive neuroscience, whether these processes occur in parallel on in serial fashion, and whether there is any interdependency among these processes. Here I put forth an argument that, in some instances, information about the neural processes underlying cognition can provide a more accurate picture of human cognition and, therefore, should be used in addition to traditional behavioral measures of cognition. A good example of an instance when such information is useful is: human operator - technology interaction. Modern military operations are increasingly complex and technology reliant. While system developers focus on incorporating emerging technological advancements into their products, they do not always consider the cognitive capabilities and limitations of the operators themselves, and how these capabilities and limitations might affect operator interactions with those technologies. In order to create a more seamless interaction between the human operators and the technologies, it is important to have a good understanding of the operator state so that the technology could adapt itself to the operator's needs. Below I provide some examples of using neural signals to improve human operator - technology interaction.

2 Benefits of Cognitive Neuroscience to our Understanding of Human Cognition

One example of using neural signals to improve human operator - technology interaction is to use them to inform decision aids about the operator state so that they do a better job supporting operator decision making. Recent advances in neuroscience allow us to reliably and unobtrusively measure cognitive states and cognitive processes involved in decision making and problem solving in operational settings [4], [5], [6], [7], [8], and [9]. For example, Luu et al. reliably identified the neural substrates of intuition in the medial Orbital Frontal Cortex (mOFC) utilizing dense-array Electroencephalography (dEEG). Intuition is often credited with helping warfighters succeed in critical situations. Research in human pattern recognition and decision-making suggest that through intuition, humans can sense unique patterns without consciously seeing them [6], [7]. Luu et al. defined intuition is an affectively charged, internal cue to the existence of meaningful information in the environment that arises rapidly and unconsciously. The development of reliable measures of intuition provides new opportunities to understand and aid human cognition. For example, this signal may enable creation of a new generation of machines to support decision making. A decision aid that senses its user's intuitive moments might serve the user best by withholding information that could conflict with the user's rising solution. For instance, automated target recognition aid typically overlays sensor imagery (e.g., a vehicle obscured by trees) with a diagrammatic template of the most likely target (e.g., a technical or pickup truck mounted with a machine gun). Given a reliable signal of intuition, the aid might withhold that template while the user's recognition of the scene resolves, with the expectation that the user's response will be more accurate and sufficiently quick. On the other hand, if the aid registers the absence of an intuition signal, it might serve its owner by rapidly cueing a solution if time is short. If time is plentiful, it might present its owner with decision analysis tools.

Another possibility is to use neural signals from the operator to inform the system about the operator intent, allowing for a more seamless interaction between the human operator and the system. This is a somewhat controversial idea because, as one would argue, we cannot and possibly will never be able to “read people’s thoughts” and gauge their intent through the analysis of the neural signals. At this point, the closest we can do, is - observe the action that followed the intent and, based on the action, infer what the intent was. Here, I define action as a behavior that is undertaken with a specific goal in mind, and we define the intent to perform an action as an aim that guides action. While actions following the intent are, for the most part, easily measured, measuring signals that represent a human’s intent to perform an action is far more complex. The neural signatures that underlie the action aim (i.e., action intent) have been investigated in monkeys since the 1990s. For example, specific cells in the ventral premotor cortex as well as the posterior parietal cortex of monkeys fire during action intent [10]. These same areas have been demonstrated to be active during action intent in the human brain [11], [12]. Even though we now know the brain regions that are activated during action intent and, therefore, could measure neural activity associated with intent, the picture is not fully complete since actions are subject to continuous action monitoring and error detection, which has a significant and continuous impact on the intent. Therefore, in order to reliably measure intent, we need to be able to measure the neural signals that underlie action monitoring and error detection as well, as they will serve as cues to possible changes in intent.

In recent years, scalp electrophysiological studies in humans have provided insight into the neurophysiological processes of action monitoring and error detection, which could be used as signals of a change in intent. Falkenstein et al. identified a negative deflection in the event-related potentials (ERPs) that is associated with an error response [13]. This negativity can be observed in both stimulus-locked and response-locked ERPs. Other researchers independently observed this effect [14] and termed it the error-related negativity (ERN). Dipole modeling of a 64-channel recording showed that the neural source of the ERN lies in the vicinity of the anterior cingulate gyrus ([15]. Luu et al. proposed that the core contribution from the anterior cingulate gyrus and associated medial frontal cortex is evaluative self-monitoring along an affective dimension [5]. When subjects are concerned with the outcome of an action, the cingulate gyrus and associated centromedial frontal lobe provide a dynamic monitoring of that action and its effects. This was evidenced by progressively-larger ERNs to responses increasingly past the deadline; this self-monitoring process appears to be a dynamic operation, providing continuous motivational evaluation, and presumably continuous executive guidance, to direct the neural mechanisms of behavior (i.e., directing action intent).

Another signal of interest for detecting intent is the lateralized readiness potential (LRP) which is a measure of asymmetric motor preparation and has been used as a measure of motor preparation in response to task demands [16]. The reason LRP is important is that it is a signal of intent to make a motor action. The LRP has also been used to study the effects of response competition arising from incongruent primes that are presented subconsciously [17]. Dehaene et al. (1998) found that the response competition indexed by the LRP can be validated using fMRI methods [17]. These researchers used a lateralized measure of the blood oxygen level-dependent (BOLD) response in motor cortices and found the effect to be identical to the effect using the LRP (i.e., incongruent primes activated the wrong motor areas). Finally, the LRP has been used to study early response activation in classical conflict paradigms, such as the Eriksen flanker task and the Stroop task, and to study incompatible response priming induced by incompatible noise [16]. In the Luu et al., study, the late responses were associated with inappropriate priming of the motor cortex ipsilateral to the eventual response hand [5]. By looking at just the incompatible trials, it could be seen that the degree of response lateness is related to the degree of priming of the inappropriate response hand. The LRP that is observed is likely caused by the effects of the incompatible stimuli, because the onset of the LRP occurs before the response.

It is evident that the neural signals of intent described above simply alert us that there is intent for action or that there is a change in intent. They do not reveal the nature of intent. However, a smart system, when it receives a signal from the brain of the human operator that there is an intent of actions, should be able to utilize other information, in addition to neural signals, to figure out what the intent actually is and either provide the operator with means to carry out the intended action (or, ideally, carry out the intended action for the operator) or advise the operator against taking that action if there is a reason for that.

3 Conclusion

In this paper, I argued that information about the neural processes underlying cognition can provide a more accurate picture of human cognition. I then provided two examples where neural signals of intuition and intent could be utilized to create a more seamless interaction between the human operators and the technologies.

References

[1] Gazzaniga, M.S. (ed.): Conversations in the Cognitive Neurosciences. The MIT Press, Cambridge (1999) ISBN 0-262-57117-X

[2] Gazzaniga, M.S., Ivry, R.B., Mangun, G.R.: Cognitive Neuroscience: The biology of the mind, 3rd edn. W.W. Norton, New York (2009)

[3] Eysenck, M.W., Keane, M.T.: Cognitive psychology: a student's handbook, 5th edn. Taylor & Francis, Abington (2005)

[4] Hackley, S.A., Valle-Inclán, F.: Automatic alerting does not speed late motoric processes in a reaction-time task. Nature 391, 786–788 (1998)

[5] Luu, P., Collins, P., Tucker, D.M.: Mood, personality and self-monitoring: negative affect and emotionality in relation to frontal lobe mechanisms of error-detection. Journal of Experimental Psychology: General 129, 1–18 (2000)

[6] Bowden, E.M., Jung-Beeman, M., Fleck, J., Kounios, J.: New approaches to demystifying insight. Trend in Cognitive Sciences 9(7), 322–328 (2005)

[7] Kounios, J., Fleck, J.I., Green, D.L., Payne, L., Stevenson, J.L., Bowden, M., Jung-Beeman, M.: The origins of insight in resting-state brain activity. Neuropsychologia 46, 281–291 (2008)

[8] Volz, K.G., von Cramon, D.Y.: What neuroscience can tell about intuitive processes in the context of perceptual discovery. Journal of Cognitive Neuroscience 18, 2077–2087 (2006)

[9] Luu, P., Geyer, A., Fedopiastis, C., Campbell, G., Cohn, J., Tucker, D.: Reentrant Processing in Intuitive Perception. PLoS One 5(3), e9523 (2010)

[10] Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., Fazio, F.: Localization of grasp representations in humans by PET: 1. Observation versus execution. Experimental Brain Research 111, 246–252 (1996)

[11] Fadiga, L., Craighero, L., Buccino, G., Rizzolatti, G.: Speech listening specifically modulates the excitability of tongue muscles: a TMS study. European Journal of Neuroscience 15, 399–402 (2002)

[12] Grezes, J., Armony, J.L., Rowe, J., Passingham, R.E.: Activations related to ”mirror” and ”canonical” neurones in the human brain: an fMRI study. Neuroimage 18, 928–937 (2003)

[13] Falkenstein, M., Hohnsbein, J., Hoormann, J., Blanke, L.: Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. Electroencephalography and Clinical Neurophysiology 78, 447–455 (1991)

[14] Gehring, W.J., Goss, B., Coles, M.G.H., Meyer, D.E., Donchin, E.: A neural system for error detection and compensation. Psycholigal Science 4, 385–390 (1993)

[15] Dehaene, S., Posner, M.I., Tucker, D.M.: Localization of a neural system for error detection and compensation. Psychological Science 5, 303–305 (1994)

[16] Coles, M.G.H., Gratton, G., Donchin, E.: Detecting early communication: using measures of movement-related potentials to illuminate human information processing. Biological Psychology 26, 69–89 (1988)

[17] Dehaene, S., Naccache, L., Le Clec, H.G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., van de Moortele, P.F., Le Bihan, D.: Imaging unconscious semantic priming. Nature 395, 597–600 (1998)

The Evolution of Human Systems: A Brief Overview

Jeff Grubb[1] and Joseph Cohn[2]

[1] Naval Air Warfare Center Training Systems Division, Orlando, USA
[2] Human and Bioengineered Systems Division, Office of Naval;
Research, Arlington VA, USA
{FLjeff.grubb,joseph.cohn}@navy.mil

Abstract. Recently, there has been a profound resurgence interest in expanding the effectiveness of human machine systems. The motivation for this interest stems not only from the growing realization that better designed systems – tailored to augment their user's innate skills and capabilities – will enable users to 'do more', but also from the fact that the world with which we interact is becoming increasingly reliant on machines. In the past, the human machine interface was bridged through engineering based principles, but, with our expanding understanding of how the human brain drives behavior it is now possible to consider, as never before, human machine design efforts that fully address human and machine needs at the same time.

Keywords: Neuroscience, Cognition, Automation, Human Systems, Cognitive Model.

1 Introduction

Traditional approaches to creating human machine systems have focused on engineering or machine learning techniques to establish couplings between humans and their machines (Cooley, 2007). For example, many of the cognitive architectures that are intended to allow the machine to infer human intention are based on computer processing metaphors, not on actual brain dynamics. This is a direct result of the levels of technology available to understand and represent the processes through which the human brain transforms information into action. Until very recently, neither the imaging technologies nor the analytic capabilities were available to truly link actual brain activity to behavior. As a result, when one wished to create human machine systems one was forced to do so by basing this integration on observed behaviors, and building predictive models of human behavior on these observed behaviors.

Of course, the ideal is to directly link high fidelity representations of human behavior with machine operating systems. These representations may be found in the neural processes leading to the actual, observed behavior. Just as understanding the equations of motion provides a much broader set of capabilities than inferring these equations from a limited set of observations (Kelso, 1995), so too understanding and modeling the dynamics of neural activity as it leads to behavior should provide a much richer and more robust set of models than those based on the actual observed

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 60–66, 2011.

behavior alone. Today, advances in neuroscience and engineering provide the basis for building these 'equations of motion' for the brain and for using brain-based techniques to create and maintain very robust human machine interactions.

1.1 The Engineering Based Approach

The notion of creating human machine systems is not new. As early as the 1940s, researchers were concerned with the question of how to represent the human element in human machine systems. Using the engineering-based terminology of the time, Bates (1947), Craik (1947/1948; 1948) and others attempted to explain human performance in control theory terms with the goal of developing engineering representations of the human that could be used to improve human machine systems (Birmingham & Taylor, 1954). Others, like Chapanais (1951), applied a similar approach for analyzing human error in human machine systems.

In all cases, the properties of the human being modeled were only at the observed behavior level. For example, Fitts' speed-accuracy tradeoff (Fitts, 1954; Fitts & Peterson, 1964) emphasizes the development of basic relationships guiding human motor planning. Stevens' power law describes the relationship between the magnitude of a physical stimulus and its perceived intensity (Stevens, 1957), and the Hick-Hyman law relates decision response time to the number of possible choices (Hick, 1952; Hyman, 1953). One of the primary applications for this line of research was to develop more effective and responsive aviation systems (Adams, 1957; McRuer & Jex, 1967; Young, 1969). More broadly, fully automated systems which could perform tasks in the absence of actual human controller inputs were also envisioned as resulting from this line of research, as were human assistive systems (i.e. symbiotic systems, Licklider, 1960).

In one sense, automation may be thought of as a means of substituting human actions with those of a machine (Parsons, 1985; Parasuraman & Riley, 1997). The reasons for automating certain tasks range from safety considerations to cost and efficiency considerations (Weiner & Curry, 1980; Weiner, 1989). Inherent to the notion of automation is the idea that of an overall pool of tasks, some may be allocated to a system or machine, while others may be allocated to a human. In the most conservative sense (e.g. Licklider, 1960) automation requires a strict parsing of tasks – those at which a machine may excel and those at which a human may excel. Along those lines, Parasuraman, Sheridan & Wickens, (2000) proposed a set of discrete levels of automation to be implemented based on overall task context. Yet, as Rouse (1977) and Woods (1996) suggest, these kinds of approaches to automation are brittle. Situations change, information changes and people change, often as a consequence of using automation (Woods, 1996) and the allocation of tasks between humans and machines should be able to change, dynamically and in real time to provide the most effective assistance.

1.2 A Turning Point

The realization that dynamic changes in both users and their systems must be accounted for opened the door for a human-centered approach to human-systems design, leading to adaptive automation. Adaptive automation is an automation scheme that allows for control of tasks to be passed in real time, between human and machine – represents one attempt to bridge the human machine gap of classic automation (Scerbo, 1996). This kind of automation seeks to optimize human machine interactions by changing task demands in response to user performance. A direct result of this kind of automation is that the task environment is restructured, dynamically, in terms of *what* tasks are automated, *how* they are automated, and *when* they are automated (Rouse, Geddes, & Curry, 1988). A key consideration in adaptive automation is the means through which the adaptation is triggered. Early attempts at creating adaptive techniques focused on a purely artificial intelligence (AI) approach (Rouse, 1977), merging expert systems with knowledge based representations of human performance (or, loosely, cognition) to detect and assess task context, and develop adaptive strategies (e.g. Hammer & Small, 1995).

Yet, a significant disadvantage with this approach is that while representations of the human operator may be gleaned from psychometric measures (speed and accuracy types of measures), the level of fidelity of these representations is typically orders of magnitude less than that of representations of the system and task environment. Furthermore, the rate at which this information may be accumulated is similarly challenged. Human behavior typically evolves over a timescale measured in seconds, while machine action may occur over a millisecond to second timescale. In other words, currently available measures of the human are rate limiting in human-machine systems, even those that are adaptive.

1.3 Enter Neuroscience

Recent attempts to get around the human representation challenge have focused on adding another dimension of measurement, based on neurophsyiologically detected processes (Morrison & Gluckman, 1994). The basic premise is that by including neurophsyiological measures, it should be possible to gain higher levels of fidelity, over a shorter time course, for the kinds of human representation needed to make adaptive automation effective, beyond than simple observationally behavior based ones. These richer metrics would therefore serve as a more effective input into a dynamic and adaptive automation system (Scerbo, 1996). Such measures include (Scerbo, et al., 2001; Cohn, et al., 2005):

- Heart rate variability
- Eye-based responses (e.g. eye blinks, pupil diameter)
- Galvanic Skin Response
- Neural based signals (Electroencephalography –EEG, functional Magnetic Resonance Imaging – fMRI; functional Near Infrared imaging – fNIR)

These measurements provide a vast improvement over traditional measures that feed into adaptive automation developed to make use of them (e.g. Schmorrow, 2005; Schmorrow, Stanney & Reeves, 2007) and, collectively, have pushed the field of adaptive automation far ahead of where it might otherwise be. At the same time while these measures provide a more effective diagnostic metric indicating when automation might be useful, they don’t provide deeper descriptions of *what* tasks should be automated, *how* they should be automated, and *when* they must be automated. These determinations, in the current approach, are left to predefined strategies that are implemented based on the triggering of these measures (Rouse, et al., 1988).

2 Neuroadaptive Systems

The brain is not a static organ. It changes with experience (Cohn, Stripling & Kruse, 2005), creating new connections and optimizing older ones; it varies with emotional and physiological state, impacting and influencing higher order processes (Glimcher & Rustichini, 2004), and it is primed by changes in environmental context (Aamodt & Wang, 2008). The end result of the brain's inherent dynamicity is often changes in observed behavior (Cohn, Stripling & Kruse, 2005). Consequently, while it should be possible to create a closed loop human machine symbiotic system that can adapt its overall performance based on representations of brain activity care must be taken to understand precisely how to transform neural activity into representations of the behavior they encode.

Human Systems that incorporate elements of the brain's activity with representations of system state to enable human machine interactions are known as neuroadaptive systems. Neuroadaptive systems use the detailed output of their human users' neural activity in order to effectively adapt their behavior to the behavior of their users. This requires more than simply taking a snapshot of brain action. Neuroadaptive systems seek to enable adaptive interactions between humans and their machines using deeper and more representative measures of human neural action underlying behavior than those used in traditional adaptive automation technologies. With these representations, high fidelity individualized models of human performance can be crafted, which can be expected to behave in a manner analogous to that in which the human brain on which they are based will behave. Although still in their infancy, neuroadaptive systems are beginning to be realized as a direct result of recent advances in neuroscience and engineering.

2.1 Basis for Neuroadaptative Systems

Neuroadaptive systems are based on the idea that representations of the human user require the integration of measures across multiple levels, including brain based measures, cognitive measures and behavioral measures. This, in turn, requires advances in three core capabilities:

- Detection Technologies
- Decoding Methodologies
- Modeling Frameworks

2.2 Detection

The notion that human behavior is the result of coordinated activity across the brain is not new (Kelso, 1995). However, access to the brain has been one of the key limiting steps in demonstrating coordinated activity across the brain as behavior develops. As new technologies, like functional Magnetic Resonance Imaging (fMIR); (Logothetis, 2001), dense array Electroencephalography (dEEG); (Junghöfer, Elbert, Tucker, &

Rockstroh, 2000), and other types of tools become increasingly refined, simplified and incorporated into the researcher's toolkit, the ability to capture neural action simultaneously across multiple regions will continue to grow. As one example, Philiastides and Sajda, (2007) used an EEG based paradigm to illustrate the integration of different neural regions over time, as participants formed and acted upon a rapid decision making task.

2.3 Decoding

Access to integrated neural data is necessary but not sufficient for interpreting it in terms of cognitive processes. New processes for analyzing these multivariate data sets must also be established and refined, and efforts to do so have led to the development, refinement and application of these multivariate analytic techniques to data captured as participants perform a range of cognitive tasks. Briefly, multivariate decoding takes into account the full spatial pattern of brain activity, measured simultaneously across many regions, and enables the decoding of the current 'cognitive state' from measured brain activity (Haynes & Rees, 2006). Using this approach, it is possible to build classifiers that can distinguish between various cognitive behaviors, provided an adequate training set can be identified. Decoding routines have recently been applied to accurately decode meaning (i.e. simple thoughts) from neural activity (Mitchell, 2004; Mitchell, 2008).

2.4 Modeling

Perhaps the greatest challenges still remain in the domain of modeling - developing cognitive models based on how information flows, and is processed, across the brain, to simulate what the content of cognition will look like, based on neural activity. One approach that continues to gain momentum is to take existing cognitive models and link them to neural data. For example one of the better known cognitive modeling approaches is ACT-R (Anderson, 1996). ACT-R is an implementable theory of how human cognition works, based on the underlying assumption that knowledge is encoded from the environment and synthesized into 'cognition,' leading to behavior. Within the ACT-R framework, modules and buffers represent knowledge and processing of that knowledge, with cognition emerging through their activation. In its executable form, the timing and sequencing of model components is based on observed behaviors, and the output is typically timing and accuracy predictions.

Acknowledgments. The author wishes to thank Ms. D. Brumer for her critical review and comments.

References

1. Adams, J.A.: Some considerations in the design and use of dynamic flight simulators. Texas: Lackland Air Force Base (Air Force research report AFPTRC-TN-57-51) (1957)
2. Anderson, J.R.: ACT: A simple theory of complex cognition. American Psychologist 51, 355–365 (1996)
3. Bates, J.A.V.: Some characteristics of a human operator. Journal of the Institute of Electrical Engineering 94, 298–304 (1947)

4. Birmingham, H.P., Taylor, F.V.: A design philosophy for man-machine control systems. Proceedings of th e I.R.E. 42(12), 1748–1758 (1954)
5. Chapanais, A.: Theory and methods for analyzing errors in man-machine systems. Annals of the New York Academy of Sciences 51, 1179–1203 (1951)
6. Cooley, M.: Cognition, communication and interaction: Transdisciplinary perspectives on interactive technology. In: Gill, S.P. (ed.) On Human-Machine Symbiosis Human-Computer Interaction Series, pp. 457–485. Springer, Heidelberg (2007)
7. Cohn, J.V., Stripling, R., Kruse, A.: Investigating the transition from novice to expert. In: Schmorrow, D. (ed.) Foundations of Augmented Cognition, pp. 946–953. Lawrence Earlbaum Associates, Mahwah (2005)
8. Craik, K.J.W.: Theory of the human operator in control systems I: The operation of the human operator in control systems. British Journal of Psychology 38, 56–61 (1947/1948)
9. Craik, K.J.W.: Theory of the human operator in control systems II: Man as an element in a control system. British Journal of Psychology 38, 142–148 (1948)
10. Fitts, P.M.: The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology 47(6), 381–391 (1954)
11. Fitts, P.M., Peterson, J.R.: Information capacity of discrete motor responses. Journal of Experimental Psychology 67(2), 103–112 (1964)
12. Glimcher, P.W., Rustichini, A.: Neuroeconomics: The consilience of brain and decision. Science 306(5695), 447–452 (2004)
13. Haynes, J.-D., Rees, G.: Decoding mental states from brain activity in humans. Nature Reviews Neuroscience 7(7), 523–534 (2006)
14. Hick, W.E.: On the rate of gain of information. Quarterly Journal of Experimental Psychology 4, 11–26 (1952)
15. Hyman, R.: Stimulus information as a determinant of reaction time. Journal of Experimental Psychology 45, 188–196 (1953)
16. Junghöfer, M., Elbert, T., Tucker, D.M., Rockstroh, B.: Statistical control of artifacts in dense array EEG/MEG studies. Psychophysiology 37(4), 523–532 (2000)
17. Kelso, J.A.S.: Dynamic patterns: The self-organization of brain and behavior. MIT Press, Cambridge (1995)
18. Licklider, J.C.R.: Man-computer symbiosis. IEEE Transactions on Human Factors in Electronics HFE-1, 4–11 (1960)
19. Logothetis, N.K.: Neurophysiological investigation of the basis of the fMRI signal. Nature 412, 150 (2001)
20. McRuer, D.T., Jex, H.R.: A Review of Quasi Linear Pilot Models. IEEE Transactions on Human Factors in Electronics HFE-3, 231–249 (1967)
21. Mitchell, T., Hutchinson, R., Niculescu, R.S., Pereira, F., Wang, X., Just, M.A., Newman, S.D.: Learning to decode cognitive states from brain images. Machine Learning 57, 145–175 (2004)
22. Mitchell, T.M., Shinkareva, S.V., Carlson, A., Chang, K.-M., Malave, V.L., Mason, R.A., Just, M.A.: Predicting human brain activity associated with the meanings of nouns. Science 320, 1191–1195 (2008)
23. Mormann, F., Fell, J., Axmacher, N., Weber, B., Lehnertz, K., Elger, C.E., Fernandez, G.: Phase / amplitude reset and theta-gamma interaction in the human medial temporal lobe during a continuous word recognition memory task. Hippocampus 15, 890–900 (2005)
24. Morrison, J.G., Gluckman, J.P.: Definitions and prospective guidelines for the application of adaptive automation. In: Mouloua, M., Parasuraman, R. (eds.) Human Performance in Automated Systems: Current Research and Trends, pp. 256–263. Erlbaum, Hillsdale (1994)

25. Parasuraman, R., Riley, V.: Humans and automation: Use, misuse, disuse, abuse. Human Factors 39, 230–253 (1997)
26. Parsons, H.M.: Automation and the individual: Comprehensive and comparative views. Human Factors 27, 99–112 (1985)
27. Philiastides, M.G., Sajda, P.: EEG-informed fMRI reveals spatiotemporal characteristics of perceptual decision making. Journal of Neuroscience 27(48), 13082–13091 (2007)
28. Rouse, W.B.: Human-computer interaction in multitask situations. IEEE Transactions Systems, Man, and Cybernetics SMC-7, 293–300 (1977)
29. Rouse, W.B., Geddes, N.D., Curry, R.E.: An architecture for intelligent interfaces: Outline of an approach to supporting operators of complex systems. Human-Computer Interaction 3, 87–122 (1988)
30. Scerbo, M., Freeman, F., Mikulka, P.J., Parasuraman, R., Di Nocero, F., Lawrence III., J.P.: The efficacy of physiological measures for implementing adaptive technology. NASA TP-2001-211018, pp. 37–63. NASA Langley Research Center, Hampton (2001); Scerbo, M.W.: Theoretical perspectives on adaptive automation. In: Parasuraman, R., Mouloua, M. (eds.) Automation and Human Performance: Theory and Applications, pp. 37–63. Lawrence Erlbaum Associates, Mahwah (1996)
31. Schmorrow, D.D.: Foundations of augmented cognition. Earlbaum, Mahwah (2005)
32. Schmorrow, D.D., Stanney, K., Reeves, L.: Foundations of augmented cognition: Past, present & future. Strategic Analysis, Arlington (2007)
33. Stevens, S.S.: On the psychophysical law. Psychological Review 64(3), 153–181 (1957)
34. Weiner, E.L., Curry, R.E.: Flight deck automation: Promises and problems. Ergonomics 23, 995–1011 (1980)
35. Weiner, E.L.: Human Factors of Advanced Technology ("Glass Cockpit") Transport Aircraft. Moffett Field, CA: NASA – Ames Research Center (NASA Technical Report - 117528) (1989)
36. Woods, D.D.: Decomposing automation: Apparent simplicity, real complexity. In: Parasuraman, R., Mouloua, M. (eds.) Automation and Human Performance: Theory and Applications, pp. 3–18. Lawrence Erlbaum Associates, Mahwah (1996)
37. Young, L.R.: On adaptive manual control. IEEE Transactions on Man-Machine Systems MMS-10, 292–331 (1969)

The Influence of Culture on Memory

Angela H. Gutchess[1], Aliza J. Schwartz[1], and Ayşecan Boduroğlu[2]

[1] Brandeis University, 415 South Street, MS 062,
Waltham, MA 02454-9110, USA
[2] Boğaziçi Üniversitesi, Psikoloji Bölümü,
34342 Bebek, İstanbul Turkey
gutchess@brandeis.edu

Abstract. The study of cognition across cultures offers a useful approach to both identifying bottlenecks in information processing and suggesting culture-specific strategies to alleviate these limitations. The recent emphasis on applying cognitive neuroscience methods to the study of culture further aids in specifying which processes differ cross-culturally. By localizing cultural differences to distinct neural regions, the comparison of cultural groups helps to identify candidate information processing mechanisms that can be made more efficient with augmented cognition and highlights the unique solutions that will be required for different groups of information processors.

Keywords: cognition, culture, memory, strategies, fMRI.

1 Introduction

To achieve the goal of understanding the ways in which computational systems and devices can interface with human cognition, the field of augmented cognition pays particular attention to bottlenecks [1]. Bottlenecks in cognition include limitations in one's ability to simultaneously attend to multiple locations or channels of information, and impairments in the ability to encode, store, and retrieve from memory veridical, detailed representations of information. Because humans are limited information processors, necessarily there are trade-offs in what is attended to and remembered, with some information being prioritized at the expense of other information.

Several studies have examined the ways in which culture can, to some extent, explain individual differences in processes such as attention and memory. Overall, much of the research to date has compared Easterners (typically operationalized as individuals from China, Japan, or Korea) and Westerners (typically including Americans, Canadians, and Western Europeans). These studies suggest that Easterners prioritize information that is holistic, considering functional relationships between elements and relating the self to the group. Westerners, on the other hand, prioritize information about individual items, focusing on objects without regard to their contexts and conceptualizing of the self as an individual entity [e.g., 2, 3-9]. This paper will provide a selective review of that literature with an eye toward identifying mechanisms and their neural correlates that can be made optimized through augmented cognition, and it will also consider some challenges to that goal.

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 67–76, 2011.

2 Mechanisms of Cross-Cultural Differences in Cognition

What individuals from different cultures prioritize and attend to in their environments likely reflects differences in strategies, which may correspond to separable underlying neural mechanisms. That is, attending to information that is object-based or self-referent, rather than context-based or group-referent, reflects the strategic deployment of cognitive resources. The selection of some aspects of information over others has been shaped by one's cultural milieu, with certain ways of processing information encouraged and reinforced through one's development in a particular cultural context. For example, attention to focal objects or relationships is reinforced through socialization and language acquisition in childhood [as discussed by 3]. Cultural differences in preferences even emerge in different styles of parental interaction during play, with American mothers emphasizing the properties of objects and Asian mothers instructing about relationships [10]. The learned aspect of culture is emphasized further through studies suggesting that cultural differences emerge late enough to reflect the products of socialization [11].

We conceptualize of strategy differences across cultures as potentially occurring due to one of three different mechanisms: 1) Cultural differences could reflect the engagement of distinct cognitive *processes*, such that individuals evoke different strategies (e.g., categorical versus relational) or process different aspects of information (e.g., object versus context). 2) Cultural differences could emerge due to differences in the underlying *content* of what information is stored and correspondingly accessed by individuals in different cultures. 3) Cultural differences could represent varying degrees of *difficulty* across cultures, such that one task is more difficult in one culture than another and therefore requires a greater commitment of cognitive and neural resources. We believe that the first mechanism, cultural differences in engagement of cognitive processes, is the most consistent with the literature thus far, and we will devote the most space in this review to considering this possibility. However, we will also consider the other candidate mechanisms, their implications for the study of cross-cultural differences in the engagement of neural resources, and the potential interface with augmented cognition.

2.1 Cross-Cultural Differences in the Engagement of Cognitive Processes

Cross-cultural differences in the use of cognitive processes have received the most support in the literature to date in explaining cross-cultural strategy differences. One process differing across cultures is a preference for analytical versus holistic processing [e.g., 2, 3-6]. Evidence suggests that Westerners tend to focus on the details, pieces, and parts of information whereas Easterners process information in terms of its context and tend to relate information together. Within this framework, Easterners attend to and remember contextual information whereas Westerners focus on focal objects on their own, without regard to the context [3].

Some of the strongest evidence that cultures differ in the strategies they employ to process information stems from the study of memory. If individuals from different cultures engage distinct information processing strategies, they should then differ in their memory for disparate aspects of the information [see 12 for a discussion of these ideas]. The strategies used to process information upon first encounter should determine what information is preserved in memory and thus most readily accessed with retrieval cues at a later point in time. Cultural differences in memory for distinct aspects of information have been shown in memory for scenes. In one study, after viewing underwater scenes, participants described what they remembered from the animated vignettes. While Americans' descriptions focused predominantly on the prominent fish, the descriptions of Japanese participants included more details about the context [9]. The same pattern also emerged in a more controlled study of recognition memory, in which Japanese participants were more affected than Americans by changing background contexts. Their memory for objects was more impaired when the objects were against novel backgrounds, as opposed to the original ones [9]. The heightened attention to context extends to the perception of emotions in others, with the emotional expressions of other faces in a crowd coloring the interpretation of a central target face for Japanese more than for Westerners [13]. While we consider in the next section how it is difficult to separate differences across cultures in the *content* of memory from the *processes* themselves, we believe that process differences are more likely to explain the findings, accounting for the initial differences in content.

Further evidence that these differences emerge at the level of strategies comes from converging methods. By analyzing measures of where people look when first encountering information and the neural regions that are engaged when processing complex information, we see evidence for what processes and aspects of information are of primary importance to the perceiver. Consistent with behavioral indicators, eye-tracking measures reveal that Americans spend more time fixating on objects and fixate to them sooner than East Asians [14, but see 15]. Neuroimaging measures of brain activity, such as fMRI, indicate that Americans modulate object processing regions more than East Asians when viewing complex scenes, perhaps reflecting visual attention and retrieval of semantic knowledge about objects [6, 16]. By identifying the neural substrates of cultural differences, we find that there is more consistent evidence that cultural differences in the processing of complex scenes may occur due to differences in the processing of objects rather than contexts. These findings from eye-tracking and fMRI measures suggest that when first viewing pictures of scenes, Americans attend quickly to objects and process them in more detail than East Asians. In contrast, there is not evidence for greater neural activation in response to background contexts for East Asians, although they fixate on backgrounds more than Americans. This type of work provides initial clues into how the field of augmented cognition would need to differently account for the influence of culture. As stimuli and information in the environment recruit unique sets of cognitive processes in individuals from different cultures, augmented cognition and human-computer interaction devices would need to account for and build upon these distinct strategies.

It is important to note that cross-cultural differences in the processing of objects and contexts also may permeate several different levels of cognition, including some lower-level attentional and perceptual processes. At the level of attention, East Asians may show more globally distributed attention than Westerners, who may attend more locally. This is demonstrated through responses to spatial configurations of shapes, with East Asians performing better than Americans when configurations were preserved but expanded in space, and worse than Americans when configurations were shrunk in space [17]. There is also evidence that priming an interdependent view of the self leads East Asians to show a stronger Flanker effect than priming with independence, consistent with the expected findings for East Asians versus Westerners (who differ on interdependence/independence) [18]. This effect seems to be linked to early visual perceptual/attentional effects, based on the modulation of the P1 effect, as assessed with ERPs [19]. Furthermore, East Asians perform more poorly on a functional field of view task than Americans [20]. These findings might indicate that East Asians distribute attention more broadly in space than Americans. It is also important to keep in mind that this might impact the resolution of their representations, such that attending more broadly necessitates a reduction in the quality of the representations. Preliminary support for this idea comes from the finding that in the useful field of view task, East Asians perform significantly lower than Westerners. Furthermore, their errors indicate that they identify "random" locations as targets more than Americans, who tend to make errors with neighboring positions [20].

Cultures also differ in their attention to categorical information. Whereas Westerners tend to focus on taxonomic categories, East Asians tend to emphasize functional relationships [e.g., 4, 5, 21, 22-25]. One paradigm employed to address this idea presents participants with triplets of words, asks participants to determine which two belong together, and provide a justification for their pairing. In a set of words such as "seagull," "squirrel," and "nut," Westerners tend to pair "seagull" and "squirrel" together because both belong to the category "animals," whereas East Asians tend to pair "squirrel" and "nut" together because the one serves as a food source for the other [22, 24]. The difference in the types of explanations is consistent with our suggestion that cultures differ in the strategies and cognitive processes they adopt, for it seems that different information is salient and useful for organizing information for each cultural group.

Some of our prior work investigated cross-cultural differences in memory for categorical information. A classic experimental paradigm [26] presents participants with words drawn from several different categories (e.g., "apple", "orange", "banana" as exemplars from the "fruit" category, and "train", "bus", "car" as exemplars from the "modes of transportation" category) and later asks them to list all of the words they remember. When analyzing the order in which the words were recalled, one finds that people tend to spontaneously organize the words by category, even though they were originally presented in an intermixed order. This finding of categorical clustering is particularly useful in identifying strategy differences, which can be separated from the amount of information remembered. Our results show that even though American and Chinese participants may recall similar amounts of information, older Chinese tend to use categorical clustering to organize their recall less than older Americans, indicating that categories provide a more useful strategy for organizing and retrieving information from memory for Americans than Chinese [21].

Recent research also identifies cultural differences in the neural regions implicated in resolving conflict when sorting triplets of words. When instructed to select a particular type of word pair, either categorical (e.g., seagull-squirrel) or relational (e.g., squirrel-nut), on each timed trial from a set of word triplets, East Asians activated more executive control processes, reflected by activity in a frontal-parietal network, while Americans activated more temporal regions, possibly reflecting conflict in the semantic content of information. Because accuracy in selecting the appropriate pair was equivalent across cultures, we interpreted differences in neural activation patterns to reflect cultural differences in conflict resolution, with East Asians adopting more domain-general processes and Americans adopting more domain-specific processes [25]. This pattern of results further indicates that the augmented cognition solutions could be very distinct for one culture compared to another, not just in terms of the types of processes engaged [6, as was the case for cultural differences in object processing, e.g., 16], but also in their implications to generalize across other functions (e.g., domain general vs. domain specific).

2.2 Cross-Cultural Differences in the Content of Cognition

Stores of semantic knowledge undoubtedly differ across cultures based on the types of experiences individuals have with their environments and what is deemed important within their culture. Educational systems may reflect these differences, for example, Chinese education is virtue-oriented and Western education is mind-oriented [27], which would maintain and perhaps extend these differences across cultures over time. Typical experiments assess the content of knowledge by asking participants to list exemplars belonging to different categories or to name pictures. There can be dramatically different responses across cultures, such as participants listing different flowers or animals based on what is native to their environment [e.g., 28]. The process of naming pictures also reveals differences in what items are familiar across cultures, as well as the specificity with which they are known. For example, using the basic level name "dog" compared to a more abstract category name such as "animal" or a more specific, detailed name such as "cocker spaniel" indicates the familiarity with which the object is known, and the level of distinction with which it is helpful to know about the concept [29-31]. Despite these differences in content, it is difficult to distinguish the *content* from the *processes* themselves. For example, we discussed attending more to objects versus contexts in the above section as representing a cultural difference in the cognitive processes engaged when encountering complex information in the environment. However, these initial differences in processing information such that Americans attend to objects and devote additional resources to processing information about their semantic and perceptual properties would then lead to cultural differences in the content of what is stored in episodic memory. While cultural differences in both content and process likely reinforce each other over time, we believe that strategy differences are what lead to eventual differences in content. However, it is admittedly difficult to distinguish strategic processes from content.

It is also notoriously difficult to study individual pieces of knowledge. Differences that have been identified in semantic content are largely based on differences in broader processes. For example, different classes of objects are associated with different properties and types of knowledge, such as the attention to functional and motor properties for tools, as opposed to the focus on perceptual properties that are important for animals [32]. Neuroimaging methods such as fMRI and ERP have traditionally relied on averaging large numbers of trials together in order to have stable and measurable signal, precluding the possibility of analyzing single trials. However, recent developments in methods, such as multivariate pattern analysis, hold promise for identifying differences at the level of the trial. This method can detect distributed patterns of neural activity, rather than being constrained by the requirement to identify activations in one focal area, and can be highly sensitive to detecting differences across conditions and trials [33].

In terms of what this means for augmented cognition approaches, it would be difficult to localize a distinct neural basis to target with neural prosthetics or other interventions, in that the pattern of neural activity is distributed. It also is challenging to separate content from process, which may make it more fruitful to focus efforts on addressing cultural differences in cognitive processes, which are likely to determine cultural differences in content through the selection and reinforcement of different types of information in the environment.

2.3 Cross-Cultural Differences in Task Difficulty

Some recent research has highlighted that one must be careful to distinguish the strategies and processes evoked by a task from the *difficulty* of that task. While it may be tempting to interpret any difficulty in regional neural activity as indicating unique processes recruited in one group compared to the other, it is necessary to interpret the activation in terms of the behavioral performance on the task, as well as in terms of the patterns of neural activity across cultures in other task conditions.

A prime example of this distinction stems from research on cultural differences in the way a focal line is interpreted relative to its frame. In line with cultural differences in attention to contexts versus objects, behavioral studies indicate cross-cultural differences in perceptual judgments of the orientation of a rod relative to its frame. Americans found it easier to ignore the frame and judge the verticality of a rod alone, whereas East Asians were more affected by the position of the frame, which could interfere with their absolute judgment of the verticality of the rod [34]. A modified version of the task, which allowed both the frame and rod to vary, revealed advantages for each culture for different types of problems. Americans were more accurate at reproducing a line independent of its frame whereas Japanese were more accurate when reproducing the line in proportion to the frame [35, but see 36]). Using fMRI to investigate the neural correlates of this task, Hedden and colleagues [37] found that the same frontal-parietal (attentional) network was implicated for Asians and Americans when each was performing their non-preferred tasks. That is, when East Asians made absolute judgments (judging the size of a line compared to a standard one while ignoring the frame), the attentional network was more active than it was for relative judgments (matching the size of a line relative to its frame, compared to a standard template). The finding was reversed for Americans, who engaged the attentional network more for relative than absolute judgments. Even though the behavioral performance was matched across cultures in this study, the neural activity was a more sensitive proxy of which information required more engagement of attentional resources.

When tasks evoke cultural differences due to greater demands for attentional or cognitive resources in one culture compared to another, distinct augmented cognition solutions would be necessary. Rather than demanding unique solutions and devices with which to overcome resource limitations or improve the efficiency of information processing, differing resource demands would seem to require the same approach but applied at different points in time, or tasks, across cultures. It will also be important to determine when tasks require a ramping up or down of the same attentional resources, compared to situations in which qualitatively different processes are recruited to aid information processing. For example, as load increases for working memory tasks such that people are asked to keep 3, 4, or 5 pieces of information in mind, dorsolateral prefrontal cortex is recruited to a greater extent for higher memory loads. In some cases, older adults may recruit these additional resources at lower levels of difficulty than young, and under-recruit the region at high levels of difficulty, compared to young [38]. However, people may adapt to greater task demands by approaching the task in a wildly different way than they did at lower levels of task difficulty, reflecting distinct patterns of neural engagement across the age groups. The literature on the cognitive neuroscience of aging has grappled with this challenge, attempting to determine whether the recruitment of additional neural regions by older adults in many studies reflects compensation, in an attempt to harness additional resources to complete challenging cognitive tasks [39].

Studies that modulate the level of difficulty, systematically increasing and decreasing the demands, will be most helpful in determining when cultural differences reflect differences in attentional difficulty, as opposed to differences in the strategies engaged. They will also be helpful in investigating whether the same processes hold for participants across varying levels of difficulty and ability. It also may be important to consider effects of difficulty over time. While initially cultures may differ in task performance or neural activity, they may eventually converge as they identify or adapt to the optimal strategy. This is consistent with our suggestion that young adults may be able to adopt culturally less-preferred strategies, whereas older adults lack the cognitive resources in order to do so [21]. With enough aid or differential amounts of practice, an augmented cognition system may similarly enhance performance across cultures, but without the initial investment in set-up and training, it may not equally benefit both cultures. In the case of a system that one will interact with daily, people may be able to learn to use it efficiently and overcome the initial costs with training. However, in the case of a system that would be used occasionally, then the match between the culturally-preferred processes and augmented cognition program may pose a larger constraint.

3 Conclusions

While we have discussed ways in which bottlenecks may vary across cultures and some challenges to applying augmented cognition solutions to address these inefficiencies, it is also important to consider the ways in which bottlenecks themselves reveal how information processing systems operate. Some of our current work investigates the ways in which memory errors may differ across cultures, offering a window into what information is distorted or lost in a given culture, reflecting its values [this framework is further discussed in 12]. In this case, it may be that the content of what is remembered per se is less important than what it reveals about the overarching values and goals that are reinforced by a culture (e.g., prioritization of context or the self). Thus, while a faulty memory system could be "patched" with augmented cognition solutions, this would not change the underlying cultural differences that had been reflected in the pattern of memory errors.

Schacter [40] raises an even larger potential concern for augmented cognition in his review of seven classes of common memory errors, concluding that these errors "are by-products of otherwise adaptive features of memory" (p. 182). That is, in addressing flaws and apparent inefficiencies in the system, augmented cognition could create additional challenges for memory. While forgetting information can be an inconvenient failing of memory, for example, remembering every piece of information ever encoded would lead to a massive store of information that would need to be searched in order to retrieve the desired information at the correct point in time. Or addressing the tendency for information in memory to be distorted to fall in line with one's existing knowledge or beliefs might reduce the usefulness of organizing principles that are based on previous experience. Of course, fully optimizing a system would address undesirable outcomes such as these, but the rich function of memory in sustaining a sense of self that varies across individuals and with culture will require complex and elegant augmented cognition solutions.

Acknowledgments. The authors gratefully acknowledge support from a Fulbright Scholar Award (awarded to A.H.G.).

References

1. Schmorrow, D., McBride, D.: Introduction. Int. J. Hum-Comp. Int. 17, 127–130 (2004)
2. Na, J., Grossmann, I., Varnum, M.E.W., Kitayama, S., Gonzalez, R., Nisbett, R.E.: Cultural differences are not always reducible to individual differences. P. Natl. Acad. Sci. USA 107, 6192–6197 (2010)
3. Nisbett, R.E., Masuda, T.: Culture and point of view. P. Natl. Acad. Sci. USA 100, 11163–11170 (2003)
4. Nisbett, R.E., Peng, K.P., Choi, I., Norenzayan, A.: Culture and systems of thought: Holistic versus analytic cognition. Psychol. Rev. 108, 291–310 (2001)
5. Nisbett, R.E.: The geography of thought: How Asians and Westerners think differently... and why. Free Press, New York (2003)
6. Gutchess, A.H., Welsh, R.C., Boduroglu, A., Park, D.C.: Cultural differences in neural function associated with object processing. Cogn. Affect Behav. Neurosci. 6, 102–109 (2006)
7. Kitayama, S., Uskul, A.K.: Culture, mind, and brain: Current evidence and future directions. Annu. Rev. Psychol. 62, 419–449 (2011)
8. Markus, H.R., Kitayama, S.: Culture and the self: Implications for cognition, emotion, & motivation. Psychol. Rev. 98, 224–253 (1991)
9. Masuda, T., Nisbett, R.E.: Attending holistically versus analytically: Comparing the context sensitivity of Japanese and Americans. J. Pers. Soc. Psychol. 81, 922–934 (2001)
10. Fernald, A., Morikawa, H.: Common themes and cultural variations in Japanese and American mothers' speech to infants. Child Dev. 64, 637–656 (1993)
11. Duffy, S., Toriyama, R., Itakura, S., Kitayama, S.: Development of cultural strategies of attention in North American and Japanese children. J. Exp. Child Psychol. 102, 351–359 (2009)
12. Gutchess, A.H., Indeck, A.: Cultural influences on memory. Prog. Brain Res. 178, 137–150 (2009)

13. Masuda, T., Ellsworth, P.C., Mesquita, B., Leu, J., Tanida, S., De Veerdonk, E.V.: Placing the face in context: Cultural differences in the perception of facial emotion. J. Pers. Soc. Psychol. 94, 365–381 (2008)
14. Chua, H.F., Boland, J.E., Nisbett, R.E.: Cultural variation in eye movements during scene perception. P. Natl. Acad. Sci. USA 102, 12629–12633 (2005)
15. Evans, K., Rotello, C.M., Li, X., Rayner, K.: Scene perception and memory revealed by eye movements and receiver-operating characteristic analyses: Does a cultural difference truly exist? Q J. Exp. Psychol. 62, 276–285 (2009)
16. Goh, J.O., Chee, M.W., Tan, J.C., Venkatraman, V., Hebrank, A., Leshikar, E.D., Jenkins, L., Sutton, B.P., Gutchess, A.H., Park, D.C.: Age and culture modulate object processing and object-scene binding in the ventral visual area. Cogn. Affect Behav. Neurosci. 7, 44–52 (2007)
17. Boduroglu, A., Shah, P., Nisbett, R.E.: Cultural Differences in Allocation of Attention in Visual Information Processing. J. Cross Cult. Psychol. 40, 349–360 (2009)
18. Lin, Z., Han, S.: Self-construal priming modulates the scope of visual attention. Q J. Exp. Psychol. 62, 802–813 (2009)
19. Lin, Z., Lin, Y., Han, S.: Self-construal priming modulates visual activity underlying global/local perception. Biol. Psychol. 77, 93–97 (2008)
20. Boduroglu, A., Lan, X., Shah, P.: Cultural differences in functional field of view, in Annual Meeting of the Psychonomics Society, Chicago, IL (2008)
21. Gutchess, A.H., Yoon, C., Luo, T., Feinberg, F., Hedden, T., Jing, Q., Nisbett, R.E., Park, D.C.: Categorical organization in free recall across culture and age. Gerontology 52, 314–323 (2006)
22. Ji, L.J., Zhang, Z.Y., Nisbett, R.E.: Is it culture or is it language? Examination of language effects in cross-cultural research on categorization. J. Pers. Soc. Psychol. 87, 57–65 (2004)
23. Unsworth, S.J., Sears, C.R., Pexman, P.M.: Cultural influences on categorization processes. J. Cross Cult. Psychol. 36, 662–688 (2005)
24. Chiu, L.H.: A cross-cultural comparison of cognitive styles in Chinese and American children. Int. J. Psychol. 7, 235–242 (1972)
25. Gutchess, A.H., Hedden, T., Ketay, S., Aron, A., Gabrieli, J.D.E.: Neural differences in the processing of semantic relationships across cultures. Soc. Cogn. Affect Neurosci. 5, 254–263 (2010)
26. Bousfield, W.A.: The occurrence of clustering in the recall of randomly arranged associates. J. Gen. Psychol. 49, 229–240 (1953)
27. Li, J.: Mind or Virtue. Curr. Dir. Psychol. Sci. 14, 190–194 (2005)
28. Yoon, C., Feinberg, F., Hu, P., Gutchess, A.H., Hedden, T., Chen, H.Y., Jing, Q., Cui, Y., Park, D.C.: Category norms as a function of culture and age: comparisons of item responses to 105 categories by american and chinese adults. Psychol. Aging 19, 379–393 (2004)
29. Yoon, C., Feinberg, F., Gutchess, A.H.: Pictorial naming specificity across ages and cultures: a latent class analysis of picture norms for younger and older Americans and Chinese. Gerontology 52, 295–305 (2006)
30. Yoon, C., Feinberg, F., Luo, T., Hedden, T., Gutchess, A.H., Chen, H.Y., Mikels, J.A., Jiao, S., Park, D.C.: A cross-culturally standardized set of pictures for younger and older adults: American and Chinese norms for name agreement, concept agreement, and familiarity. Behav. Res. Methods Instrum. Comput. 36, 639–649 (2004)
31. Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., Herron, D., Lu, C.-C., Pechmann, T., Pléh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., Orozco-Figueroa, A., Tzeng, A., Tzeng, O.: Timed picture naming in seven languages. Psychon. B Rev. 10, 344–380 (2003)

32. Martin, A., Chao, L.L.: Semantic memory and the brain: structure and processes. Curr. Opin. Neurobiol. 11, 194–201 (2001)
33. Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V.: Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn. Sci. 10, 424–430 (2006)
34. Ji, L.J., Peng, K.P., Nisbett, R.E.: Culture, control, and perception of relationships in the environment. J. Pers. Soc. Psychol. 78, 943–955 (2000)
35. Kitayama, S., Duffy, S., Kawamura, T., Larsen, J.T.: Perceiving an object and its context in different cultures: a cultural look at new look. Psychol. Sci. 14, 201–206 (2003)
36. Zhou, J.W., Gotch, C., Zhou, Y.F., Liu, Z.L.: Perceiving an object in its context-is the context cultural or perceptual? J. Vision 8 (2008)
37. Hedden, T., Ketay, S., Aron, A., Markus, H.R., Gabrieli, J.D.E.: Cultural influences on neural substrates of attentional control. Psychol. Sci. 19, 12–17 (2008)
38. Cappell, K.A., Gmeindl, L., Reuter-Lorenz, P.A.: Age differences in prefontal recruitment during verbal working memory maintenance depend on memory load. Cortex 46, 462–473 (2010)
39. Park, D.C., Reuter-Lorenz, P.A.: The adaptive brain: Aging and neurocognitive scaffolding. Annu. Rev. Psychol. 60, 173–196 (2009)
40. Schacter, D.L.: The seven sins of memory - Insights from psychology and cognitive neuroscience. Am. Psychol. 54, 182–203 (1999)

Using Computational Modeling to Assess Use of Cognitive Strategies

Michael J. Haass and Laura E. Matzen

Sandia National Laboratories,
P.O. Box 5800, MS 1188,
Albuquerque, NM, USA 87185
{mjhaass,lematze}@sandia.gov

Abstract. Although there are many strategies and techniques that can improve memory, cognitive biases generally lead people to choose suboptimal memory strategies. In this study, participants were asked to memorize words while their brain activity was recorded using electroencephalography (EEG). The participants' memory performance and EEG data revealed that a self-testing (retrieval practice) strategy could improve memory. The majority of the participants did not use self-testing, but computational modeling revealed that a subset of the participants had brain activity that was consistent with this optimal strategy. We developed a model that characterized the brain activity associated with passive study and with explicit memory testing. We used that model to predict which participants adopted a self-testing strategy, and then evaluated the behavioral performance of those participants. This analysis revealed that, as predicted, the participants whose brain activity was consistent with a self-testing strategy had better memory performance at test.

Keywords: Memory, computational modeling, electroencephalography.

1 Introduction

Memory underlies and supports all forms of high-level cognition and accurate memory is essential to good decision making. However, human memory is extremely fallible. Although there are many factors that can improve memory performance, such as selecting appropriate memory strategies, people are poor at predicting what they will or will not remember and tend to choose strategies that are suboptimal or counterproductive. In our research, we are investigating patterns of brain activity associated with good and poor memory performance. We are examining methods for improving human performance by identifying cases where learners are using suboptimal memory strategies. Through this effort, we hope to lay the foundation for closing the loop between recording brain activity and using those recordings to augment performance.

As a part of this effort, one of our goals is to create a model of brain activity that can be used in a predictive fashion. Using brain activity recorded from participants who tried to memorize words under a variety of study and test conditions, we selected two conditions where the brain's response to stimuli should be similar across all

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 77–86, 2011.

participants. We used those two conditions to develop a computation model and then tested the model on a third condition in which participants' brain activity should depend on their choice of study strategies. The model was used to predict which participants were using a more effective memory strategy. We then tested the predictions of the model by assessing the participants' behavioral memory performance.

1.1 Metamemory and Memory Strategies

The term *metamemory* refers to a person's judgments about the state of his or her own memory. Successful encoding and retrieval of information requires a number of metamemory decisions, such as deciding what information is worth remembering, what strategies should be used to encode the information, and whether or not information retrieved from memory is accurate. People typically develop metamemory skills over time, through experience with different kinds of learning situations. For example, after practice with sequences of study and test questions, people tend to get better at predicting which items they will remember later and which items need additional study [1]. Through experience, people learn memory strategies such as spending more time studying items that seem difficult to remember or using different study strategies depending on when and how the information will need to be remembered. However, the strategies that learners develop are often affected by cognitive biases and may not be optimal. Numerous studies have shown that people often fail to use appropriate memory strategies [2,3,4,5,6].

One memory strategy that can improve performance is self-testing, or retrieval practice [7]. A common example of retrieval practice is studying with flashcards. If a language student studies new vocabulary by quizzing herself with flashcards, she will be more likely to remember the new words than if she skimmed over the words and their definitions in a textbook. Retrieval practice is beneficial because it gives learners experience with retrieving the needed information from memory. It also provides learners with a more accurate sense of what they do and do not remember. The effectiveness of retrieval practice increases as the practice becomes more difficult [3].

Although retrieval practice is a highly effective strategy, it is not a strategy that learners are likely to adopt on their own. Studying with this strategy can be frustrating because learners feel that they are performing poorly and progressing slowly. In reality, they are developing accurate assessments of how well they have learned the material. However, learners tend to prefer study strategies that make them feel successful at the time of study, even when those strategies are less effective in the long run [2,3].

1.2 Event-Related Potentials

When a person's brain activity is recorded using electroencephalography (EEG), different patterns of activity emerge for passive study and for active retrieval of information from memory. In EEG research, a participant's brain activity is recorded using sensors placed on his or her scalp. The EEG data provide an ongoing record of the brain's electrical activity with very high temporal resolution. To separate out the brain activity related to a particular type of processing, the EEG data are time-locked to the presentation of events of interest and events of the same type are averaged together. The averaging process should average out any ongoing processing that was not related to the experimental stimuli, leaving only the activity that was elicited by the events of interest. These averaged waveforms are called event-related potentials (ERPs).

Researchers have mapped the relationships between different ERP waveforms and different types of processing in the brain. The ERP component of interest in the present study is the late positive component (LPC). The LPC is thought to be related to explicit processing, such as the process of deliberately searching memory for a particular piece of information [8, 9, 10, 11, 12].

When a learner is presented with an item to study, the LPC elicited by that item will be small. However, when the learner is tested on an item and has to retrieve it from memory, the presentation of that item will elicit a large LPC. In the absence of an explicit memory test, a participant's self-induced retrieval practice should also produce a larger LPC. It is likely that use of retrieval practice as a memory strategy will be reflected in participants' brain activity during study.

1.3 Modeling Event-Related Potentials

The EEG data selected for modeling was taken from a study in which participants were presented with a list of words and asked to remember them for a later memory test. Some of the words were studied once, some words were studied twice, and some were studied once and then quizzed once during the study session. All of the words appeared again on a subsequent memory test, intermixed with an equal number of new words. We hypothesized that participants would have the worst memory for the words that were studied only once and the best memory for the words that were quizzed during the study sessions. The quizzes provide an opportunity for retrieval practice that should benefit subsequent memory performance.

For the words that were studied twice but not quizzed, we hypothesized that some participants would engage in retrieval practice on their own. Even though the words were not explicitly tested, participants might recognize them as previously studied words and retrieve the first presentation of the word from memory. This self-testing should benefit subsequent memory performance much like explicit testing. Since, as discussed above, most people are unlikely to adopt a strategy such as retrieval practice on their own, we expected that the average performance across all participants would be lower for the twice-studied items than for the quizzed items. However, we expected that a subset of the participants would use more effective memory strategies and would perform better on this condition than their peers.

The design of the EEG experiment allowed us to model each participant's brain activity in two "known" conditions: the first presentation of each studied word, which should not elicit an LPC, and the words that were quizzed during the study block, which should elicit a large LPC. We applied the model to ERPs from an unknown condition, the second presentation of repeated study words. The words in that condition should elicit an LPC only for the participants who engaged in retrieval practice. We used the model to classify the ERPs from the unknown condition as being more like passively studied words or more like explicitly tested words. We then tested the predictive power of the model by comparing the subsequent memory performance for participants in those two groups.

2 Experimental Methods

Participants. Twenty-four University of Illinois students participated in this study and were paid for their participation. Half of the participants were male and half were female. The average age of the participants was 21.

Materials. The materials used in the experiment consisted of 320 common nouns that served as study items, and 320 nouns that were matched in terms of length and frequency and served as new items at test. The average frequency was 57.6 for the study items and 50.9 for the new items; the average word length was 4.6 letters for both sets of words (frequency data was taken from the Kucera and Francis, 1967; norms included in Balota et al., 2002; a frequency value of zero was assumed for items not appearing in the database).

The study words were divided into eight counterbalanced lists. The experimental lists were subdivided into four study blocks and four test blocks. Each study block contained 80 of the experimental items. Of those items, 20 were studied once, 20 were studied twice, 20 were studied and then tested within the block, and 20 were paired with a synonym. For the items that were studied twice or studied and then tested, half of the items were repeated at a short lag, defined as one intervening item, and half were repeated at a long lag, defined as nine intervening items. For the items that were paired with synonyms, half of the synonyms were presented at a short lag and half were presented at a long lag. In addition, half of the synonym items were tested at each lag.

Each study block was followed by a test block in which all of the nouns from the block were re-tested, intermixed with an equal number of new, unstudied items.

Procedure. The participants were instructed that they would be tested on their memory for a list of study words. They were not given any information about different types of memory strategies and were not asked to use a particular memory strategy. As discussed above, the study list was broken into four parts in order to make the task easier for the participants. Each study block contained a total of 140 study words and each test block contained a total of 160 test words.

Throughout the experiment, there was a white fixation cross in the center of the computer screen. The participants were asked to keep their eyes on the fixation cross at all times during the experiment. All of the study words were presented immediately above the fixation cross in white 38-point Helvetica font on a black background. Within the study blocks, each word was preceded by a pound symbol (#) that was presented above the fixation cross for one second. Participants were instructed that they could blink or move their eyes while the pound symbol was on the screen, but that when it disappeared they should refrain from blinking and prepare to see the next study word. For the tested words, the pound symbol was red, indicating that the next word would be tested. For the words that were only studied, the pound symbol was white. The study word was presented 500 ms after the pound symbol disappeared and remained on the screen for one second. The tested words were followed by a red question mark that remained on the screen until the participants pressed a response button to indicate whether or not that word had appeared earlier in the study block. The same test procedure was used in the test blocks that followed each study block. In the test blocks, all of the words from the study block were tested or retested, intermixed with an equal number of new words. The participants took short breaks before starting each new study block in order to reduce interference from the preceding blocks.

The electroencephalogram (EEG) was recorded from 26 silver/silver-chloride electrodes embedded in a geodesic arrangement in an elastic cap (EASY-cap). Five additional free electrodes were placed on the left and right mastoids, on the outer canthus of each eye, and below the left eye. The three free electrodes near the eyes were used to record blinks and horizontal eye movements (vertical and horizontal EOG). The scalp electrodes were referenced on-line to the left mastoid. Following the experiment, the scalp electrodes were re-referenced off-line to an average of the left and right mastoids. All of the electrodes were tested before recording begins to ensure that their impedance was below 3 KOhms. During the experiment, the EEG from all electrodes was amplified through a bandpass filter of 0.02-100 Hz and recorded at a sampling rate of 250 Hz.

ERPs were computed at each electrode for each experimental condition by averaging the EEG data from 100 ms before the onset of a word until 920 ms after word onset. Trials containing blinks were corrected using the blink correction procedure described by Dale (1994) and trials containing artifacts such as excessive eye movement, signal drift or muscle activity were excluded from the averages. The mean amplitude of the ERPs within time windows of interest was calculated using data digitally filtered off-line using a bandpass filter of 0.2 to 20 Hz.

3 Experimental Results

Behavioral Results. Memory accuracy was assessed using the percentage of correct answers on the memory tests. Only the data relevant to the computational model will be discussed here. On average, participants were 39% correct for words that were studied only once, 50% correct for items that were studied twice with a long lag between the repetitions, and 66% correct for items that were studied once and quizzed at a long lag during the study block. These results were consistent with the prediction that retrieval practice during study would benefit subsequent memory performance. The difference in performance between the twice-studied words and the quizzed words also supports the hypothesis that most participants would not use retrieval practice when presented with repeated study words.

ERP Results. The LPC was measured by computing the mean amplitude of the ERPs in a time window from 500-900 ms post stimulus onset. Repeated measures ANOVAs were used to test the results, with degrees of freedom adjusted using the Greenhouse-Geisser correction. All effects are significant at or above the $p = 0.05$ level unless otherwise specified.

The LPC was significantly larger for the words that were quizzed during the study block than for those that were not, as shown in Figure 1. As predicted, this indicates that participants actively searched their memory for the words that were explicitly quizzed. However, for the words that were studied twice, most (if not all) of the participants studied the words passively. They did not search their memory to retrieve the previous presentation of the words, so the second presentation of the words did not elicit and LPC.

Although the majority of the participants did not engage in retrieval practice when they were not explicitly tested, we developed a model to identify whether or not there were any subgroups of participants who did employ that strategy.

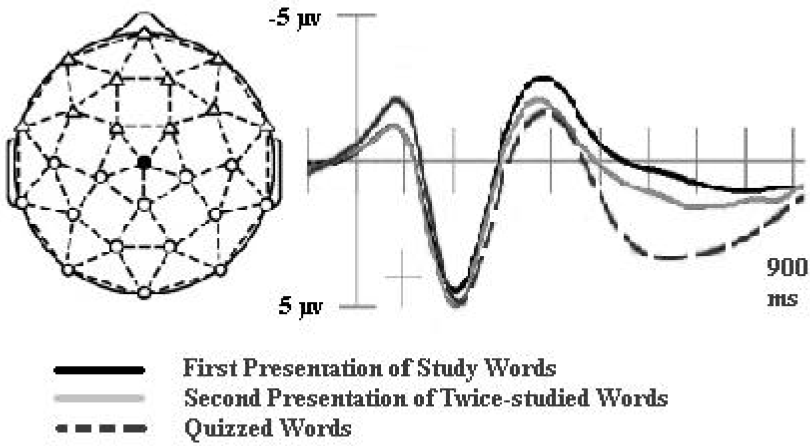


Fig. 1. Grand average ERPs to first presentation of studied words (black line), second presentation of twice-studied words (gray line), and quizzed words (dotted line). ERPs are shown at the midline central (MiCe) electrode.

4 Computational Modeling

Our goal was to construct a computational model that would classify ERPs elicited by the words in the twice-studied condition based on the brain activity associated with a particular study strategy (retrieval practice or passive study). This was achieved by constructing a naive Bayes classifier trained on the known study and test conditions and applying this classifier to the unknown ERPs. A significant challenge faced when constructing computational models from EEG signals is the low signal-to-noise ratio due to the presence of simultaneously recorded brain activity that is unrelated to the event of interest. This is often addressed by averaging all single-trial EEG recordings to form a grand average ERP. However this approach removes most of the trial-to-trial variability and can result in the formation of a classifier that is not robust to variances present in the ERPs from the "unknown" condition.

To overcome these obstacles, we developed an approach that better balances variability and signal averaging. Our approach combines ensembling classification results from multiple models and randomized signal averaging of individual trial ERPs. Randomized signal averaging was accomplished using an n-choose-k approach to create a new set of ERPs for use in the classifier training step. We examined maximized signal averaging by using k=39 to select and average single trial EEG recordings in a time window from 100 ms pre-stimulus to 900 ms post-stimulus to create 40 ERP samples for each of the two known conditions (study and test). For the study condition (the first presentation of all studied words), there were 278 single trial EEG recordings available from which to choose and for the test condition (the words that were quizzed during the study block) there were 40 single trial EEG recordings. The resulting ERP samples were then transformed via principal component analysis and the scores of the first five principal components were used as an uncorrelated feature set to train a naive Bayes classifier. The classifier was implemented by using MATLAB's [13] classify function provided in the Statistical Toolbox with the "diaglinear" discriminant function. This process was then repeated 50 times using a new random seed to randomize the single trial EEG recordings chosen for signal averaging from the n-choose-k trial selection process. In this way, each model was exposed to different signal averaging in the unknown condition ERP samples while maintaining a balanced number of training examples across the two known conditions.

5 Modeling Results

The performance of the classifier was estimated using sample-out cross validation. For this work, single trial data was available from twenty-three of the participating subjects.

The mean area under the receiver-operator curve (AUC) for sample-out cross validation over all models and all subjects was 0.99. The standard deviation of the mean sample-out cross validation AUC for each subject was 0.01. These cross validation results provide confidence that the feature extraction and classification methods are well suited to model the brain activity related to passive study or retrieval practice strategies.

For classification of the unknown ERPs, a full model was constructed with all samples from each of the 50 randomly constructed training sets described in section 4. This model was then used to classify the unknown ERPs as belonging to the study or retrieval groups. Examination of the number of models classifying the unknown ERPs as belonging to the study group identified eighteen subjects whose brain activity was consistent with their previously used study strategy. For this group of eighteen subjects, more than 97% of the models for each subject identified the unknown ERPs as belonging to the study class. Another group of five of the subjects exhibited brain activity that was consistent with the retrieval practice elicited by the quizzed words. The number of models identifying the unknown ERPs as belonging to the test class varied with subject and ranged from 22% to 80% of the 50 models constructed for each subject (Table 1). This variation is indicative of individual differences and may indicate that the retrieval practice strategy was employed with different frequency by each subject.

Table 1. Percentage of models indicating a retrieval practice strategy for each subject

Subject	Percentage of models indicating retrieval strategy	Percentage of twice-studied words remembered at test
27	80%	73%
15	78%	78%
2	54%	58%
12	52%	88%
3	22%	63%
22	2%	33%
5	0%	43%
7	0%	18%
8	0%	13%
9	0%	58%
10	0%	45%
11	0%	58%
13	0%	45%
14	0%	80%
16	0%	68%
18	0%	28%
19	0%	38%
20	0%	63%
21	0%	58%
26	0%	38%
28	0%	23%
29	0%	55%
30	0%	58%

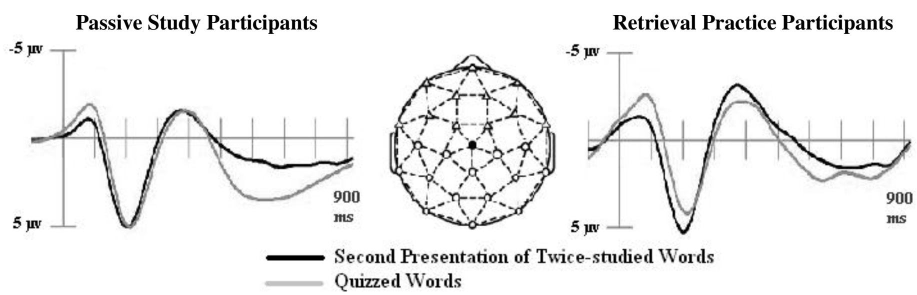


Fig. 2. Grand average ERPs to the unknown condition, the *second presentation of twice-studied words (black line)*, and the test condition, the *quizzed words (gray line)*. ERPs are shown at the midline central (MiCe) electrode. The participants whose brain activity was consistent with passive study in the unknown condition are shown on the left and the participants whose brain activity was consistent with a retrieval practice strategy are shown on the right.

To test the model's classification performance, we compared the behavioral memory performance across the two groups of participants. As predicted, the participants whose brain activity in the unknown condition was consistent with their brain activity in the retrieval practice condition had better memory for the twice-studied items than participants whose brain activity was consistent with passive study. On average, the participants in the former group correctly recognized 28.6 out of 40 words (71.5%) from the twice-studied condition, while the participants in the latter group correctly recognized 18.1 out of 40 words (45.2%). Welch's t-test showed that the performance of the two groups was significantly different [$t(9.4) = 3.82$, $p < 0.01$]. Figure 2 shows the grand average ERPs for the unknown condition and the test condition for the two groups.

6 Discussion

The results of this experiment indicate that ERPs elicited under known conditions can be modeled and used to classify ERPs from an unknown condition. In this experiment, the known conditions included a passive study condition and a condition in which participants were quizzed on previously studied words, leading the participants to engage in retrieval practice. The unknown condition was the second presentation of repeated study items. For those items, participants might retrieve the first presentation of the word from memory, adopting a retrieval practice strategy on their own. Previous research on study strategies and cognitive biases led us to predict that few participants would spontaneously engage in retrieval practice, but those that did would outperform the other participants for the words in that condition.

As we predicted, the average memory performance across all participants was lower for the words that were studied twice than for the words that were studied and then quizzed. Using the model, we identified a group of five participants whose brain activity was consistent with use of a retrieval practice strategy. That small subset of participants performed significantly better than the other participants on the subsequent memory test.

The experiment and model described in this paper represent the first steps toward using recorded brain activity to improve human memory performance. We have identified patterns of brain activity that are associated with the use of an effective memory strategy and developed a model that can predict which participants are using that strategy and which are not. In future research, we hope to expand on these results and investigate ways to coach people on the effectiveness of their study strategies as they attempt to learn new information.

Acknowledgments. The authors would like to thank Dr. Kara Federmeier for her invaluable advice and assistance with this project. We also thank Sean Cusick, Bridget Milligan, and Charlotte Laguna for their assistance with data collection. This work was supported by the Laboratory Directed Research and Development (LDRD) Program at Sandia National Laboratories. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

References

1. Benjamin, A.S.: Memory is more than just remembering: Strategic control of encoding, accessing memory, and making decisions. In: Ross, B.H., Benjamin, A.S. (eds.) The Psychology of Learning and Motivation: Skill and Strategy in Memory Use, vol. 48, pp. 175–223. Academic Press, London (2008)
2. Benjamin, A.S., Bjork, R.A., Schwartz, B.L.: The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. Journal of Experimental Psychology: General 127, 55–68 (1998)
3. Bjork, R.A.: Assessing our own competence: Heuristics and Illusions. In: Gopher, D., Koriat, A. (eds.) Attention and Performance XVII: Cognitive Regulation of Performance: Interaction of Theory and Application, pp. 435–459. MIT Press, Cambridge (1999)
4. Koriat, A., Bjork, R.A.: Illusions of competence in monitoring one's knowledge during study. Journal of Experimental Psychology: Learning, Memory, and Cognition 31, 187–194 (2005)
5. Koriat, A., Bjork, R.A., Sheffer, L., Bar, S.K.: Predicting one's own forgetting: The role of experience-based and theory-based processes. Journal of Experimental Psychology: General 133, 643–656 (2004)
6. Schwartz, B.L., Benjamin, A.S., Bjork, R.A.: The inferential and experiential bases of metamemory. Current Directions in Psychological Science 6, 132–137 (1997)
7. Landauer, T.K., Bjork, R.A.: Optimum rehearsal patterns and name learning. In: Gruneberg, M.M., Morris, P.E., Sykes, R.N. (eds.) Practical Aspects of Memory, pp. 625–632. Academic Press, London (1978)
8. Allan, K., Rugg, M.D.: An event-related potential study of explicit memory on tests of cued recall and recognition. Neuropsychologia 35, 387–397 (1997)
9. Neville, H.J., Kutas, M., Chesney, G., Schmidt, A.: Event-related brain potentials during the initial encoding and subsequent recognition memory of congruous and incongruous words. Journal of Memory and Language 25, 75–92 (1986)
10. Paller, K.A., Kutas, M.: Brain potentials during memory retrieval provide neuropsychological support for the distinction between conscious recollection and priming. Journal of Cognitive Neuroscience 4, 375–391 (1992)
11. Rugg, M.D., Doyle, M.C.: Event-related potentials and stimulus repetition in direct and indirect tests of memory. In: Heinze, H.J., Munte, T.F., Mangun, G.R. (eds.) Cognitive Electrophysiology, pp. 124–138. Birkhauser, Boston (1994)
12. Van Petten, C., Senkfor, A.J.: Memory for words and novel visual patterns: Repetition, recognition, and encoding effects in the event-related brain potential. Psychophysiology 33, 491–508 (1996)
13. MATLAB version 7.11.0.584 (R2010b). The MathWorks Inc., Natick (2010)

Advances and Challenges in Signal Analysis for Single Trial P300-BCI

Kun Li[1], Vanitha Narayan Raju[1], Ravi Sankar[1],
Yael Arbel[2], and Emanuel Donchin[3]

[1] Dept. of Electrical Engineering, University of South Florida
[2] Dept. of Communication Sciences and Disorders, University of South Florida
[3] Dept. of Psychology, University of South Florida
{Li,vanitha,yarbel,donchin}@mail.usf.edu,
sankar@eng.usf.edu

Abstract. In this paper a brief introduction to some of the goals, recent developments, and open problems in BCI research are given. We mainly focus on presenting our research work in signal processing for single-trial P300-BCI and discuss our current plans for improving the BCI method.

1 Introduction

Amyotrophic lateral sclerosis (abbreviated *ALS*, also referred to as *Lou Gehrig's disease*) is a progressive, fatal, neurodegenerative disease caused by the degeneration of motor neurons, the nerve cells in the central nervous system that control voluntary muscle movement. Affected individuals may ultimately lose the ability to initiate and control all voluntary movement; bladder and bowel sphincters and the muscles responsible for eye movement are usually, but not always, spared. It is especially for these people's benefit that researchers are eager to design a brain–computer interface (BCI) system to restore their communication ability which depends on peripheral nerves and muscles, brain signals such as electroencephalographic (EEG) activity. It provides a direct communication pathway between the brain and an external device.

BCI research is a multidisciplinary field involving collaboration of researchers from neuroscience, physiology, psychology, engineering, computer science, rehabilitation, and other technical and health-care disciplines. As a result, there have been several varied approaches to the design of BCIs reported over the last three decades. The major goal of this research is to create a specialized interface that will allow an individual with severe motor disabilities to have effective control of devices such as computers, speech synthesizers, assistive appliances, and neural prostheses. This type of interface would increase an individual's independence, leading to an improved quality of life and reduced social costs [1].

The idea underlying BCIs is to measure electric, magnetic, or other physical manifestations of the brain activity and to translate these into commands for a computer or other devices [2]. An activity in a normal human brain can generate various responses including electrical, magnetic, and metabolic responses. These signals can be detected by appropriate sensors and they can be used for controlling a

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 87–94, 2011.

BCI system. For example, brain activity can produce magnetic fields that can be recorded using magentoencephalographic (MEG) activity. Brain activity can also result in some metabolic consequences in terms of changes in the blood flow and metabolism.

Two basically different methods exist to control BCI for users to acquire conscious control over the brain activity. In the first method, subjects perceive a set of stimuli displayed by the BCI system and can control their brain activity by focusing onto one specific stimulus. The changes in neurophysiologic signals resulting from perception and processing of stimuli are event-related potentials (ERPs) [2]. In the second method, users control their brain activity by concentrating on a specific mental task. For example, imagination of hand movement can be used to modify activity in the motor cortex. In this approach feedback signals are often used to let subjects learn the production of easily detectable patterns of neurophysiologic signals [2].

There are various ways to record the electrical activities of the brain. Non-invasive BCI techniques mostly use the EEG signals as the source of information. EEG signals are recorded by means of electrodes placed on the scalp. Invasive technique, on the other hand, use electrocorticography (ECoG) signals recorded from the surface of the brain or action potentials of single neurons in the cerebral cortices, using implanted microelectrodes. However, they required a heavy surgical intervention. EEG signals have good temporal resolution, but their spatial resolution is not that good compared to other recording methods [3]. A recent study showed that only 12% of published BCI studies use implanted electrodes, 5% use microelectrode arrays, and more than 80% use EEG signals [4]. The main reason is that the EEG recording equipment is commercially produced and their cost is lower than other brain signal recording technologies. Also, it is non-invasive without any implanted electrodes (no surgery is necessary for placing electrodes) and so more individuals are willing to participate in such BCI experiments.

The aim of every brain computer interface is to translate simulated brain activity into a relevant computer command. The non-invasive methods like the P300 speller are enough to give back a communication potential emulating a keyboard or a mouse. The P300-Speller proposed by *Farwell and Donchin* in 1988 [5] provided researchers a practical way to accomplish this aim. To allow actual control of a BCI, the neurophysiologic signals have to be mapped to values that allow discriminating different classes of signals, i.e. the neurophysiologic signals have to be classified. In BCIs, statistical pattern recognition techniques and machine learning algorithms are applied to learn how to classify the signals of a specific user from a training dataset. As is well known, the signal processing methods and pattern recognition involve pre-processing of the raw data, feature extraction, and classification [2].

The P300 speller is based on event related potential (ERP) which are cerebral waves propagated in the cortex after stimulation (visual, hearing or tactile) [6]. ERP can be categorized into two types:

- Exogenous potentials, corresponding to non-cognitive activity. They appear after luminous flash, a noise or a sudden action. Their location on the cortex depends on stimulation type. For example, a visual stimulation causes a decrease of electrical potential 100 ms after the stimulation (called N100) in the occipital area.

- Endogenous potential corresponding to cognitive activity. For example, a patient is asked to differentiate two visual stimulations X and O. The X is less common than the O. The patient has to count the number of appearing X. At each X, an endogenous ERP appears 300 ms after the stimulation on central and parietal area (called P300) which is detected by the P300 speller.

The P300 is a positive deflection in the EEG, appearing approximately 300 ms after the presentation of rare, task-relevant stimuli [7]. To evoke the P300, subjects are asked to observe a random sequence of two types of stimuli. One stimulus type (the oddball or target stimulus) appears only rarely in the sequence, while the other stimulus type (the normal or nontarget stimulus) appears more often. Whenever the target stimulus appears, a P300 can be observed in the EEG. This principle was exploited by Farwell and Donchin in a BCI system which allowed spelling words by sequentially selecting symbols from a matrix of symbols [5].

2 Available Signal Processing Techniques for P300-Bci Speller

P300-Speller is widely used by researchers and many signal processing techniques have been developed to work with P300-Speller. Various signal processing techniques including SWLDA (Step Wise Linear Discriminant Analysis), ICA (Independent Component Analysis), Matched filter, Wavelet, and SVM (Support Vector Machine) have been developed and designed for the construction of a reliable BCI system with high accuracy and processing speed. It has been shown that some of the techniques can be effective in practical BCI systems, such as P300-Speller, and have been successfully applied to restore the communication ability of ALS patients.

In the past, P300-based BCI system has been successfully implemented using simple signal processing techniques such as signal averaging and LDA (Linear Discriminant Analysis). By using SWLDA as the classification method, Krusienski [8] achieved at least 60% accuracy for all participants. Three of the five participants performed above 90% accuracy with fewer than 15 sequences. This indicates that the classification can be performed on a minimal number of sequences without compromising accuracy and increase the communication rate. These results are consistent with those reported by Sellers and Donchin [9]. However, SWLDA has a drawback (increase in processing time since it has to undergo several trials to remove the background noise and magnitude of P300 response is to be enhanced before applying P300 classifier on EEG signal) which makes it not suitable for online P300 classification with single trial.

In signal processing, a matched filter is obtained by correlating a known signal, or template, with an unknown signal to detect the presence of the template in the unknown signal. It is actually the convolution of the unknown signal and the conjugated time-reverse version of the template. Matched filter is used to maximize the signal to noise ratio (SNR) when there are background stochastic noises. In frequency domain, matched filter can be considered as applying the greatest weighting to spectral components that have the greatest signal to noise ratio. Serby *et al.* [10] used match filtering with independent component analysis (ICA) to separate the P300 source from the background noise. A matched filter was used together with averaging and threshold techniques for detecting the existence of P300s. The processing method was evaluated offline on data recorded from six healthy subjects. The method achieved a communication rate of 5.45 symbols/min with an accuracy of 92.1% compared to 4.8 symbols/ min with an accuracy of 90% reported by Donchin *et al.* using SWLDA and Discrete Wavelet Transform (DWT) [11]. When the detection was made in real-time by online testing with the same six subjects, the average communication rate achieved was 4.5 symbols/min with an accuracy of 79.5% as opposed to the 4.8 symbols/min with an accuracy of 56% reported by Donchin *et al.* [11].

Wavelet is a mathematical function used to decompose a given function or continuous-time signal into different scale components that have been assigned a frequency range. A wavelet transform use the wavelet functions to represent the signal. Wavelet transform has advantages over traditional Fourier transform for representing functions that have discontinuities and sharp peaks, and for accurately deconstructing and reconstructing finite, non-periodic and non-stationary signals. Wavelet transforms are classified into continuous wavelet transforms (CWTs) and discrete wavelet transforms (DWTs). The difference is that CWTs operate over every possible scale and translation whereas DWTs use a predefined subset of scale and translation values. By using the CWT based on a modified Mexican Hat function and two-sample t-test, Bostanov [12] designed a feature extraction algorithm that works with P300-Speller. Classification accuracies of 82.6% and 54.4% for two different data sets provided by BCI Competition 2003 were reported.

In our earlier research [13][14], we have developed a blind tracking based ICA method [15] and another based on variance analysis for a single trial P300 classification to detect a chosen character in real-time in the P300-BCI speller. The key problem of ICA based P300 classification is finding the optimal IC set so that the algorithm guarantees a clear feature and IC mapping. Even though the proposed methods can be considered to be promising and reasonable solutions for single trial EEG signal classification, we are focusing our investigation on further improving the accuracy and the processing speed.

Machine learning methods have been recently investigated by researchers since the neurophysiology of the mental states that are used in BCIs are well-known. For example, the intention for a hand movement is reflected by the so called lateralized readiness potential (LRP): a negative shift of the brain potentials contralateral to the hand. And also, it seems possible to extract simple features that very well distinguish between the mental states. A simple but efficient method for supervised machine learning, appropriate for use in BCIs, is Fisher's discriminant analysis (FDA) which is related to LDA. The main advantages of FDA are that it is a computationally and conceptually simple method and that very good classification accuracy can be achieved. A possible drawback of FDA is that a squared error loss function is used which makes the method vulnerable to outliers in the training data. Furthermore, a precondition for using FDA is that the number of training examples is higher than the number of dimensions of the training data. In BCI applications it can happen that this precondition is not fulfilled.

Another method that is used in BCIs is the support vector machine (SVM) [16] [17] [18] [19]. The main advantages of the SVM are that it allows achieving very good classification accuracy and that nonlinear classification functions can be easily implemented by using kernels. The original problem may be stated in a finite dimensional space; it often happens that in that space the sets to be discriminated are not linearly separable. For this reason it was proposed that the original finite dimensional space be mapped into a much higher dimensional space presumably making the separation easier in that space. SVM schemes use a mapping into a larger space so that cross products may be computed easily in terms of the variables in the original space making the computational load reasonable. The cross products in the larger space are defined in terms of a kernel function $K(x,y)$ which can be selected to suit the problem. The hyperplanes in the large space are defined as the set of points whose cross product with a vector in that space is constant. The vectors defining the hyperplanes can be chosen to be linear combinations with parameters α_i of images of feature vectors which occur in the data base. A drawback is that training SVMs is computationally complex because regularization constants and kernel parameters are typically estimated with a cross-validation procedure. A second issue is that the loss function used in the SVM is designed for problems in which only binary yes/no outputs are needed. The problem with binary yes/no outputs is that no information is given about the confidence the system has in those outputs. Besides FDA and SVM, other machine learning algorithms have been tested in the context of BCI systems. An overview of these algorithms can be found in [20].

We need to study more cases and also further optimize these algorithms by involving statistical models to solve the non-stationary problem [21] which will be researched in our future work.

3 Comparison of Results

From 70s to 90s, researcher have done a tremendous work in single trial ERP analysis, they have discussed the feasibility of single trial ERP analysis, investigated the factors that may affect the analysis, suggested possible denoising and classification techniques and conducted different experiments on single trial ERP signal analysis. Their work directly or indirectly proved that single trial P300 classification on P300-Speller is possible and promising. According to the definition of Wolpaw [22], the maximal bitrate, computed was approximately 25 bits/min.

It has also been stated that accuracy more than 70% allows communication and device control [23]. It is argued in [24] that it may be frustrating if a BCI system does not reach at least a 70% accuracy level. But the validity of such claim still needs to be verified. It has also been shown that the performance of subjects during online systems may be significantly lower than their performance when evaluated offline (probably because of lack of focus) [25]. Recently more advanced techniques have been used to process the P300 signal but none were able to successfully accomplish the detection and classification in a single trial with high accuracy and suitable for practical implementation. More results and possible factors causing the differences between the present study and other studies are discussed in [26], [27].

In our work [13][14][15], attempts have been made to develop single trial P300 classification methods to detect a chosen character in real-time in the P300-BCI speller. The results demonstrate that the proposed methods dramatically reduce the signal processing time, improve the data communication rate, and achieve overall accuracy of 79.1% for ICA based method and 84.8% for variance analysis based method in single trial P300 response classification task. It provided 34.1% increase in accuracy and 139% more effective in communication speed over single trial SWLDA. And the variance analysis based classification method demonstrate 39.8% increase in accuracy and 60% more effective in communication speed over single trial SWLDA.

Moreover, in the comparison of bit rate according to Wolpow's definition [28], our method achieved 129.4 bits/min while SWLDA achieved 33.8 bits/min. The proposed methods are promising and can be considered to be reasonable solutions for single trial P300 signal detection and classification problem in BCI.

4 Open Issues and Challenges in BCI Research

As for trends in BCI stimulus presentation paradigms, the P300 Speller has proven to be successfully used by various patient groups with relatively high accuracy for typing emails and other similar tasks, yet it has the drawback of being quite slow compared to non-BCI technology such as eye gaze tracking.

In almost all current BCI systems, subjects first have to go through a training phase, in which they concentrate on prescribed mental tasks or prescribed stimuli. After the training phase a classifier is learned and used to classify new, unseen data. [2] A drawback of this setup is that for many disabled users a long training phase is an insurmountable obstacle due to cognitive impairments and concentration problems. Another problem is caused by the fact that patterns of cerebral activity are constantly changing, and hence new training sessions have to be performed periodically to adapt classification rules. [2] One approach to overcome these problems is to develop machine learning algorithms, with which subjects can immediately start using a BCI, without training. At present, the system applicability has several limitations such as low detection rates of mental states, slow response times, slow command speed, low number of possible decisions per command, and bulky preparation. A new user interface will emerge when these limitations are overcome.

5 Conclusion and Future Work

In this paper we discuss the recent advances in BCI research including our recent research work on signal processing for P300-based BCI system. We plan on extending our past research to develop and design a single trial P300 response classification methods based on support vector machine (SVM). The ultimate goal is to further improve the accuracy of our single trial P300 analysis algorithms to make them more suitable for real-world applications and clinical use. We see that there are large varieties of BCI systems but yet many challenging and interesting questions are still waiting to be explored.

References

[1] Mason, S.G., Birch, G.E.: A general framework for brain–computer interface design. IEEE Trans. Neural Syst. Rehab. Eng. 11(1), 80–95 (2003)

[2] Hoffmann, U., Vesin, J.-M., Ebrahimi, T.: Recent advances in brain-computer interfaces. In: Proc. IEEE 9th Workshop on Multimedia Signal Processing (MMSP), October 1-3 (2007)

[3] Vaughan, T.M., Heetderks, W.J., Trejo, L.J., Rymer, W.Z., Wienrich, M., Moore, M.M., Kubler, A., Dobkin, B.H., Birbaumer, N., Donchin, E., Wolpaw, E.W., Wolpaw, J.R.: Brain-computer interface technology: a review of the second international meeting. IEEE Trans. Neural Syst. Rehab. Eng. 11(2), 94–109 (2003)

[4] Mason, S.G., Bashashati, A., Fatourechi, M., Navarro, K.F., Birch, G.E.: A comprehensive survey of brain interface technology designs. Ann. Biomed. Eng. 35(2), 137–169 (2007)

[5] Farwell, L.A., Donchin, E.: Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalography and Clinical Neurophysiology 70, 510–523 (1988)

[6] Samuel, B.: Event related potential (March 2011), `http://www.samuelboudet.com/en/EEG/P3speller.php`

[7] Sutton, S., Braren, M., Zubin, J., John, E.: Evoked-potential correlates of stimulus uncertainty. Science 150(700), 1187–1188 (1965)

[8] Krusienski, D.J., Sellers, E.W., McFarland, D.J., Vaughan, T.M., Wolpaw, J.R.: Toward enhanced P300 speller performance. Journal of Neuroscience Methods 167, 15–21 (2008)

[9] Sellers, E.W., Donchin, E.: A P300-based brain-computer interface: initial tests by ALS patients. Clin. Neurophysiol. 117, 538–548 (2006)

[10] Serby, H., Yom-Tov, E., Inbar, G.F.: An improved P300-based brain-computer interface. IEEE Trans. Neural Syst. Rehab. Eng. 13, 89–98 (2005)

[11] Donchin, E., Spencer, K.M., Wijesinghe, R.: The mental prosthesis: Assessing the speed of a P300-based brain-computer interface. IEEE Trans. Rehabil. Eng. 8, 174 (2000)

[12] Bostanov, V.: BCI competition 2003-data sets Ib and IIb: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram. IEEE Trans. Biomed. Eng. 51, 1057–1061 (2004)

[13] Li, K., Sankar, R., Arbel, Y., Donchin, E.: P300 based single trial independent component analysis on EEG signal. In: Schmorrow, D.D., Estabrooke, I.V., Grootjen, M. (eds.) FAC 2009. LNCS, vol. 5638, pp. 404–410. Springer, Heidelberg (2009)

[14] Li, K., Sankar, R., Arbel, Y., Donchin, E.: Single trial independent component analysis for P300 BCI system. In: 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 4035–4038 (September 2009)

[15] Li, K.: Advanced signal processing techniques for single trial electroencephalography signal classification for brain computer interface applications, Ph.D. Dissertation, University of South Florida (September 2010)

[16] Hastie, T., Tibshirani, R., Friedman, J.: The elements of statistical learning - data mining. Springer, NY (2001)

[17] Kaper, M., Meinicke, P., Grossekathoefer, U., Lingner, T., Ritter, H.: BCI competition 2003-data set IIb: Support vector machines for the P300 speller paradigm. IEEE Trans. Biomed. Eng. 51, 1073 (2004)

[18] Meinicke, P., Kaper, M., Hoppe, F., Huemann, M., Ritter, H.: Improving transfer rates in brain computer interface: A case study. In: NIPS (2002)

[19] Thulasidas, M., Cuntai, G., Wu, J.: Robust classification of EEG signal for brain-computer interface. IEEE Trans. Neural Syst. Rehab. Eng. 14(1), 24–29 (2006)

[20] Bashashati, A., Fatourechi, M., Ward, R.K., Birch, G.E.: A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. Journal of Neural Engineering 4(2), R32–R57 (2007)

[21] Hyvärinen, A., Oja, E.: Independent component analysis: algorithms and applications. Neural Networks 13, 411–430 (2000)

[22] Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M.: Brain-computer interfaces for communication and control. Clinical Neurophysiology 113(6), 767–791 (2002)
[23] Kübler, A., Mushahwar, V.K., Hochberg, L.R., Donoghue, J.P.: BCI meeting 2005-workshop on clinical issues and applications. IEEE Trans. Neural Systems and Rehab. Eng. 14(2), 131–134 (2006)
[24] Sellers, E.W., Kübler, A., Donchin, E.: Brain-computer interface research at the University of South Florida cognitive psychophysiology laboratory: the P300 speller. IEEE Trans. Neural Syst. Rehab. Eng. 14(2), 221–224 (2006)
[25] Muller-Putz, G.R., Scherer, R., Neuper, C., Pfurtscheller, G.: Steady-state somatosensory evoked potentials: suitable brain signals for brain-computer interfaces? IEEE Trans. Neural Syst. and Rehab. Eng. 14(1), 30–37 (2006)
[26] Hoffmann, U., Vesin, J.-M., Ebrahimi, T., Diserens, K.: An efficient P300-based brain-computer interface for disabled subjects. Journal of Neuroscience Methods 167, 115–125 (2008)
[27] Hoffmann, U.: Bayesian machine learning applied in a brain-computer interface for disabled users, Ph.D. dissertation, Ecole Polytechnique Federale de Lausanne, Switzerland (2007)
[28] Wolpaw, J.R., Ramoser, H., McFarland, D.J., Pfurtscheller, G.: EEG-based communication: Improved accuracy by response verification. IEEE Trans. Neural Syst. Rehab. Eng. 6(3), 326–333 (1998)

Characterizing the Performance Limits of High Speed Image Triage Using Bayesian Search Theory

Santosh Mathan[1], Kenneth Hild[2], Yonghong Huang[2], and Misha Pavel[2]

[1] Honeywell Laboratories, 15001 NE 36th St., Redmond, WA 98052, USA
[2] Oregon Health and Science University, 3303 SW Bond Ave, Portland, OR 97239, USA
santosh.mathan@honeywell.com, k.hild@ieee.org, huang@csee.ogi.edu, pavel@bme.ogi.edu

Abstract. The rapid serial visual presentation (RSVP) modality has been used in conjunction with neurophysiological and behavioral responses to identify targets within large volumes of imagery efficiently. The research reported here uses optimal search theory to characterize the limits of this approach. Search theory is used to inform the estimation of detection functions. These functions provide a principled basis for selecting presentation parameters that balance search efficiency and accuracy. Detection functions are also used to characterize individual differences in search performance and to assess the extent to which the RSVP presentation modality generalizes across a class of complex targets.

Keywords: EEG, Search Theory, Rapid Serial Visual Presentation, Visual Psychophysics, Detection Functions, Target Detection.

1 Introduction

The challenge of finding information in large volumes of imagery has few good solutions. Both automated and manual approaches to target detection have limitations. Computer vision-based solutions often can't deal with novelty or variability, nor can they exploit contextual information and prior knowledge to the extent that humans can. On the other hand, manual image analysis is slow—requiring careful, methodical scrutiny of an image to identify potential targets. Manual analysis tools typically engage slow and deliberate, top-down cognitive processes that operate on a time scale of seconds or tens of seconds. These limitations of conventional search techniques have practical implications in a wide variety of domains; from military image analysis, to geospatial and medical imagery analysis. The volume of imagery available in these domains far exceeds the resources available to process them.

1.1 Tapping into Split-Second Perceptual Judgments

One avenue for raising the efficiency of the manual search process is to exploit the fast, automatic, bottom-up perceptual judgments that people make routinely. Specific examples include the perceptual processes engaged in returning a tennis serve, hitting a baseball, or reacting to an obstacle on the highway while driving—we can detect critical events and initiate physical responses to them in a couple of hundred

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 95–103, 2011.

milliseconds. Yet, these processes can solve complex perceptual recognition problems, including those that appear to require cognitive interpretation [1]. In recent years, researchers have attempted to take advantage of these processes to boost the efficiency of manual search. Research has shown that a combination of the rapid serial visual presentation (RSVP) technique and the event-related potential (ERP) signal detected using electroencephalograph (EEG) sensors can provide a way to identify targets in imagery using split-second perceptual judgments [2]. RSVP is a presentation paradigm under which a sequence of images is flashed to users at very high rates—where each image is presented for durations spanning just a few tens or hundreds of milliseconds (Figure 1).

Our research has shown that the RSVP presentation modality, when employed in the context of a multi-stage search process, can help professional image analysts identify a broad range of complex targets in high-resolution satellite imagery efficiently and accurately [4]. In the first stage of our approach, broad area images, spanning tens of thousands of pixels in width and height are decomposed into chips a few hundred pixels wide and tall. Each image chip must be scaled appropriately to the dimensions of the target of interest, as the high presentation rates used in the RSVP paradigm preclude eye saccades to search an image. In the first stage of the search process, these chips are presented to users in high speed bursts—anywhere from 2 to 15 chips per second. EEG sensors record neural responses to each chip. Images that elicit an ERP signal are classified as potential targets. In the second stage of the search process, users examine the subset of images identified as being targets in the RSVP-based search and eliminate false positives.

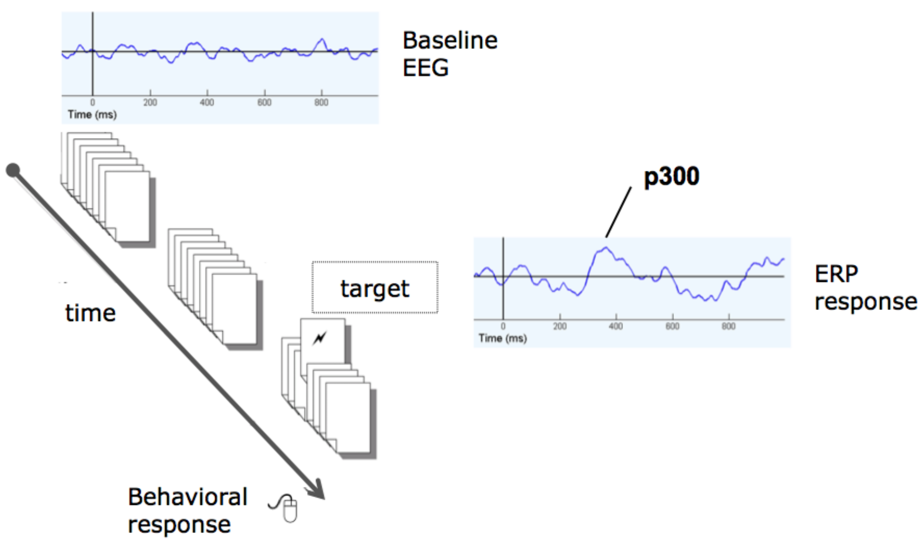


Fig. 1. Large high-resolution images are decomposed into chips and presented to users in high-speed bursts lasting 2 to 5 seconds. Patterns of EEG and other responses are used to detect likely targets in the image sequence.

Empirical evaluations of the approach outlined above show that the RSVP presentation modality in conjunction with response signals ranging from EEG [2, 4, 5, 7] to button clicks [7, 4] and pupil dilation [6] can provide an efficient and accurate basis for identifying a broad range of targets in natural imagery. For instance, in an experiment involving experienced military image analysts, an RSVP-based search for surface-to-air missile sites contributed to a six-fold reduction in processing time relative to conventional search [4]. Participants were able to scan images at an average rate of .86 sq km/sec (s.d = 0.11) in the RSVP condition, compared to 0.15 sq km/sec (s.d = 0.04) in the conventional search condition. This increase in processing efficiency was achieved without compromising target detection accuracy. On average, participants in the RSVP condition detected 96% (s.d = 7%) of targets, compared to 87% (s.d = 17%) in the baseline condition. The false positive rate in both conditions was low (average of 3.9 false positives in the triage condition and 0.8 in the baseline). Neither the differences in detection rate, nor the false positive rate were statistically significant.

While demonstrations of the efficacy of an RSVP-based triage approach are promising, important questions about the generality of the observed results must be addressed in order for this technique to be broadly adopted. For example, it is important to determine whether the observed efficiency gains extend beyond the specific targets used in a given experiment to other targets of a particular class. Additionally, it is essential to determine whether the observed results generalize across individuals. Effective generalization of these findings to practical task contexts also requires knowledge of appropriate presentation parameters—such as the presentation rate—to maximize processing efficiency without compromising accuracy. To address these issues, we turned to Bayesian search theory.

1.2 Bayesian Search Theory

Statistically optimal search theory was developed during World War II, by the US Navy's Anti Submarine Warfare Operations Research group to optimize the use of limited search resources for a range of search problems: from rescue missions, to locating the wreckage of aircraft and ships. This work, based on maximizing expected utility using a Bayesian framework, has been very influential and continues to form the theoretical basis for modern search operations. The focus of the work reported here has been to apply techniques informed by search theory to estimate the expected performance of a human observer for a given class of targets. We start with a brief review of optimal search theory to illustrate the importance of the detection function, which is the cornerstone of the proposed estimation process. The following explanation is a simplification of the normative theory of optimal search presented by Koopman [3] and Stone [8].

For the sake of simplicity, we assume a discrete search space, i.e., that the targets could be located at one of K locations and that the prior probability of a target at the i^{th} location is given by p_i. The effectiveness of a search process is often determined in part by the search resources that can be allocated to different search locations. Examples of resources that can be distributed include time or the resolution at which the search area is traversed (e.g. coarse vs. fine search grid) under the constraint by which the total resources are limited. The goal of the optimization is to determine the best distribution of search resources over locations. This computation requires knowledge of the relationship between specific resources allocated to locations and the probability of detecting targets, if present. To express this mathematically we denote the resources allocated to the i^{th} location by h_i and define a detection function u(h) as the conditional probability of finding a target at that location, given that the target is at the i^{th} location by the detection function: . Since the resources available to search in most situations are limited, the goal of the search optimization process is to allocate resources to maximize expected detection performance, for example, the number of detected targets,

$$E\{Utility\} = U\{\vec{r}, \vec{h}\} = \sum_{i=1}^{K} u_i\,(h_i)\,p_i, \tag{1}$$

where $\vec{h} = \{h_1, h_2, ...h_n\}$ is the vector of allocated resources constrained by

$$H = \sum_{i=1}^{K} h_i \tag{2}$$

and $\vec{r} = \{p_1, p_2, ...p_n\}$ is the vector of prior probabilities Assuming that the detection function u, has certain convexity properties, it is possible to determine the optimal resource allocations by constrained optimization of Equation (1). In particular, using the method of Lagrange multipliers, we can show that the optimal allocation of resources to the i^{th} location must obey the following

$$\frac{d}{dh_i} u(h_i) = \frac{\lambda}{p_i} \tag{3}$$

where λ is a Lagrange multiplier, typically calculated using the constraint expressed by Equation (1). The result in Equation (3) means that the resource allocated to a location should be proportional to the instantaneous rate of detection, combined with prior probability.

In the case of the RSVP search technique outlined here, the resource to be optimized is the time allocated to processing each chip. The idea is to make the search as efficient as possible without compromising detection accuracy. The detection function is, therefore, a fundamental component that will allow us to estimate the tradeoff between efficiency and accuracy. Detection functions are typically empirically derived because of the complex interaction between a target and features of the search area [9]. The next section illustrates a parametric approach to estimation of detection functions.

2 Method

As mentioned above, the estimation of an empirical detection function is of central importance in search theory. In many search domains, proxy targets are embedded in specific search contexts. Empirical search performance data is used to construct functions that help approximate the probability of detecting targets under various scenarios. For example, proxy hikers wearing different types of clothing may be placed in terrain where search and rescue missions are common [9]. Detection performance is assessed as a function of various combinations of search resources and parameters (for example, search using airplanes or helicopters from different altitudes). Fits of psychophysical functions to this data can help answer questions about the optimal combination of resources and parameters to search for the target of interest.

We adapted this approach to characterize detection performance as a function of the RSVP presentation rate based on empirical data. The targets used in our study consisted of surface-to-air missile (SAM) sites (Figure 2). We chose SAM sites because they represent a common type of target that a large population of image analysts (military) are trained to detect. Additionally, they have a well-defined set of visual features that are easy to describe to both experienced image analysts and inexperienced participants. SAM sites also vary considerably in complexity—some targets have features that are prominent enough to pop out with exposures of a few tens of milliseconds, while detection of other SAM sites requires careful reasoning based on prior knowledge and contextual information.

We employed an RSVP experimental paradigm in order to develop an empirically derived detection function. Images were presented to users in blocks that lasted two-seconds. Half the blocks contained targets drawn from the 105 targets extracted from several broad area satellite images. To assure that each image was given equal exposure for processing, a pattern mask followed each stimulus chip. The presentation duration of each image varied between 25, 50, 100, 150, and 200 milliseconds. Ten participants recruited from the population of researchers at Honeywell Laboratories were asked to respond to each target with a button press.

Fig. 2. Example of surface to air missile sites

3 Results

The behavioral data gathered in conjunction with the RSVP study described above were summarized in terms of the probability of detection at each image presentation rate. These estimates correspond closely to the definition of the detection function in the search theory framework. The detection function in this case was similar to a psychometric function in a standard signal detection paradigm. In contrast to a standard psychophysical function, the objective signal strength was not known and its effect had to be estimated from the data. The analyses were based on the assumption that the detection function can be well-approximated by parametric fits to a Gumbel distribution of the form:

$$Pr\{Correct\ Detection \mid T\} = u_i(T) = \exp(-e^{B_0+B_1T})\,, \tag{4}$$

where B_0 and B_1 are constants to be estimated and T is the presentation rate. In our study, these constants, estimated on data averaged over subjects, were directly related to the objective strength of the target. This family of functions was selected in order to capture long tails of the detection function for some targets. Instead of reporting these abstract quantities, however, it seemed more natural to report the presentation rate at which the detection function reached a fixed point, e.g., $u_i(T_{75}) = 0.75$. These thresholds are then used to represent the target difficulty.

Figure 3 illustrates a few examples of psychometric functions empirically derived from the SAM site data. Each point (blue circle) on the four plots summarizes detection performance averaged across subjects at a particular rate. The green curve fitted through these points is a psychometric function that describes detection performance as a function of performance rate. The vertical red line in each plot represents the detection threshold, i.e. the rate at which detection performance reaches 0.75. The images and associated plots in Figure 3 show how the detection threshold rises as discriminating features of SAM sites (radial arrangement of missiles and launchers, service paths, central RADAR etc.) become less prominent. As described below, detection thresholds estimated in this way provide a way to make inferences about numerous issues that could influence detection performance—including the optimal presentation rate for a given class of targets and individual differences in detection performance.

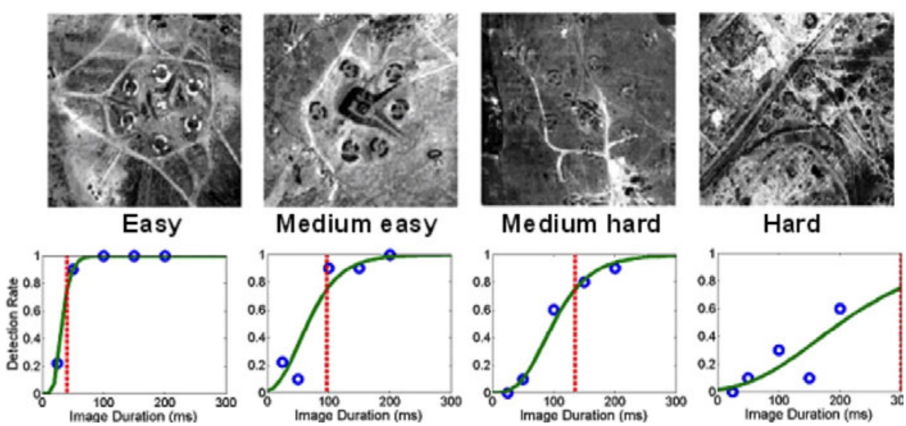


Fig. 3. SAM sites of varying complexity with associated detection thresholds. Detection threshold (red line) estimated from performance data rises as a function of target complexity.

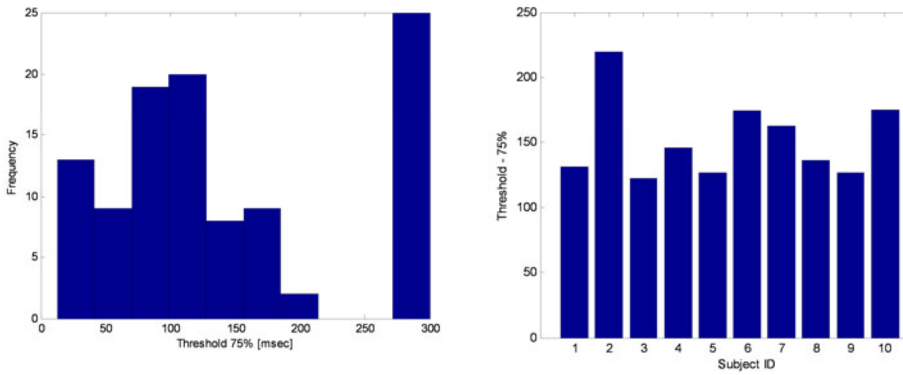


Fig. 4. Distribution of detection threshold across all targets included in study (left). Distribution of average detection threshold across individuals (right).

To assess the overall efficacy of the RSVP technique for the targets of interest, we pooled detection thresholds associated with each of the 105 targets included in the experiment. A histogram based on these thresholds (Figure 4, left) suggests that over 75% of targets could be detected with an accuracy of 75% or more at rates of 200 ms/image or less. Approximately 25% of targets could not be detected with 75% accuracy at the presentation rates used in our study. It is likely that a substantial proportion of the targets that could not be detected are perceptible at rates above 200 ms; however, a subset of these targets may not be well-suited for RSVP-based triage. Whether or not detecting 75% of targets is acceptable depends on the nature of the missions—in particular, based on considerations that balance the risk of missed targets against processing efficiency. Nevertheless, these results suggest that the RSVP modality can provide the basis for efficiently and accurately detecting complex targets of practical relevance.

Our analysis also examined individual differences in performance by pooling average detection thresholds for each participant across presentation rates. Results show that the detection threshold for six of the ten participants was150 ms or lower (Figure 4, right). Nine of ten participants had average detection thresholds below 200 ms. Only one participant had a detection threshold exceeding 200 ms. This result has several implications. First, detection performance is consistent across individuals—suggesting that the RSVP modality may be a viable image screening modality for a wide range of users. Second, empirically derived detection functions can provide a principled way to optimize presentation rates for an individual or a group of individuals. Third, detection functions can also provide a way to screen individuals who may be more effective than others at detecting targets using the RSVP modality.

4 Discussion

The research described here builds on previous work demonstrating the potential to enhance the efficiency of image analysis by combining the RSVP presentation modality with response modalities ranging from EEG-based event related potentials to pupil and motor responses. The analysis described here employs Bayesian search theory to investigate the extent to which these results generalize across a complex class of targets. Additionally, we demonstrate how an empirically derived detection function, a central component of optimal search theory, can provide a principled basis for estimating the optimal presentation rate for an individual or a group and to identify individual differences in search efficacy. The analysis presented above suggests that most SAM sites can be detected at rates of 200ms/chip. Additionally, we find a fairly consistent distribution of detection thresholds across subjects.

The work described here has focused on a detection function based on a single parameter: RSVP presentation rate. However, this approach could be extended to include other parameters that affect search performance, from the spatial scale at which an image is processed to the eccentricity of targets within each image chip.

The described detection function is just one component necessary to optimize search. Optimal search theory emphasizes the importance of considering prior probabilities of targets; search resources, characterized by empirically derived detection functions, should be allocated to regions in proportion to the prior target probability associated with specific regions of the search space. Additionally, the presentation rate for a given area of an image should be determined not just by the detection function, but a joint consideration of the prior probability of a target in the region and the cost associated with the risk of missing a target, if present.

Acknowledgements. This work was supported by the Defense Advanced Research Projects Agency under contract N10PC20048. The views, opinions, and/or findings contained in this article/presentation are those of the author/presenter and should not be interpreted as representing the official views or policies, either expressed or implied, of DARPA or the Department of Defense. The authors thank James Carciofini and Karen Feigh for their contributions to this work. Satellite images included in this paper were produced by DigitalGlobe Inc., Longmont, CO 80501, USA. (c) 2003

References

1. Chase, W.G., Simon, H.A.: Perception in Chess. Cognitive Psychology 4, 55–81 (1973)
2. Huang, Y., Erdogmus, D., Pavel, M., Mathan, S., Hild, K.E.: A Framework for Visual Image Search Using Single-Trial Brain Responses. Accepted by the Journal of Neurocomputing
3. Koopman, B.O.: Search and Its Optimization. The American Mathematical Monthly 86(7), 527–540 (1979)
4. Mathan, S., Whitlow, S., Erdogmus, D., Pavel, M., Ververs, P., Dorneich, M.: Neurophysiologically Driven Image triage: a Pilot Study. In: CHI 2006 Extended Abstracts on Human Factors in Computing Systems, CHI EA 2006, pp. 1085–1090. ACM, New York (2006)
5. Parra, L., Christoforou, C., Gerson, A., Dyrholm, M., Luo, A., Wagner, M., Philiastides, M., Sajda, P.: Spatiotemporal Linear Decoding of Brain State. IEEE Signal Processing Magazine 25(1), 107–115 (2008)

6. Qian, M., Aguilar, M., Zachery, K.N., Privitera, C., Klein, S., Carney, T., Nolte, L.W.: Decision-Level Fusion of EEG and Pupil Features for Single-Trial Visual Detection Analysis. IEEE Transactions on Bio-medical Engineering 56(7), 1929–1937 (2009)
7. Sajda, P., Gerson, A.D., Philiastides, M.G., Parra, L.C.: Single-Trial Analysis of EEG During Rapid Visual Discrimination: Enabling Cortically-Coupled Computer Vision. MIT Press, Cambridge (2007)
8. Stone, L.D.: Theory of Optimal Search. Academic Press, New York (1975)
9. Koester, R.J., Cooper, D.C., Frost, J.R., Robe, R.Q.: Sweep Width Estimation for Ground Search and Rescue. US Department of Homeland Security, United States Coast Guard Operations (G-OPR), Washington DC (2004)

Facial Recognition: An Enabling Technology for Augmented Cognition Applications

Denise Nicholson[1], Christine Podilchuk[2], and Kathleen Bartlett[1]

[1] DSCI, Human Systems Engineering, Orlando, Florida
[2] DSCI, STAR Technologies, Eatontown, New Jersey
{Dnicholson,cpodilchuk,kbartlett}@dsci.com

Abstract. Research in Augmented Cognition (AugCog) investigates computational methods, technologies, and non-invasive neurophysiological tools to adapt computational systems to the changing cognitive state of human operators to improve task performance. Closed-loop AugCog systems contain four components: 1) operational or simulated environment, 2) automated sensors to monitor and assess cognitive state via behavior and/or physiology, 3) adaptive interface, and 4) computational decision architecture that directs AugCog adaptations. Since cognitive state is influenced by environment, a critical challenge for AugCog systems is capture of situational awareness (SA) within the decision architecture. Previously, AugCog systems have been demonstrated within simulated environments that provide SA and ground truth data to drive intelligent decision architecture. In live operating environments, electronic C4 systems (i.e., communications), provide a limited model of operator "state," but emerging facial recognition/analysis technology can provide detection, identification, and tracking of humans in the environment to increase the accuracy of the AugCog system's SA.

Keywords: Augmented Cognition, facial recognition, situation awareness, biometrics, environmental monitoring.

1 Augmented Cognition

Building on advances in the fields of neuroscience, cognitive science, and computer science, Augmented Cognition research focuses on the real-time cognitive state of the operator [1]. Current AugCog methods, techniques, and applications range from academic research to industrial/ military operational and training systems to computing and entertainment devices [2]. To enhance human performance, Aug Cog uses physiological and neurophysiological sensors in a closed loop system (see Fig. 1) to detect when the human's cognitive capacity, which fluctuates due to fatigue, stress, overload, or boredom, cannot meet mission demands [3]. Neurophysiological- and physiological-based assessment of cognitive states relies upon a variety of data, including cardiac measures, electroencephalogram (EEG), and functional near-infrared (fNIR) imaging, to evaluate cognitive ability in diverse environments [3]. These real-time, non-invasive measures of an operator's cognitive state can be used to trigger adaptive automation that enhances human performance [2].

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 104–111, 2011.

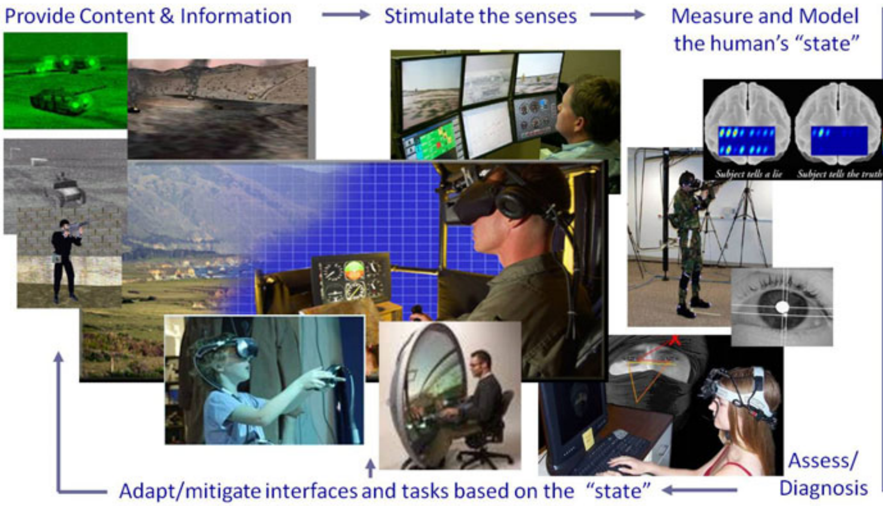


Fig. 1. Model for Adaptive Interaction in Closed Loop System (CLHS) [4]

Regardless of the field of use, the core components of the CLHS, including skill assessments, physiological and behavioral measurements, and mitigations, are extendable concepts. Coupled with field specific task modeling, a flexible system design can identify how users can best achieve and retain mastery of multiple tasks / skills through individualized adaptive interaction. In situations of overload, when the demand for speed or attention exceeds human ability, technology may be activated to compensate for human performance degradation [5] in simulated training and live environments. In tests at a base in Orlando, for example, researchers observed the degraded performance caused by information overload: when soldiers operated a tank while monitoring remote video feeds, they often failed to see nearby targets [6]. Augmented reality (AR) systems, which combine real and artificial stimuli, prepare trainees to operate successfully in a wide range of dangerous, unpredictable environments by monitoring operator state and providing technology-assisted support, via mitigation strategies, when performance overload is indicated. In live situations, mitigation strategies have the potential to save lives.

A mitigation strategy is an intervention technique (i.e., adaptive interaction, operational performance support, task cooperation, and individualized embedded training strategies) driven by the task analysis, automated measures, and diagnostic assessment, that significantly improves human-system performance and enhances skill proficiency and retention. The iterative mitigation process can be mapped to Norman's (1988) "seven stages of action" model, which was developed to represent the human action cycle as people accomplish goals (Fig. 2). Within the adaptive design environment, mitigation strategies are triggered in stages one to three of the "seven stages" model by *perceiving* a user's behavioral or physiological activity through automated measures and diagnostic assessment. Mitigation covers stages four to six [4].

Mitigation strategies can only be effectively triggered if measurement of operator state reflects a complete picture of the operating environment. If the operator registers stress, for example, the digital audio (C4 channels) may reflect voices, but the number and type of individuals in the environment cannot be represented in the AugCog SA in order to trigger the supporting mitigation strategy. Facial recognition technologies may be useful to monitor and report on humans in the environment, thus providing a missing puzzle piece in the monitoring of operator state.

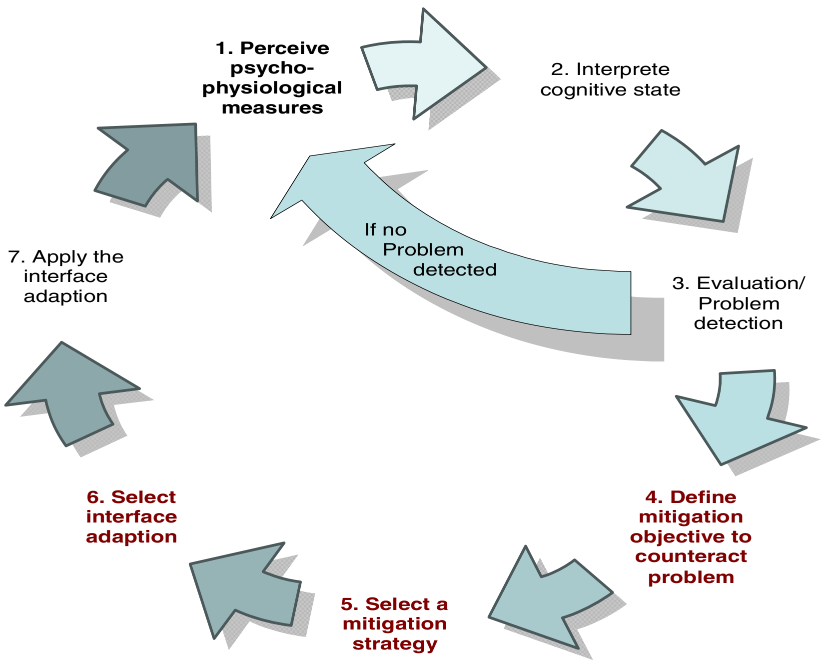


Fig. 2. Norman's "7 Stages of Action" modified for mitigation strategy design [4]

2 Visual Environment and Cognitive State

A complimentary definition of this domain space is provided by Neuroergonomics, which "postulates that the human brain, which implements cognition and is itself shaped by the physical environment, must be examined in interaction with the environment in order to understand fully the interrelationships of cognition, action, and the world of artifacts" [7]. Therefore, collection of comprehensive environmental information is essential when monitoring cognitive state. However, a critical challenge in developing a full closed-loop AugCog system is the accurate representation of the operational environment or situational awareness (SA) within the decision architecture's expert model used to drive system adaptations. Accurate SA data is critical to achieve correct interpretation of the physiological sensors for cognitive state determination of the operator and to ensure appropriate recommendations for system adaptations which optimize human-system performance, under specific environmental conditions.

To address this challenge, most of the previous AugCog closed-loop systems have been demonstrated within fully characterized simulated environments that provide the necessary situational awareness and ground truth data about the operating state to drive the intelligent decision architecture. Other demonstrations have explored live operating environments which provide an opportunity to monitor electronic C4 ISR systems (i.e., communications) to supply the necessary SA. However, the C4 channels don't contain all information influencing the human operator. For example, operators collect additional information about objects, personnel, and targets via their visual field of view, which is not represented in the C4 data. Consequently, a decision system that doesn't have access to the visual information in the operating space is limited in achieving a fully accurate model of the operator's "state."

In any operating environment, the presence (or absence) of other persons represents one of the most fundamental elements of operator state and frequently influences whether or not operator action is required. In a simulated AugCog environment, the entities populating any given situation are programmable, and the manipulation and measurement of operator response can be based on predictable encounters with anticipated threats. In live situations relying on AugCog support, measurement of SA lacks input about the numbers and types of humans in the live environment, although these elements exert significance influence on the cognitive state of the individual. Emerging video surveillance and facial recognition/analysis technology has the potential to provide detection, identification, and tracking of the presence of human targets in the operator's live environment. With this information, diagnosis of the incongruency between the ideal expert state and the actual human state of the operator might result in appropriate adaptations to ensure optimum operating state. Facial recognition technology can be a critical solution for providing the necessary environmental SA for effective diagnosis and driving of AugCog adaptations.

3 Facial Recognition

A suite of facial recognition applications currently being developed at DSCI provides real-time face recognition in uncontrolled environments using novel algorithms for pattern recognition that are robust for differences in facial expression, pose, illumination, camera angle, and facial occlusions. This technology can locate, analyze, and provide information about the faces of personnel located within a live environment and can specify or determine the number of faces in an image. This technology can also be applied to images extracted from a video stream where faces are detected, identified, and tracked through a video clip or across cameras, detecting and responding to occlusions. The approach consists of a general mapping between two images, followed by a measurement of different properties of the two-dimensional vector field representing the mapping task. Since it does not depend on domain specific features, this general approach can be applied to any object or target recognition.

Working in tandem with the target recognition algorithm is a fast search technique for rapid identification via large databases. This fast search algorithm can be applied to any algorithm that computes similarity scores for the purpose of pattern recognition. This method allows for quick recognition across a large database of stored faces or targets by computing only a subset of scores. This technique is particularly useful for applications with a large database or list, as well as applications that require processing large amounts of data, as in video surveillance applications. The face recognition and tracking algorithm, along with the fast search approach, can be used for video surveillance applications such as detecting, identifying, and tracking individuals and objects (such as suitcases or weapons), face recognition for physical and logical access control, watch list identification, suspicious behavior detection, and other applications, including providing real-time information about the people within a given operating environment. It could also be used to enhance SA in simulated AugCog environments by providing detection, identification, and tracking of humans in the visual environment.

4 A New Approach for Biometric Monitoring

The framework for this approach is similar to the edit distances used in text searches (and other searches) where the data can be represented by a one dimensional string of symbols, such as letters and numbers. Unlike traditional image recognition algorithms, the features extracted from the data are not used for classification. Instead, a mapping is found between the image to be identified and the database of faces, and properties of this mapping are used for classification. The mappings represent the edit distances previously only associated with one-dimensional data such as text [8, 9] and other data such as DNA [10] that is typically represented by a string of symbols. Once the mappings are found between the test and train images, properties associated with edit distances, namely – insertion, deletion, and substitution errors [8, 9] – are used to measure the degree of similarity between the images.

This approach involves the innovative mapping of two-dimensional image data into insertion, deletion, and substitution errors traditionally associated with one-dimensional strings [11,12 and 13]. Also, techniques used to solve the problem in one dimension, such as dynamic programming, cannot be extended to two-dimensional problems. In this unique approach, properties of the two-dimensional mapping between two images are found which represent deletions, insertions, and substitution (or match) distances in the one-dimensional problem. This technique is robust to variations in lighting and poses, which is critical for applications where surveillance cameras are used and the capture of video and still image data occurs in an uncontrolled environment. An early version of this algorithm was applied to the problem of face recognition [14] as well as image preprocessing and registration for face recognition [15].

The block matching algorithm used for motion estimation in current video coding standards such as MPEG [16, 17, 18, and 19] is the basic framework for the mapping between the database or train image and the unknown or test image. The mapping between test and train images can be applied in a "forward" direction where the mapping is found which converts the test image into the train image, as well as in a "backward" direction where the mapping is found that converts the train image into the test image. The forward and backward mappings are not simply the inverse of one another. The block matching algorithm was first introduced to perform motion estimation and compensation for video compression in order to take advantage of temporal correlations between video frames by estimating the current frame from the previous frame. In a hybrid video coder based on the traditional motion compensation scheme, motion estimation is performed by matching blocks [16] between the original frame and the previously reconstructed frame. An estimate of the current block can be obtained by searching similar blocks in the previous encoded (or original image) frame in a predetermined search area. The block matching algorithm is used for motion estimation between two video frames for compression and, in our case, the block matching algorithm is used for disparity estimation between the test image and each train image. In addressing face recognition, the key differentiator is that we expect the disparity map between the correctly matched faces to have significantly different properties than the disparity maps found for mismatched faces. We use the properties of the disparity fields found in mapping the test image and train images and how they relate to the traditional edit distances used in text for optical character recognition (OCR).

5 Applications of the P-Edit Distance to Face Recognition

While current applications of this face recognition technology include physical and logical access control and face detection, identification, and tracking for surveillance applications, STAR Face system has the potential to enhance the in SA in simulated AugCog environments by providing detection, identification, and tracking of the human element in the visual environment. Figures 3 and 4 illustrate typical surveillance video footage and face detection results for the STAR Face Recognition System.

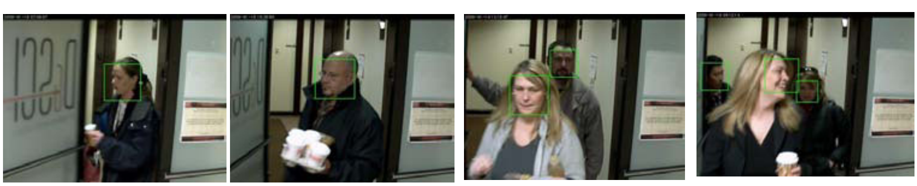


Fig. 3. Face detection results for surveillance data

Star Technologies Surveillance System detects faces from surveillance video to enroll individuals into a database and to recognize them in later footage. When an individual is detected for the first time and a match is not found in the current stored database, the database is updated with a new entry for that individual. Later occurrences of the same individual, either from the same sensor or other sensors, can then be identified. Other information about the individual can also be stored in the database for future reference.

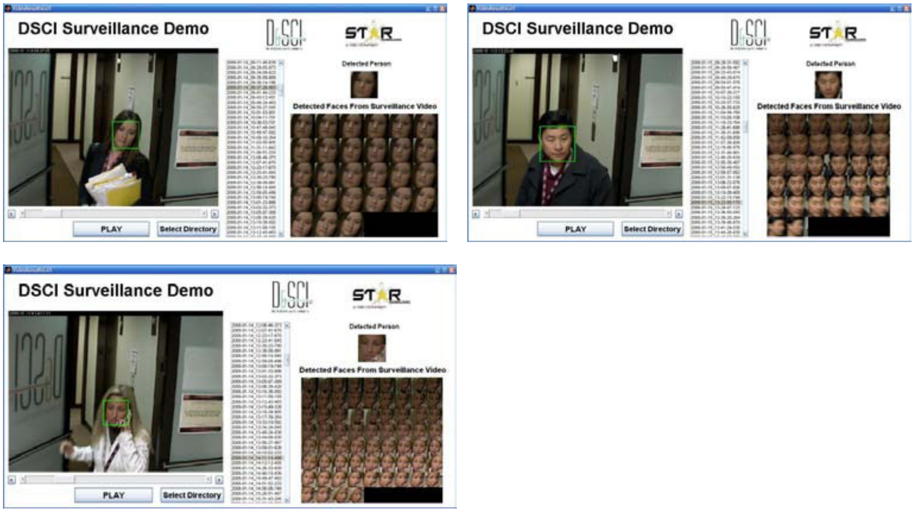


Fig. 4. User interface for the surveillance system and face detection results

6 Summary

Since human cognition is influenced by interaction with the physical environment, in order to measure user state for AugCog intervention, tools to enhance environmental monitoring abilities are needed. Emerging video surveillance and facial recognition/analysis technology has the potential to provide detection, identification, and tracking of targets in the operator's live environment to increase the accuracy of the AugCog system's automated SA. A new methodology for image-based face or object recognition based on extending the concept of edit distances for one-dimensional signals can be extended to the detection, identification, and tracking of individuals in a video sequence and includes a method for detecting and recovering from occlusions. A fast search method allows for fast recognition when the database of individuals to be identified is large or if the amount of data to be processed is large such as in video surveillance applications. The searches can be performed on still images, videos, or a combination of both. Boolean operators are supported so that the user can narrow down the search and filter out unwanted results. Facial recognition technology can be a critical solution for providing the necessary environmental SA for effective diagnosis and driving of AugCog mitigation, particularly in live operating environments.

References

1. Kruse, Schmorrow: Foundations of Augmented Cognition, 1st edn., pp. 441–445. Lawrence Erlbaum Associates, Mahwah (2005)
2. Schmorrow, Reeves: Foundations of Augmented Cognition, 2nd edn., p. XIII. Springer, Germany (2007)

3. Dorneich, et al.: Supporting real-time cognitive state classification on a mobile individual. JCEDM 1(3), 241–242 (2007)
4. Nicholson, et al.: An adaptive system for improving and augmenting human performance. In: Schmorrow (ed.) Foundations of Augmented Cognition, 2nd edn., pp. 215–222. Lawrence Erlbaum Associates, Mahwah (2006)
5. van Maanen, et al.: Closed-loop adaptive decision support based on automated trust assessment. In: Foundations of Augmented Cognition, 2nd edn., p. 267. Lawrence Erlbaum Associates, Mahwah (2007)
6. Shanker, T., Richtel, M.: In: new military, data overload can be deadly. New York Times (January 16, 2011)
7. Parasuraman, Rizzo: Introduction to neuroergonomics. In: Neuroergonomics: The Brain at Work, p. 6. Oxford University Press, New York (2007)
8. Wagner, R.A., Fisher, M.J.: The string-to-string correction problem. Journal of ACM 21, 168–178 (1974)
9. Gusfield, D.: Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology. Cambridge University Press, New York (1997)
10. Altschul, S.: Amino acid substitution matrices from an information theoretic perspective. J. Mol. Biol. 219, 555–565 (1991)
11. Podilchuk, C., Barinov, L., Hulbert, W., Jairaj, A.: Face Recognition in a Tactical Environment. In: IEEE Proceedings of MILCOM (2010)
12. Podilchuk, C., Hulbert, W., Flachsbart, R., Barinov, L.: Face Recognition for Uncontrolled Environments. In: Proceedings of SPIE Defense, Security and Sensing, Biometric Technology for Human Identification VII, Orlando, Florida, April 5-9 (2010)
13. Hulbert, W., Podilchuk, C., Mammone, R.J.: Face Recognition Using a Pictorial-Edit Distance. In: Proceedings of the IEEE International Conference on Image Processing, ICIP 2008, pp. 1908–1911 (2008)
14. Podilchuk, C., Patel, A., Harthattu, A., Anand, S., Mammone, R.: A New Face Recognition Algorithm using Bijective Mappings. In: Proceedings IEEE International Conference on Computer Vision and Pattern Recognition (PAMI 2005), June 20-26, vol. 3, pp. 165–175 (2005)
15. Savvides, M., Xie, C., Chu, N., Kumar, B.V.K.V., Podilchuk, C., Patel, A., Harthattu, A., Mammone, R.: Robust face recognition using advanced correlation filters with bijective-mapping preprocessing. In: Kanade, T., Jain, A., Ratha, N.K. (eds.) AVBPA 2005. LNCS, vol. 3546, pp. 607–616. Springer, Heidelberg (2005)
16. Sikora, T.: MPEG digital video-coding standards. IEEE Signal Processing Magazine 14(5), 82–100
17. Servetto, S.D., Podilchuk, C.I.: Stochastic modeling and entropy constrained estimation of motion from image sequences. In: Proceedings of the IEEE International Conference on Image Processing (ICIP 1998), vol. 3, pp. 591–595 (1998)
18. Han, S.-C., Podilchuk, C.I.: Efficient encoding of dense motion fields for motion-compensated video compression. In: Proceedings of the IEEE International Conference on Image Processing (ICIP 2000), vol. 1, pp. 84–88 (2000)
19. Han, S.-C., Podilchuk, C.I.: Modeling and coding of DFD using dense motion fields in video compression. IEEE Transactions on Image Processing 10(11), 1605–1612 (2001)

Analysis of Multiple Physiological Sensor Data

Lauren Reinerman-Jones[1], Grant Taylor[1], Keryl Cosenzo[2], and Stephanie Lackey[1]

[1] University of Central Florida, Institute for Simulation and Training
3100 Technology Pkwy, Orlando, FL 32826
[2] Army Research Laboratory
Aberdeen Proving Ground, MD
lreinerm@ist.ucf.edu

Abstract. Physiological measures offer many benefits to psychological research including objective, non-intrusive assessment of affective and cognitive states. However, this utility is limited by analysis techniques available for testing data recorded by multiple physiological sensors. The present paper presents one set of data that was attained from a repeated measures design with a nominal independent variable for analysis. Specifically, the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2008), a series of images known to convey seven different emotions, was presented to participants while measures of their neurological activity (Electroencephalogram; EEG), heart rate (Electrocardiogram; ECG), skin conductance (Galvanic Skin Respond; GSR), and pupillary response were taken. Subsequently, a discussion of statistics available for analyzing responses attained from the various sensors is presented. Such statistics include correlation, ANOVA, MANOVA, regression, and discriminant function analysis. The details on design limitations are addressed and recommendations are given for employing each statistical option.

Keywords: EEG, ECG, Eye Tracking, Statistical Analyses, Emotion.

1 Introduction

Psychological research benefits from the implementation of physiological measurement techniques as a way to assess and predict performance. Physiological measures index cognitive state and resources used. Along those lines, resource theory is supported and cognitive states indexed include workload, stress, fatigue, and emotion. Unlike surveys and questionnaires that require task interruption to be administered, physiological measures are continuous and tend to be relatively non-intrusive. The advantage is that state is assessed throughout an entire task, not just before or after task completion. Thus, dynamic changes are recorded and a detailed understanding of the phenomenon under investigation is provided. Additionally, physiological measures provide an objective method for evaluating state, unlike questionnaires that fall subject to bias. Ultimately a multi-dimensional approach employing physiological, subjective, and performance measures is best to account for the most variance, but a couple challenges need to be addressed with regard to physiological assessment.

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 112–119, 2011.

Physiological measures are currently limited by two factors. First, different physiological measures tend to not strongly correlate with one another. For example, results attained using Electroencephalogram (EEG) typically have low correlations with Electrocardiogram (ECG) responses recorded simultaneously on the same task. This makes it difficult to determine which measure to discuss for interpretation and which is accurate. Thus, there is no standardization throughout literature and meta-analysis is challenging. Second, individual difference is an influential factor across physiological measures and is the reason for taking baseline readings. Eye tracking is the exception to the rule of needing a baseline. The restriction is that a system employing physiological measures must be calibrated to the individual user and every time the operator changes, this calibration process must occur. Physiological response during a task is compared to the baseline reading taken before a task to determine the amount of resources utilized and change in state. The goal is to find an analysis technique that allows researchers to create a physiological profile of overall state that can then be used as a model for all individuals, enabling a percentage or type of change to inform a closed-loop system for any person entering the system. To clarify, a physiological profile for workload might be described as an increase in certain EEG activity, perspiration, heart rate, and pupil dilation.

Keeping with that effort, the current paper provides a review of statistics as applied to one physiological data set. Knowledge certainly is gained by analyzing physiological variables separately, but their greatest potential for usefulness lies in the ability to perform multivariate analyses. Univariate analysis determines the effect of the independent variable (IV) on the dependent variable (DV) in isolation, but multivariate analysis lends itself to investigating potentially complex interactions that occur between the many variables, thus moving closer to identifying state profiles. Easily accomplished with some study designs, other data structures do not lend themselves to multivariate analyses. The aim for the present discussion is to investigate the potential solutions to analyzing data collected from one experimental design – a repeated measures design with a nominal independent variable.

2 Method

2.1 Participants

Forty-six participants ranging in age from 18-40 years volunteered from the University of Central Florida.

2.2 Procedure

The present study examined the influence of emotion on various physiological responses. The International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2008), a series of images known to convey seven different emotions, was presented to participants while measures of their neurological activity (Electroencephalogram; EEG), heart rate (Electrocardiogram; ECG), skin conductance (Galvanic Skin Respond; GSR), and pupillary response were taken. Specifically, participants were required to watch a computer monitor as 42 images were presented, meaning that participants saw six images of each of the emotional categories – amusement, awe, contentment, disgust, excitement, fear, and sadness. The presentation of the images was randomized, with each participant observing a uniquely randomized order. No physical or verbal response was required. Each image was presented for six seconds with a six second inter-stimulus interval.

Before the study began, each participant completed a series of baseline tasks to account for individual differences in physiological activity. The baseline for EEG was Advanced Brain Monitoring's (ABM) three task baseline battery consisting of a visual stimulus-response eyes open task, auditory stimulus-response eyes closed task, and a short vigil. Analyses used data computed from the change in power measured during the task compared to the power measured during baseline. A five minute resting baseline (with eyes closed) was recorded for ECG and GSR for use as a comparison of the recordings obtained during the experimental task. A baseline was not required for pupil diameter.

3 Results and Discussion

To attack the problem of analyzing physiological data, several statistical analyses were conducted. The two most common methods for revealing trends and effects were performed: correlation and Analysis of Variance (ANOVA). As previously mentioned these often provide interesting results, but are limited. Additional analyses were explored for comparing the multiple DVs present when using physiological measures. Multivariate Analysis of Variance (MANOVA), regression, and discriminant function analysis (DFA) were discussed as potential options with the understanding that the application of such analyses would need to be performed using a different experimental design.

Table 1. Intercorrelations between physiological measures

*correlation is significant at $\alpha = .05$.

	IBI	HRV	GSR
HRV	$r = 0.944$* $p < 0.001$		
GSR	$r = 0.014$ $p = 0.547$	$r = 0.010$ $p = 0.660$	
Pupil	$r = 0.003$ $p = 0.885$	$r = 0.040$ $p = 0.084$	$r = 0.179$* $p < 0.001$

3.1 Correlation

Correlation data is often a valuable starting point for analysis. However, this approach is unlikely to reveal insight from the physiological measures, as week correlations often result. In this particular study a total of 31 data points were obtained for each stimulus presentation [27 EEG values (Nine Sensor Sites: F3, Fz, F4, C3, Cz, C4, P3, P0, P4), interbeat interval (IBI), heart rate variability (HRV), galvanic skin response (GSR), and pupil diameter]. The average intercorrelation between these variables, taking the absolute value of all 31 points so as to eliminate a possible artifact of averaging positive and negative numbers, was found to be moderately weak (r=0.2870). Even the intercorrelations within the EEG data produced a moderately weak average (r=0.346). The remaining correlations can be seen in *Table 1*.

3.2 Analysis of Variance (ANOVA)

Repeated measures ANOVA is arguably the most practical analysis to conduct for the given dataset because the design and assumption requirements are met. By conducting a series of ANOVAs with each of the measurements (EEG values, interbeat interval (IBI), heart rate variability (HRV), galvanic skin response (GSR), and pupil diameter) obtained by the physiological sensors as the dependent variable and emotion type as the independent variable, a basic understanding of how the different physiological systems respond to emotions is obtained.

ECG. The ECG data was processed to form two separate variables: inter-beat interval (IBI, the inverse of heart rate) and heart rate variability (HRV). As evidenced by the very strong correlation between these two variables (r = .944, p < .001), they are roughly equivalent. This equivalence is likely due to the brief time periods over which each data point is sampled. Over longer time periods, these two methods of processing ECG data would likely produce two distinct values, but it seems six seconds is not sufficient to adequately differentiate between the two. The ANOVA results show virtually identical results for both variables. Significant main effects for emotion were found for both IBI [F(6, 264) = 7.435, p < .001] and HRV [F(6, 264) = 7.406, p < .001]. Subsequent pairwise comparisons found identical patterns for both variables, the only difference being the direction of the group differences, which is a result of the algorithms used to process the two values. The direction of results of the IBI pairwise comparisons are presented in *Figure 1*.

- Sadness > Amusement, Awe, Contentment, Excitement, and Fear
- Disgust > Awe, Contentment, Excitement, and Fear
- Amusement > Excitement
- Amusement < Sadness
- Awe, Contentment, and Fear < Disgust and Sadness
- Excitement < Amusement, Disgust, and Sadness

Fig. 1. IBI pairwise comparison results. All listed comparisons are significant at α = .05.

GSR. GSR data was computed as the average skin conductance measured over the six seconds of each image presentation. A repeated measures ANOVA found no significant effect of emotion category on GSR [F(6, 264) = 0.958, p = .454].

Pupillary Response. The pupil response was measured using a head-mounted camera, which recorded the average pupil diameter for each image presentation. The ANOVA showed that pupil diameter was significantly influenced by emotion category [F(6, 270) = 19.903, p < .001]. Details from the pairwise comparisons are presented in *Figure 2*.

- Awe < Amusement, Contentment, Disgust, Excitement, Fear, and Sadness
- Excitement < Amusement, Disgust, Fear, and Sadness
- Excitement > Awe
- Contentment < Amusement, Disgust, and Fear
- Contentment > Awe
- Sadness < Disgust and Fear
- Sadness > Awe and Excitement
- Amusement > Awe, Contentment, and Excitement
- Disgust and Fear > Awe, Contentment, Excitement, and Sadness

Fig. 2. Pupil diameter pairwise comparison results. All listed comparisons are significant at $\alpha = .05$.

EEG. The large amount of EEG data required a much more complex analysis. Each image generated 27 separate EEG values (power within three separate wavelengths across nine separate sensors), therefore simple one-way ANOVAs were not sufficient. A 7 (emotion category) x 3 (anterior/posterior sensor position) x 3 (lateral sensor position) x 3 (wavelength) repeated measures ANOVA was conducted. Anterior/posterior sensor position indicates the location of the sensor on the head with three sensors each covering frontal, central (prefrontal), and parietal areas. Lateral sensor position further defines the sensor's location, with three sensors each covering left, central (mid-sagital), and right areas. This analysis matrix enabled testing complex interactions, which could potentially determine how different wavelengths within separate brain areas were affected by various emotions. Unfortunately increasing the complexity of the analysis, while improving its potential insight, also increases the difficulty of interpretation.

The results of the EEG ANOVA showed a significant main effect for emotion [$F(6, 150) = 4.399$, $p < .001$], such that Fear resulted in less EEG power than Amusement, Awe, Contentment, Excitement, and Sadness, while Disgust was less than Amusement, Contentment, and Sadness. A significant interaction was shown between emotion, wavelength, and lateral position [$F(24, 600) = 1.792$, $p = .012$]. To further analyze this interaction, an emotion x lateral position ANOVA was conducted for each wavelength. Within the alpha band, there was a significant main effect for emotion [$F(6, 192) = 4.116$, $p = .001$], such that Fear had significantly less power than all other emotions and there was no significant emotion by lateral position interaction. Significant emotion by lateral position interactions were found for both beta [$F(12, 360) = 2.709$, $p = .002$] and theta activity [$F(12, 408) = 2.143$, $p = .014$]. These significant interactions required the completion of one-way ANOVAs for emotion within each category of lateral position for both beta and theta activity. No significant effect was found for emotion on either left or central sensors for either beta or theta activity. However, as shown in *Figure* 3, there was a significant main effect for emotion on the right sensors within both beta [$F(6, 216) = 9.018$, $p < .001$], and theta [$F(6, 252) = 4.307$, $p < .001$].

- Beta
- Fear and Disgust < Amusement, Awe, Contentment, Excitement, and Sadness
- Theta
- Fear < Amusement, Contentment, Excitement, and Sadness
- Disgust and Awe < Amusement, Contentment, and Sadness

Fig. 3. EEG activity recorded from the right hemisphere. All listed comparisons are significant at $\alpha = .05$.

To summarize the complexity of these results, EEG power does reflect changes in emotion. Specifically, Fear, Disgust, and Awe tend to result in less EEG power, but this effect only occurs in the right hemisphere and only within the beta and theta wavelengths with the effect varying slightly between the two wavelengths.

3.3 Multivariate Analysis of Variance (MANOVA)

At a glance, MANOVA appears to be the best option for analyzing multiple physiological measures. However, the use of MANOVA does not yield a great amount of additional information over the use of individual ANOVAs. MANOVA creates a linear combination of all of the dependent variables and then a traditional ANOVA is conducted on this newly calculated DV. Therefore, the analysis only reveals whether the IV (emotion) has an effect on the combination of all of the DVs (physiological response from each sensor). Assuming this analysis is significant, the only real knowledge gained is that emotion has some effect on some aspect of the physiological measures, but individual ANOVAs (as discussed above) must still be conducted to determine the specifics of this effect.

3.4 Regression

Regression could be capable of providing a great deal of information about the data, but was unable to be used to analyze the current dataset because emotion is a nominal variable. Regression could still be used with emotion as a variable by dummy coding it into a series of dichotomous variables acting as predictors and the physiological measures as dependent variables. However, analysis would be limited to one DV at a time, which does not provide an understanding of how the physiological responses interact or vary together for a given state. The true potential for regression would only be possible if emotion were comprised of interval or ratio data. This would allow emotion to be entered into the model as if it were the dependent variable (to be predicted) with the many physiological variables included simultaneously as predictors. In addition to showing the relationships between each physiological measure and emotion, this analysis would enable the evaluation of complex interactions between physiological measures (though these terms are not computed automatically and so would need to be generated manually by the researcher) and would test for mediation between variables. Given the extent of the additional analyses possible when employing regression, it is unfortunate that it is not an option for the current dataset. Nevertheless, any study in which the independent variable can be quantified on some continuous scale, for example if the same study were conducted by varying image size, regression should be considered as one of the primary analysis techniques.

3.5 Discriminant Function Analysis

Thus far, regression affords the greatest potential for understanding the true underlying physiological processes of emotion. However, DFA is likely to provide the most benefit toward more practical applications because the result of DFA is a simple classification algorithm, which weights each of the predictor variables according to their relationship with the predicted value (emotion in this case). This algorithm could then be used to predict the emotion a person is experiencing based solely on their physiological responses. As with regression, DFA is again incapable of being used to analyze the current dataset, this time being restricted by the repeated-measures design of the study. Traditional DFA assumes that each observation is independent of all others, an assumption that is clearly violated by the current study. Fortunately, Roy and Khattree (2005) have begun adapting traditional DFA methods for analyzing repeated-measures data. No single set of standards has yet been agreed upon and as such, no statistical analysis software has adopted this new implementation of DFA. As a result, it would be both difficult and inadvisable for researchers to attempt to utilize these new methods at this time, but their continued development should be monitored. DFA for a repeated-measures design will truly be a powerful tool.

4 Conclusion

The endeavor to effectively capitalize on the potential offered using physiological measures for assessing human state and performance is complicated. Correlation and ANOVA provide direct methods for analyzing physiological data. However, the limitations for explaining in greater depth overall physiological response recorded by multiple sensors led to the search for more comprehensive analyses. Regression and DFA appear to provide the greatest utility for analyzing physiological measures, but are also limited by design and scale requirements. Therefore, a carefully designed study should be conducted that addresses evaluating the effects of multiple physiological measures used to classify overall physiological response for phenomena such as emotion, workload, stress, and other states. The rise of physiological assessment implementation into all types of human research demands effort put forth to discovering and using the best analyses. The present paper is the start of that journey and provides one study example that clarifies the types of analyses available for physiological measures and when to use each. The challenge to researchers utilizing physiological measures, whether for Brain-Computer Interface (BCI), Augmented Cognition, or Neuroergonomics, is to stretch the limits and attain deeper insight through the best analyses possible.

Acknowledgments. This work was funded by Alion IDIQ Subcontract Agreement No. 8005.00X.10 in support of Army Research Laboratory Prime Contract DAAD19-01-C-0065, Delivery Order #119 along with Alion P.O. #STM1196782.

References

1. Lang, P.J., Bradley, M.M., Cuthbert, B.N.: International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8. University of Florida, Gainesville, FL (2008)
2. Roy, A., Khattree, R.: On discrimination and classification with multivariate repeated measures data. Journal of Statistical Planning and Inference 134, 462–485 (2005)

Exploring New Methodologies for the Analysis of Functional Magnetic Resonance Imaging (fMRI) Following Closed-Head Injuries

Peter B. Walker[1] and Ian N. Davidson[2]

[1] Naval School of Aviation Safety
[2] University of California – Davis
peter.b.walker@navy.mil, davidson@cs.ucdavis.edu

Abstract. An increasing amount of research has focused on the use of newer and alternative data analytic approaches to multi-dimensional data sets. The primary aim of this paper is to introduce two data analytic approaches as they have been applied to image scans from functional Magnetic Resonance Imaging (fMRI). The first approach involves loading data from fMRI scans into multi-dimensional cubes and performing tensor decomposition. In addition, we introduce a second approach involving the use of network modeling that attempts to identify stable networks in fMRI scans across time. Discussion will be focused on the application of these approaches to the modeling and rehabilitation following closed-head injury.

Keywords: fMRI, Tensor Decomposition, Graph/Network Modeling.

1 Introduction

There has been an increasing need over the past twenty years for the military services to develop research programs dedicated to the understanding and application of neurosciences in operational settings. This "revolution" in the neurosciences, now known as Operational Neuroscience, has resulted in the emergence of an entire discipline dedicated to the application of those principles to warfighters in the field [1, 2]. However, as research programs continue to seek to integrate both basic and applied sciences to maximize the effectiveness of these warfighters, there continues to be a need to explore new and/or alternative approaches to analyze human performance and physiological data more effectively.

Recently, there has been a great deal of interest in the application of exploratory data analysis techniques such as Singular Value Decomposition (SVD) and Principle Components Analysis (PCA) to the analysis of data from functional imaging studies such as functional Magnetic Resonance Imaging (fMRI). Data analytic approaches such as SVD and PCA, when applied to fMRI, are limited in that they are based on matrix calculations where the data may be defined in only two dimensions (i.e., time and location) [3, 4]. However, this form of data analysis, by making the data two dimensional, abstracts out important details such as the identification of activation patterns over time.

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 120–128, 2011.

Techniques such as SVD and PCA attempt to portray the data from a single imaging study as a two dimensional data matrix (*X*). Accordingly, data matrix *X* can be further decomposed into a sum of R outer products of individual factors by:

$$X = \sum_{r}^{R} a_r \otimes b_r + E. \tag{1}$$

During the decomposition of this data matrix, spatiotemporal properties from the fMRI scan are encoded as vectors a_r (spatial properties) and b_r (temporal properties). Noise from the functional image is represented as a constant E. The relationships between spatio and temporal data from the fMRI scan can be discovered using a variable number of data analytic processes. Regardless of the data analytic process implemented, both SVD and PCA apply matrix computations and factorize a single two-dimensional data matrix into time courses and spatial maps [3].

A more thorough explanation of a particular dataset might require the simultaneous analysis of three or more dimensions of data (i.e., time, location, and stimulus). Within the past ten years, there has been a proliferation of research attempting to explore newer and alternative data analytic approaches to multi-dimensional data [5, 6, 7]. With respect to Operational Neuroscience, these approaches offer a great deal of promise due to their ability to efficiently identify patterns in very dense data sets.

A promising new approach for the analysis of multi-dimensional data involves the use of tensors. For example, the use of tensors allows the analysis of the fMRI data, without compromise, by representing the data as a four dimensional object (x y z locations on a fMRI scan over time). Simply put, a tensor is a generalization of a matrix (or scalar or vector) to more than two dimensions. Multi-dimensional data can be viewed such that each dimension of a particular dataset might represent a different aspect or characteristic. In the case of the fMRI of a single person: the four dimensions correspond to location and time, or in the case of more than one person, five dimensions and so forth. Figure 1 shows a simple order-three tensor where location has been simplified to be viewed in three dimensions.

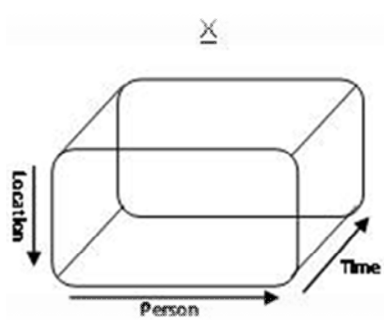


Fig. 1. Order-three tensor illustrating a functional image. This figure illustrates how data from an fMRI scan can be extended to an order-3 (or greater) tensor. Dimensions on the tensor include dimensions for (i) individuals, (j) locations, and (k) activation over time.

Tensors Defined. As mentioned previously, tensors can be viewed as a data cube whereby each dimension of the cube represents some aspect or characteristic of the dataset. For functional imaging scans, a 3rd order tensor can be used to represent three aspects of the dataset: individual, location of activation, and how the activation pattern has changed over time (see Figure 1). A single entry (i, j, k) in the tensor then, corresponds to a single individual (i), location of activation for that individual (j), and the change in activation of that individual's scan over time (k).

2 BOLD Analysis

Brain imaging techniques such as MRI and, more recently, fMRI have been used in a variety of experimental and clinical studies to investigate phenomena ranging from working memory to traumatic brain injury [5]. The most likely explanation for the continued use of these imaging approaches is the ability of these paradigms to portray information processing in the brain as it occurs in real time.

Perhaps the most common approach to measuring brain activity is through the use of fMRI and Blood Oxygen Level Dependence (BOLD). Measurement of BOLD level assumes that increased neuronal activity requires more glucose and oxygen to be rapidly delivered through the blood stream. The ratio of oxygenated and deoxygenated blood in a particular area is therefore presumed to represent brain activation during a specific task. Functional Magnetic Resonance Imaging has revolutionized the behavioral sciences by offering spatial and temporal resolutions far exceeding brain imaging techniques available in the past. Since its introduction during the early 1990s, thousands of studies have been conducted examining a range of issues including structure, pathology, and processing.

While BOLD measurements are commonly viewed as the 'gold standard' in neuroscience today, there are growing concerns over the reliability of fMRI findings and the interpretation of their results. For example, BOLD fMRI is often referred to as a relative technique in that it attempts to compare images taken during one mental state to different scans of the same individual in another state. Series of fMRI scans are aggregated to measure the relative differences between two states to perform a statistical analysis within a single individual.

Similarly, neuroimaging studies usually involve the analysis of scans from several individuals taken from several different sessions. For analysis techniques such as SVD and PCA, this results in the aggregate of data across individuals and/or time. Therefore, these types of data analytic approaches may result in the inability to identify specific individual differences across different imaging scans. Therefore, for the purposes of identifying individual differences across scans, a more suitable data analytic approach is one that involves the analysis of multiple data sources all at once [3].

3 Tensor Applications to fMRI Analysis

Due to many of the aforementioned limitations in fMRI, there has been an increased interest in the application of multi-dimensional data analytic tools for functional imaging analysis. Here, we describe the application of PARAFAC decomposition to fMRI. Other more complex decompositions exists but we shall use this simplest of decompositions to illustrate our points.

3.1 PARAFAC Decomposition

Our primary criticism to previous approaches to the analysis of fMRI data is the aggregation across various data sources to limit the data to two dimensions. However, often times it is more meaningful to identify patterns in the dataset across more than two dimensions. For example, suppose that there is a group of functional images that identify a pattern of activation for a specific cognitive task (i.e., spatial rotation task). In addition, there may be one or more patients in that population that performs poorly on that task due to some preexisting trauma. One goal of the data analytic approach might be to then identify those individuals that performed poorly on the cognitive task and those that performed equivalent to the normal group. Moreover, we might be interested in identifying how the pattern of activation for the two groups different across time.

The approach we outline here, constrained PARAFAC decomposition, overcomes many of the aforementioned limitations by using a Low-rank tensor approximation. This process involves loading images from an fMRI scan into a multi-dimensional tensor. After the image is loaded into the tensor, a PARAFAC decomposition (see Figure 2) is performed such that each slice may be analyzed independently at separate locations and at different times [7, 8].

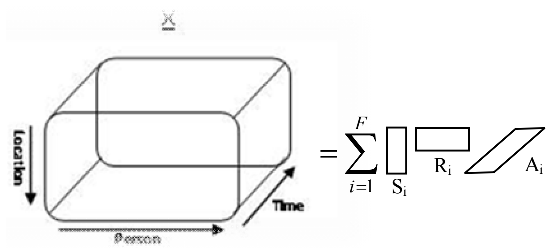


Fig. 2. PARAFAC Decomposition of FMRI. This figure illustrates the decomposition of F factors (i), each of which describes a time, individual, and location of activation. The decomposition of the tensor varies depending upon the relationship between the different dimensions in the data cube.

The PARAFAC decomposition is accomplished by representing three (or more dimensions) of data by a trilinear combination of three outer products:

$$X = \sum_{r}^{R} a_r \otimes b_r \otimes c_r ... \otimes i_r + E. \tag{2}$$

For the purposes of fMRI, properties such as spatial location (a_r), the individual (b_r), and/or the activation pattern as it occurs over time (c_r) are each encoded as vectors.

However, other properties of the functional image may also be projected within the multi-dimensional data cube. Regardless of the number of dimensions, decomposition of the tensor allows for the identification of specific relationships between any or all vectors of the tensor.

A novel computation of our work is to explore constrained tensor decomposition. Regular decompositions will find the mathematical optimal decomposition but this may yield non-actionable results. For example, the outer product of a_i and b_i may yield a non-contiguous activation area or the activation level (the t dimension) may be non-smooth or multi-modal. In our work we explore constraining the decomposition so that these and other issues which may make the decomposition difficult to interpret are constrained not to occur

3.2 fMRI Interpretation

For the present study, we used a rank 10 tensor to approximate the original tensor. The tensor approximation allows us to view brain activation as a multi-dimensional process. In the case presented below, we are representing BOLD activation in a three dimensional space as it unfolds over time. However, the tensor decomposition approach allows us to model activation for any number of dimensions.

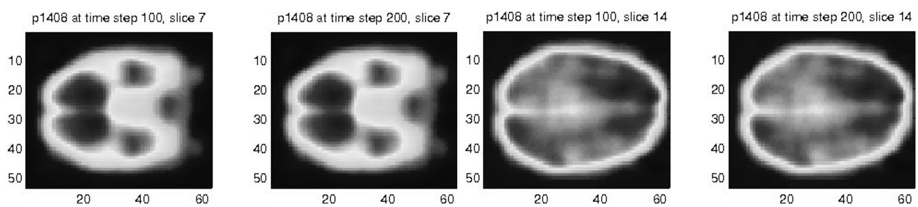


Fig. 3. Visualization of functional images from the same patient at different time steps. Note that similar activation patterns are clearly identified for different time series.

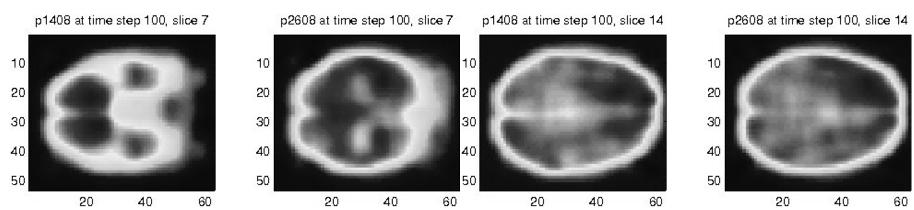


Fig. 4. Visualization of functional images from different patients at the same time steps. Note that very different activation patterns are clearly identified for different image slices.

To further illustrate this process, the tensor decomposition process was applied to several fMRI images from patients while at rest. Since a 53×63×28×235 tensor can also be considered as 53×63×28 tensors over 235 time steps, we compared these tensors (from the same patient) over time. Not surprisingly, when this process was applied to the same patient at rest we found that the tensors do not differ much. However, when comparisons from two tensors from two different patients were made, the resulting tensors were quite noticeably different (see Figures 3and 4). These conclusions were consistent for both the original tensor and its approximation. Together, these results provide converging evidence that the decomposition approach does provide an alternative analytic approach to brain imaging techniques.

4 Network/Graph Analysis of fMRI

Though PARAFAC analysis has many advantages, it is limited in that it does not (in its basic form) consider the spatial relationships between the different locations from an fMRI scan. When searching for interactions, PARAFAC treats different but adjacent locations from the functional scan the same as locations that are far apart. This can, in turn, lead to the inability to identify factors which are spatially diverse and not contiguous. Furthermore, tensor analysis also has other inherent limitations such as requiring a symmetrical distance which is implicitly defined. A symmetrical distance function requires the distance from a to b to be the same as b to a and this is not often the case. Therefore, we have also applied principles of network/graph analysis to fMRI data in an attempt to overcome some of these limitations. The benefit of a network/graph analysis of fMRI data is that these spatial relationships and preferences can be directly encoded.

Network/graph analysis attempts to identify both symmetrical and asymmetrical relationships between discrete objects. A graph can be viewed as an abstract representation of a network consisting of nodes and a set of edges (or connections). An edge that connects two nodes suggests there is a relationship between both nodes in the graph (see Figure 5) and the weight of the edge indicates a measure of distance or similarity between the nodes.

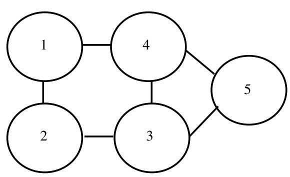


Fig. 5. Visualization of a graph or network. Each node in the network may refer to the activation of a particular area during functional imaging. Edges connecting various nodes in the network suggest that different locations may have a similar pattern of activation.

Formally, graphs can be represented as an adjacency matrix A. Edges that connect two nodes in the graph are represented in the adjacency matrix as $A_{i,j} = 1$. However, if no relationship exists between the two nodes, then the adjacency matrix represents that relationship as $A_{i,j} = 0$. Furthermore, the number of edges connecting to a particular node is described as the degree k. Therefore, the probability P(k) that a randomly chosen node will have degree k is given by the degree distribution [9].

The edges in any particular network can be represented as either nondirectional, where the relationship between each node is homogenous, or directional, where the relationship between the nodes may be heterogeneous (one node influences the other). This concept is important with regard to modeling fMRI activation. Specifically, this allows us to model patterns of activation such that some areas of activation may facilitate the activation of other areas in the brain. However, that relationship need not be reciprocal.

The network model we use is an example of an enhanced 2-Graph. In this network/graph, each voxel from an fMRI scan is treated as a separate node in the network. Similarly, there exist edge weights between nodes indicative of their spatial distance or some other measure of similarity. However, the analyst can set these edge weights to reflect what-ever relationship they wish the analysis to emphasize. At each node/voxel is the behavior of that node over time as given by the fMRI scan. In this way, the network analysis can be viewed as a tensor with encoded spatial (or other) information in the form of edge weights.

The benefits of edge weights as a tool to allow analysts to emphasize their domain expertise cannot be over-emphasized. The relationship between voxels need not be symmetrical for a pair of voxels or even provided if it is not known. Furthermore, graphs are a more natural interpretation (extension) of the way in which we theorize processing to occur in the brain. That is, it is common to view the brain as being composed of very discrete structures separated by geographic boundaries within the brain. Therefore, it would seem apparent of the need to apply a discrete modeling process. Similarly, this discrete modeling process takes into account many of the spatial properties that were alluded to earlier.

The study of graph theory and its application to neuroscience is an important area of focus. For example, it has been suggested that this approach can be used as a methodology for identifying functional clusters of brain activity during experimental tasks. However, it is critical to first identify under what boundary conditions this form of data analytic approach might be used.

The analysis of such a complex graph is an area of ongoing research. Recently, Davidson and collaborators showed how to analyze such graphs so as to segment them [10], project them into lower dimensional space [11] and perform multilabel prediction [12]. However, very little research has been conducted to examine the application of these approaches to fMRI data [13, 14].

5 Diagnosis and Rehabilitation Following Closed-Head Trauma

FMRI is commonly recognized as a premier modality for imaging brain physiology and tracking neural correlates of plasticity. The increase in popularity of functional imaging paradigms is due to the ability of these imaging technologies to view the activated brain during specific tasks. However, functional imaging modalities are not immune to their own criticisms. As discussed previously, these imaging techniques are often limited to very coarse data analytic techniques. The techniques outlined in this paper are an attempt to formalize alternative data analytic techniques that might identify more granular patterns in the data set. Specifically, the use of tensor decomposition or network/graph analysis is able to efficiently analyze interactions in the data that might occur across individuals, localized activation from the scan, or activation patterns across time.

However, there has been little progress in the use of imaging technology on patients following closed head trauma [15]. Unfortunately, it is these types of functional imaging modalities that offer the most promise with respect to the diagnosis and rehabilitation of patients following closed-head trauma. Ultimately, it is our hope to establish a paradigm for the diagnosis of and eventual rehabilitation of patients following traumatic brain injury. Specifically, it is believed that the approaches highlighted above can be used to identify intact neural pathways for patients following traumatic brain injury. In addition, these approaches may also allow for the identification of those neural pathways that might bypass neural pathways affected by cortical injury or strengthen those pathways that promote neural plasticity.

6 Concluding Remarks

Currently, there is a gap that exists between the basic science of brain imaging technology and the implementation of that science in the operational environment. Operational Neuroscience, to be effective, must continue to utilize recent innovations in both basic and applied research. Ultimately, operational neuroscience should continue to challenge the boundaries that define the limits of human performance. The primary aim of this paper was to introduce alternative data analytic approaches to fMRI. These approaches, if proven to be efficient and scalable, might be used to supplement the use of neuroimaging tools in operational settings.

Acknowledgements. The research reported in this paper was supported by Office of Naval Research Grants NAVY 00014-09-1-0712 and NAVY 00014-11-1-0108. The opinions of the authors do not necessarily reflect those of the United States Navy. We would also like to acknowledge Owen Carmichael from the Alzheimer's Disease Institute at the University of California – Davis for his assistance in data collection.

References

1. Kollmorgan, L.S.: A Case For Operational Approach in Advanced Research Projects – The Augmented Cognition Story. Aviat. Space Environ. Med. 78(1), 1–3 (2007)
2. Schmorrow, D., Reeves, L.M.: 21st Century Human Systems Computing: Augmented Cognition for Improved Human Performance. Aviat. Space Environ. Med. 78(1), 7–11 (2007)
3. Beckmann, C.F., Smith, S.M.: Tensorial extensions of independent component analysis for multisubject FMRI analysis. Neuroimage, 294–311 (2005)
4. Sun, J., Tao, D., Faloutsos, C.: Beyond streams and graphs: Dynamic tensor analysis. In: KDD 2006 (2006)
5. Sun, J., Charalampos, T.E., Hoke, E., Faloutsos, C., Tina, E.R.: Two heads better than one: Pattern discovery in time evolving multi-aspect data. Data Mining & Knowledge Discovery, 111–128 (2008)
6. Belliveau, J.W., Kennedy, D.N., McKinstry, R.C., Buchbinder, B.R., Weisskoff, R.M., Cohen, M.S., Vevea, J.M., Brady, T.J., Rosen, B.R.: Functional mapping of the human visual cortex by magnetic resonance imaging. Science 254, 716–719 (2006)
7. De Vos, M., Vergult, A., De Lathauwer, L., De Clercq, W., Van Huffel, S., Dupont, P., Palmini, A., Van Paesschen, W.: Canonical decomposition of ictal scalp EEG reliably detects the seizure onset zone. Neuroimage 37, 844–854 (2007)

8. Morup, M., Hansen, L.K., Hermann, C.S., Parnas, J., Arnfred, S.M.: Parallel Factor Analysis as an exploratory tool for wavelet transformed event-related EEG. Neuroimage, 938–947 (2006)
9. Stam, C.J., Reijneveld, J.C.: Graph Theoretical Analysis of Complex Networks in the Brain. Nonlinear Biomedical Physics 1, 3 (2007)
10. Wang, X., Davidson, I.: Flexible constrained spectral clustering. In: KDD 2010, pp. 563–572 (2010)
11. Davidson, I.: Knowledge driven dimension reduction for clustering. In: IJCAI, pp. 1034–1039 (2009)
12. Qian, B., Davidson, I.: Semi-supervised dimension reduction for multi-label classification. In: AAAI (2010)
13. Salvador, R., Suckling, J., Coleman, M.R., Pickard, J.D., Menon, D., Bullmore, E.: Neurophysiological architecture of functional magnetic resonance images of human brain. Cereb Cortex 15, 1332–1342 (2005)
14. Salvador, R., Suckling, J., Schwarzbauer, C., Bullmore, E.: Undirected graphs of frequency-dependent functional connectivity in whole brain networks. Phil. Trans. R Soc. Lond. B 360, 937–946 (2005)
15. Ricker, J.H., Hillary, F.G., DeLuca, J.: Functionally Activated Brain Imaging (O-15 PET and fMRI) in the Study of Learning and Memory after Traumatic Brain Injury. J. Head Trauma Rehabil. 16(2), 191–205 (2006)

Part II

Neuroscience and Brain Monitoring

EEG Knows Best: Predicting Future Performance Problems for Targeted Training

Gwendolyn E. Campbell[1], Christine L. Belz[1], Charles P.R. Scott[2], and Phan Luu[3]

[1] Naval Air Warfare Center Training Systems Division, Code 4651, 12350 Research Parkway, Orlando, FL 32826, USA
[2] Kaegan Corporation, 12000 Research Parkway #164, Orlando, FL 32826, USA
[3] Electrical Geodesics, Inc., 1600 Millrace Drive, Suite 807, Eugene, OR 97403, USA
{Gwendolyn.Campbell,Christine.Belz}@navy.mil,
CScott@Kaegan.com, PLuu@EGI.com

Abstract. Many uses for neurophysiological data in training have been proposed in the literature [6], [10], and [11]. However, to date it has not been demonstrated that the use of EEG yields performance diagnoses that are actually more accurate. The current study investigated the capability of EEG to accurately diagnose performance difficulties by examining the predictive ability of an accurate diagnosis on future performance. The data from this study suggests that using EEG to filter a trainee's performance data prior to analysis on a computer based tank identification task yields a more accurate diagnosis than analyzing the data with the traditional statistical methods.

Keywords: electroencephalography, training, neurophysiology.

1 Introduction

Recent advances in our ability to measure and interpret brain activity have sparked great enthusiasm and optimistic speculation about the potential benefits this technology may bring to a diverse collection of domains. Training and education, in particular, appear to be well suited to reap the benefits of the developments in both the theory and the measurement capabilities of the brain sciences. For example, Poulsen, Luu, & Tucker [1] argue that a more accurate and detailed articulation of a theory of the neurophysiology of learning and memory is almost certain to provide insights into the causal mechanisms of effective instructional interventions and may inspire the development of new strategies for facilitating and motivating learning processes.

Even without utilizing the theoretical advances, the improved measurement capability could have a significant impact in supporting training and educational goals. Across the last decade, a large variety of potentially relevant, physiologically-based indices have been proposed and investigated. These can be loosely grouped into three categories [2,3]: affective state, cognitive state and expertise. Stevens, Galloway & Berka [4], for example, propose that EEG-based indices can be used to determine a learner's position within Shriffin's three stages of skill acquisition. They provide preliminary evidence of systematic changes in brain activity with increasing expertise

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 131–136, 2011.

at a specific problem solving task. Other researchers [e.g., 5] have begun searching for task-independent indices related to a student's level of expertise.

While measurement is typically the final stage of an instructional episode, it can also be used to support instructional decisions that are tailored to the needs of the learner throughout the course of instruction. In fact, much of the recent literature proposes that we can maximize the value of a training experience by adapting some aspect of the instruction based on a real-time assessment of a student's affective state, cognitive state and/or level of expertise. DuRousseau, Mannucci & Stanley [6] state that the "form, timing and content of an individual's training protocol" could be shaped by using EEG to monitor changes in aspects of the individual's cognitive state, such as arousal, attentional capacity, executive workload and coordination of cognitive resources. That's not to say that researchers are proposing that neurophysiological measures alone are always sufficient to drive instructional adaptation. Mathan & Dorneich [7] and Stevens, Galloway & Berka [8] provide some explanation as to how these measures could be integrated with performance data and student models to conduct a complete diagnosis of a student's level of expertise.

Early, broad claims that instructional adaptations can be driven by neurophysiological assessment have given way to the specification of candidate pairings between individual indices and manipulations. Examples include controlling the presentation speed of textual information based on an assessment of how well the student is processing the content [9], breaking a task into smaller steps as a student's working memory capacity appears to be overloaded [7], and modifying the modality of feedback to leverage underutilized processing channels [2].

What is missing in all of this work is clear empirical evidence that the incorporation of neurophysiological measurement into a training system can yield some significantly improved capability over and above what can be accomplished without that measurement. As a first step towards this goal, Campbell & Luu [10] demonstrated that EEG could be used to distinguish intentional, learned responses made during training from other types of responses such as guesses and slips (i.e., unintentional actions). They argued that different types of errors should be remediated in different ways – an accidental slip, for example, should not be treated the same as an incorrect action that the student believes to be appropriate. Thus, being able to distinguish intentional responses (whether correct or incorrect) from other responses should allow for more focused and appropriate training responses.

Of course, showing that it is possible to distinguish different types of responses doesn't ensure that distinguishing between those responses will actually lead to a different diagnosis of a trainee's underlying knowledge, skills and abilities. It is possible, for example, that guesses and slips could be so rare as to not significantly impact an interpretation of a performance pattern. In a follow-up study, however, Campbell, Belz & Luu [11] took this work a step further and demonstrated that in at least one domain, removing slips and guesses from a trainee's performance data set did change the performance pattern that emerged.

In summary, to date it has been shown that EEG data is capable of distinguishing intentional responses from guesses and slips made during training, and that when it is used in conjunction with performance data it may yield different diagnoses of student competence than would be inferred based on performance data alone. The objective of the current study is to test the hypothesis that diagnoses based on a combination of EEG data and performance data are more accurate than diagnoses based solely on performance data. Our approach to evaluating the accuracy of a diagnosis is to evaluate the extent to which that diagnosis predicts future performance patterns. If a problem occurs because of an actual error in the trainee's underlying knowledge, then it is very likely to occur again in the future. However, if a mistake occurs as the result of a bad guess or accidental slip, then that mistake is not highly likely to occur in the future. In other words, if using EEG truly improves diagnostic accuracy, then it should improve our ability to predict future problems. This basic argument is the foundation for this study.

2 Method

2.1 Participants

Thirty-two volunteers, 18 men and 14 women, all 18 years or older were given financial compensation for participating in this study. The mean age of the participants was 23 years (SD = 5; range: 18-35).

2.2 Apparatus

The EEG data for each participant was acquired using a 256-channel HydroCel Geodesic Sensor Net (Electrical Geodesics, Inc., Eugene, OR). The recordings were referenced to Cz and all of the electrodes were kept below 70 KΩ. The EEG was sampled with a 16-bit analog-to-digital converter at 250 s/s and was bandpass filtered at 0.1- to 100-Hz. Stimulus control was maintained by Eprime© (Psychology Software Tool, Pittsburgh, PA).

2.3 Materials

A computer-based, speeded, flash-card style program, created in Eprime©, was used to train participants on identification of military vehicles. The images used were selected from the United States Marine Corps unclassified anti-armor training materials. They consisted of six tank illustrations: the BMP-2, BMP-3, M1A1, T-72, T-80, and the ZSU-23-4.

2.4 Procedure

Each participant took part in two 90 minute experimental sessions separated by 48 hours. The first was a training session in which they were taught to recognize the six tank silhouettes from three different viewing angles (head on, side view and rear view). During this session each participant first completed the informed consent paperwork, and was then fitted with a 256-channel sensor array. The computer based learning task then began with 132 familiarization trials in which participants learned to associate each of the six tank names with a particular key press. Immediately following, the tank identification training began. Each of the six tanks was presented at each of the three angles 23 times for a total of 414 trials during the training session. Each trial began with the presentation of a tank image in the center of the computer screen. Participants had 2000 milliseconds to identify the tank by pressing the corresponding response key. For a random selection of one-third of the trials, knowledge of correct feedback was shown on the screen for 2000 milliseconds immediately following a response or if the response time ended before a key press was made. For the other two-thirds, the screen went blank for 100 milliseconds and then the next trial began. Reaction times, responses, and EEG data were recorded for each trial.

Following session one, each participant's performance was evaluated in two different ways. First, we computed a confusion matrix using all of their performance data, and flagged the errors that were the most problematic for that participant. Second, we used single trial analyses of their EEG data to identify and remove guesses and slips from the performance data and then computed a confusion matrix on the filtered data to identify the errors that were most problematic. While there was some degree of overlap between these two methods, for approximately two-thirds of the participants, these two diagnoses diverged on at least one tank image. We then randomly assigned the participants to one of two groups, which identified which problems to highlight: in one group we used the confusion matrix generated by all of the data and in the other we used the confusion matrix generated by the EEG filtered data. We then created feedback for the participant that provided cues that would help them better identify the tanks they had the most trouble with (See Figure 1).

Two days later, the participants returned for the second session. They again completed the informed consent and were fitted with the EEG apparatus. Then they were given the feedback to review before completing the testing trials of the tanks they saw during session one. During these trials participants were again asked to identify the tank images in a brief period of time (1000 milliseconds) by pressing the appropriate key. This time, however, no feedback was provided.

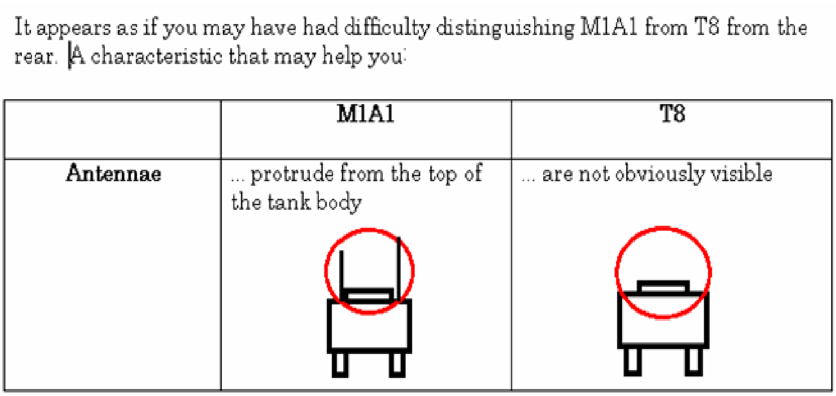


Fig. 1. An example of a cue used to provide feedback to participants

3 Results

Our hypothesis was that the use of EEG data to filter the performance data set prior to computing a confusion matrix would yield a more accurate assessment of each participant's true underlying confusions, and thus would flag the tank images that would be most likely to cause problems during the second session. In order to test this hypothesis, we calculated average improvement scores from session one to session two for the subsets of images on which the two assessment methods diverged (feedback was suggested but not given in one of the two conditions). For those images that were flagged as problematic using the entire data set, but not after the EEG-filtering, the average improvement score was 40% (SD = 0.38). For those images that were flagged as problematic in the EEG-filtered data set, but not in the complete data set, the average improvement score was -9% (SD = 0.66). This difference represented a statistical trend, $t(20) = 1.92$, $p = 0.07$. It is important to note, that for the purpose of this paper, we were only interested in, and thus only report data on the trials from session 2 in which participants did not receive feedback. This is because we wanted to assess future performance on items that did not receive feedback in the condition a participant was in, but would have received feedback if they had been in the other condition.

4 Discussion

There are many claims in the research community that neurophysiological data will revolutionize training; however, there are few studies that actually provide direct evidence to back up these claims. Even people who show neurophysiological data could be used don't show that it gives you an advantage over not using it. This study is one of a series attempting to bridge this gap. Our approach to evaluating the accuracy of a diagnosis is to evaluate the extent to which that diagnosis predicts future performance patterns.

It is something of a truism in psychology that the best predictor of future performance is past performance. In this case, our data suggest that this is not necessarily the case. When we examined all of the past performance data, the "problems" that people experienced in the past showed some amount of spontaneous improvement in the future, suggesting that some of those "problems" did not represent true faults in an underlying knowledge representation. However, when we focused on the subset of past performance problems that were identified by EEG data as representing true misconceptions we saw that, when left untreated instructionally, these same problems reoccurred with approximately the same frequency in the future.

While our results only represent a statistical trend, when taken in combination with other studies of the use of EEG to distinguish intentional responses made during training [10, 11], we believe that the data are converging on the conclusion that this is a viable use of neurophysiological data in at least some training systems. Of course, work remains to be done. Attention needs to be paid to the question of which training systems are most likely to benefit from this type of data. More importantly, trainers need to demonstrate that they can use a more accurate diagnosis of a trainee's true problems as the basis to deliver more efficient training.

References

1. Poulson, C., Luu, P., Tucker, D.: The neurophysiology of learning and memory: Implications for training. In: Schmorrow, D., Cohn, J., Nicholson, D. (eds.) The PSI Handbook of Virtual Environments for Training and Education, vol. 1, pp. 7–30. Praeger Security International, Westport (2009)
2. Bolton, A., Campbell, G.E., Schmorrow, D.D.: Towards a closed-loop training system: Using a physiological-based diagnosis of the trainee's state to drive feedback delivery choices. In: Schmorrow, D., Reeves, L. (eds.) Foundations of Augmented Cognition, 3rd edn., pp. 409–414. Springer, Berlin (2007)
3. Oertel, K., Kaiser, R., Voskamp, J., Urban, B.: AFFectIX – An affective component as part of an E-Learning system. In: Schmorrow, D., Reeves, L. (eds.) Foundations of Augmented Cognition, 3rd edn., pp. 385–393. Springer, Berlin (2007)
4. Stevens, R., Galloway, T., Berka, C.: Allocation of time, EEG-engagement, and EEG-workload resources as scientific problem solving skills are acquired in the classroom. In: Schmorrow, D., Nicholson, D., Drexler, J., Reeves, L. (eds.) Foundations of Augmented Cognition, 4th edn., pp. 128–137. Strategic Analysis Incorporated, Arlington (2007)
5. Dickson, B., Belyavin, A.: The use of electrophysiological markers of expertise to configure adaptive training. In: Schmorrow, D., Nicholson, D., Drexler, J., Reeves, L. (eds.) Foundations of Augmented Cognition, 4th edn., pp. 138–144. Strategic Analysis Incorporated, Arlington (2007)
6. DuRousseau, D.R., Mannucci, M.A., Stanley, J.P.: Will augmented cognition improve training results? In: Schmorrow, D.D. (ed.) Foundations of Augmented Cognition, vol. 11, pp. 956–963. Lawrence Erlbaum Associates Publishers, Mahwah (2005)
7. Matham, S., Dormeich, M.: Augmented tutoring: Enhancing simulation based training through model tracing and real-time neurophysiological sensing. In: Schmorrow, D.D. (ed.) Foundations of Augmented Cognition, vol. 11, pp. 964–973. Lawrence Erlbaum Associates, Mahwah (2005)
8. Stevens, R., Galloway, T., Berka, C.: Exploring neural trajectories of scientific problem solving skill acquisition. In: Schmorrow, D., Reeves, L. (eds.) Foundations of Augmented Cognition, 3rd edn., pp. 400–408. Springer, Berlin (2007)
9. Palmer, E.D., Kobus, D.A.: The future of augmented cognition systems in education and training. In: Schmorrow, D., Reeves, L. (eds.) Foundations of Augmented Cognition, 3rd edn., pp. 373–379. Springer, Berlin (2007)
10. Campbell, G.E., Luu, P.: A preliminary comparison of statistical and neurophysiological techniques to assess the reliability of performance data. In: Schmorrow, D., Nicholson, D., Drexler, J., Reeves, L. (eds.) Foundations of Augmented Cognition, 4th edn., pp. 119–127. Strategic Analysis Incorporated, Arlington (2007)
11. Campbell, G.E., Belz, C.L., Luu, P.: What was he thinking?: Using EEG to facilitate the interpretation of performance patterns. In: Schmorrow, D., Estabrooke, I., Grootjen, M. (eds.) Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience, pp. 339–347. Springer, Heidelberg (2009)

Computational Cultural Neuroscience: Implications for Augmented Cognition

Joan Y. Chiao

Department of Psychology, Northwestern University,
2029 Sheridan Rd. Evanston, IL 60208 USA
jchiao@northwestern.edu

Abstract. From perceiving objects in space to recognizing emotions at a distance, culture affects how people think, feel, reason as well as the neurobiological mechanisms underlying these processes. Here I review recent evidence from cultural neuroscience, introduce the notion of computational cultural neuroscience – the development of computational and formal models of how culture affects neurobiological mechanisms and vice versa – and finally, discuss the implications of computational cultural neuroscience for research in augmented cognition.

Keywords: cultural neuroscience; computational cultural neuroscience; augmented cognition.

1 Introduction

Over the past two decades, researchers in the field of augmented cognition have worked to develop novel technologies that can both monitor and enhance human cognition and performance. Much of this research in augmented cognition has relied on research findings from cognitive science and cognitive neuroscience, fields which seek to illuminate how the mind and brain work. Seminal findings from these fields, such as resource limitation capacities in working memory and attention, have enabled augmented cognition researchers to identify potential bottlenecks in human achievement and to develop technological solutions that overcome such limitations.

While notable advances have been made in the field of augmented cognition, recent advanced in the fields of cultural psychology and cultural neuroscience suggest that across a range of cognitive and perceptual abilities vary across cultures leading to the need for researchers across disciplines to determine ways to model and implement culturally-diverse technologies that can monitor and enhance human cognition and performance with efficacy across cultural groups.

In this paper, I will review recent empirical evidence in cultural psychology and cultural neuroscience demonstrating cultural variation in perceptual, cognitive and socioemotional processing. Next, I will describe ways in which formal computational models of cognition across cultures may facilitate the ability of augmented cognition researchers to design technologies that enhance how the human mind and brain work in individuals across diverse cultures.

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 137–142, 2011.

1.1 Cultural Influences on Behavior and Brain Function

A fundamental dimension that cultures vary on is *individualism* and *collectivism* [1-3]. Individualistic cultures encourage thinking of people as independent of each other. By contrast, collectivistic cultures endorse thinking of people as highly interconnected to one another. Individualistic cultures, such as the West, emphasize self-expression and pursuit of individuality over group goals, whereas collectivistic cultures, such as the East, favour maintenance of social harmony over assertion of individuality [1-3]. Cultural psychological research has shown that cultural variability in self-construal style affects a wide range of psychological processing, from how people perceive objects in the environment to how they think about the world around them and recognize the mental states of others.

Recent evidence from cultural neuroscience is demonstrating that culture affects not only behavior, but also brain function. Cultural neuroscience is an emerging research discipline that investigates cultural variation in psychological, neural and genomic processes as a means of articulating the bidirectional relationship of these processes and their emergent properties. Research in cultural neuroscience is motivated by two intriguing questions of human nature: how do cultural traits (e.g., values, beliefs, practices) shape neurobiology (e.g., genetic and neural processes) and behavior and how do neurobiological mechanisms (e.g., genetic and neural processes) facilitate the emergence and transmission of cultural traits?

The idea that complex behavior results from the dynamic interaction of genes and cultural environment is not new [4]; however, cultural neuroscience represents a novel empirical approach to demonstrating bidirectional interactions between culture and biology by integrating theory and methods from cultural psychology [5], neuroscience [6] and neurogenetics [7-9]. Cultural neuroscience aims to explain a given mental phenomenon in terms of a synergistic product of mental, neural and genetic events. Cultural neuroscience shares overlapping research goals with social neuroscience, in particular, as understanding how neurobiological mechanisms facilitate cultural transmission involves investigating primary social processes that enable humans to learn from one another, such as imitative learning. However, cultural neuroscience is also unique from related disciplines in that it focuses explicitly on ways that mental and neural events vary as a function of culture traits (e.g., values, practices and beliefs) in some meaningful way. Additionally, cultural neuroscience illustrates how cultural traits may alter neurobiological and psychological processes beyond those that facilitate social experience and behavior, such as perception and cognition.

For instance, cultural variation between Westerners and East Asians affects how people think about themselves and their relation to the environment not only affects human behavior, but underlying neurobiological processes. For instance, Westerners engage brain regions associated with object processing to a greater extent relative to East Asians who are less likely to focus exclusively on objects within a complex visual scene [10]. Westerners show differences in medial prefrontal activity when thinking about themselves relative to close others, but East Asians do not [11]. Activation in frontal and parietal regions associated with attentional control show greater response when Westerners and East Asians are engaged in culturally preferred judgments [12]. Evolutionarily ancient limbic regions, such as the human amygdala, respond preferentially to fear faces of one's own cultural group [13]. Brain regions associated with social cognition, or thinking about what others are thinking, such as the superior temporal sulcus, show greater response when inferring the mental states of members of one's own cultural group [14]. Taken together, these findings show cultural differences in brain functioning across a wide variety of psychological domains and demonstrate the importance of comparing, rather than generalizing, between Westerners and East Asians at a neural level.

2 Computational Cultural Neuroscience

The existence of cultural variation in neural systems presents a novel opportunity and challenge for the development of a *computational cultural neuroscience*. Computational modeling of human brain and behavior provide a potent way to develop and test formal theories of the multilayered, complex and dynamic relation between cellular and network properties of neurons to mental representations that guide how people think and behave [15].

While computational modeling, in principle, allows for the development and testing of a multitude of theories regarding the relation between neural and behavioral systems, growing evidence from cultural neuroscience regarding how culture affects the human brain and vice versa represents a key advance allowing for the emergence of a computational cultural neuroscience. By knowing which neural systems show modulation of activation by cultural values, practices and beliefs and how systematic cultural modulation of neural systems alters human behavior, we gain important insights into fundamental structural and functional constraints underlying formal mathematical models of cultural influences on mind, brain and behavior.

Ultimately, by combining cultural neuroscience evidence with computational modeling, we may uncover an array of distinct formal models that capture the basic informational processing mechanisms underlying how people think, feel and behave across a diversity of human cultures. One important challenge in this endeavor is determining guiding principles that help to constrain and characterize the range of influence of culture on neurobiological systems (e.g., what aspects of the informational processing systems are universal or distinct across cultures and why). Another important challenge is distinguishing formal models of cultural influences on brain and behavior from formal models of individual differences in brain and behavior (e.g., what aspects of the informational processing systems represent individual versus group differences in brain-behavior relations).

2.1 Three Starting Points for a Computational Cultural Neuroscience

Here I describe three examples of a computational cultural neuroscience approach to understanding how and why structure-function mapping may vary across cultures. The first structure-function mapping that may vary across cultures exists within subregions of the occipitotemoral cortex, which is responsible for the learning and representation of complex visual recognition, including objects, places and faces. Recent evidence from cultural neuroscience indicates that cultures vary in the extent to which neural responses within occipitotemporal cortex vary within fusiform gyrus and occipitotemporal gyrus, but not parahippocampal gyrus. For instance, Goh and colleagues [16] showed that activity within object-processing and object-scene binding in the ventral visual area varies across cultures. While encoding complex visual scenes consisting of objects embedded within a background, Caucasian-Americans and East Asians varied in the extent to which they showed increased neural response within ventral visual regions. More specifically, recent neuroimaging evidence shows that visual processing of complex visual scenes in Caucasian-Americans is more object-focused compared to East Asians and to facilitate this variability in attention and encoding towards objects compared to the background, neural response within object processing brain regions, such as lateral occipital cortex, is significantly heightened in Caucasian-Americans compared to East Asians. This variability in the degree to which occipitotemporal cortex is recruited during scene encoding reflects group differences in neural connectivity underlying attentional and memory processes. Future work in computational cultural neuroscience is needed to test formal models of how cultural variation in engagement of ventral visual regions during object processing arises from cultural differences in the type or kind of neural representation of objects and scenes within this region or merely reflects differential recruitment of core object processing regions universal across both groups.

Additionally, Gutchess and colleagues [10] recently showed that brain regions within fronto-parietal regions, including the right angular gyrus and right middle frontal gyrus, reveal cultural differences in neural response during semantic categorization. For instance, East Asians show increased neural response when categorizing semantic relations as a function of relation and category compared to Caucasian-Americans. Additionally, during semantic conflict trials compared to match trials, East Asian participants, showed greater response within a frontal-parietal network previously implicated in controlled executive function, whereas Caucasian-Americans, who showed increased response within the temporal lobe and cingulate gyrus, a brain region engaged during cognitive conflict. This variability in neural response suggests that cultural modulates the strength of network connectivity of fronto-parietal and temporo-cingulate connectivity during semantic categorization. Future work in computational cultural neuroscience is needed to test formal models of the extent cultural variation in network connectivity between these two cortical circuits arises from differences in feedforward or bidirectional neural connectivity.

Cultural variation exists not only with cortical regions underlying perception and cognition, but also socioemotional processes associated with emotion recognition and mental state inference. For instance, the human amygdala is a subcortical brain region which dense innerconnectivity with cortical regions and is specialized for recognition of emotional and social information [17]. Prior affective neuroscience studies have shown that the amygdala is critical to fear recognition and in particular the inferences of emotional states from the eye region of the face [17]. The ability to infer the mental states of others also recruits lateral brain regions such as the superior temporal sulcus that decodes social information from perceptual information within the face, including the eye, nose and mouth region. Recent evidence from cultural neuroscience suggests that Chiao and colleagues [13] showed that evolutionarily ancient limbic regions, such as the human amygdala, respond preferentially to fear faces of one's own cultural group. Additionally, brain regions associated with social cognition, or thinking about what others are thinking, such as the superior temporal sulcus, show greater response when inferring the mental states of members of one's own cultural group [14]. In both instances, these brain regions engaged during socioemotional processing show heightened response to culturally-familiar social signals. One important question that computational modeling can provide insight on is the extent to which this heightened processing for own-culture social stimuli is a function of cultural differences in perceptual information within the stimulus (e.g., template arrangement of facial signals) that activates a fixed or template-like neural representations within these regions. Alternatively, increased neural recruitment of amygdala and superior temporal sulcus for culturally-congruent perceptual information could simply reflect a learned neural response to a specific stimulus that remains flexible or dynamic to a wide range of perceptual input at certain period of development but then tunes with experience.

3 Implications for Augmented Cognition

While much work lies ahead in developing a computational cultural neuroscience approach to human brain and behavior, the implications of such an approach are numerous, particularly for the field of augmented cognition. A chief concern in developing novel technologies that increase human performance and decision-making is determining whether or not such technologies will adapt readily, with similar efficiency and accuracy, to human users who vary in cognitive styles due to cultural differences. To address this problem, augmented cognition researchers may use models from computational cultural neuroscience to engineer culturally-flexible technologies, achieving the goal of enhancing human cognition and socioemotional processing across culturally diverse populations.

Acknowledgments. I thank members of the NU Social Affective and Cultural Neuroscience Lab and Steve Franconeri for helpful discussion.

References

[1] Markus, H.R., Kitayama, S.: Culture and the self: implications for cognition, emotion and motivation. Psychol. Rev. 98, 224–253 (1991)
[2] Triandis, H.C.: Individualism and collectivism. Westview, Boulder (1995)
[3] Nisbett, R.E., Peng, K., Choi, I., Norenzayan, A.: Culture and systems of thought: Holistic versus analytic cognition. Psychol. Rev. 108(2), 291–310 (2001)
[4] Caspi, A., Moffitt, T.: Gene-environment interactions in psychiatry: joining forces with neuroscience. Nat. Rev. Neurosci. 7, 583–590 (2006)
[5] Kitayama, S., Cohen, D.: Handbook of Cultural Psychology. Guilford Press, New York (2007)
[6] Gazzaniga, M.S., Ivry, R., Mangun, G.R.: Cognitive neuroscience: The biology of the mind. Norton, New York (2002)
[7] Canli, T., Lesch, K.P.: Long story short: the serotonin transporter in emotion regulation and social cognition. Nature Neuroscience 10, 1103–1109 (2007)

[8] Green, A.E., Munafò, M., DeYoung, C.G., Fossella, J., Fan, J., Gray, J.R.: Using genetic data in cognitive neuroscience: From growing pains to genuine insights. Nature Reviews Neuroscience 9, 710–720 (2008)
[9] Hariri, A.R., Drabant, E.M., Weinberger, D.R.: Imaging genetics: perspectives from studies of genetically driven variation in serotonin function and corticolimbic affective processing. Biological Psychiatry 59(10), 888–897 (2006)
[10] Gutchess, A.H., Welsh, R.C., Boduroglu, A., Park, D.C.: Cross-cultural differences in the neural correlates of picture encoding. Cognitive, Affective, and Behavioral Neuroscience 6(2), 102–109 (2006)
[11] Zhu, Y., Zhang, L., Fan, J., Han, S.: Neural basis of cultural influence on self representation. Neuroimage 34, 1310–1317 (2007)
[12] Hedden, T., Ketay, S., Aron, A., Markus, H.R., Gabrieli, J.D.E.: Cultural influences on neural substrates of attentional control. Psychological Science 19(1), 12–16 (2008)
[13] Chiao, J.Y., Iidaka, T., Gordon, H.L., Nogawa, J., Bar, M., Aminoff, E., Sadato, N., Ambady, N.: Cultural specificity in amygdala response to fear faces. Journal of Cognitive Neuroscience 20(12), 2167–2174 (2008)
[14] Adams Jr., R.B., Rule, N.O., Franklin Jr., R.G., Wang, E., Stevenson, M.T., Yoshikawa, S., Nomura, M., Soto, W., Kveraga, K., Ambady, N.: Cross-cultural reading the mind in the eyes: An fMRI Investigation. Journal of Cognitive Neuroscience (in press)
[15] O'Reilly, R.C., Munakata, Y.: Computational Neuroscience and Cognitive Modeling. In: Nadel, L. (ed.) Encyclopedia of Cognitive Sciences. Macmillan, London (2003)
[16] Goh, J.O., Chee, M.W., Tan, J.C., Venkatraman, V., Hebrank, A., Leshikar, E.D., Jenkins, L., Sutton, B.P., Gutchess, A.H., Park, D.C.: Age and culture modulate object processing and object-scene binding in the ventral visual area. Cognitive Affective & Behavioral Neuroscience 7(1), 44–52 (2007)
[17] Adolphs, R.: Fear, faces and the human amygdala. Curr. Opin. Neurobiol. 18(2), 166–172 (2008)

Enhancing Team Performance Using Neurophysiologic Synchronies in a Virtual Training Environment

Marianne Clark[1], Kimberly Cellucci[1], Chris Berka[2], Daniel J. Levendowski[2], Jonny Trejo[2], Amy Kruse[3], and Ron Stevens[4]

[1] Scientific Research Corporation, 1101 Remount Rd, Suite 500, Charleston, SC, 29406
[2] Advanced Brain Monitoring, Inc. 2237 Faraday Avenue, Suite 100, Carlsbad, CA, 92008
[3] Total Immersion Software 2001 Jefferson Davis Highway, Suite 703, Arlington, VA 22202
[4] IMMEX Project/UCLA Los Angeles, CA 90030
mclark@scires.com, {chris,dan,jtrejo}@b-alert.com, akruse@totimm.com, ron@immex.com

Abstract. A study was conducted to investigate the use of neurophysiologic synchronies as a measurement of team cognition (1) in a military-style virtual environment simulation. Neurophysiologic synchronies (NS), defined as the second-by-second quantitative co-expression of the levels of cognitive measures by individual members of a team (8), were found to be useful in monitoring the quality of teamwork and to be a means to identify more optimal patterns of team interaction which can be used to provide feedback during training. In the current study, findings showed promise for further research in the collection of NS. A framework is also proposed to support the research and training of team cognition.

Keywords: Team performance, team cognition, shared mental models, collaboration, neurophysiologic synchronies, electroencephalography (EEG), virtual environments, mission rehearsal training, RealWorld.

1 Introduction

Team performance is critical to the success of many organizations whether the teams are co-located or virtual; hierarchical or decentralized. Today, team performance is equally critical for our warfighter. Mission success is dependent upon the coordinated effort of those teammates executing the mission side-by-side as well as the efforts of the extended team responsible for planning and supporting the mission.

Our service men and women are well trained before they deploy. They are trained in how to operate their equipment, trained in operations and other requisite skills. However there is an identified gap between their formal training and the skills needed during a mission in a forward location. Sometimes what they need to know for a successful mission is only discovered the day prior.

This gap is being addressed by the Services via deployed mission rehearsal training. During these training exercises, specific, current tactics, threats and situations can be rehearsed and practiced prior to the mission. Critique of the training rehearsal can be

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 143–152, 2011.

given prior to the mission and a level of mission expertise can be attained by the team prior to the beginning of the mission.

Deployed mission rehearsal training provides an opportunity to address the complexity of team performance. One aspect of team performance is experience. Dyer (2) suggests that when a team has more experience working together, they are better able to coordinate their efforts, and have better team processes and performance. Klein, Zsambok and Thordsen (3) suggest that teams may have maturity levels as they move from entry-level to expert-level team. Mission rehearsal provides an opportunity to execute scenarios of potential situations to allow the teams to move from entry-level to expert.

A component of team expertise is addressed in the construct of team cognition or sometimes referred to as team level macrocognition (4, 5, and 1). This construct can be seen as the integrated thinking of a team that includes cognition, behavior and attitudes that contribute to team performance. It includes shared knowledge or mental models and shared processes. Blickensderfer et al (6) distinguish between pre-performance shared knowledge (mental model) and the dynamic shared knowledge, such as shared situation awareness, that forms among team members while performing a task. Klein (7) lists real-time team processes whose quality contributes to team performance, such as controlling the flow of information, forming shared situation awareness, applying strategies for decision making and problem solving and monitoring team performance.

Klein (7) and Blickensderfer et al. (6) suggest cognitive task analysis techniques to identify the knowledge requirements, including processes, decisions, barriers, errors, cues and strategies that need to be shared for effective team performance. Team cognitive task analysis techniques can be used to define the content of the shared mental model which will be the basis for mission training, including: mission objective, unique mission situation, roles and responsibilities, possible threats, and modifications to procedures for response to threats based upon latest intelligence and intent.

The use of a virtual training environment simulation of the mission provides the rehearsal needed to build the expertise of the current team and to create a team shared mental model when directed feedback can be provided. Mathieu et al. (5) used a networked simulation with six scenarios to investigate the development of shared mental models and their effect on team performance. They did *not* provide detailed analysis and feedback, such as an After Action Review, after each scenario because they were looking at the effects of experience alone. They found some improvement to coordination and cooperation, but it did not result in a greater shared mental model. Their findings support the idea that guided experience and developmental feedback is needed to support this type of learning. Given that team cognition addresses both internal and external processes of the team, a challenge to providing the most appropriate feedback to enhance training becomes identifying a measurement of team performance that allows for timely feedback to the team to guide them in their *internal* process of building their shared mental model and improving their team processes. Stevens et al. (8) suggests that internal processes can be studied via surrogate quantitative measures such as EEG metrics, pupil size, and heart rate variability.

2 Neurophysiologic Measurement of Team Performance

Previous research examined the utility of collecting team members' electroencephalography (EEG) and electrocardiography (ECG) to measure cognitive components such as attention, workload, engagement, and stress for the modeling of team cognition (8). In this study, five three- person- teams of college students were asked to do a problem solving task to make a determination if the person shown in a simulated reality show should seek help for drug abuse. EEG data collected using the B-Alert wireless headset and software from Advanced Brain Monitoring, Inc (ABM). first decontaminates and then performs real-time calculations of cognitive states changes. The result is a value for mental work load (WL) and engagement (E) ranging from 0.1 to 1.0 for each 1 second epoch. The WL and E values are then normalized, values for each team are provided to a self organizing artificial neural network which then provides patterns of WL and E for each team, called neurophysiologic synchronies (NS). Neurophysiologic synchronies (NS) are the second-by-second quantitative co-expression of the levels of cognitive measures by individual members of a team. Stevens et al. (8) report unique neurophysiologic synchronies that were commonly found across the teams. Examples are patterns where all team members are actively engaged with moderate to high workload values to ones where one or more team members were not engaged and had low workload values.

Each team was tested against the identified patterns and it was determined that each team was different in the number of NS and which NS were most frequent. For example, Group 3 had fewer NS frequently repeated and those were ones where all teammates were engaged and working. At the other end of the spectrum was Group 2 with a high number of NS, with more of them representing team members who were not engaged or not working. Group 3 performed better based upon subjective ratings of video logs and objective measure of time to solution and correct solution. Group 2 took the longest to complete the task. Stevens et al. (8) further aligned the NS to tasks the operators were performing and how those tasks map to three processes: mental model formation, mental model sharing and integration, and mental model convergence and revision. They identified three NS whose patterns of reoccurrence aligned well to each process.

These findings suggest a promising approach for monitoring the quality of team internal processes that relate to team performance. We wanted to apply this approach to small teams of warfighters and to hopefully identify points to recommend more optimal synchronies after a mission rehearsal in order to improve team cognition and performance.

3 Pilot Study

A preliminary study was designed to further investigate the nature of team cognition, in particular, aspects of neurophysiologic synchronies as applied to small groups of warfighters. While the end objective is to look at team cognition at multiple levels: 1) the small team and 2) the multi-team, such as platoon level team, and to identify a method for providing directed feedback on internal team processes during a mission rehearsal session, the initial pilot study was designed to begin more simply.

Our hypothesis is that as members of a team perform a collaborative task, each team member will generate varying degrees of cognitive components such as attention, workload and engagement and that the levels of these will vary depending upon the task and the level of expertise of the team. Specifically, we expect to see changes in the levels of the cognitive component from first execution of the scenario to the sixth execution. We expect to see improvements in team performance in terms of time to execute the scenario and ability to recognize and avoid threats from the first execution of the scenario to the sixth execution. We expect to see patterns in neurophysiologic synchronies change from the first execution of the scenario to the sixth execution.

3.1 Tasks and Methods

For this study, the training scenario was built to investigate the fundamentals of three-person team collaboration in a dismounted engagement in a small village. Total Immersion Software, Inc. RealWorld software was used to construct the scenario. RealWorld is a PC-based simulation platform that enables non-programmers to rapidly build 3D, geo-specific simulations. The primary team objective is to eliminate a foe and complete the mission, but to do that they must traverse a village and overcome unexpected obstacles. The scenario was designed to create cause-and-effect across team members. For example, if one team member did not obtain and communicate intelligence in a timely manner, the other team member would advance too far. This would trigger a hostile attack, a detonation of an IED, etc. Figure 1 shows a view of the village with potential threats identified. Team member A (route shown in red) and B (route shown in blue) were on the ground with a choice of routes to achieve their objective and meet at the rendezvous point(s). Team member C acted in a command and control role from a remote location while using a UAV (free camera) feed to gain his tactical oversight of the village and mission. His second role is to be the communication link for the team. He has full access to beyond-line-of-sight communications to provide intelligence reports of late breaking news to the team and to pass on requests for support from the team.

Six subjects were recruited to participate in the study. Testing was conducted at the ABM offices located in Carlsbad, Ca. Three gaming systems were set side-by-side in a test room. Subjects were assigned to role of team member A, B, or C.

The three team members were fitted with the ABM B-Alert EEG 9-channel wireless headset that has the following sensor site locations: F3, F4, C3, C4, P3, P4, Fz, Cz, and POz. An individual EEG baseline was collected. Subjects were given a 30- minute practice session to allow them to become familiar with the RealWorld gaming controls. The subjects were briefed on their mission and a provided map of the village. The scenario was then initiated. Duration of each scenario was approximately five minutes. Subjects participated in 6 sessions running the scenarios with a 10 minute break between each session. Subject's sessions were video recorded. Scenario start and end times and key event times (such as the triggering of an IED) were logged to RealWorld event log.

The EEG data was processed by the software from ABM which first decontaminates and then performs real-time calculations of cognitive state changes.

Values for mental workload (WL) and engagement (E) for each individual team member was provided in a range from 0.1 to 1.0 for each 1 second epoch. Please refer to Berka et al. (9, 10 and 11) for a detailed description of the development of the gauges and this process. WL and E values were then normalized for each individual team member, and then combined at each epoch for every team member into a vector representing the state of the team as a whole. The epoch-by-epoch team vectors were then presented to a self organizing artificial neural network developed by the IMMEX Project, which results in a series of neurophysiologic synchronies (NS).

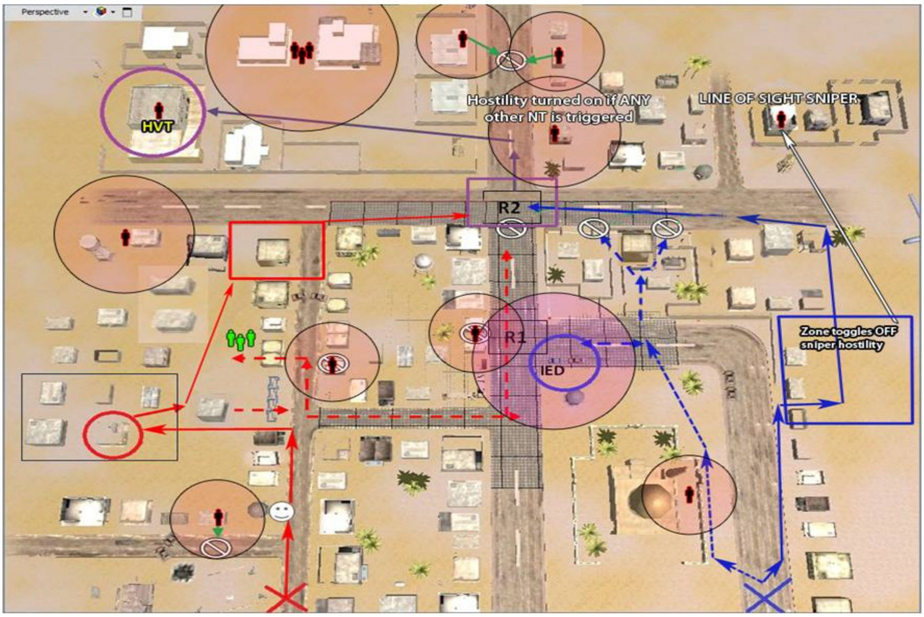


Fig. 1. Map of Village with Threats and Possible Routes

3.2 Results and Discussion

Preliminary analysis of the data was both encouraging and disappointing. Time to complete the scenario was reduced by 59% from session 1 to session 2, but all subsequent sessions were no different from session 2. This suggests that although the scenario had a dynamic quality, given that many of the threats could be triggered by proximity, the actual execution of the scenario lacked the complexity necessary to distinguish or allow for improvements of the team after the second trial.

The EEG data analysis was encouraging. Six neurophysiologic synchronies were frequently identified among the two teams. These can be characterized as the following:

- Team member A highly engaged, but B and C had low engagement
- Team members A and B were highly engaged

- Team member A and C were highly engaged
- Team member B and C were highly engaged
- No team members were highly engaged

These NS occurrences are displayed over the course of a scenario in Figure 2. These invite the opportunity to investigate a relationship among operator tasks and team processes to NS as was done by Stevens et al. (8).

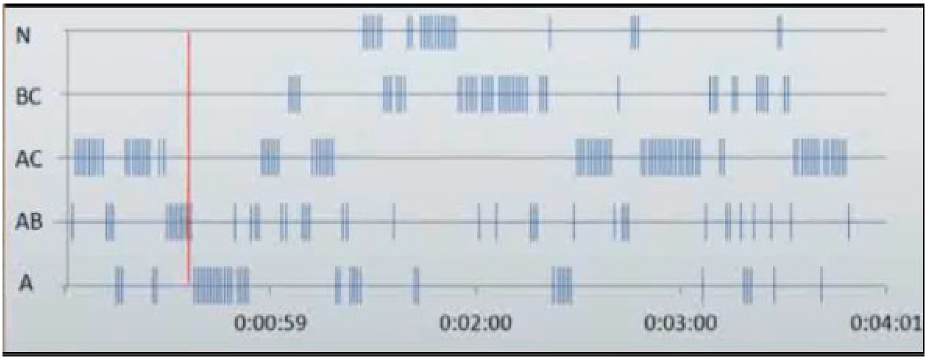


Fig. 2. Neurophysiologic Synchronies during a Scenario

Team engagement values were compared from session 1 to session 2. All three team members had a 6% increase in the occurrence of low levels of engagement (this is the None NS described above) in session 2 when compared to session 1. Said another way, team members spend 6 % more time not engaged in session 2 compared to session 1. Also, the amount of time Team member C was highly engaged with either Team Member A or Team member B decreased by 6% from session 1 to session 2.

While the difficulties encountered with the scenario resulted in limitations in the data collected, two conclusions can comfortably be reached. First, neurophysiologic synchronies can be identified in a military-type scenario using a virtual environment simulation. This suggests that with further refinement of methodologies and future research, it may be possible to measure NS, correlate them to tasks and internal processes, identify optimal and suboptimal NS demonstrated by performance, and provide this feedback in a form for training during deployed mission rehearsal.

A second conclusion that can comfortably be reached is that as new tools and techniques are discovered, a frequent challenge for cognitive researchers is to establish an experimental system that integrates these tools and techniques in a way that acknowledges the complexity of human cognition while allowing for experimental manipulation. Team cognition relies on a sophisticated understanding of shared mental models and the real-time monitoring and analysis of internal cognitive processes which must be tied to external performance. The number and types of human internal and external data points necessary to represent this complex phenomenon is in itself complex. In an attempt to address this fundamental research, a framework is proposed in the following section.

4 Proposed Experimental Framework

In studying team cognition, as well as other complex phenomenon involving human cognitive abilities, the idea of an experimental platform has started to take form (12). Previous researchers in the field have developed systems to manage the research process, as well as the training process. This framework hopes to capitalize on these prior systems by use of modular plug-in components and defined interfaces. This will allow the "best of breed" tools and systems to be integrated into the framework.

The EXpertise transferred to novice Performance for Evaluation, Research and Training (EXPERT) framework is designed to be scalable and flexible which could provide a working platform for Cognitive Research, Training as well as Adaptive User Interfaces and Human Factors/User Interface evaluation.

As discussed in sections 1 and 2 above, both the delivery of a deployable mission rehearsal training system with access to targeted feedback for both internal and external process contributing to team performance, as well as the research platform needed to reach the point of such a deployable system requires the flow and integration of data that is easy for non-programmers to operate. The components shown in Figure 3 would make up such a system.

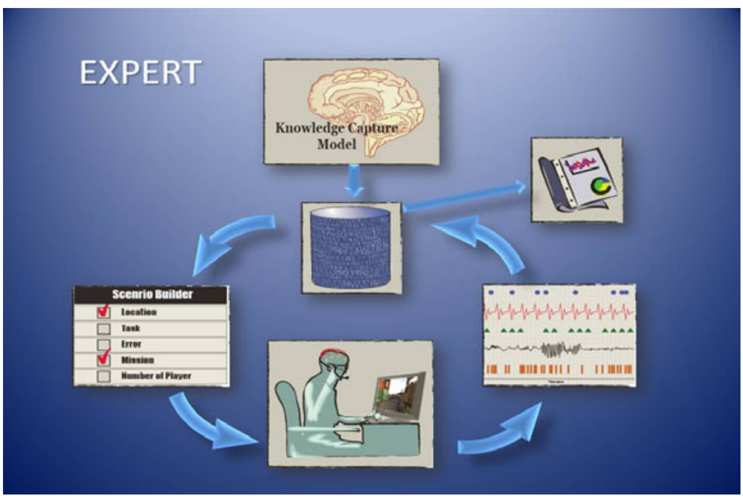


Fig. 3. EXPERT Framework Components

4.1 Knowledge Capture and Cognitive Model

Desired or expert knowledge, processes, and procedures need to be captured, edited regularly and understood by the expert and training staff. The ability to systematically capture and represent this expert model allows the model to become a foundational piece to the EXPERT framework to track variations from the model during scenario execution and provide a basis for establishing “success” criteria for a study. Variations may represent errors or at least events of interest to the study at hand whereas adherence to the model may represent confirmation or achievement of cognitive milestones.

4.2 Database

The database will provide the storage for the expert cognitive model as well as the operator/subject testing data that is stored in SCORM compliant format. This will allow the tracking of user progress toward objectives as well as the training record that could interface with a Learning Management System, if desired. The database will also provide the content required by the Mitigation component.

4.3 Scenario Builder

This component is the intersection of the cognitive model, the presentation system and the training objectives. It will contain a user interface that allows a trainer, tester or experimenter to specify the characteristics of the scenario to be run. Multiple scenarios can be created, stored and accessed by the testing component (for sequencing, etc). The builder will pass those values to the presentation system and also link to the knowledge model to draw the knowledge and behaviors appropriate for target participants and the OPFOR avatars. Scenario characteristics, such as the following need to be extracted and presented in a “checklist” format: pre-conditions, equipment load out, location and related map to be brought into presentation component, mission or tasking order details, personnel players, OPFOR personnel, and errors, threats, and situations of interest.

4.4 Presentation System

Presentations can include prototypes or already developed systems, training presentations such as simulations or serious games, or specially developed presentations for research studies. Outputs presented to the participant include sound, smell, facial features, virtual 3-D, and haptic sensations. Input devices to the system will include keyboard, mouse, joystick, touch, and gesture.

4.5 Data Collection

The framework will allow the collection of a variety of participant and system outputs for analysis to form the basis for understanding the current student/team model.

- Neurophysiologic sensors data from systems such as EEG, EKG, GSR
- Eye movement, fixations and pupilometry via eye trackers
- Participant facial responses
- Communication both verbal and chat
- Critical system events such as the presentation of cues as well as user’s actions taken within the system

- Performance measures. This may require an input module to prompt the participant for a response or may require correlation to the model to monitor performance.
- Synchronized logging of system events, sensor data, and user performance is stored and made available to the framework.

4.6 Mitigation

This component will compare the student model with the expert model. It will contain the rules for when a presentation intervention/change is required based upon participant and system data. It will also point to the content (remediation information, displays, etc) needed to address the mitigation.

As our research continues our goal is to build out this framework.

Acknowledgements. The development of the IMMEX ANN was supported by The Defense Advanced Research Projects Agency under contract number(s) NBCHC070101, NBCHC090054. The views, opinions, and/or findings contained in this article/presentation are those of the authors and should not be interpreted as representing the official views or policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the Department of Defense.

References

1. Warner, N., Letsky, M., Cowen, M.: Cognitive Model of Team Collaboration: Macro-Cognitive Focus. In: Proceedings of the 49th HFES Annual Meeting, Orlando, FL, September 26-30 (2005)
2. Dyer, J.L.: Team research and team training: A state-of-the-art review. In: Muckler, F.A. (ed.) Human Factors Review, pp. 285–323. Human Factors Society, Santa Monica (1984)
3. Klein, G.A., Zsambok, C.E., Thordsen, M.L.: Team decision training: Five myths and a model. Military Review, 36–42 (1993)
4. Cooke, N., Kiekel, P., Salas, E., Stout, R.: Measuring team knowledge: A window to the cognitive underpinnings of team performance. Journal of Applied Psychology 7, 179–199 (2003)
5. Mathieu, J., Heffner, T., Goodwin, G., Salas, E., Cannon-Bowers, J.: The influence of shared mental models on team process and performance. Journal of Applied Psychology 85(2), 273–283 (2000), doi:10.1037/0021-9010.85.2.273
6. Blickensderfer, E., Cannon-Bowers, J.A., Salas, E., Baker, D.P.: Analyzing Knowledge Requirements in Team Tasks. In: Schraagan, J., Chipman, S., Shalin, V. (eds.) Cognitive Task Analysis, Mahwah, NJ, pp. 431–447 (2000)
7. Klein, G.: Cognitive Task Analysis of Teams. In: Schraagan, J., Chipman, S., Shalin, V. (eds.) Cognitive Task Analysis, Mahwah, NJ, pp. 417–429 (2000)
8. Stevens, R.H., Galloway, T., Berka, C., Sprang, M.: Can Neurophysiologic Synchronies Be Detected during Collaborative Teamwork? In: Proceedings: HCI International 2009, San Diego, CA, July 19-24, pp. 271–275 (2009)
9. Berka, C., Levendowski, D., Cventinovic, M., Petrovic, M., Davis, G., Lumicao, M., Zivkovic, V., Popovic, M., Olmstead, R.: Real-time Analysis of EEG Indices of Alertness, Cognition, and Memory with a Wireless EEG Headset. International Journal of Human-Computer Interaction 17(2), 11–170 (2004)

10. Berka, C., Levendowski, D.J., Ramsey, C.K., Davis, G., Lumicao, M.N., Stanney, K., Reeves, L., Regli, S., Tremoulet, P.D., Stibler, K.: Evaluation of an EEG-Workload Model in an Aegis Simulation. In Biomonitoring for Physiological and Cognitive Performance during Military Operations. In: Caldwell, J.A., Wesensten, N.J. (eds.) Proceedings of the International Society for Optical Engineering, vol. 5797, pp. 90–99 (2005)
11. Berka, C., Levendowski, D.J., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T., Olmstead, R.E., Tremoulet, P.D., Craven, P.L.: EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning and Memory Tasks. Aviation, Space, and Environmental Medicine 78(5) (2007)
12. Clark, M., Skinner, A., Kruse, A., Berka, C., Fidopiastis, C.: The Transfer of Learning from Virtual Reality Environments and Immersive Games to the Real World: Approaches and Challenges (2010) (unpublished paper)

Theoretical Transpositions in Brain Function and the Underpinnings of Augmented Cognition

Cali M. Fidopiastis

University of Alabama-Birmingham
1530 3rd Avenue South
Birmingham, Alabama 35294
cfidopia@uab.edu

Abstract. Augmented Cognition (AugCog) explores behavior in real-time and in real world settings. This research avenue is a departure from standard experimental approaches such as those accepted in the fields of Cognitive Psychology and the Neurosciences. AugCog as a field of study, therefore, has the potential to up-end some of the tried-and-true laboratory based findings on such topics as learning and transfer of learning. Steeped in history from both the biological systems perspective and the cognition neuroscience vantage, the future of AugCog seems contingent on its success at merging these paradigms and concurrently producing analysis tools with which to keep peering into the brain as it functions in operational environments. In this paper, we review the theories that drive Augmented Cognition approaches and evaluate their capacity to keep the field moving forward.

1 Introduction

Augmented Cognition (AugCog) was borne out of the Defense Advanced Research Projects Agency's (DARPA) push for technologies that enhanced the Warfighter's communication skills and those technologies that improved biosensing for medical applications [1]. The significance of combining these technologies into a capability that proposes to improve human cognitive abilities within high stress operational environments utilizing cognitive state information and a seamless human-computational interface cannot be understated. Technology, methods and, constructs of cognitive neuroscience found their way into the dynamically changing, and at times, highly stressful reality of the combat soldier. However, AugCog is an applied research paradigm, and thus necessitates appropriate brain theories, sensitive measurement tools, as well as valid, reliable metrics to support its continued success.

Kruse [2] identified several research gap areas that encompass AugCog protocols. These gap areas included validating baseline cognitive state measures and separating task-independent cognitive states and functions. However, the research gaps may be more fundamental. For example, the choice of brain theory guides the experimental approach and subsequently affects the results. More specifically, AugCog accepts an information-processing approach to brain function. This theoretical base determines the types of response variables and overall metrics needed to create a closed-loop adaptive human-computer interface. What evidence is there to suggest the type of

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 153–158, 2011.

information-processing approach appropriate for reliably classifying cognitive states such that state information can in turn become a reliable response variable to the system?

From a historical perspective, Bartlett [3] in his studies of memory also questioned how well we understood the internal state of the individual, especially at the start of an experiment (p. 10). He was concerned that the use of statistics would lend to confirmation bias and false support for theories based solely on results from laboratory experiments, which may be far from real world responses. In his view, we can never know the internal starting point of a participant; therefore, individual differences plagued experimental design despite randomization. Bartlett chose meaningful stimuli and explored how people passed information on to others (serial reproduction) or remembered the information themselves over time (repeated reproductions). Highly reliable or reproducible behavior patterns across participants emerged after repeating the experiment over several sessions. Repeating the trials with the same participants controlled for individual difference, while using two related response variables converged outcomes to suggest that memory was not veridical or stored in a single place or memory trace. While memory involved processing, it was not linear or statically stored information.

Bartlett predicted that the interface between environment and mental process produced emergent properties of the memory "trace" not found in the foundational interaction [4]. Thus, memory formation and its subsequent recollection are active processes whereby there is a "top-down" or personal influence when interpreting the world. With respect to memory, Bartlett's active construction process based upon prior knowledge is very different from a stimulus driven associative process that theorists such as William James proposed [5]. This latter constrained single direction of brain processing stages seems inherent in the current AugCog approach. Moving to the 21st century, do current dynamical neural systems approaches and their more explicit relation to cognitive processes provide a better foundation from which to operationally define important constructs of AugCog?

In this review, we outline the historical attributes of information-processing theory that currently influence the AugCog approach. Multiple resource theory, which drives underlying constructs pertaining to adaptive system design, is also discussed [6]. We further present alternative theories such as cybernetics and dynamic neural systems that may change the outcome of not only the system design, but also how we measure operator cognitive states.

2 Historical Review of the Information Processing (IP) Approach

Neisser [7] defined 'cognition' as "all the processes by which the sensory input is transformed, reduced, elaborated, stored, recovered, and used". The computer metaphor acted as an organizing principle from which to guide appropriate avenues of study. Humans as information-processing systems actively acquired knowledge and skills through interaction with the environment, as well as through inner reflection. Information was then processed, stored, and subsequently retrieved when needed for responding. Foundational units of cognition are internal mediating states called representations that can be symbolic (image or words), enduring, and amodal (sensory independent) in nature [8]. Learning became a process, then that involved the forming, integrating, and organizing of mental representations in a series of sequences and stages. What a person knows, therefore, is not necessarily observable by their behavior as suggested by other schools of thought (i.e. Behaviorism).

Mental representations are similar to data structures of a computer whereby they allow for efficient manipulation and storage of information. More specifically, a representation defines the type of input and output (information) a process or more generally, a system handles as opposed to the environment (i.e., behaviorism). Computations are the logical rules that organize and operate on representations, for example, to create new representations. In general, an information processing approach determines the sequence of metal operations (cognitive processing) and defines the products (representations) that delineate the cognitive constructs (e.g., memory systems) of interest.

David et al. [9] described the fundamental constructs of cognitive psychology as: information processing, cognitive structures, and cognitive architectures. Accordingly, IP provides the rules necessary to transform mental representations. A cognitive structure (e.g., memory systems) is the organized outcome of these computations. A cognitive architecture then integrates and unifies all related cognitive structures, processes, and representations such that a model (e.g., symbolic or network) can explain and predict cognition [10]. Further, the IP theory afforded analytic methods that allowed for mathematical definitions that provided an objective means of characterizing, quantifying, and qualifying underlying brain processes [11]. IP, as a framework, provided the constructs and analysis tools for a rich exploration into the nature of information processing (e.g., continuous versus discrete transmission, limited-capacity versus capacity free processing, etc.). However, these experiments remained laboratory based and contrived in the sense that they matched assumptions to a model and not directly to the human operator. Would the IP model hold in operational environments?

3 Multiple Resources Theory and the Human Operator

Multiple Resource Theory (MRT) extends IP theory into the multi-tasking domain of real world human performance [6, 12]. Resources are loosely defined as "commodities" possibly "pools of energy" used during information processing. Wickens [12] postulated the existence of separate resources for the different stages of processing. These stages include perception, cognition, and responding. In addition, codes of processing or algorithms used to transform information are different across modalities such as vision and audition. The model suggests that time-sharing will be more efficient when tasks do not share the same resources within a stage. The strength of this theory is its predictive power when determining performance outcomes of dual-task experiments [6, 13]. The MRT has become a widely used heuristic in human-system interface design [14]. While cited as a theoretical premise for AugCog research, the MRT may not be detailed enough to predict performance results from dynamically changing, complex tasks.

Wickens [6] outlined weaknesses of the MRT such as not considering tactile input to the modalities dimension and the inability of the model to discern resources allocation. Szalma and Hancock [15] suggested that the term resources itself is not operationally defined meaningfully. Further, the MRT fails to distinguish between bottlenecks due to structural attributes (e.g., brain architecture) and those that are truly resource related. While decreased performance is the common outcome of either a structural or a resource limitation, from the AugCog perspective, these underlying mechanisms may lead to different approaches to mitigating the performance decrement. Thus, a better understanding of how the brain works may be necessary to continue the forward progress of AugCog.

4 Behavioral Cybernetics and Automated Systems Intelligence

Current neuroimaging data support that the brain performs information-processing tasks. However, it is also true that the human system provides feed-forward and feed-backward controls such as those proposed by behavioral cybernetics. According to a cybernetics framework, motor processing modified through sensory feedback underlies all behavior [16]. Theoretically, this closed-loop system may provide a more natural paradigm from which to guide interface design. Human-system interactions should determine the types of sensory feedback and subsequent motor control, both static and dynamic, necessary for optimal task performance. Thus, task-specific factors supported more or less by the design features within the human-system setup affects cognitive learning, transfer of training, and other individual difference based outcomes. Smith et al. [16] suggested that task design components affect cognitive performance of the operator more than internal information-processing bottlenecks (p 284). Smith and Henning [17] further contended that information-processing theories are "nonrefutable" and provide little explanatory power; therefore, theories such as the MRT are incapable of driving the field of AugCog.

The failure of information-processing approaches to account for motor-sensory control issues of the operator potentially oversimplifies the human to input/output responses that do not account for feedback based control or behavioral adaptation in response to changing environmental variables. The AugCog closed-loop system would lack appropriate transfer of control between the automated system and the human operator. A system that does not allow the operator to predict next actions readily or flexibly may be a deterrent to consistent task performance and learning [17]. The adaptive system within the AugCog closed-loop design should be intelligent enough to support the operator in a cooperative manner. In this bi-directional information exchange between the adaptive system and the operator, how then does the system initialize the starting point of the operator and subsequently identify state changes (machine and human) that may be detrimental to the symbiotic relationship needed during task performance? The lack of specificity of either behavioral cybernetics or IP theories are not convincingly powerful enough alone to address the critical design issues of an AugCog system.

5 Dynamical Neural Systems and Cognitive State

Cognitive state implies the existence of a measurable global neurocognitive state and additionally the potential to measure the status of different brain systems that mediate response behavior [18]. The IP approach suggests that the brain manipulates or processes brain representations via specified rules or in the case of a connectionist model, learning algorithms [19]. Therefore, we must localize the brain area that facilitates processing of task related representations to assess operator cognitive status. Bressler [18] points out that the brain is highly interconnected via long and short-range connections, thus cognitive state may not be measurable through a localization or input/output approach.

A state reflects a change over time that can dynamically shift based upon internal and external forces. Coordinated brain patterns and not the sequential sense-think-act model may apply better when considering meaning in the context of a particular environment or situation [19]. The theoretical position of coordination dynamics posits that there is a dynamic coupling between brain areas, as well as the world, that functions to convey cognition [20]. Electroencephalography (EEG) may capture the neural signatures related to brain state changes better than other neurosensing tools (e.g., functional magnetic resonance imaging-MRI). A dynamical systems approach may provide the flexibility and measurability needed for successful AugCog research.

6 Conclusion

In his historical review of skill acquisition, Adams [21] noted that, "experiments enriched by history could contribute to the science rather than only brightening an inconsequential corner" (p. 41). Additionally, grounding theoretical explorations within a historical framework protects against fad research that detracts from the forward progress of a field of study. AugCog, from its inception, based itself in cognitive psychology theory. Translating information-processing theories into a viable applied approach whereby reliable measures of operator cognitive state within operational environments are questionable. A more sustaining and fruitful approach to AugCog research may lie in understanding dynamical neural systems.

References

1. Schmorrow, D.D., Kruse, A.A.: Augmented Cognition. In: Bainbridge, W.S. (ed.) Berkshire Encyclopedia of Human-Computer Interaction, pp. 54–59. Berkshire Publishing Group, Great Barrington (2004)
2. Kruse, A.A.: Operational neuroscience: Neurophysiological measures in applied environments. Aviation, Space, and Environmental Medicine 78(5), B191–B194 (2007)
3. Bartlett, F.C.: Remembering: A study in experimental and social psychology. Cambridge University Press, Cambridge (1932)
4. Brewer, W.F., Nakamura, G.V.: The nature and functions of schemas. In: Wyler Jr., R.S., Srull, T.K. (eds.) Handbook of Social Cognition, vol. 1, pp. 119–160. Erlbaum, Hillsdale (1984)
5. James, W.: The principles of psychology, vol. I. Dover Publications, New York (1890); Reprint edition (June 1, 1950)
6. Wickens: Multiple resources and mental workload. Hum Fac. 50(3), 449–455 (2008)
7. Neisser, U.: Cognitive Psychology, p. 4. Appleton-Century-Crofts, New York (1967)
8. Markman, A.B., Dietrich, E.: Extending the classical view of representation. T. in Cog. Sci. 4(1), 470–475 (2000)

9. David, D., Milcea, M., Opre, A.: The information-processing approach to the human mind: Basics and beyond. J. of Clin. Psych. 60, 353–368 (2004)
10. Newell, A.: Unified theories of cognition. Harvard University Press, Cambridge (1990)
11. Massaro, D.W., Cowan, N.: Information processing models: Microscopes of the mind. Ann. Rev. of Psych. 44, 383–425 (1993)
12. Wickens: Multiple resources and performance prediction. Theor. Iss. in Erg. Sci. 3(2), 159–177 (2002)
13. Boles, D.B., Law, M.B.: A simultaneous task comparison of differentiated and undifferentiated hemispheric resource theories. J. of Exp. Psych. 24(1), 204–215 (1998)
14. Hancock, P.A., Oron-Gilad, T., Szalma, J.L.: Elaborations of the multiple-resource theory of attention. In: Kramer, A.F., Wiegmann, D.A.,, K. (eds.) Attention: From Theory to Practice, pp. 45–56. Oxford University Press, Oxford (2007)
15. Szalma, J.L., Hancock, P.A.: On mental resources and performance under stress. White paper, MIT2 Laboratory, University of Central Florida (2002), http://www.mit.ucf.edu
16. Smith, T.J., Henning, R.A., Smith, K.U.: Sources of performance variability. In: Salvendy, G., Karwowski, W. (eds.) Design of Work and Development of Personnel, ch. 11, pp. 273–330. Wiley & Sons, New York (1994)
17. Smith, T.J., Henning, R.A.: Cybernetics of augmented cognition as an alternative to information processing. In: Proceedings of the 1st International Conference on Augmented Cognition, Las Vegas, November 22-27 July (2005)
18. Bressler, S.L.: The formation of global neurocogntive state. In: Perlovsky, L.I., Koma, R. (eds.) Neurodynamics of Cognition and Consciousness, pp. 61–72. Springer, Berlin (2007)
19. Beer, R.D.: Dynamical approaches to cognitive science. Tr. in Cog. Sci. 4(3), 91–99 (2000)
20. Bressler, S.L., Kelso, J.A.: Cortical coordination dynamics and cognition. Tr. in Cog. Sci. 5(1), 26–36 (2001)
21. Adams, J.A.: Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. Psych. Bull. 101(1), 41–74 (1987)

Non-invasive Functional Brain Biomarkers for Cognitive-Motor Performance Assessment: Towards New Brain Monitoring Applications

Rodolphe J. Gentili[1,2]

[1] Cognitive Motor Neuroscience Laboratory
Department of Kinesiology
School of Public Health University of Maryland
College Park, MD 20742, USA
[2] Neuroscience and Cognitive Science Program University of Maryland
College Park, MD 20742, USA
rodolphe@umd.edu

Abstract. Along with theoretical advances in neuroscience research, recent neurotechnological developments provide portable recording and processing systems that can be employed for real-time assessment in applied military environments. This article provides a brief overview of research related to non-invasive brain biomarkers derived from brain signals that can track brain dynamics during cognitive-motor performance. Potential applications of such brain biomarkers for military personnel such as neurofeedback for accelerated learning as well as brain monitoring for performance assessment and rehabilitation are discussed.

Keywords: Cognitive-motor performance, EEG/fNIRS biomarkers, alpha power, phase synchronization, brain monitoring, neurofeedback, rehabilitation.

1 Introduction

In parallel with the continuous advancements of neuroscience, the recent developments of neurotechnology are able to provide wearable sensors and portable recording systems that can be employed for real-time assessment in applied environments [1],[2]. Thus, the idea to transfer some current techniques from laboratories to the field is increasingly being considered. Among the various avenues of applied neurosciences, applications in clinical settings and in the operational realm could be highly beneficial to the military as well as commercial/industrial communities [2]. In particular, when considering the military environment, it is widely accepted that "the Soldier is the acknowledged centerpiece of the Army's warfighting system, and success largely depends on the mental status of these individuals." ([3], p.1). Therefore, there is a need to develop rigorous neuropsychological/neurological assessment techniques, which would represent a major advance in Soldier performance research [3],[4]. Specifically, the training of individuals, as well as the monitoring of individual cognitive-motor states during performance, appear to be two important avenues that need to be considered in operational settings [2].

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 159–168, 2011.

Training and monitoring of cognitive-motor performance can be based on the combination of behavioral analyses as well as on the investigation of brain dynamics. Brain activity assessment can provide sensitive indicators to evaluate the brain/mental status of the performer. For instance, while behavioral outcomes are informative on the quality of performance, the performance can be influenced by multiple confounded factors (e.g., fatigue, workload, level of expertise, etc.), which could be revealed by accessing brain/mental status.

Such assessment is possible by employing non-invasive embedded brain monitoring systems based on specific indicators or brain biomarkers to track brain dynamics. Additionally, such monitoring tools should allow for the tracking of brain dynamics in ecological situations in which humans learn new tasks, master novel tools, and/or adapt to changing environments. Thus, these brain biomarkers should not only be non-invasive (i.e., no surgical intervention needed), but also simple to record and analyze in addition to being robust and sensitive to specific changes in brain function in natural situations. Electroencephalography (EEG) is well suited for such assessment in situations that require non-invasive recording of the dynamic brain activity with high temporal resolution (e.g., millisecond). In addition, hemodynamic signals such as functional near infra-red spectroscopy (fNIRS) can also be recorded by portable devices and provide a good complementary method to EEG [5]. Over the years, multiple research efforts (e.g., [6],[7],[8],[9],[10],[11],[12],[13],[14],[15],[16]) proposed methods and provided such biomarkers that could be employed to track the brain status during cognitive-motor performance.

This article, will first overview elements related to brain biomarkers of cognitive-motor performance. Then, potential applications in a military context such as neurofeedback for accelerated learning as well as brain monitoring for veteran rehabilitation and performance assessment will be discussed.

2 Elements of Overview

This section will focus mainly on brain biomarkers associated with cognitive-motor performance derived from EEG and to a lesser extent from fNIRS. As mentioned above, both can be recorded by portable systems and, therefore, are particularly well suited for brain assessment during cognitive-motor performance (e.g., aiming, reaching) in the field.

2.1 Spectral Power

Since the seminal work conducted by Hatfield and colleagues [10] close to three decades ago, a growing body of evidence suggests that it is possible to assess the cortical dynamics of cognitive-motor skills in expert performers during visuomotor tasks. These investigations revealed progressive changes in EEG during skill learning and also differences in the level of EEG power between novice and expert sport performers [8],[9],[11],[14],[13],[14],[15],[16]). Particularly, alpha (~ 8-13 Hz) and theta (~ 4-7 Hz) power were positively related to the level of cognitive-motor performance [8],[11],[12],[13]. For the sake of clarity and conciseness, only the results related to alpha power are presented here. However, theta oscillations also appear to be related to performance enhancement and, specifically, an increasing body of evidence supports that theta oscillations could be functionally related to error monitoring [9],[17],[18].

Namely, previous investigations reported that experts demonstrated an overall increase in EEG alpha power compared to novices in the left temporal region during a precision aiming task [11],[12]. The results suggest that differences in EEG alpha power are related to the differences in the level of mastery of the cognitive-motor task. In general, such EEG changes are indicative of high levels of skill and associated with a cortical refinement leading to reductions of nonessential cortical resources to perform the task [11]. This is consistent with the idea that a high level of alpha power corresponds to a reduced activation of a given cortical region indicating a reduction of the recruitment of neural resources [11]. Also, these differences in cortical dynamics between novices and experts mirror important performances discrepancies. Namely, the experts scored higher and exhibited lower performance variability compared to the novices. Thus, these studies can contribute to the development of brain biomarkers (e.g., changes in EEG alpha power) capable of identifying a high level of cognitive-motor performance resulting from an extensive practice period. However, this research did not focus on the evolution of such brain biomarker throughout the training period itself.

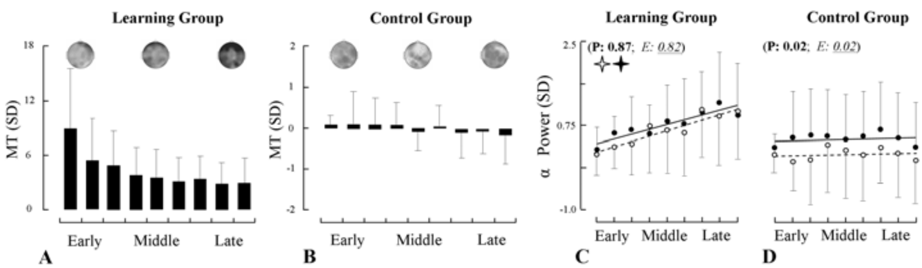


Fig. 1. Concomitant EEG and kinematic changes throughout adaptation for the learning and control groups. Changes in movement time (MT) for the learning (A) and the control (B) groups for early, middle, and late adaptation. The scalp plots represent the changes in alpha power through the same practice period for the planning stage. Changes in the magnitude of the standardized (sd units) EEG alpha power during planning (black circles) and execution (empty circles) for the learning (C) and control (D) groups. (Adapted from [9]).

Cortical changes during learning were examined in a recent investigation that assessed changes in brain dynamics throughout a marksmanship intensive training of novice performers from the US Naval Academy during three months [13]. The findings revealed that throughout the training period the shooting performance was enhanced along with an increase in alpha power at the (contralateral) left temporal site while such a result was not observed when participants were at rest. Recently, this research has been extended by employing movements related to daily activities (e.g., reaching, grasping; [9],[16]). This research employed a shorter learning timeframe (hours instead of weeks/years) of a new cognitive-motor skill and assessed interferences with previously acquired motor experiences [9]. This latter aspect is important since, in practice, Soldiers often need to switch/adapt rapidly from one task to another, potentially inducing interference between tasks. In this study, EEG brain signals were analyzed from participants divided into two groups: i) a learning group that had to learn a new visuomotor transformation while performing drawing movements requiring suppression of familiar cognitive-motor responses, and ii) a control group for whom no visual transformation had to be learned [9]. The findings revealed that as participants of the learning group enhanced their cognitive-motor performance, EEG frontal alpha power during both movement planning and execution was progressively increased. It was suggested that such changes reflected initial involvement of frontal executive functioning to suppress established visuomotor mappings followed by a progressive idling [9]. No change in EEG alpha power nor performance was observed in the control group.

Thus, altogether this research suggests that changes in EEG alpha power can be used as non-invasive functional brain biomarkers either to assess the level of mastery of a particular cognitive-motor skill and/or to track the evolution of brain dynamics throughout cognitive-motor training.

2.2 Phase Synchronization

Besides biomarkers derived from EEG power, phase synchronisation (coherence, phase locking value (PLV)) of EEG signals that reflect the degree of cortico-cortical communication also provide brain biomarkers for cognitive-motor performance.

Although relatively less common than studies investigating EEG power, several recent investigations focusing on the degree of cortico-cortical communications in relation to changes in cognitive-motor performance have been conducted [6],[7],[16],[19]. Generally, these studies have provided convergent evidence that skilled individual exhibited a reduced level of coherence between cortical areas. For instance, EEG coherence between frontal and several other cortical regions in two groups of highly skilled marksmen who were similar in expertise, but who differed in competitive performance history were compared. One group performed consistently better in competition and exhibited significantly lower coherence between the left temporal region and the premotor area in the high-alpha (10–13 Hz) and low-beta (13–22 Hz) bandwidths during the aiming period [6]. Similarly, it was shown that coherence could assess brain dynamics in relation to the level of mastery of a motor task. Particularly, expert marksmen exhibited lower coherence over the whole scalp compared to novices, with the effect most prominent in the right hemisphere. Coherence was positively related to aiming movement variability in experts [7]. A reduction in coherence is generally interpreted as a refinement of cortical networks in experts reflecting a reduction of nonessential functional communications among the cortical regions of interest associated to a superior performance [11].

Recently, (similar to the EEG power, see section 2.1) this research has been extended by considering a short period of training during which participants had to learn a new visuomotor task while inhibiting prepotent cognitive-motor responses [20]. The findings revealed a decrease of phase synchronization (PLV) for both movement planning and execution as participants adapted throughout training. These changes were correlated with enhanced kinematics as the task progressed.

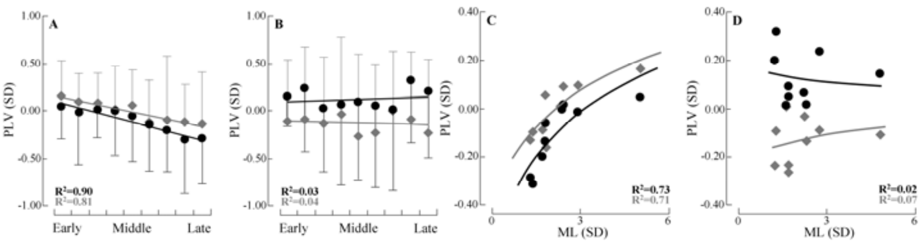


Fig. 2. Changes in PLV for the learning (A) and control (B) groups throughout practice during the planning stage. PLV versus movement length (ML) for the learning (C) and the control (D) groups. Standard deviation (SD) is used as unit. The black and gray color represent the PLV values for the pair of electrodes Fz-F4 (high beta band) and Fz-C4 (gamma band), respectively.

Therefore, as a whole, changes in EEG coherence can be used as additional functional brain biomarkers to assess the level of a particular cognitive-motor skill and/or to track brain dynamics during cognitive-motor training.

2.3 Functional Near Infra Red Spectroscopy and Hemodynamic Responses

Besides EEG it is also important to investigate hemodynamic changes by using alternative approaches such as fNIRS to derive brain biomarkers related to cognitive-motor performance. While EEG measures electrical activity, fNIRS measures blood oxygenation levels in the brain, providing a different, possibly complementary, source of information about brain functioning that is accessible with portable recording systems [1],[5]. fNIRS is an emerging optical brain imaging modality that measures hemodynamic response in order to provide biologically relevant indicators for brain functions such as cognitive workload [21] and cognitive-motor performance [22],[23]. These qualities make fNIRS well-suited for assessing brain status during ecological situation in the field [21],[22]. Specifically, several studies investigated the brain activity by employing fNIRS when comparing cognitive-motor skills in novice and expert performers during training [22],[23]. For instance, by analyzing fNIRS signals recorded in prefrontal regions and manual dexterity, technical skill of expert and novice surgeons while performing a surgical knot-tying task were compared [22]. The findings revealed a decrease in relative changes of total and oxygenated hemoglobin as well as an increase in deoxygenated hemoglobin throughout training. It was concluded that learning-related refinements in performance were mediated by reductions in prefrontal activation. Recently, by using a task where participants had to learn a novel visuomotor transformation while suppressing familiar motor plan (see section 2.1), the changes in fNIRS recorded at the prefrontal regions along with changes in performance were investigated [23]. Preliminary results revealed that throughout training the performance was enhanced along with a simultaneous and progressive reduction in prefrontal oxygenated hemoglobin.

The high prefrontal activation during early learning may reflect a primary role of the inhibitory processes to suppress familiar motor responses of inappropriate actions while such a role is reduced during late learning reflected by a smaller prefrontal activation [23]. It must be noted that the enhancement of the performance along with a reduction of the prefrontal activity are in accordance with results revealing an increase in prefrontal alpha power previously reported with the same task [9].

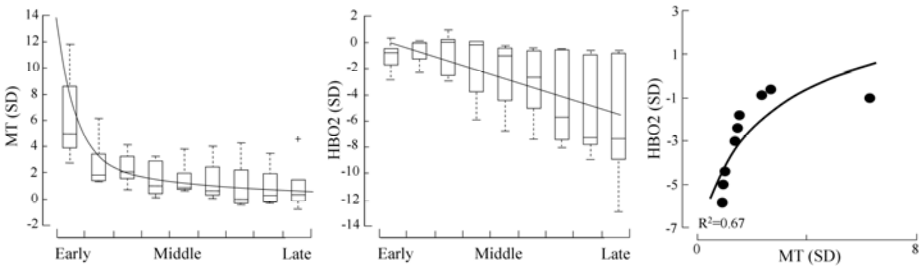


Fig. 3. Changes in movement time (MT; left panel); in oxygenated hemoglobin (HBO2; middle panel) throughout adaptation for the learning group. Oxygenated hemoglobin versus movement time (MT) for the learning groups (right panel). Standard deviation (SD) is used as unit (Adapted from [23]).

Therefore, although additional research is needed, changes in oxygenation, deoxygenation, total haemoglobin extracted from fNIRS signals seem to be promising to provide (hemodynamic-based) non-invasive functional brain biomarkers to assess the level of a particular cognitive-motor skill and/or to track the evolution of brain status during cognitive-motor practice.

3 Future Potential Applications

3.1 Neurofeedback and Accelerated Learning

A possible area of application of these EEG/fNIRS biomarkers would be the training of military personnel. “Today's warfighter is required to master a large number of diverse skills spanning the range of cognitive and motor domains in increasingly rapid timeframes” [24]. Thus, the training/learning of cognitive-motor performance is often performed under time pressure and may require the acquisition of several tasks at the same time. Thus, there is a growing interest to develop methods to enhance/accelerate the learning.

A possible method to enhance/accelerate learning would be to use augmented feedback with neurofeedback (or biofeedback) systems. Neurofeedback (or biofeedback) is a training technique that measures processes and provides feedbacks (e.g., visual, auditory) based on the neural activity (e.g., EEG) of individuals to help these individuals learn to control/change their brain activity. Although, further research seems to be needed, one possible application of neurofeedback with healthy individuals is enhancement of cognitive-motor performance [25]. For instance, based on previous results associating the changes in EEG alpha power at specific scalp sites and performance outcomes, it was shown that neurofeedback could significantly improve shooting performance in twenty four skilled archers [26]. The effect of EEG neurofeedback training was also examined in golfers, comparing performance enhancement with and without neurofeedback [27]. It was shown that the overall percentage of successful putts was significantly greater once neurofeedback was administered. Moreover, combined with additional findings (e.g., performance related to memory, attention, creativity and mood [28],[29]) it appears that neurofeedback can elicit positive changes in performance and thus has promising implications to enhance/accelerate cognitive-motor performance.

Therefore, since the functional biomarkers previously described were able to track changes in cerebral cortical dynamics during learning of a new cognitive-motor skill, it would seem reasonable to integrate them in a neurofeedback system to provide augmented feedback to individuals and thus analyze how this would enable enhanced/accelerated learning. Thus, this augmented (brain-based) feedback could be combined with feedbacks from the behavior that should allow the participants to learn faster/better compared to the situation when only 'classical' feedbacks are available [27]. When considering the augmented feedback, (without overloading the information processing capacity of individuals) multimodal augmented feedback could be considered including EEG (e.g., power, coherence) and fNIRS (e.g., oxygenated hemoglobin) biomarkers to provide a more robust, accurate and efficient feedback during the learning of new cognitive-motor skills. However, general (e.g., directionality of the changes in power and coherence) as well as specific (e.g., brain sites) features of these biomarkers depend on several factors (such as the nature of the task; task requiring inhibition of prepotent cognitive-motor responses or not, etc.) and thus should be taken into account in neurofeedback systems design.

3.2 Brain Monitoring for Cognitive-Motor Performance Assessment

Military Personnel

Another application of these brain biomarkers would be to monitor brain activity for cognitive-motor performance assessment in military personnel performing a specific task. An application that seems particularly well-suited would be to monitor brain activity when behavioral performance may be not directly accessible. This is typically the case of the remote warfighters where Soldiers are unseen from other team members, operating in remote outpost (e.g. inside building in urban conflict) [3]. Thus, the recording and processing of objective brain biomarkers for such individuals would allow brain monitoring related to performance, fatigue, workload or stress/emotional status assessment [3],[30]. Thus, biomarkers related to cognitive-motor performance could also be combined with others related to specific mental states such as fatigue [31], workload [32] or stress/emotional responses [33]. For instance, it has been recently suggested that changes in EEG coherence would provide a metric that could monitor workload/task-related mental demand and assess cognitive-motor learning [32]. Moreover, it was also suggested that the commonly observed high level of alpha power this is generally associated to the mastery of cognitive-motor skills in the left temporal region appears to be reduced when individuals perform under pressure [33]. In the same vein, theta power increase can be related to fatigue and reduced alertness [31] as well as performance enhancement [17],[9]. Interestingly, changes in theta power related to drowsiness are reflected by the low component (~4-5 Hz) located at the central and posterior sites of the frontal regions [31],[34] while those observed for cognitive-motor performance enhancement are reflected by the high component (~ 6-7 Hz) located at the anterior and lateral sites of the frontal regions [23]. Therefore, when combining these various biomarkers it would possible to disentangle and thus assess, to some extent, separately the level of cognitive-motor performance, stress, fatigue and workload. It is important to note that even if the behavioral performance is available, the monitoring of brain states is still important since a poor cognitive-motor performance can result from several factors (e.g., stress, workload, fatigue, etc.) that will affect the behavior. Thus, such brain status assessment could help to provide a more accurate and relevant decision making tool related to Soldier and task force management. For instance, if the level of fatigue or stress is becoming too high, the Soldier should be replaced before he/she makes any mistakes. In the future, brain monitoring systems based on multiple biomarkers (e.g., EEG power, coherence) could allow identifying when Soldiers are incapacitated and predict forthcoming human failure. By providing robust brain status assessments, such multimodal brain monitoring systems could be particularly significant as a part of a more general system that can automatically take over from the human in the loop in order to prevent any critical failure [3].

Rehabilitation of Veterans

Another interesting avenue for brain monitoring applications would be the rehabilitation of veterans and specifically to provide assessment tools for diagnostic, prediction and recovery. This seems particularly relevant for veterans having cognitive-motor problems (e.g., post-traumatic stress disorder, traumatic brain injury) such as the "Invisible Wounded Soldier" where the integrity of brain functions has been compromised. Thus, in addition to the information provided by the usual behavioral performance, brain biomarkers specifically engineered for a particular impairment could provide critical metrics for diagnosis (e.g., the Soldier is able to return to duty if desired, impairment identification, etc.). Similarly, the combination of behavior and brain biomarkers could be employed for cognitive-motor restoration by providing complementary information that would allow identifying potential cerebral compensation mechanisms. One promising direction would be to employ such biomarkers to assess the dynamics of the cognitive workload capacity recovery in veterans throughout the rehabilitation process. In such a context, heightened mental workload could be indicative of compromised brain processes resulting from cognitive impairments. In addition to the EEG power and coherence previously mentioned [32], metrics such as event related potentials (ERPs) could also be employed as reliable indices of mental workload [35].

Therefore, the combination of brain biomarkers from (e.g., EEG power, coherence, ERPs, and fNIRS) different brain imaging modalities could provide a higher degree of confidence in assessment tools that would enable efficient and robust brain monitoring systems to enhance cognitive-motor performance.

References

1. Parasuraman, R.: Neuroergonomics: research and practice. Theor. Issues Ergon. Sci. 4, 5–20 (2003)
2. Kruse, A.: Operational neuroscience: Neurophysiological measures in applied environments. Aviation, Space and Environmental Medicine 78(5), 191–194 (2007)
3. Friedl, K.E., Grate, S.J., Proctor, S.P., Ness, J.W., Lukey, B.J., Kane, R.L.: Army research needs for automated neuropsychological tests: monitoring soldier health and performance status. Arch. Clin. Neuropsychol. 22, 7–14 (2007)

4. Letz, R.: Continuing challenges for computer-based neuropsychological tests. Neurotoxicology 24, 479–489 (2003)
5. Coyle, S.M., Ward, T.E., Markham, C.M.: Brain-computer interface using a simplified functional near-infrared spectroscopy system. J. Neural Eng. 4(3), 219–226 (2007)
6. Deeny, S.P., Haufler, A.J., Saffer, M., Hatfield, B.D.: Electroencephalographic coherence during visuomotor performance:a comparison of cortico-cortical communication in experts and novices. J. Mot. Behav. 41, 106–116 (2009)
7. Deeny, S.P., Hillman, C.H., Janelle, C.M., Hatfield, B.D.: Cortico–cortical communication and superior performance in skilled marksmen: An EEG coherence analysis. J. Sport and Exercise Psychology 25, 188–204 (2003)
8. Del Percio, C., Rossini, P.M., Marzano, N., Iacoboni, M., Infarinato, F., et al.: Is there a ”neural efficiency” in athletes? A high-resolution EEG study. Neuroimage 42(4), 1544–1553 (2008)
9. Gentili, R.J., Bradberry, T.J., Oh, H., Hatfield, B.D., Contreras-Vidal, J.L.: Cerebral cortical dynamics during visuomotor transformation: Adaptation to a cognitive-motor executive challenge. Psychophysiology (in press)
10. Hatfield, B.D., Landers, D.M., Ray, W.J.: Cognitive processes during self-paced motor performance: an electroencephalographic profile of skilled marksmen. J. Sport Psychol. 6, 42–59 (1984)
11. Hatfield, B.D., Haufler, A.J., Hung, T.M., Spalding, T.W.: Electroencephalographic studies of skilled psychomotor performance. J. Clin. Neurophysiol. 21(3), 144–156 (2004)
12. Haufler, A.J., Spalding, T.W., Santa Maria, D.L., Hatfield, B.D.: Neurocognitive activity during a self-paced visuospatial task: comparative EEG profiles in marksmen and novice shooters. Biol. Psychol. 53(3), 131–160 (2000)
13. Kerick, S.E., Douglass, L.W., Hatfield, B.D.: Cerebral cortical adaptations associated with visuomotor practice. Med. Sci. Sports Exerc. 36(1), 118–129 (2004)
14. Landers, D.M., Han, M.W., Salazar, W., Petruzzello, S.J., Kubitz, K.A., et al.: Effects of learning on electroencephalographic and electrocardiographic patterns in novice archers. Int. J. Sport Psychol. 25, 313–330 (1994)
15. Slobounov, S., Ray, W., Cao, C., Chiang, H.: Modulation of cortical activity as a result of task-specific practice. Neurosci. Lett. 421(2), 126–131 (2007)
16. Kranczioch, C., Athanassiou, S., Shen, S., Gao, G., Sterr, A.: Short-term learning of a visually guided power-grip task is associated with dynamic changes in EEG oscillatory activity. Clin. Neurophysiol. 119(6), 1419–1430 (2008)
17. Caplan, J.B., Madsen, J.R., Schulze-Bonhage, A., Aschenbrenner-Scheibe, R., Newman, E.L., et al.: Human theta oscillations related to sensorimotor integration and spatial learning. J. Neurosci. 23(11), 4726–4736 (2003)
18. Yordanova, J., Falkenstein, M., Hohnsbein, J., Kolev, V.: Parallel systems of error processing in the brain. Neuroimage 22(2), 590–602 (2004)
19. Bell, M.A., Fox, N.A.: Crawling experience is related to changes in cortical organization during infancy: evidence from EEG coherence. Dev. Psychobiol. 29(7), 551–561 (1996)
20. Gentili, R.J., Bradberry, T.J., Hatfield, B.D., Contreras-Vidal, J.L.: Brain Biomarkers of Motor Adaptation Using Phase Synchronization. In: Proceedings of the IEEE International Conference of the Engineering in Medicine and Biology Society, Minneapolis, Minnesota, USA, September 2-6, vol. 1, pp. 5930–3 (2009)
21. Izzetoglu, M., Bunce, S.C., Izzetoglu, K., Onaral, B., et al.: Functional brain imaging using near-infrared technology. IEEE Eng. Med. Biol. Mag. 26(4), 8–46 (2007)
22. Leff, D.R., Orihuela-Espina, F., Atallah, L., Athanasiou, T., et al.: Functional prefrontal reorganization accompanies learning-associated refinements in surgery: a manifold embedding approach. Comput. Aided Surg. 13, 325–339 (2008)

23. Gentili, R.J., Hadavi, C., Ayaz, H., Shewokis, P.A., Contreras-Vidal, J.L.: Hemodynamic Correlates of Visuomotor Adaptation by Functional Near Infrared Spectroscopy. In: IEEE EMBS Proceedings, Buenos Aires, Argentina, pp. 2918–2921 (2010)
24. Defense Advanced Research Projects Agency - Defense Science Office, `http://www.darpa.mil/dso/thrusts/trainhu/accelerated/index.htm`
25. Thompson, T., Steffert, T., Ros, T., Leach, J., Gruzelier, J.: EEG applications for sport and performance. Methods 45, 279–288 (2008)
26. Landers, D.M., Petruzzello, S.J., Salazar, W., Crews, D.L., Kubitz, K.A., Grannon, T.L., Han, M.: The influence of electrocortical biofeedback on performance in pre-elite archers. Medicine and Science in Sports and Exercise 23, 123–129 (1991)
27. Arns, M., Kleinnijenhuis, M., Fallahpour, K., Breteler, R.: Golf Performance Enhancement and Real-Life Neurofeedback Training Using Personalized Event-Locked EEG Profiles. J. Neurother. 11(4), 11–18 (2009)
28. Vernon, D., Egner, T., Cooper, N., Compton, T., Neilands, C., Sheri, A., Gruzelier, J.: The effect of training distinct neurofeedback protocols on aspects of cognitive performance. Int. J. Psychophysiol. 47(1), 75–85 (2003)
29. Egner, T., Gruzelier, J.H.: Ecological validity of neurofeedback: Modulation of slow wave EEG enhances musical performance. NeuroReport 14, 1221–1224 (2003)
30. Hoyt, R.W., Reifman, J., Coster, T.S., Buller, M.J.: Combat medical infomatics: Present and future. In: Proceedings of AMIA Symposium, pp. 335–339 (2002)
31. Oken, B.S., Salinsky, M.C., Elsas, S.M.: Vigilance, alertness, or sustained attention: physiological basis and measurement. Clin. Neurophys. 117(9), 1885–1901 (2006)
32. Miller, M.W., Rietschel, J., McDonald, C.G., Pangelinan, M., Bush, L., Hatfield, B.D.: EEG assessment of incremental changes in cognitive workload during an ecologically valid visuo-motor task. In: 40th SFN Meeting, San Diego, CA, USA, November 13-17 (2009)
33. Costanzo, M.E., Oh, H., Bulkley, B., Contreras-Vidal, J.L., Goodman, R., Haufler, A., Lo, L.C., et al.: Independent component analysis of brain processes under psychological stress during motor performance. In: 39th SFN Meeting, Chicago, IL, USA, October 17-21 (2009)
34. Makeig, S., Jung, T.P.: Changes in alertness are a principal component of variance in theEEG spectrum. NeuroReport 7, 213–216 (1995)
35. Miller, M.W., Rietschel, J., McDonald, C.G., Hatfield, B.D.: A Novel Approach to the Physiological Measurement of Mental Workload. International Journal of Psychophysiology (in press)

Estimating the Level of Motion Sickness Based on EEG Spectra

Li-Wei Ko[1], Chun-Shu Wei[1], Tzyy-Ping Jung[1,2], and Chin-Teng Lin[1]

[1] Brain Research Center, Department of Electrical Engineering, National Chiao-Tung University, Hsinchu, 30010, Taiwan
[2] Swartz Center for Computational Neuroscience, Institute for Neural Computation, University of California San Diego, La Jolla, CA, 92093-0961 USA
{lwko,ctlin}@mail.nctu.edu.tw, treeseert@gmail.com, jung@sccn.ucsd.edu

Abstract. Motion sickness (MS) is a normal response to real, perceived, or even anticipated movement. People tend to get motion sickness on a moving boat, train, airplane, car, or amusement park rides. Although many motion sickness-related biomarkers have been identified, but how to estimate human's motion sickness level (MSL) is a big challenge in the operational environment. Traditionally, questionnaire and physical check are the common ways to passively evaluate subject's sickness level. Our past studies had investigated the EEG activities correlated with motion sickness in a virtual-reality based driving simulator. The driving simulator comprised an actual automobile mounted on a Stewart motion platform with six degrees of freedom, providing both visual and vestibular stimulations to induce motion-sickness in a manner that is close to that in daily life. EEG data were acquired at a sampling rate of 500 Hz using a 32-channel EEG system. The acquired EEG signals were analyzed using independent component analysis (ICA) and time-frequency analysis to assess EEG correlates of motion sickness. Subject's degree of motion-sickness was simultaneously and continuously reported using an onsite joystick, providing non-stop psychophysical references to the recorded EEG changes. We found that the parietal, motor, occipital brain regions exhibited significant EEG power changes in response to vestibular and visual stimuli. Based on these findings and experimental results, this study aims to develop an EEG-based system to estimate subject's motion sickness level upon the EEG power spectra from motion-sickness related brain areas. The MS evaluation system can be applied to early detection of the subject's motion sickness and prevent its uncomfortable syndromes in our daily life. Furthermore, the experiment results could also lead to a practical human-machine interface for noninvasive monitoring of motion sickness of drivers or passengers in real-world environments.

Keywords: EEG, ICA, motion-sickness, estimation, time-frequency, driving cognition.

1 Introduction

Motion-sickness is a common experience to everybody, and it has provoked a great deal of attentiveness in plenty of studies. The sensory conflict theory that came about

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 169–176, 2011.

in the 1970's has become the most widely accepted theorem of motion-sickness among scientists [1]. The theory proposed that the conflict between the incoming sensory inputs could induce motion-sickness. Accordingly, new research studies have appeared to tackle the issue of the vestibular function in central nervous system (CNS). In the previous human subject studies, researchers attempt to confirm the brain areas involved in the conflict in multi-modal sensory systems by means of clinical or anatomical methods. Brandt et al. demonstrated that the posterior insula in human brain was homologous to PIVC in the monkey by evaluating vestibular functions in patients with vestibular cortex lesions [2]. In agreement with previous clinical studies, the cortical activations during caloric [3] and galvanic vestibular stimulation [4] had been studied by functional imaging technologies such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI). To overcome the temporal limitation of the two imaging modalities, some studies have investigated the vestibular information transmission in time domain. Monitoring the brain dynamics induced by motion-sickness because of its high temporal resolution and portability De Waele et al., for example, applied current pulse stimulation to patients' vestibular nerve to generate vestibular evoked potentials [5].

The EEG studies related to motion-sickness can be divided into two groups according to the types of stimuli: vestibular and visual. Vestibular stimuli were normally provided to the subjects with rotating chair [6-7], parallel swing [8], and cross-coupled angular stimulation [9] to induce motion-sickness. Theta power increases in the frontal and central areas were reported to be associated with motion-sickness induced by parallel swing [8] and rotating drum [6-7]. Chelen et al. [9] employed cross-coupled angular stimulation to induce motion-sickness symptoms and found increased delta- and theta-band power during sickness but no significant change in alpha power. Visually induced motion-sickness is also commonly studied in previous studies. Visually induced sickness can be provoked with an optokinetic drum rotating around the yaw axis. This situation can cause a compelling sense of self-motion (called vection). Vestibular cues indicate that the body is stationary, whereas visual cues report the body is moving. Hu et al. investigated MS triggered by the viewing of an optokinetic rotating drum and found a higher net percentage increase in EEG power in the 0.5-4 Hz band at electrode sites C3 and C4 than in the baseline spectra. [10]. This study employees ahe driving simulator comprised an actual automobile mounted on a Stewart motion platform with six degrees of freedom, providing both visual and vestibular stimulations to induce motion-sickness and accompanied EEG dynamics.

Our past studies [11-13] had investigated the EEG activities correlated with motion sickness in a virtual-reality based driving simulator. We found that the parietal and motor brain regions exhibited significant alpha power suppression in response to vestibular stimuli, while the occipital area exhibited motion sickness related power augmentation in mainly theta and delta bands; the occipital midline region exhibited a broad band power increase. Based on these results, we think that both visual and vestibular stimulations should be used to induce motion sickness in brain dynamic research. Hence, we attempt to implement an EEG-based evaluation system to estimate subject's motion sickness level (MSL) upon the major EEG power spectra from these motion sickness related brain area in this study. The evaluation system can be applied to early detect the subject's MSL and prevent the uncomfortable syndromes occurred in advance in our daily life.

2 Materials and Methods

2.1 Experimental Paradigm

Unlike the previous studies, we provided both visual and vestibular stimuli to participant through a compelling VR environment consisting of 360^{o} projection of VR scene and a motion platform with six degree-of-freedom to induce motion-sickness. With such a setup, we expected to create motion-sickness in a manner that is close to that in daily life.. During the experiment, the subjects were asked to sit inside an actual vehicle mounted on a motion platform, with their hands holding a joystick to report their sickness level continuously. The VR scenes simulating driving in a tunnel were programmed to eliminate any possible visual distracter and shorten the depth of visual field such that motion-sickness could be easily induced. A three-section experimental protocol (shown in Fig. 1) was designed to induce motion-sickness.

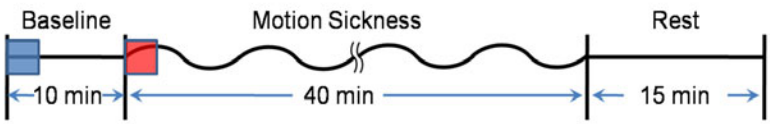


Fig. 1. Experimental design of motion-sickness experiments

First, the baseline section contained a 10-minute straight road to record the subjects' baseline state. Then, a 40-minute motion-sickness section, which consisted of a long winding road, was presented to the subjects to induce motion-sickness. Finally a 15-minute rest section with a straight-road condition was displayed for the subjects to recover from their sickness. The level of sickness was continuously reported by the subjects using a joystick with continuous scale on its side. The experimental setting successfully induced motion sickness to more than 80% of subjects in this study.

2.2 Subjects

Sixteen healthy, right-handed volunteers with no history of gastrointestinal, cardiovascular or vestibular disorders or of drug or alcohol abuse, taking no medication and with normal or corrected-to-normal vision participated in this experiment. EEG signals were recorded with 500 Hz sampling rate by 32-channel NuAmps (BioLink Ltd., Australia). Simultaneously, during EEG recording, the level of sickness was continuously reported by each subject using a joystick with a continuous scale ranging 0 – 5. The subjects were asked to raise/lower the scale to a higher/lower level if they felt more motion sick comparing to the last condition. This continuous sickness level was reported in real time without interrupting the experiment rather than the traditional motion-sickness questionnaire (MSQ).

2.3 EEG Data Acquisition and Signal Processing

EEG signals were recorded with 500 Hz sampling rate by 32-channel NuAmps (BioLink Ltd., Australia). Simultaneously during EEG recording, the level of sickness was continuously reported by each subject using a joystick with a continuous scale ranging 0 – 5. The subjects were asked to raise the scale to a higher level if they felt more sick comparing to the last condition. This continuous sickness level was reported in real time without interrupting the experiment rather than the traditional motion-sickness questionnaire (MSQ).

The acquired EEG signals were first inspected to remove bad EEG channels and then down-sampled to 250 Hz. A high-pass filter with cut-off frequency at 1 Hz and transition band width 0.2 Hz was used to remove baseline-drifting artifacts, and a low-pass filter with cut-off frequency at 60 Hz and transition band width 7 Hz was to remove muscular artifacts and line noise. After the preprocessing procedures, the clean EEG signals will feed into the proposed evaluation system for further analysis.

3 Proposed MS Level Estimation System

The proposed evaluation system to estimate subject's motion sickness level can be divided into five parts: independent component analysis (ICA), component clustering, time-frequency analysis, Feature Extraction by Principle Component Analysis (PCA), and Estimation part by applying linear regression, RBF Neural Network and Support Vector Regression with leave one out (LOO) cross validation. Figure 2 shows the system flowchart of the proposed motion sickness evaluation system.

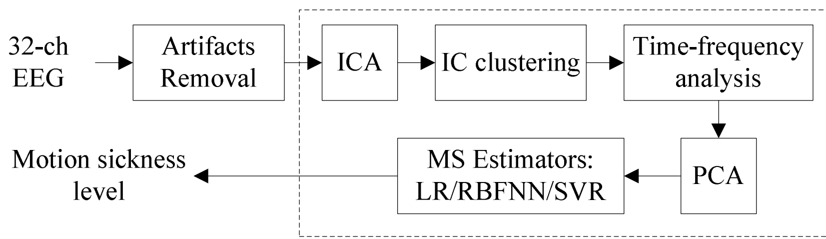


Fig. 2. System flowchart of the proposed motion sickness evaluation system

Independent Component Analysis (ICA) was applied to EEG recordings to remove various kinds of artifacts, including blink artifact and indoor power-line noise, and to extract features of human's cognition. Among components from all subjects, those with similar scalp topographies, dipole locations and power spectra were grouped using *k*-means clustering.

After doing ICA process, component clustering was analyzed using DIPFIT2 routines, a plug-in in EEGLAB, to find the 3D location of an equivalent dipole or dipoles based on a four-shell spherical head model. Among components from all subjects, those with similar scalp topographies, dipole locations and power spectra were clustered. Ten component clusters recruited more than 10 components from multiple subjects with similar topographic maps were regarded as robust component clusters. In component clustering results, we find that not all subjects have every motion sickness related components because the level of motion sickness induced by vestibular and visual stimuli to each subject had the significant individual difference. According to MSQ results and subject's self-response of motion sickness, we can confirm that each subject indeed felt sickness during the whole experiment session. Consequently, these extracted components are correlated with motion sickness. Then we can feed the ICA signals into the system and do time-frequency analysis.

Time-frequency analysis was used to investigate the dynamics of the ICA power spectra. In order to provide a temporal resolution of 30 seconds, the spectra of ICA activations were calculated using 7500-point non-overlapping window, and subdivided into several 125-point sub-windows with 25-point overlaps. Each 125-point sub-window was zero-padded to 512 points for using 512-point fast Fourier transform (FFT) with 1 Hz resolution in frequency. The linear power spectrum density (PSD) was then converted into a logarithmic scale (dB power).

The data set of each subject was combined with all the PSD of the subject's desired component. Since each subject provided 2 – 5 components for MS level estimation and each PSD was in 50 dimensions (from 1-50 Hz), the data set of each subject was between 100 – 250 dimensions.

PCA was then used to summarize the variances and extract first few principal components of the components' PSD's. In this study, the number of eigenvectors/components to retain was set to the number of first principal components that are necessary to explain 80% of the variances in the data. In addition, the EEG-based motion sickness evaluation system proposed in this study is including three different estimators: 1) Linear regression (LR), 2) Radial basis function neural network (RBFNN), and 3) Support vector regression (SVR). The MS level of the subject was estimated with leave-one-out cross-validation (LOOCV), where each observation was took as the validation data while the remaining as the training data.

4 Experimental Results and Discussion

In this study, we totally selected 16 subjects that were analyzed and applied to the modeling the estimation of our proposed MS level evaluation system. Figures 3 and 4 were shown the correlation coefficients (CC) results and root mean square errors (RMSE) in comparison with the actual MS level and the estimation performance of our proposed system. To summary the all estimation performance from 16 subjects, the average correlation coefficients were about 0.6467, 0.6738, and 0.7123 in corresponding to linear regression (LR), RBF neural networks (RBFNN), and support vector regression (SVR), respectively. As for the RMSE performance, the average estimation results were 0.2256, 0.2134, and 0.2199 in corresponding to LR, RBFNN, and SVR, respectively. According to the estimation results in Figure 3, we can see that the proposed MS level estimation system using SVR is better than using RBFNN or LR models. Subject 11's performance is an outlier subject because he/she reported not feel sickness during the experiment period according the self-reported MSQ.

We then select two subjects (S1 & S13) to show their comparison results between the actual MS level and estimation results from LR, RBFNN and SVR in Figure 5.

The black line shows the subject's actual (self-response) MS level and the black, blue and red dotted lines plot the estimation level from LR, RBFNN, and SVR evaluation systems, respectively. We can clearly see that the estimated performance followed the trend of the actual MS level, especially at the MS level increasing in the curving road.

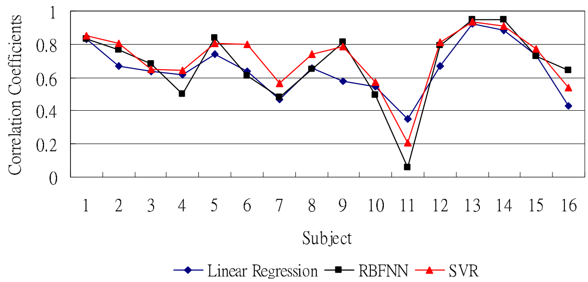


Fig. 3. Motion sickness estimation via using correlation coefficients results

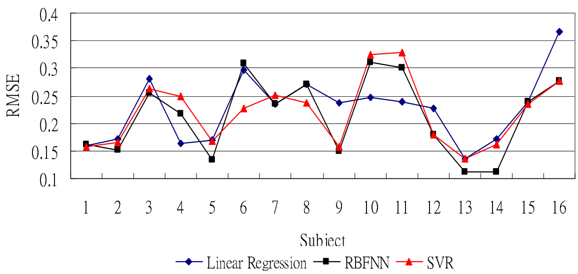


Fig. 4. Motion sickness estimation results via using correlation coefficients

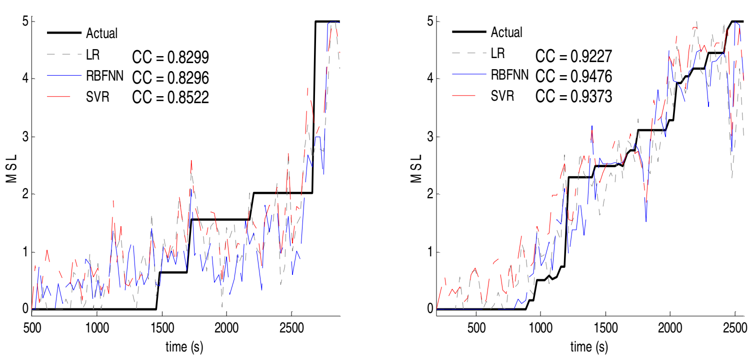


Fig. 5. Subjects 1's (left) and 13's (right) motion sickness estimation performance via using LR, RBFNN and SVR

Through the experimental results on the system performance under different conditions, we find that 9 subjects out of 15 subjects except subject 11 had the better CC estimation result via using SVR, and 6 subjects had better performance via using RBFNN. In conclusion, this study demonstrated that our proposed EEG-based evaluation system could successfully estimate the motion sickness level reported by individual subject, we suggest that SVR model can be utilized to estimate the motion sickness level in the operational environment. Since the potential of real-time application is emerging and desired, nevertheless, we need to consider more about the complexity, instantaneity, and robustness of the system. These results let us open an emerging sight on the potential of real-time application. Nevertheless, the complexity, instantaneity, and robustness of the system still have to be considered for the implementation.

Acknowledgment. This work was supported in part by the National Science Council, Taiwan, on Establishing "International Research-Intensive Centers of Excellence in Taiwan" (I-RiCE Project) under Contract NSC 99-2911-I-010-101, in part by the Aiming for the Top University Plan of National Chiao Tung University, the Ministry of Education, Taiwan, under Contract 99W906, and supported in part supported by the National Science Council, Taiwan, under Contracts NSC 99-3114-E-009-167 and NSC 99-2628-E-009-091. Doctors Li-Wei Ko and Chin-Teng Lin are sponsored in part by grants from the US Army Research Laboratory (W911NF-10-2-0022).

References

1. Reason, J.T., Brand, J.J.: Motion-sickness. Academic Press, London (1975)
2. Brandt, T., Dieterich, M., Danek, A.: Vestibular cortex lesions affect the perception of verticality. Annals of Neurology 35(4), 403–412 (1994)
3. Fasold, O., von Brevern, M., Kuhberg, M., Ploner, C.J., Villringer, A., Lempert, T., Wenzel, R.: Human vestibular cortex as identified with caloric stimulation in functional magnetic resonance imaging. NeuroImage 17(3), 1384–1393 (2002)
4. Lobel, E., Kleine, J.F., Le Bihan, D., Leroy-Willig, A., Berthoz, A.: Functional MRI of galvanic vestibular stimulation. The Journal of Neurophysiology 80(5), 2699–2709 (1998)
5. De Waele, C., Baudonniere, P.M., Lepecq, J.C., Tran Ba Huy, P., Vidal, P.P.: Vestibular projections in human cortex. Experimental Brain Research 141, 541–551 (2001)
6. Wood, C.D., Stewart, J.J., Wood, M.J., Struve, F.A., Straumanis, J.J., Mims, M.E., Patrick, G.Y.: Habituation and motion-sickness. Journal of Clinical Pharmacology 34, 628–634 (1994)
7. Wood, S.J.: Human otolith-ocular reflexes during off-vertical axis rotation: effect of frequency on tilt-translation ambiguity and motion-sickness. Neuroscience Letters 323(1), 41–44 (2002)
8. Wu, J.P.: EEG changes in man during motion-sickness induced by parallel swing. Space Medicine and Medical Engineering 5(3), 200–205 (1992)
9. Chelen, W.E., Kabrisky, M., Rogers, S.K.: Spectral analysis of the electroencephalographic response to motion-sickness. Aviation, Space, and Environmental Medicine 64(1), 24–29 (1993)

10. Hu, S., McChesney, K.A., Player, K.A., Bahl, A.M., Buchanan, J.B., Scozzafava, J.E.: Systematic investigation of physiological correlates of motion-sickness induced by viewing an optokinetic rotating drum. Aviation, Space, and Environmental Medicine 70(8), 759–765 (1999)
11. Chen, Y.C., Duann, J.R., Chuang, S.W., Lin, C.L., Ko, L.W., Jung, T.P., Lin, C.T.: Spatial and Temporal EEG Dynamics of Motion Sickness. NeuroImage 49(3), 2862–2870 (2010)
12. Lin, C.T., Chuang, S.W., Chen, Y.C., Ko, L.W., Liang, S.F., Jung, T.P.: EEG Effects of Motion Sickness Induced in a Dynamic Virtual Reality Environment. In: Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS 2007), Cité Internationale, Lyon, France, August 23-26 (2007)
13. Yu, Y.H., Lai, P.C., Ko, L.W., Chuang, C.H., Kuo, B.C., Lin, C.T.: An EEG-based Classification System of Passenger's Motion Sickness Level by using Feature Extraction/Selection Technologies. In: Proceedings of the 2010 IEEE World Congress on Computational Intelligence (WCCI 2010), Barcelona, Spain, July 18-July 23 (2010)

Combining fNIRS and EEG to Improve Motor Cortex Activity Classification during an Imagined Movement-Based Task

Darren J. Leamy, Rónán Collins, and Tomas E. Ward

Biomedical Engineering Research Group,
Department of Electronic Engineering,
NUI Maynooth,
Ireland. Adelaide & Meath Hospital, Tallaght,
Dublin 24, Ireland
dleamy@eeng.nuim.ie,
ronan.collins2@amnch.ie,
tomas.ward@eeng.nuim.ie

Abstract. This work serves as an initial investigation into improvements to classification accuracy of an imagined movement-based Brain Computer Interface (BCI) by combining the feature spaces of two unique measurement modalities: functional near infrared spectroscopy (fNIRS) and electroencephalography (EEG). Our dual-modality system recorded concurrent and co-locational hemodynamic and electrical responses in the motor cortex during an imagined movement task, participated in by two subjects. Offline analysis and classification of fNIRS and EEG data was performed using leave-one-out cross-validation (LOOCV) and linear discriminant analysis (LDA). Classification of 2-dimensional fNIRS and EEG feature spaces was performed separately and then their feature spaces were combined for further classification. Results of our investigation indicate that by combining feature spaces, modest gains in classification accuracy of an imagined movement-based BCI can be achieved by employing a supplemental measurement modality. It is felt that this technique may be particularly useful in the design of BCI devices for the augmentation of rehabilitation therapy.

1 Introduction

A brain-computer interface (BCI) is a system for generating computer control signals based on changes in monitored brain activity [1], [2]. BCIs have been used for many diverse reasons, such as for allowing tetraplegics to interact with computers [3], amputees to control prosthetic robotic limbs [4] and healthy subjects to control computer interfaces through thought alone [5]. Our research interest however is in the use of a BCI for stroke rehabilitation. We aim to use this system to circumvent the stroke affected area of a patient's brain by encouraging the neuroplastic process.

Neuroplasticity is the process by which the human brain physically alters neuronal connections within itself in order to adapt to sensory input. During the 20th century it was widely believed that physical changes in the adult brain as a response to sensory

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 177–185, 2011.

input were impossible. Research articles challenging this consensus appeared during the past decade and began changing this belief about the brain. Researchers thus began exploring the capabilities and possibilities of a brain that can physically adapt to changing sensory inputs. Research into neuroplasticity has found that dyslexia in children can be treated [6], discovered that blind people use the visual cortex to help process other information [7] and that a musician can improve their musical abilities through mental rehearsal [8] - each of these as a result of a physically changing brain.

In cases where a stroke sufferer has lost the use of a limb, the neuroplastic process is capable of reassigning a different area of the brain to take over from the stroke-damaged area [9]. In certain stroke cases, it may still be possible to record a patient's attempt to move a stroke-affected limb in the motor cortex. A stroke patient's attempts to move a stroke-affected limb may be similar to a healthy subject imagining limb movement when the the motor cortex is still intact (as can be the case for lacunar strokes). For this reason, this paper investigates imagined movement-related activity in the motor cortex of healthy subjects.

BCIs require a method for monitoring brain activity, from the analysis of which external control signals are generated. Our research is in stroke rehabilitation so our subjects may be weak, have low mobility and may move during measurement. We therefore use two measurement modalities that are portable and relatively inexpensive: functional near infrared spectroscopy (fNIRS) and electroencephalography (EEG). Both of these modalities have unique advantages and they do not interfere with each other. EEG has very high temporal resolution whereas fNIRS is not affected by electromagnetic interference and is not as susceptible to movement artefact as EEG. By using both modalities on the same area of cortex, extra information about the cortical activity can be recorded. As this implements a combined electrical and hemodynamic recording of cortical activity, we are making direct observations of neurovascular coupling. Such information may prove to be vital for our research into stroke rehabilitation.

1.1 fNIRS

fNIRS is a measurement modality based on changing concentrations of oxy-haemoglobin (HbO) and deoxy-haemoglobin (HbR) in cortical areas of the brain. Multiple wavelengths of light in the red to near-infrared range of the electromagnetic spectrum (620 nm - 1200 nm) are emitted into the scalp of a subject from the surface of the head from an fNIRS "source". Light incident on the head disperses through the biological tissue, a portion of which exits the head again after passing through cortical areas of the brain, where the chromophores HbO and HbR are present. For a given entry and exit point on the scalp, the photons are known to have followed a roughly banana-shaped path through the head, known as the "photon path" [10]. The mean depth of the photon path is related to the physical distance between the points of entry and exit on the scalp. The intensity of the wavelengths of light transmitted through the head is measured with a fNIRS "detector", which is then used to infer time-changing concentrations of HbO and HbR along the photon path. This is done using the modified Beer-Lambert law, which describes optical attenuation in a highly-scattering medium [11]:

$$Attenuation\ (OD) = log_{10}\frac{I_0}{I} = \alpha cBL + G \tag{1}$$

where OD is the optical densities, I_0 is the incident light intensity, I is the transmitted intensity, α is the absorption coefficient of the chromophore, c is the concentration of the chromophore, B is the differential pathlength factor, L is the distance between the source and detector and G is a term to account for scattering loss. If measurements are only made of the changes in light attenuation then B, L and G remain constant and changes in HbO and HbR concentration can be derived from the expression:

$$\delta c = \frac{\delta OD}{\alpha BL} \tag{2}$$

A typical hemodynamic activation response is an initial decrease in HbO and increase in HbR followed by a large increase in HbO and a decrease in HbR while the cortex is active. When the cortex is at rest, HbO and HbR concentrations return to baseline levels. These changes in HbO and HbR concentration are used to determine hemodynamically whether an area of cortex is active or not.

1.2 EEG

Non-invasive EEG is the measurement of the spatially integrated dendritic activity of similarly oriented neurons near the surface of the brain. EEG features a spectral structure which changes locally in response to neuronal activity. Spectral power changes in the EEG which occur in temporal relation to subject engagement with a task are known as Event Related Synchronisation (ERS) and Event Related Desynchronisation (ERD). The particular ERS/ERD responses in the motor cortex to motor tasks have been detailed elsewhere [12]. Immediately before a subject engages with a motor task, the motor cortex EEG exhibits a suppression of power in the μ frequency range (8-12 Hz), known as Pre Movement Mu Desynchronisation (PMMD) [13]. Similarly, when a subject rests from motor activity, an increase of power in the β frequency range (12-30 Hz) is observed shortly after, known as Post Movement Beta Synchronisation (PMBS) [14]. These changes in spectral power are used to determine electrical changes in motor cortex activity associated with movement and imagined movement.

2 Methodology

2.1 Subjects

Two healthy individuals participated in the study. Subject A was 38 years old and left-handed. Subject B was 26 years old and right-handed. Both participants had normal or corrected vision. Neither participant had consumed any stimulant prior to the experiment. Each participant gave informed oral consent. The experiment was approved by the ethics board of the National University of Ireland Maynooth.

2.2 Experimental Procedure

The subjects were seated in a comfortable chair facing a computer monitor with their feet flat on the floor and their arms resting on armrests. Instructions were delivered visually via a computer monitor diagonally measuring 43 cm and positioned 80 cm from the subject's eyes, centred at eye-level. Instruction presentation and trigger signal generation were carried out with custom software written using C# and the .NET framework. Trigger signals were recorded simultaneously by both the fNIRS and EEG systems.

Each subject completed 40 experimental trials, during which instructions were presented to the subject. Before instruction periods began there was a 10 second wait during which the computer screen was blank. Two types of instruction period were used - an activity period during which the screen displayed the message "Imagine Movement" and a rest period during which the screen displayed the message "Relax". Instruction periods lasted 10 seconds, facilitating a total experimental time of 410 seconds.

Prior to the commencement of the experiment, the subject was handed a fist-sized soft ball and asked to practice squeezing the ball with their dominant hand for one minute. The subject was instructed to imagine squeezing the ball during an activity instruction period. Subjects were told to not make any actual movements. Subjects were instructed to stop imagining the movement during a rest instruction period.

2.3 Signal Acquisition

Multichannel fNIRS and EEG systems were implemented concurrently to record both hemodynamic and electrical responses in the motor cortex of the subjects during experiments. Three fNIRS sources, three fNIRS detectors and seven EEG electrodes were arranged in a unique montage on a custom-made head mount. The head mount was made of low-density polythene backed by polyurethane foam. fNIRS sources and detectors were positioned with 3 cm spacing, with the EEG electrodes positioned at the mid-way point of each pair of adjacent fNIRS sources and detectors. fNIRS acquisition was performed with a TechEn CW6 system (TechEn Inc., MA, USA) using wavelengths of 690 nm and 830 nm. fNIRS data was digitally sampled at 25 Hz. EEG acquisition was performed with a BioSemi Active Two system (BioSemi Inc., The Netherlands). DC-coupled EEG data was digitally sampled at 2048 Hz.

Figure 1 shows the layout of the fNIRS-EEG montage. fNIRS source positions are labelled S1-S3, fNIRS detector positions are labelled D1-D3 and EEG electrode positions are labelled E1-E7. Seven channels of fNIRS-EEG were recorded. The fNIRS sources and detectors and EEG electrodes used for each channel are displayed in Table 1. During experiments, the fNIRS-EEG montage was centred over the subject's dominant-side motor cortex (C4 of the 20/20 system for left-handed Subject A and C3 for right-handed Subject B). Figure 2 shows the orientation of the montage in place on Subject A's head.

2.4 Signal Processing and Feature Extraction

fNIRS. The fNIRS system records raw light intensity signals. These are first converted to changes in optical density (ΔOD) and then, using the modified Beer-Lambert law, to changes in concentration (ΔHbO and ΔHbR).

Table 1. Channel fNIRS and EEG designations

Channel Num	fNIRS		EEG
	Src	Det	
1	S1	D1	E1
2	S1	D2	E2
3	S2	D1	E3
4	S2	D2	E4
5	S3	D2	E5
6	S2	D3	E6
7	S3	D3	E7

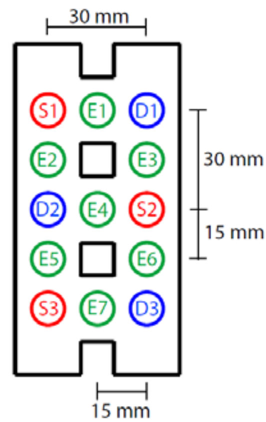


Fig. 1. fNIRS-EEG montage

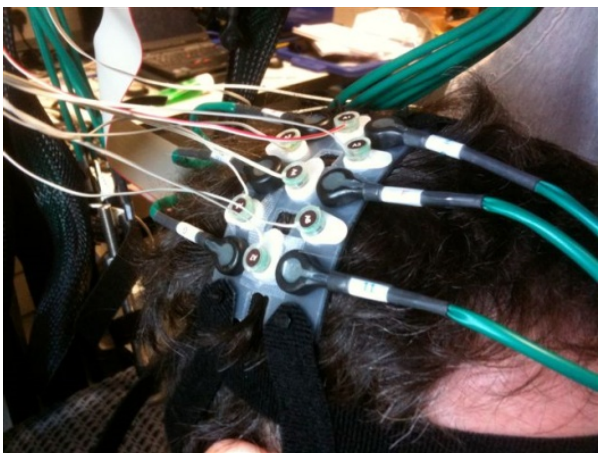


Fig. 2. Dual fNIRS-EEG module over subject's motor cortex

The raw intensity signals were initially low-pass filtered with a cut-off frequency of 0.5 Hz to remove high-frequency components, such as those associated with the cardiac cycle. Next, the intensity signals are normalised using their mean amplitude over the entire recording. The normalised intensity signals are then high-pass filtered with a cut-off frequency of 0.01 Hz to remove baseline drift and Mayer wave interference. ΔOD signals are obtained by taking the negative logarithm of this filtered signal. 4th order Butterworth filters were used for all filtering steps.

The modified Beer-Lambert law is then used to find the ΔHbO and ΔHbR signals from the ΔOD signals for each channel. Once the source-detector separation for the channel, the extinction coefficients for both the 690 nm and 830 nm light in HbO and HbR and the differential pathlength factor are known, the operation involves a simple matrix inversion and multiplication to obtain the ΔHbO and ΔHbR signals. A differential pathlength factor of 5.93 was used (in accordance with the literature [15]).

The ΔHbO and ΔHbR data was then separated into individual 10 second trials. During each trial, the change in average amplitude of the ΔHbO and ΔHbR signals from the initial 5 seconds to the subsequent 5 seconds was used as the feature to train and test the classifier.

EEG. The EEG data was first high-pass filtered with a cut-off frequency of 1 Hz to remove DC and near-DC components. The EEG data was then analysed to identify the frequencies at which PMMD and PMBS occurred for the different events. The events were "imagined movement onset" and "imagined movement offset", which coincide with a change in on-screen instruction to the subject. The frequencies at which ERD and ERS occurred were identified by visual inspection of the average FFT plots for the reference and activity periods for both events.

The reference period was chosen to be between 4.5 and 3.5 seconds before both types of event. For an imagined movement onset event, the activity window was chosen to be from between 0 and 1 second after the event. For an imagined movement offset event, the activity window was chosen to be from between 0.5 and 1.5 seconds after the event. The activity windows were chosen because of the expected timings of PMMD and PMBS. The raw EEG is filtered for the identified ranges of PMMD and PMBS. These μ-range and β-range signals are then squared to get power signals.

The change in μ-range power and β-range power from reference window to activity window were used as features of EEG activity for classification.

2.5 Classification

The goal of classification is to decode the subject's current state based on the features extracted from the fNIRS and EEG responses. The classifier attempted to classify features into one of two classes: 'active' and 'rest'. We employed the Linear Discriminant Analysis (LDA) classifier and calculated classification accuracy via Leave-One-Out Cross Validation (LOOCV), as we had employed previously [16]. In particular, for a total of N trials of data, N-1 trials were used for training the classifier and the remaining trial was used for testing. This was repeated N times with each trial used for testing once. Accuracy was measured as the amount of correct classifications over N trials. Following classification of the fNIRS and EEG 2-dimensional feature spaces, the individual feature spaces were combined into an all-encompassing 4-dimensional feature space. Every trial of data thus had 4 features available for classification - change in HbO over the trial, change in HbR over the trial, change in μ-range power at the start of the trial and change in β-range power at the start of the trial. This combined 4-dimensional feature space was also classified using LDA and LOOCV.

2.6 Results

The classification results are presented in Table 2. Shown are the classification accuracies of the classifier when operating on fNIRS features alone, EEG features alone and combined fNIRS/EEG features.

Table 2. LDA classification results for fNIRS, EEG and combined features

	Subject A			Subject B		
Channel	fNIRS	EEG	Dual	fNIRS	EEG	Dual
1	59%	51%	64%	64%	46%	62%
2	56%	59%	67%	51%	54%	59%
3	56%	54%	64%	61%	41%	56%
4	69%	67%	72%	64%	59%	67%
5	61%	51%	72%	41%	36%	46%
6	56%	77%	64%	74%	59%	69%
7	56%	59%	62%	15%	43%	49%
Average	59%	60%	66%	53%	48%	58%

3 Discussion

Our results demonstrate that through combining fNIRS and EEG features into a single fNIRS-EEG feature space, an increase in classification accuracy of imagined movement can be obtained. The two experimental subjects had very different classification accuracies of the cortical activity. For Subject A, all but one channel made gains in classification accuracy by combining feature spaces. For Subject B, however, combining feature spaces often results in a classification accuracy result intermediate to those of fNIRS and EEG alone.

From these results, it appears that when fNIRS and EEG classification is reasonably accurate, the combined classification result tends to be higher than both. When either fNIRS or EEG classification accuracy is good and the other is not, combining feature spaces seems to result in intermediate classification accuracy. It is also worth pointing out that when both fNIRS and EEG classification was poor, combining feature spaces resulted in better performance.

These results show that combining fNIRS and EEG feature spaces can results in higher classification accuracy. This is of more importance to improving an fNIRS system than an EEG system. Improving EEG classification accuracy can be accomplished by increasing the density of electrodes over a cortical area. However, fNIRS has a limit to the proximity of source-detector pairs. An fNIRS detector can suffer from "source-blinding" if sources are placed too close, even if that detector is not intentioned to record light from that source. Therefore, for the sole purposes of improving classification accuracy, supplementing an fNIRS system with EEG data is more useful than adding fNIRS to an EEG system.

An advantage of a dual-modality system such as this one is that for the same measurement space on the head, more information about the underlying neurovascular relationship is being recorded. An EEG of fNIRS alone system can only record the electrical or hemodynamic response in an area of cortex. Our system records fNIRS and EEG but also records information about the relationship between them, even if we do not yet fully understand that relationship.

We expect a similar improvement in classification accuracy when using dry EEG electrodes instead of the standard wet electrodes used here. Dry electrodes have much lower signal-to-noise ratio (SNR), so combining a dry EEG set-up with fNIRS could help offset the poor SNR. We expect a completely dry fNIRS-EEG system would significantly reduce set-up time, reduce subject discomfort and have similar classification performance to a wet electrode EEG system.

4 Conclusion

Investigation into dual-modality measurement is of importance to BCI researchers due to the potential gains in classification accuracy while utilising the same area of cortex. This work has demonstrated that improvements to imagined-movement based BCIs are possible by implementing multi-modal measurements. We believe this research will lead to more accurate BCIs and smaller measurement devices.

Acknowledgements. The authors gratefully acknowledge the contribution of Science Foundation Ireland: Research Frontiers Program 2009, Grant No. 09/RFP/ECE2376.

References

1. Angelakis, E., et al.: Brain-computer interface: a reciprocal self-regulated neuromodulation. Acta Neurochir. Suppl. 97, 555–559 (2007)
2. Wolpaw, J.R., et al.: Brain-computer interfaces for communication and control. Clinical Neurophysiology 113(6), 767–791 (2002)
3. Conradi, J., et al.: Brain-Computer Interfacing in Tetraplegic Patients with High Spinal Cord Injury. International Journal of Bioelectromagnetism 11(2), 65–68 (2009)
4. Hochberg, L.R., et al.: Neuronal ensemble control of prosthetic devices by a human with tetraplegia. Nature 442, 164–171 (2006)
5. Scherer, R., et al.: The Self-Paced Graz Brain-Computer Interface: Methods and Applications. Comput. Intell. Neurosci., 79826 (2007)
6. Meyler, A., et al.: Modifying the brain activation of poor readers during sentence comprehension with extended remedial instruction: A longitudinal study of neuroplasticity. Neuropsychologia 46(10), 2580–2592 (2008)
7. Burton, H.: Visual Cortex Activity in Early and Late Blind People. The Journal of Neuroscience 23(10), 4005–4011 (2003)
8. Pascual-Leone, A., et al.: Modulation of muscle responses evoked by transcranial magnetic stimulation during the acquisition of new fine motor skills. J. Neurophysiol. 74(3), 1037–1045 (2005)
9. Johansson, B.B.: Brain Plasticity and Stroke Rehabilitation: The Willis Lecture. Stroke 31, 223–230 (2000)
10. Okada, E., et al.: Theoretical and experimental investigation of near-infrared light propagation in a model of the adult head. Appl. Opt. 36(1), 21–31 (1997)
11. Sassaroli, A., Fantini, S.: Comment on the modified Beer–Lambert law for scattering media. Phys. Med. Biol. 49(14), 255 (2004)
12. Pfurtscheller, G., Lopes da Silva, F.H.: Event-related EEG/MEG synchronization and desynchronization: basic principles. Clinical Neurophysiology 110, 1842–1857 (1999)

13. Pfurtscheller, G., Aranibar, A.: Evaluation of event-related desynchronization (ERD) preceding and following voluntary self-paced movements. Electroenceph. Clin. Neurophysiol. 46, 138–146 (1979)
14. Pfurtscheller, G., Stancák Jr., A., Neuper, C.: Post-movement beta synchronization. A correlate of an idling motor area? Electroenceph. Clin. Neurophysiol. 98, 281–293 (1996)
15. van der Zee, P., et al.: Experimentally measured optical pathlengths for the adult head, calf and forearm and the head of the newborn infant as a function of inter optode spacing. Adv. Exp. Med. Biol. 316, 143–153 (1992)
16. Leamy, D.J., Ward, T.E.: A novel and concurrent fNIRS/EEG measurement system: design and initial results. In: 2010 Annual International Conference of the IEEE on Engineering in Medicine and Biology Society (EMBC), pp. 4230–4233 (2010)

The Frustration Status and Noise Proof Feature During Perception of the Auditory Images

Sergey Lytaev and Yuliaj Surovitskaj

Saint Petersburg State Pediatric Medical Academy, Litovskaya, 2,
194100, Saint Petersburg, Russia
{Sergey.Lytaev,Yuliaj.Surovitskaj,slytaev}@spiiras.nw.ru

Abstract. Tests for modeling of the human status at recognition of target and non target stimulus with auditory evoked potentials (AEPs) registration; emotional neutral and significant information-psychological influences with EEG registration and analysis of fractal dynamics (AFD) were applied. From the moment of signal presentation the greatest difference of AEPs at target stimulation is marked in frontal areas of the left hemisphere through 16-18 ms. Emotionally-neutral and emotionally-significant psycho-informational influences provided the most conclusive AFD EEG data. Essentially, personal frustration is activated when the subject perceives situations to be threatening to his or her self-esteem and self-evaluation. Individuals with high levels of frustration are inclined to perceive a wide range of situations as threatening and therefore, will respond according to what they think the situation dictates.

Keyword: Auditory Evoked Potentials (AEPs), Brain Mapping, EEG, Information-Psychological Influences.

1 Introduction

During last more than 100 years human brain is classical model of “the black box”. We can estimate parameters of acoustic signals acting to the brain. From other side it is possible to register some target data of the acoustic information analysis. Development of means and algorithms of multivariate dynamic biological measurements creates predictors for expansion of a circle of problems for biometric systems [4], [5]. One of perspective directions of development biometrics is the analysis of a human functional status in remote research. For example, in such context the problem of remote recognition of a human mental condition and likelihood forecasting is solved at processing the acoustical information [5], [8], [9].

The answer reactions of the brain, namely, auditory evoked potentials (AEPs) (event-related potentials) are the target data. The modern methods of brain mapping allow to investigate the spatial-temporary characteristics of reactions of the brain in reply to acoustic signals and also to estimate mechanisms of processing of the auditory information [2], [3], [9], [10].

The present research was aimed for modeling of human frustration status and noise proof feature at perception of the auditory images.

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 186–193, 2011.

2 Methods

86 healthy examinees (man, age 20-27 years) were surveyed. Audio tests were performed with using an original hardware-software complex with neurosystems for auditory evoked potentials (AEPs) and EEG registration. Tests for modeling of the human status at recognition of target and non target stimulus with AEPs registration; emotional neutral and significant information-psychological influences (IPI) with EEG registration and analysis of fractal dynamics (AFD); recognition of

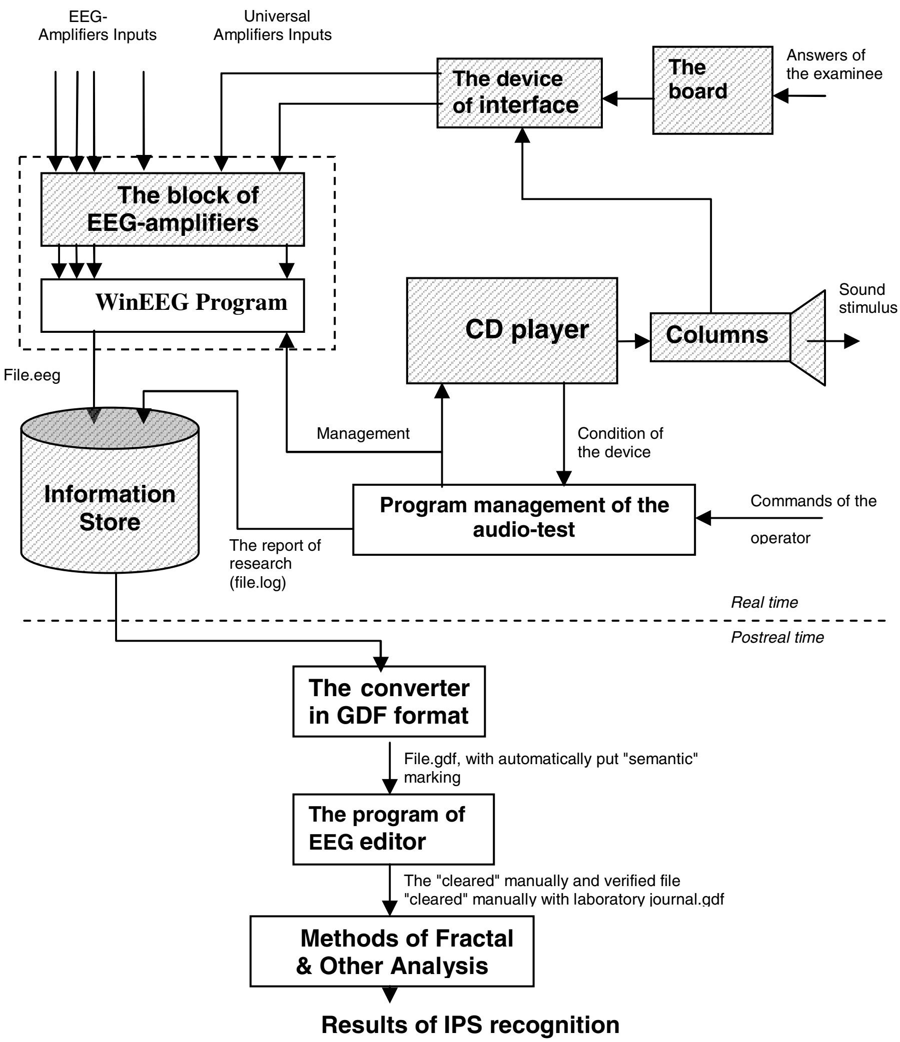


Fig. 1. Software and hardware complex for support of audio test

homogeneous verbal significant and insignificant stimulus (with EEG); with the compelled reaction to verbal stimulus, overcoming steady stereotypes (with EEG) and psychological testing were applied.

For realization of audio of testing the hardware-software complex has been created. Hardware devices are allocated by shading (fig. 1). The complex contains in the structure standard electroencephalograph with regular software (in fig. are led round by a dotted rectangular) and also originally programs and the devices developed by us especially for realization of the test. These are programs of management of the audio test, devices of interface and for processing data of research in a post real time. Fragments of presentation of information-psychological influences at simultaneous EEG registration are presented in table 1.

Table 1. Text codes of EEG fragments

Code	Fragment
1	Background prior to the beginning of presentation of stimulus
20	The incoherent text
21	The neutral intelligent text
22	The incoherent text after the neutral text
23	The prevention of painful stimulus
24	The incoherent text up to stimulus
25	The incoherent text after stimulus
26	The disturbing text-attitude
27	The incoherent text after the disturbing text

3 Results

The data of the brainstem (short-latency) auditory evoked potentials registration testify that the acoustic signal is involved in processing by brain structures in 3 ms from moment of its presentation. Usually on this analyzed epoch (3-10 ms) different authors allocate 6-8 components [6], [7], in our case – 6 waves (Fig. 2).

Brain mapping was performed on the analyzed epoch from 10 to 400 msec. From the moment of signal presentation the greatest difference of potentials at target stimulation is marked in frontal areas of the left hemisphere through 16-18 msec, while in occipital and parietal sites the insignificant symmetric activation is marked.

On the diagram the statistics of peak-time characteristics of potentials are presented at target and non-target stimulation (fig. 3).

Among middle-latency auditory evoked potentials the most stable in both tests are 4 waves – P_{16}, N_{30}, P_{40} and N_{60}. The amplitude-temporary characteristics of the Pa/Nb complex don't differ significantly in variants of testing. Some features are observed in formation of waves Po and Na at relevant stimulation. The amplitude of Po in the parietal (P_3 and P_4), and also in the left hemispheric C_3 and F_3 sites ($F > 4.0$) grows significantly. The similar tendency is kept and on the time interval of formation of a wave Na in left (P_3 and F_3, $F > 4.0$) points of registration.

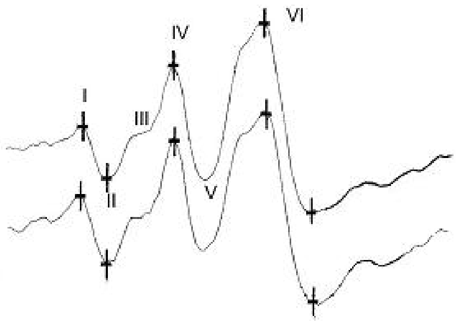


Fig. 2. The components of the BSAEPs, I-VI. Epoch of the analysis – 10 ms; the ordinate's line – in 1 cm 1 uV.

The temporary parameters of late waves (200-400 ms) are leveled in both tests. Only peak latency of the wave P_{250} is significantly lower at action of the relevant signals (F > 4.0). Simultaneously the amplitude of this component in conditions of a target task is reduced. Significant differences are observed in left P_3 and C_3 sites and in vertex (fig. 3).

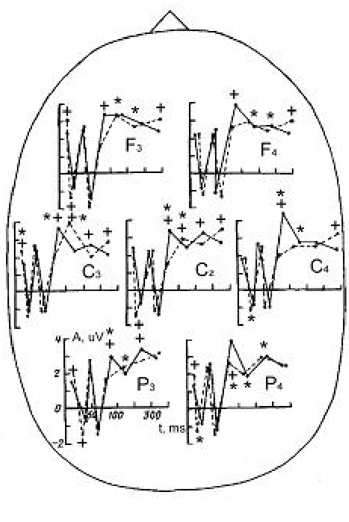


Fig. 3. The average meanings of amplitudes and of PL AEPs waves at non-target (continuous) and target (dotted line) stimulation. Badges (+) and (*) – significant (F > 4.0) accordingly for amplitudes and PL. P_4, P_3, …, F_3 - sites on system 10/20.

The marked facts are reflected in the brain mapping images (a, 220 ms; b, 226 ms). At performance of the target task the hemispheric symmetry of processes of excitation pays attention appreciable.

In this research healthy subjects and also persons with boundary frustration were investigated. For subjects with a high level of personal and situational frustration low percent of identification of acoustical images was characteristic. 8.3 % did not have associations. The identification had guessing character. The long identification of an auditory image was marked in 15.3 % of cases. If the average duration of the latent period of an identification of healthy subjects was 3-5 sec, persons with high levels of frustration took 10 to 30 sec. to recognize the acoustical images (table 2).

The splitting of perception, i.e. loss of the ability to form a complete image of object was registered in 12.5 % of cases. Some examinees correctly perceived separate details of an acoustical image, but could not connect them into a complete structure. Late recognition was marked in 13.8 % of cases, false recognition an acoustical image – in 22 % and unreasonably certain character of identification – in 8.3 % (table 2).

Affective illusions are established in 8.3 % of cases. Such subjects instead of traditional sounds heard a shutter, a gun, shots, steps and the breath of persecutors, agonizing groans and shouts of people familiar to them. These features of perception correlate with a high level situational and an average level of personal frustration. The long identification of an acoustical image (15.3 %), propensity to jam the same images (8.3 %) met at high situational and low personal frustration (table 2).

Table 2. Recognition of Auditory Images Depending on a Level of Situational and Personal Frustration

Features of Auditory Images Perception	N, (% from total)	Level of situational frustration (Points)	Level of personal frustration (Points)
False recognition	16 (22,2%)	49	33
Illegibility of the recognition	8 (11,1%)	30,5	28,75
Unreasonably certain identification and/or affective illusions	6 (8,3%)	50	37,5
Late recognition	10 (13,8%)	48,6	31
The prolonged recognition	11 (15,3%)	45, 6	28,4

In conditions of strong white noise healthy subjects correctly identified 82 % of acoustical images and people with a high level of frustration – 52 % of all masked acoustical images (fig. 4). Overall, 30 acoustical images were shown in random order over a period of 20 sec. Simultaneously EEG was registered.

With a decrease in the level of white noise the distinction between examinees was reduced. However parameters of an identification of acoustical images at healthy were approximately 10 % higher in comparison with to subjects’ high level of frustration (fig. 4).

We have applied 3 variants of subtests at EEG registration.

S u b t e s t 1 simulates steady human status during the presentation of emotionally-neutral and emotionally-significant information-psychological influences. To emotionally-neutral influence carried the incoherent text (a casual set of neutral words) and the neutral intelligent text on the abstract topic. For emotionally-significant influences we have used texts that force the subject to listen and induce feelings of alarm.

S u b t e s t 2 simulates the human status at recognition of a line of homogeneous verbal audio stimulus. One of the stimuli was more significant for the subject.

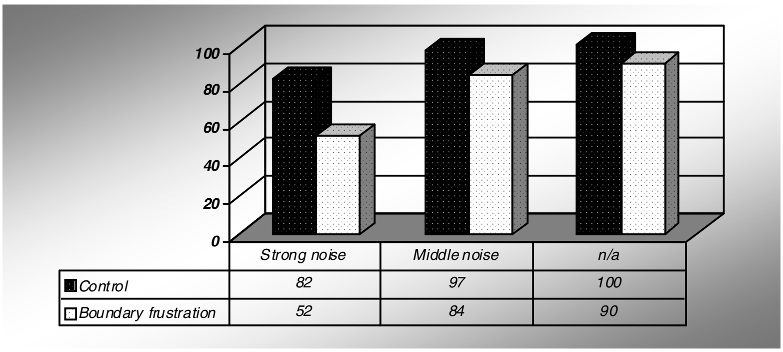


Fig. 4. Recognition of Images on a Background of Acoustic Noise

S u b t e s t 3 simulates the status of human reaction on a verbal stimulus in an unusual image, overcoming stereotypes. When the subject during conversation intentionally deforms true, which to it for a long time and well-known, similar status take place.

The registered EEG has divided into patterns corresponding with fragments from the audio test. To each fragment a code is assigned. EEG fragment codes and variants of their sequences are presented in the table 1.

Further EEG processing was carried out in post real time after the removal of artifacts from each file. This is a manual procedure.

In the EEG files, software has allocated analyzed epoch according to semantics of the subtest. For each of epoch are settled an invoice a vector from 79 informative attributes. It is automatic procedure.

For the analysis and recognition of the EEG patterns were automatically allocated of values of informative attributes. Thus values of attributes of fragments of different EEG type minimally depend on specific EEG features and mainly reflect the EEG change at transition from a fragment to a fragment. That is they reflect transition from one test to another.

The subsequent automatic processing of EEG patterns was carried out from complex to simple – for allocation of the most significant parameters with a reduction of less significant results.

4 Conclusion

I. Brain mapping images of middle latency AEPs (26 ms) during target stimulation show distinct symmetry with maximum activation of the occipital-parietal areas, to a lesser degree – the central departments and, practically no electrical dipoles in the frontal areas of the cortex.

A higher level of symmetric excitation with a latent period ranging from 300 to 350 ms was observed when the components of the experiment were mapped. The analysis of the spatial–temporary characteristics in this task marks significant simplification of amplitude of this wave in all sites, except for the parietal.

In the first case, it is a question of activating unrealized attention processes, in the second – a question of indicating the decision-making process.

II. Subjects with high levels of personal and situational frustration respond differently to the acoustic images when compared to healthy individuals.

It was characteristic for subjects with high levels of personal and situational frustration, to achieve low percentage scores when attempting to identify acoustic images. Their responses to the images differed as well, when compared to healthy individuals. These responses include: making no association; guessing; taking an extended period of time to respond; the splitting of perception; incorrect identification; and affect illusions.

High emotional intensity, anxiety, concern and propensity to test alarm, fears, fear even in habitual for them to conditions were observed in individuals with higher levels of personal frustration.

III. Emotionally-neutral and emotionally-significant psycho-informational influences provided the most conclusive ADF EEG data (subtest 1). Decreased, or a complete absence of motivation and emotionally neutrality account for the low recognition scores in subtest 2.

This demands the formulation of an action plan (the reply) and simultaneously, an initial cognitive analysis of the information received, regardless of the individual's attitude towards the question or situation.

Essentially, personal frustration is activated when the subject perceives situations to be threatening to his or her self-esteem and self-evaluation. Individuals with high levels of frustration are inclined to perceive a wide range of situations as threatening and therefore, will respond according to what they think the situation dictates.

Unlike personal frustration, the reactive (situational) alarm is dynamical, changes over time and influences how one would react to a potentially stressful situation.

References

1. American Electroencephalographic Society: Guidelines for clinical evoked potentials studies. J. Clin. Neurophysiol. 1, 3–54 (1984)
2. Khil'ko, V.A., Lytaev, S.A., Ostreiko, L.M.: Clinical Physiological Significance of Intraoperative Evoked Potentials Monitoring. Human Physiology 28, 617–624 (2002)
3. Lytaev, S., Shevchenko, S.: VEPs and AEPs Mapping of Occlusive Lesions in Cerebral Vessels. Ann. New York Acad. Sci. 821, 524–528 (1997)

4. Lytaev, S.: Brain Topography of Perception of Target and Non-Target Acoustic Signals. In: Tolstoy, A., Teng, Y.-C., Shang, E.C. (eds.) Theoretical and Computational Acoustics 2003, pp. 291–297. World Scientific, New Jersey (2004)
5. Lytaev, S.A.: Brain Mapping during Perception of the Auditory Information. In: Proceedings of the SPIIRAN, vol. 2, Nauka, St.Petersburg, vol. 2, pp. 469–477 (2005) (Russ.)
6. Picton, T.W., Hillyard, S.A., Krausz, H.I., et al.: Human auditory evoked potentials. 1. Evaluation of components. EEG Clin. Neurophysiol. 36, 179–190 (1974)
7. Picton, T.W.: Human Event-Related Potentials. In: Handbook of EEG and Clin. Neurophysiol. Revised Ser., vol. 3, p. 3. Elsevier, Amsterdam (1988)
8. Shostak, V., Lytaev, S., Golubeva, L.: Topography of Afferent and Efferent Flows in Auditory Selective Attention. Neurosci. & Behav. Physiol. 25, 665–673 (1997)
9. Sukov, W., Barth, D.S.: Cellular Mechanisms of Thalamically Evoked Gamma Oscillations in Auditory Cortex. J. Neurophysiol. 85, 1235–1245 (2001)
10. Woldorff, M.G., Hackley, S.A., Hillyard, S.A.: The effects of channel-selective attention on the mismatch negativity wave elicited by deviant tones. Psychophysiol. 28, 30–42 (1991)

Cultural Neuroscience and Individual Differences: Implications for Augmented Cognition

Laura E. Matzen

Sandia National Laboratories
P.O. Box 5800, MS 1188,
Albuquerque, NM, USA 87185
lematze@sandia.gov

Abstract. Technologies that augment human cognition have the potential to enhance human performance in a wide variety of domains. However, there are a number of individual differences in brain activity that must be taken into account during the development, validation, and application of augmented cognition tools. A growing body of research in cultural neuroscience has shown that there are substantial differences in how people from different cultural backgrounds approach various cognitive tasks. In addition, there are many other types of individual differences and even changes in a single individual over time that have implications for augmented cognition research and development. The aim of this session is to highlight a few of those differences and to discuss how they might impact augmented cognition technologies.

Keywords: Cultural neuroscience, individual differences.

1 Introduction

Augmented cognition technologies use physiological measures recorded from humans to direct human-systems interactions and improve human performance [1]. A major challenge in developing augmented cognition systems stems from the variability of physiological measures across individuals. Differences in age, fitness, cultural background, use of cognitive strategies, and numerous other factors can affect the performance of augmented cognition systems. A tool that works for one group of people may not work for another group. A technology that improves one person's performance may hinder another's performance. Even a tool designed for a single individual may become less effective as he or she changes over time. In order to develop effective augmented cognition tools, researchers and designers must take cultural and individual differences into account. Although these differences can be problematic, research on cultural and individual differences also provides information that designers could leverage to improve their systems.

2 Cultural Neuroscience

A growing body of research shows that a person's cultural background influences his or her cognitive processes in fundamental and pervasive ways. Researchers have

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 194–198, 2011.

argued that several common behavioral findings thought to be universal may not generalize to groups other than the narrow demographic from which they were collected [2]. For example, people from different cultural backgrounds are differentially sensitive to simple visual illusions such as the Müller-Lyer illusion [2,3]. Neuroimaging research faces the same problems; patterns of brain activity observed for one group of participants may be very different from the patterns observed for participants from a different culture [4].

Researchers studying cultural neuroscience have already found many differences in neural processing between groups of people from different cultures. Gutchess and colleagues describe culture as a lens through which people attend to and process their environment [5]. This lens can have a profound effect on how people perceive the world [6].

A framework for understanding cultural differences in brain activity and behavior comes from the work of Nisbett and colleagues [7,8]. They propose that values and beliefs that are central to different cultures bias the ways in which people process their environment. Western cultures value individualism, biasing Westerners to focus on central objects and categorical relationships. East Asian cultures value collectivism, biasing East Asians to use more holistic processing and to focus more on relationships than on categories [8].

A great deal of experimental evidence supports this framework. Westerners and East Asians often have different patterns of eye movement when viewing scenes [9, 10] and faces [11]. Cultural background also influences emotional processing and responses to social information [12,13]. Westerners and East Asians have different patterns of brain activity that correspond with these processing biases [12,13,14,15]. Perhaps because of these fundamental differences in perceptual processing, Westerners and East Asians also tend to categorize information differently [16,17]. This leads the two groups to use different memory strategies [18] and also makes them susceptible to different types of memory errors [5].

Cultural differences may also be reflected in the physical structure of the brain. Functional magnetic resonance imaging (fMRI) studies have found that certain brain regions related to language processing are larger for Chinese speakers than for English speakers [19,20]. Other studies have found differences in cortical thickness between Westerners and East Asians [21].

The numerous differences in neural processing between people of different cultural backgrounds could impact augmented cognition technologies in a variety of ways. For example, a visual display that is optimized for Western users may be less effective for East Asian users. If a system is designed to support memory performance, the characteristics of that system may need to be customized for different groups. A system intended to help users avoid errors might use different types of error detection for people of different cultural backgrounds. Although cultural differences might be problematic in some cases, designers can also take advantage of the growing literature on cultural neuroscience to make augmented cognition systems as effective as possible.

3 Individual Differences

There are also important individual differences between people of the same cultural group. Fitness levels can play a major role in brain activity and cognitive function, differentiating people within the same cultural and age groups. Erickson and colleagues have found that an exercise intervention can effectively reverse age-related losses in brain volume [22]. In one study, older adults with higher levels of aerobic fitness were found to have better spatial memory performance and greater hippocampus volume. In a second study, older adults were assigned to participate in aerobic or non-aerobic exercise three times per week for a year. The participants in the aerobic exercise group benefited from a 2% increase in hippocampus volume. These studies indicate that factors related to lifestyle can have a substantial impact on individuals' brain volume and cognitive function. Human cognitive performance can be augmented simply by a change in lifestyle, such as beginning a moderate exercise regimen. In addition, this research indicates that human brains retain some plasticity throughout the lifespan. Neurogenesis can occur even for older adults.

Brain plasticity can also affect an individual's neural responses over very short time frames, such as the duration of a single experiment. Weisend and colleagues have conducted experiments to measure the variability in individuals' responses to stimuli over the course of an experimental session. Using magnetoencephalography (MEG) recordings of an oddball paradigm, they have found that participants' responses to the stimuli become less variable over time. This stabilization may to be related to synaptic plasticity. As participants gain experience with the experimental paradigm, changes may occur in the timing, phase, or frequency of the neural response [23]. These results indicate that a person's expertise with a system or a set of stimuli may be an important factor in augmented cognition technologies. As an individual interacts with an augmented cognition system, that system may need to adapt to the user's changing neural response as he or she gains expertise.

Like cultural differences, differences in personality and previous experience can also influence individuals' strategy choices and their corresponding brain activity. Forsythe and colleagues have shown that differences in working memory capacity and processing speed, as assessed by standard cognitive tests, correlate with individuals' strategy choices and their willingness to change strategies [24]. These correlations have been observed for extremely simple tasks such as drawing a figure eight under time or accuracy pressure. The individual differences that correlate with performance are fundamental aspects of cognitive processing. These differences are likely to affect how people approach any type of task, from the very simple to the very complex. It may prove useful to adapt augmented cognition systems to particular users by assessing each user's abilities with a battery of individual difference measures. While ignoring such differences could hinder the usability of a system, taking them into account and leveraging them could optimize the system's ability to improve human performance.

References

1. Schmorrow, D., Stanney, K.M., Wilson, G., Young, P.: Augmented cognition in human system interaction. In: Salvendy, G. (ed.) Handbook of Human Factors and Ergonomics, 3rd edn. John Wiley, New York (2005)
2. Henrich, J., Heine, S.J., Norenzayan, A.: The weirdest people in the world? Behavioral and Brain Sciences 33, 61–135 (2010)
3. Segall, M.H., Campbell, D.T., Herskovits, M.J.: The influence of culture on visual perception. Bobbs-Merrill, Indianapolis (1966)
4. Chiao, J.Y., Cheon, B.K.: The weirdest brains in the world. Behavioral and Brain Sciences 33, 88–90 (2010)
5. Gutchess, A.H., Schwartz, A.J., Boduroğlu, A.: The influence of culture on memory. In: Schmorrow, D.D., Fidopiastis, C.M. (eds.) FAC 2011, HCII 2011. LNCS (LNAI), vol. 6780, pp. 67–76. Springer, Heidelberg (2011)
6. Park, D.C., Huang, C.-M.: Culture wires the brain: A cognitive neuroscience perspective. Perspectives on Psychological Science 5, 391–400 (2010)
7. Nisbett, R.E., Masuda, T.: Culture and point of view. Proceedings of the National Academy of Sciences, USA 100, 11163–11170 (2003)
8. Nisbett, R.E., Peng, K., Choi, I., Norenzayan, A.: Culture and systems of thought: Holistic versus analytic cognition. Psychological Review 108, 291–310 (2001)
9. Chua, H.F., Boland, J.E., Nisbett, R.E.: Cultural variation in eye movements during scene perception. Proceedings of the National Academy of Sciences, USA 102, 12629–12633 (2005)
10. Goh, J.O., Tan, J.C., Park, D.C.: Culture modulates eye-movements to visual novelty. PLoS One 4, e8238 (2009)
11. Blais, C., Jack, R.E., Scheepers, C., Fiset, D., Caldara, R.: Culture shapes how we look at faces. PLoS One 3, e3022 (2008)
12. Chiao, J.Y., Harada, T., Komeda, H., Li, Z., Mano, Y., Saito, D., et al.: Neural basis of individual and collectivistic views of self. Human Brain Mapping 30, 2813–2820 (2009)
13. Chiao, J.Y., Iidaka, T., Gordon, H.L., Nogawa, J., Bar, M., Aminoff, E., et al.: Cultural specificity in amygdale response to fear faces. Journal of Cognitive Neuroscience 20, 2167–2174 (2008)
14. Gutchess, A.H., Welsh, R.C., Boduroğlu, A., Park, D.C.: Cultural differences in neural function associated with object processing. Cognitive, Affective, & Behavioral Neuroscience 6, 102–109 (2006)
15. Goh, J.O., Park, D.C.: Culture sculpts the perceptual brain. Progress in Brain Research 178, 95–111
16. Chiu, L.H.: A cross-cultural comparison of cognitive styles in Chinese and American children. International Journal of Psychology 7, 235–242 (1972)
17. Ji, L.J., Zhang, Z., Nisbett, R.E.: Is it culture or is it language? Examination of language effects in cross-cultural research on categorization. Journal of Personality and Social Psychology 87, 57–65 (2004)
18. Gutchess, A.H., Yoon, C., Lou, T., Feinberg, F., Hedden, T., Jing, Q., Nisbett, R.E., Park, D.C.: Categorical organization in free recall across culture and age. Gerontology 52, 314–323 (2006)
19. Green, D.W., Crinion, J., Price, C.J.: Exploring cross-linguistic vocabulary effects on brain structures using voxel-based morphometry. Bilingualism: Language and Cognition 10, 189–199 (2007)

20. Kochunov, P., Lancaster, J., Tan, L.H., Amunts, K., Zilles, K., et al.: Localized morphological brain differences between English-speaking Caucasians and Chinese-speaking Asians: New evidence of anatomical plasticity. NeuroReport 14, 961–964 (2003)
21. Chee, M.W., Zheng, H., Goh, J.O., Park, D.C.: Brain structure in young and old East Asians and Westerners: Comparisons of structural volume and cortical thickness. Journal of Cognitive Neuroscience 23, 1065–1079 (2010)
22. Erickson, K.: Augmenting brain and cognition by aerobic exercise. In: Schmorrow, D.D., Fidopiastis, C.M. (eds.) FAC 2011, HCII 2011. LNCS (LNAI), vol. 6780, pp. 30–38. Springer, Heidelberg (2011)
23. Weisend, M.P.: Inter and intrasubject variability in time-frequency analyses of responses to complex audiovisual stimuli. This volume
24. Trumbo, M., Stevens-Adams, S., Hendrickson, S.M.L., Abbott, R., Haass, M., Forsythe, C.: Individual differences and the science of human performance. In: Schmorrow, D.D., Fidopiastis, C.M. (eds.) FAC 2011, HCII 2011. LNCS (LNAI), vol. 6780, pp. 46–54. Springer, Heidelberg (2011)

Towards a Software Toolkit for Neurophysiological Data Collection and Analysis

James Niehaus and Peter Weyhrauch

Charles River Analytics, USA
{Jniehaus,pweyhrauch}@cra.com

Abstract. Modern devices such as cell phones, handheld computers, and technical equipment enable professional users to communicate, understand, and act more efficiently and effectively. However, these new systems often increase cognitive workload, and may even introduce performance errors. System analysts can decrease these errors by identifying a users cognitive performance deficits and addressing them through training, improved performance support, and redesigned operational systems. To identify these deficits, neurocognitive measurements of indicators such as cognitive workload and attention can be approximated with high accuracy by using non-invasive sensors to measure brain activity and other physiological indicators. Thus, we are designing and demonstrating the feasibility of a toolkit for system analysts to use neurocognitive measurements to recommend additional training for individual users, performance support for all users of the system, and the redesign of system interfaces or components. This research addresses a clear need for an extensible, general-purpose, stand-alone neurocognitive assessment toolkit that can be incorporated into new and existing technology development with little to no integration effort.

1 Introduction

Modern devices such as cell phones, handheld computers, and technical equipment enable professional users to communicate, understand, and act more efficiently and effectively. However, these new systems often increase cognitive workload, and may even introduce performance errors. System analysts can decrease these errors by identifying users' cognitive performance deficits and addressing them through training, improved performance support, and redesigned operational systems.

The neuroscience of cognition has the potential to offer these kinds of insights. By using non-invasive sensors to measure brain activity and other physiological indicators, neurocognitive measurements such as cognitive workload and attention can be approximated with high accuracy. What is needed is an extensible, general-purpose, stand-alone suite of sensors that can incorporate rapidly and affordably into new and existing technology development. To be practical, such stand-alone sensors must incur little to no development overhead, cannot require re-engineering or re-architecting of the operational system, and must be relatively inexpensive. Similarly, they should not overly burden the user; sensors should be portable, light, non-invasive, low-power, and relatively easy to equip and remove.

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 199–202, 2011.

To understand the cognitive properties of system use, the system analyst must first use these sensors to collect neurocognitive data on system use (i.e., user neurological data, user physiological data, and available system state data). Once the neurocognitive sensors have been deployed in a test or training situation, the analyst must have tools to help analyze the collected data to identify key neurocognitive measurements for assessment. In particular, the analyst must identify the cognitive causes of performance deficits. For example, the analyst of a radar display may discover that high operator cognitive workload during target identification severely degrades the accuracy of target tracking. Once the analyst identifies the combination of neurocognitive measures that predict performance, he or she must implement performance measures to identify training needs, performance support, and system interface or component redesign. The implemented performance measures eliminate the need for the analyst to reanalyze the neurocognitive measures when analyzing the system users. Continuing the example, the analyst may develop a “cognitive workload during target identification” performance measure; this measure indicates which operators are experiencing this particular difficulty so the analyst can recommend additional training for those users. Once implemented, these custom performance measures can be used to recommend additional training for individual users, performance support for all users of the system, and the redesign of system interfaces or components.

2 Related Work

The field of neuroergonomics is the study of the brain and behavior at work, a combination of human factors design and cognitive neuroscience [1]. Neurocognitive measurements are an active field of research, and identified measures include cognitive workload [1-5], situational awareness [1], change detection [1], vigilance [6], and operator fatigue [7]. Research by Parasuraman and Wilson [8] [9] has implemented neuroergonomics techniques for the design of unmanned aerial vehicle (UAV) interfaces and automated support, and [10] discussed ramifications of neuroergonomics and human error. The design and development of embeddable agents for cognitive assessment, however, is a novel task.

3 System Design and Architecture

Fig. 1 shows the workflow for using the designed toolkit, termed Cognitive Readiness Agents for Neural Imaging and Understanding Models (CRANIUM). The system analysis proceeds in two stages. In the first stage, the analyst performs analysis of the group to determine which neurocognitive measures correlate with performance, and the analyst saves these measurements for the second stage. In the second stage, the analyst uses these measures to collect performance data on individual users, grade the performance according to the measures, and make recommendations for performance support. Each stage includes each of the components in the toolkit architecture.

In the first stage, the system evaluator uses the **Agent Editor** to refine the data collection specification that specifies which system events (i.e., log file changes, user interface screens, or other output) trigger information collection and what sensor and

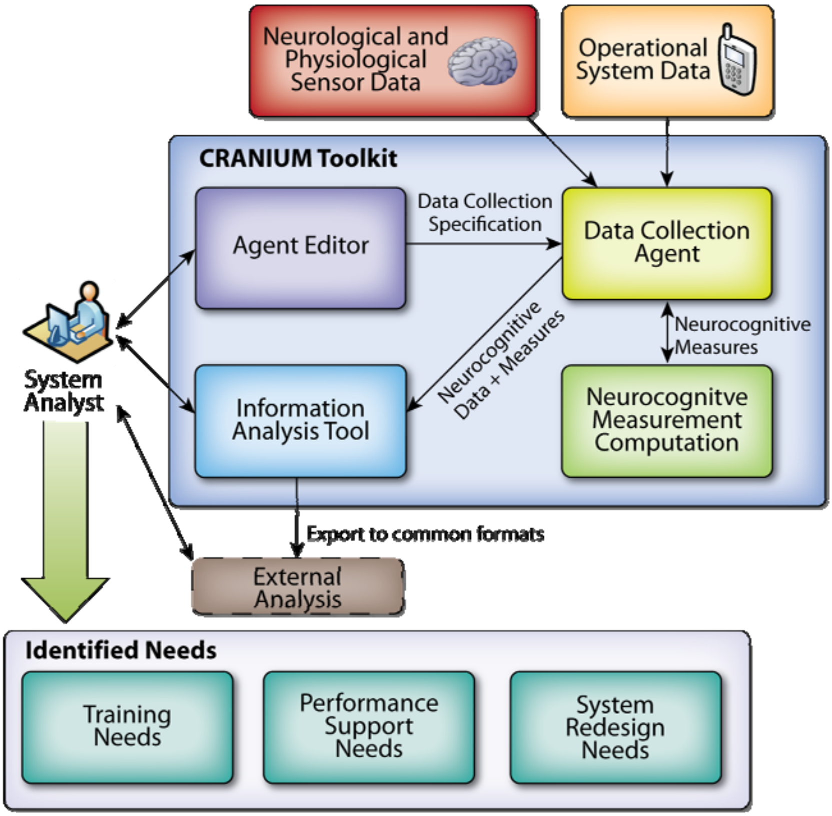


Fig. 1. CRANIUM toolkit workflow

system information about those events is collected. This specification is deployed through **Data Collection Agents** that collect the specified neurocognitive data from the neurological and physiological sensors and from the operational system. The data is then used in the **Neurocognitive Measurement Computation,** and both the neurocognitive measurements and data is passed to the **Information Analysis Tool**. The system analyst works with the **Information Analysis Tool** (or exports the data for external analysis) to identify the neurocognitive measurements that best correlate with performance. The analyst can repeat this first stage to refine the data collection specification or add or remove sensors. The results of this analysis are used to construct performance measures for user assessment.

In the second stage, once training begins on the operational system, the system analyst uses CRANIUM to implement the identified performance measures and collect performance data on individual users. The analyst uses the **Agent Editor** to define performance measures as combinations of neurocognitive measures and deploy the **Data Collection Agents** to collect them. The **Data Collection Agents** either compute the performance assessment directly—consulting the **Neurocognitive Measurement Computation**—or, if the analysis requires more processing than is available on the device, save the relevant data for offline analysis. The performance assessments are returned to the **Information Analysis Tool**, and the system trainer works with this tool to review user performance and identify training needs. If the performance measures show that the majority of the user population is experiencing significant difficulties using the system, the system analyst may instead recommend performance support or system redesign to better meet the needs of the users.

4 Conclusions and Future Work

This paper presents the first steps towards an extensible, general-purpose, stand-alone neurocognitive assessment toolkit that can be incorporated into new and existing technology development with little to no integration effort. With a toolkit such as this, system analysts can investigate the impact of their operational system on user cognitive workload and overall performance, and the analyst can address potential performance issues early in the system development cycle. The end result is a system that is more neuroergonomic, augmenting the cognitive abilities of the user instead of overloading them.

Future work includes the design of user interfaces for system analysts to perform this function, incorporation of existing neurophysiological sensors, and evaluation with user populations.

References

1. Parasuraman, R.: Neuroergonomics: Research and Practice. Theoretical Issues in Ergonomics Science 4(1), 5–20 (2003)
2. Wickens, C.D.: Multiple Resources and Performance Prediction. Theoretical Issues in Ergonomics Science 3(2), 159–177 (2002)
3. Parasuraman, R., Caggiano, D.: Neural and Genetic Assays of Human Mental Workload. Quantifying Human Information Processing (2005)
4. Posner, M.I.: Cognitive Neuroscience of Attention. The Guilford Press, New York (2004)
5. Sirevaag, E.J., Kramer, A.F., Wickens, C.D., Reisweber, M., Strayer, D.L., Grenell, J.F.: Assessment of Pilot and Mental Workload in Rotary Wing Aircraft. Ergonomics 36(9), 1121–1140 (1993)
6. Byrne, E.A., Parasuraman, R.: Psychophysiology and Adaptive Automation. Biological Psychology 42(3), 249–268 (1996)
7. Hillman, C.H., Belopolsky, A.V., Snook, E.M., Kramer, A.F., McAuley, E.: Physical Activity and Executive Control: Implications for Increased Cognitive Health During Older Adulthood. Research Quarterly for Exercise and Sport, 176–185 (2004)
8. Parasuraman, R., Wilson, G.F.: Putting the Brain to Work: Neuroergonomics Past, Present, and Future. Human Factors 50(3), 468 (2008)
9. Wilson, G.F., Russell, C.A.: Performance Enhancement in an Uninhabited Air Vehicle Task Using Psychophysiologically Determined Adaptive Aiding. Human Factors 49(6), 1005 (2007)
10. Fedota, J.R., Parasuraman, R.: Neuroergonomics and Human Error. Theoretical Issues in Ergonomics Science 11(5), 402–421 (2009)

From Sound to Meaning: Changes in EEG Source-Localized Brain Activity with Foreign-Language Training

Catherine Poulsen, Phan Luu, Colin Davey, Don Tucker, and Joey Nelson

Electrical Geodesics, Inc., Eugene, OR 97403, USA
cpoulsen@egi.com

Abstract. Learning a foreign language is a complex human task, involving multiple processes and a dynamic network of brain activity. The present study used 256-channel dense-array electroencephalography (dEEG) and linear-inverse source analysis (sLORETA) to identify changes in brain activity during the early stages of language training. Twenty native English speakers attended two 50-minute sessions of computer-assisted, virtual-reality Dari language instruction. Training-specific changes in neural activity were observed in both articulatory-motor and semantic processing regions, including increases in left posterior inferior temporal gyrus and left lateral inferior frontal regions. Also observed was increasing left lateralization, and an increase in mediotemporal regions suggestive of memory reconsolidation. These findings illustrate the ability to track changes with training in recognized language-processing brain regions using source-localized EEG recorded while listening to continuous, naturalistic speech. Subsequent research will explore individual differences and the development of adaptive training based on neural indices.

Keywords: language learning, training, dense-array EEG, linear-inverse source analysis, electroencephalography, event-related potentials.

1 Introduction

When acquiring a foreign language, a stream of initially unintelligible sounds slowly emerges into distinct words and meaningful phrases. This transition includes the ability to perceive speech sounds (phoneme perception), parse the phonetic stream into words (speech segmentation), encode and retrieve word meanings (lexical and semantic access), and combine individual words into comprehensible sequences (syntactic processing and sentence integration). Underlying this acquisition, which is typically experienced in behavior or self-observation of learning, are changes in the learner's brain responses during exposure to the new language. Tracking these brain responses during the learning process may guide the development of more efficient and effective training, and the formation of real-time, customized, adaptive training regimes.

Event-related potential (ERP) studies of bilingualism and foreign language learning have revealed interesting differences in brain responses to language acquired

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 203–211, 2011.

as a native versus non-native speaker. For example, Sanders and Neville [1] showed that when native English speakers listened to continuous speech, they evidenced large, auditory N1 and N200-300 responses to the onset of each content word, accompanied by a distinct absence of such responses to the onset of syllables internal to a word, indicative of their ability to parse a continuous stream of speech into individual words. In contrast, the brain responses of relatively proficient Japanese non-native English speakers did not differentiate between word-initial and word-medial syllables, presumably due to less efficient speech segmentation processes [2]. These findings by Sanders and Neville revealed that differences in the brain responses of native versus non-native speakers can persist even after several years of exposure to the foreign language. Within-subject longitudinal studies of foreign language learning by Osterhout and colleagues, however, have also revealed striking changes in brain responses to a foreign language as a function of learning, even very early in acquisition. For example, Osterhout et al. [3] examined changes in the N400, an ERP index of lexico-semantic access that is typically larger for words whose meanings are more difficult to retrieve (e.g., words that are low-frequency, less predictable, or semantically anomalous). After only 14 hours of instruction, learners evidenced a significantly larger N400 to pseudowords than words, even though they performed at chance when distinguishing real words from pseudowords. Importantly, this study shows that brain responses can reveal a learner's acquired knowledge, here about the lexical status of these words, even before the learners are aware or able to demonstrate this learning behaviorally. Thus, ERP studies have been shown to be sensitive to both native versus non-native contrasts in brain activity, as well as longitudinal changes within non-native brain responses as a function of language training.

The majority of ERP studies of language processing to date, however, have relied on mapping the brain's detection of, or differential response to, linguistic anomalies, such as pseudowords, semantically incongruous words, or syntactic violations, as compared to well-formed linguistic input. Furthermore, these studies typically describe ERP responses on the scalp surface, measured by relatively sparse electrode arrays (e.g., 16, 32, or 64 sensor locations), and rely on a limited number of well-established scalp ERP components, such as the N1, N400, and P600. Ideally, one would like to track changes in brain responses while the learner is actively engaged in training, and listening and responding to natural language for comprehension and communication. Not only would this better characterize the brain's processing of authentic language processing, it would also provide the neural signatures that would be required for aligning language training with the brain's current learning state. In addition, electrical source analysis of dense-array (i.e., 128 or 256-channel sensor arrays) EEG (dEEG) recordings would allow the ERP researcher to relate the scalp data to brain regions and networks previously revealed by fMRI and lesion research on language.

Prior research on the functional neuroanatomy of speech [e.g., 4, 5] has identified a widely distributed network of regions engaged during speech comprehension. These include the classic Broca (inferior frontal gyrus) and Wernicke (posterior superior temporal gyrus) areas, as well as extensive processing along the superior temporal gyrus, lexical and semantic processing along middle and inferior temporal gyri and temporal pole, and sensorimotor processing in frontal premotor and parietal regions.

Hickok and Poeppel [4] proposed these regions can be divided into two processing networks, a dorsal network processing the articulatory sounds of speech perception, and a ventral stream focusing on meaning.

The present study aimed to track processing associated with early language acquisition by measuring changes in brain activity across language training sessions using dense-array electroencephalography (dEEG) and linear-inverse electrical brain source analysis. This study is distinguished from traditional event-related potential (ERP) language research by the incorporation of several features. First, brain activity was recorded as subjects listened to naturalistic samples of continuous speech for comprehension. Thus, it measured the brain's natural response to auditory language as communication, a more ecologically valid task than typical ERP studies that rely on the brain response to linguistic violations as a measure of language processing. Second, source analysis was conducted to identify the underlying brain regions engaged during speech comprehension. This made it possible to draw upon fMRI and lesion literature and relate observed regional changes to specific aspects of language processing. Third, repeated measurements at multiple time points across early stages of language acquisition was better suited to documenting the actual learning process than cross-sectional contrasts and will permit more effective exploration of individual differences.

2 Method

2.1 Participants

Twenty right-handed, native English speakers (12 male, mean age = 25 years) were recruited from the University of Oregon. All participants gave informed consent prior to participation and received $60 in remuneration. The study was approved by the institutional review boards at Electrical Geodesics and the University of Oregon.

2.2 Materials

Language Training. Training consisted of two identical 50-minute sessions of Tactical Dari (Alelo, TLT, Los Angeles, CA), a computer-assisted, foreign-language training software for Dari, an Afghani language similar to Persian. These training sessions covered greetings and introductions through interactive dialogues and exercises controlled by the learner with a mouse. Integrated voice recognition software was used to enable the computer characters to interact with the learner utterances and to provide automated feedback to the learner.

Pretest-Posttest. The 30-minute pretest and posttest were identical and consisted of 76 mini-dialogues in Dari read aloud by a single native-speaker of Dari. Half the dialogues were composed of phrases from the lesson in which they then received training (Trained), and half were composed of Dari phrases from a lesson in which they received no training (Untrained). The Trained and Untrained dialogues were intermixed across the length of the test. Participants actively listened to each dialogue and rated after each one, on a scale from 1 to 4, both how well they recognized the words in the dialogue and how well they understood the dialogue.

2.3 Procedure

Participants attended two training sessions, separated by a day with no instruction. At the start of each day, participants were fitted with a 256-channel HydroCel Geodesic Sensor Net for recording of EEG data during the Dari pretest, training, and posttest. Participants were seated in front of a computer monitor and listened to all dialogues and the Tactical Dari lesson via air canal insert earphones (Etymotic Research, Elk Grove Village, IL). During the pretest and posttest, stimulus presentation was controlled by E-Prime Software, Version 1.2.1.795 (Psychology Software Tools, Pittsburgh, PA), and synchronized with EEG acquisition via the E-Prime Extension for Net Station. During the pretest and posttest dialogues, participants maintained fixation on a cross (+) presented in the center of the computer monitor in order to reduce eye movement artifact during EEG recording. Following the 30-minute pretest on Day 1, participants completed a module on how to use the Tactical Language Training software in which French was substituted as the foreign language so as not to confound exposure to Dari. They then completed a 40-minute Lesson 1 in Tactical Dari, Greetings and Introductions, followed by a 10-minute quiz. The session concluded with the 30-minute posttest. Day 2 was identical to Day 1, except that participants did not complete the module on how to use the software on Day 2. Each session, including net application, pretest, language training, and posttest, lasted approximately three hours.

2.4 EEG Recording, Preprocessing, and Source Analysis

The EEG was recorded during the pretest and posttest using the 256-channel HydroCel Geodesic Sensor Net, Net Amps 300 amplifier, and Net Station, Version 4.4.1, software (Electrical Geodesics, Eugene, OR). Electrode impedances were maintained below 100 kΩ [6]. The EEG was recorded with a 100 Hz low-pass filter, amplified at a gain of 1,000, with a 250 Hz sampling rate, and digitized with a 24-bit A/D converter. All channels were referenced to Cz during data acquisition.

After acquisition, the continuous EEG was filtered with a 0.1-Hz to 30-Hz bandpass filter, and segmented by word onset into 1,000-ms epochs with a 100-ms baseline period. Epochs contaminated by eye or movement artifact were identified by computer algorithm and eliminated. Individual bad channels were identified and interpolated on a segment-by-segment basis using spherical spline interpolation. For each subject file, the EEG was averaged into event-related potentials (ERPs), time-locked to the onset of each word, for each of eight categories: Test (Pre, Post) x Day (1, 2) x Lesson (Trained, Untrained). Following visual inspection, the data from one participant was eliminated due to a bad reference electrode. A grand average file was computed from the remaining 19 participant ERP averages.

Linear-inverse source analysis was performed on the grand-averaged scalp ERPs using GeoSource, Version 2.0, software (Electrical Geodesics, Eugene, OR). A finite difference model (FDM) of a typical human head was computed using a segmented, high-resolution T1-weighted MRI scan of the brain and CT of the skull in order to model realistic head tissue geometry and conductivity. Conductivity values used in the FDM model were 0.25 S/m (Siemens/ meter) for the brain, 1.8 S/m for the cerebral spinal fluid, 0.018 S/m for the skull, and 0.44 S/m for the scalp. These conductivity values reflect the more accurate 14:1 ratio between brain and skull conductivity rather than the traditional 80:1 estimate [7, 8]. Dipole triples (x, y, z orthogonal orientations) were placed in 2447 7-mm voxels distributed throughout cortical gray matter based on the MNI305 probabilistic map. This FDM lead field matrix was inverted using the standardized low-resolution electromagnetic tomography analysis (sLORETA) method [9]. The resulting voxel intensities at each time point for each condition were displayed on MRI slice views of a typical Talairach-transformed brain.

3 Results

Rated comprehension increased significantly from pretest to posttest on each day of training for dialogues composed of trained, but not untrained, words (see Figure 1), $p < 0.0001$. For the trained-word dialogues, rated comprehension also decreased a small, but significant, amount from the posttest immediately following training on Day 1 to the time of the pretest on Day 2, $p < 0.0002$.

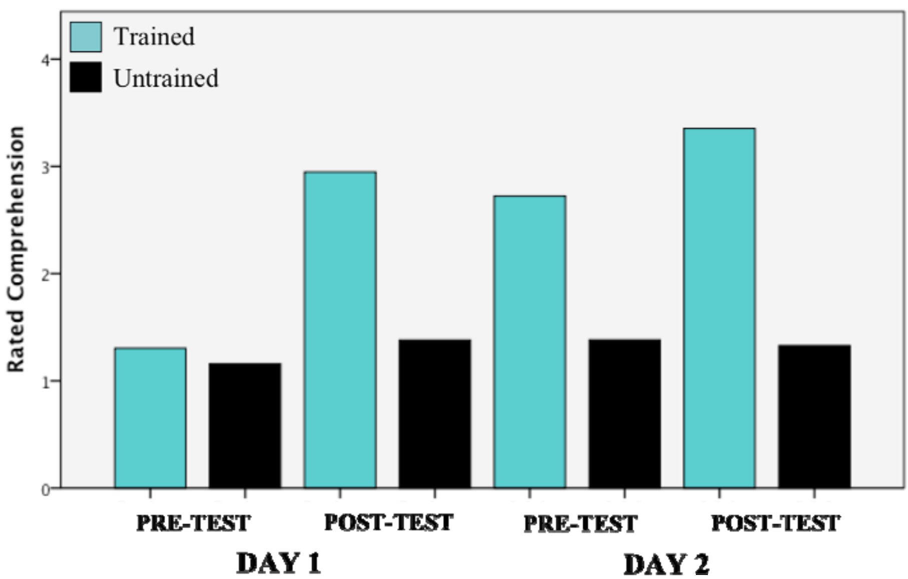


Fig. 1. Rated comprehension of pretest and posttest dialogues composed of trained or untrained words

Figure 2 illustrates training-specific changes in brain activity, changes that were observed in both articulatory-motor and semantic processing networks. Premotor activity was larger on Day 1, both before and after training, than on Day 2. Activity in Broca's area, however, was observed only after training on Day 1, and persisted through Day 2. Activity in semantic processing regions, including the left posterior inferior temporal lobe (anterior fusiform gyrus) and left temporal pole increased with training, especially on Day 2.

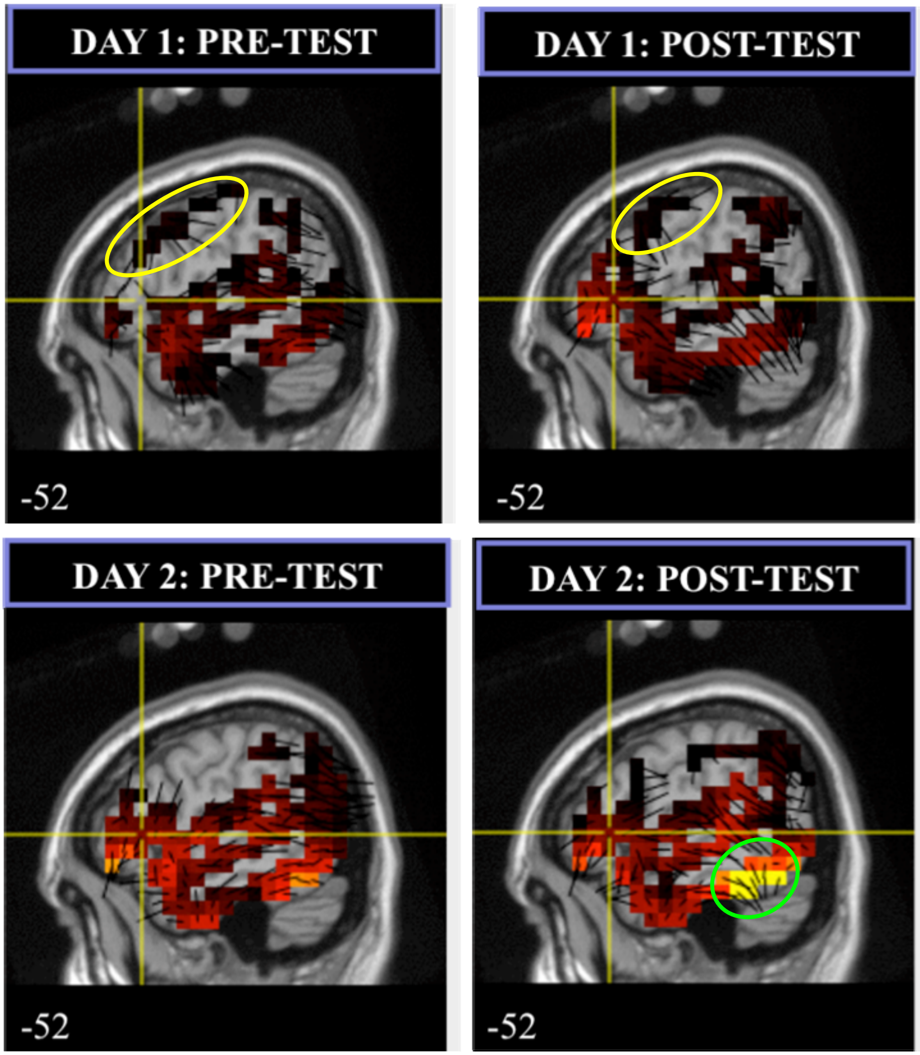


Fig. 2. Grand-average source solution for the left hemisphere for trained words at pretest and posttest on Days 1 and 2. Yellow circles highlight articulatory-motor activation present on Day 1. Crosshair indicates Broca's Area, which is engaged only after training on Day 1 and continues through Day 2. The green circle marks the fusiform gyrus of the semantic network, which increases in activity with training.

Figure 3 shows the source localized brain activity for both trained and untrained words superposed on sagittal (x = -45) and axial MRI views. As can be observed in the axial slice view, there is increasing left lateralized activity on Day 2, particularly for the trained words. Finally, a large increase in medial temporal regions of the left hemisphere also obtains on Day 2 in both the pre- and posttest for trained, but not untrained, words.

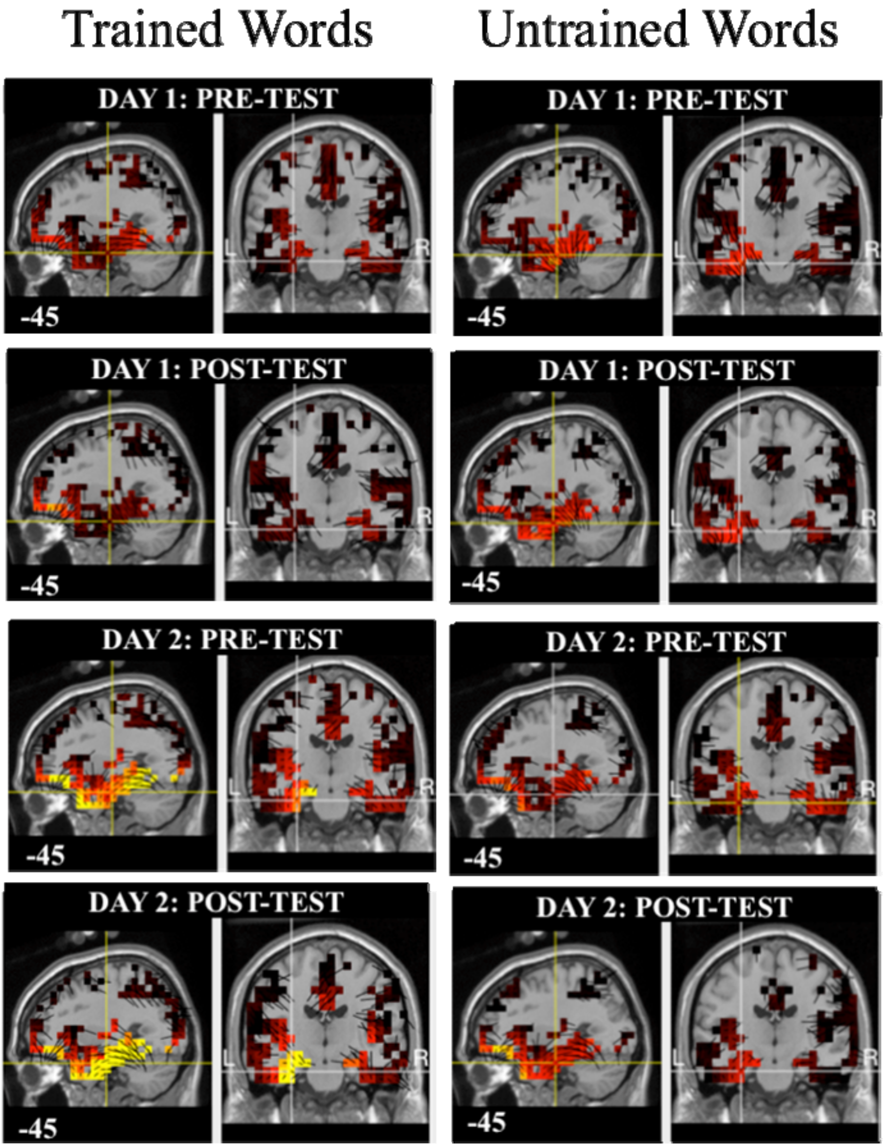


Fig. 3. Sagittal and axial views of source-localized grand-averaged electrical activity for trained and untrained words showing increased left-lateralized activity with training, and a large increase in medial temporal activity on Day 2 pretest and posttest for trained words only

4 Discussion

As expected, behavioral results for self-ratings of comprehension on the pretests and posttests indicated a strong effect of learning from language training on each day.

Furthermore, the small decrease in perceived comprehension from the Day 1 posttest to Day 2 pretest suggests that learners may have experienced some forgetting over the one-day break between sessions, but largely retained their learning from the first day of training.

The pattern of changes in source-localized brain activity also indicated striking training-specific changes that are highly consistent with fMRI-based models of language processing [e.g., 4, 5]. The decrease, from the first to the second day of training, in articulatory-motor regions accompanied by an increase in ventral semantic processing regions for trained words suggests a shift away from focusing on the sound and articulation of Dari to focusing on the meanings of the Dari words. This is reinforced by the observed increase in left lateralized activity, stronger for trained words, but also present to a lesser degree for the untrained words by the Day 2 posttest. Such a pattern suggests that the learners' brains are increasingly processing the Dari input within a left-lateralized language-specific network. Finally, the large increase in activity in medial temporal regions following the one-day break between training sessions suggests that important memory reconsolidation processes may have occurred such that on Day 2, even at the time of the pretest before any additional training had been given, the learners' medial temporal memory regions were actively engaged, retrieving the word meanings from consolidated semantic memories. This effect is clearly absent during the processing of dialogues composed of untrained words for which no consolidated memory traces can exist.

In summary, these findings illustrate the ability to measure changes associated with language learning in recognized language-processing brain regions using source-localized EEG recorded while listening to continuous, naturalistic speech. Future analyses will explore statistical analysis of these source activations and individual differences in performance and brain indices of learning. Subsequent studies will attempt to identify and track these same changes during engagement in Tactical Dari training itself, rather than examining changes from pre- to posttest. This should, ultimately, permit the development of adaptive language training based on real-time monitoring of neural indices of language acquisition.

References

1. Sanders, L.D., Neville, H.J.: An ERP study of continuous speech processing. I. Segmentation, semantics, and syntax in native speakers. Brain Res. Cogn. Brain Res. 15(3), 228–240 (2003)
2. Sanders, L.D., Neville, H.J.: An ERP study of continuous speech processing. II. Segmentation, semantics, and syntax in non-native speakers. Brain Res. Cogn. Brain Res. 15(3), 214–227 (2003)
3. Osterhout, L., McLaughlin, J., Pitkanen, I., Frenck-Mestre, C., Molinaro, N.: Novice learners, longitudinal designs, and event-related potentials: A means for exploring the neurocognition of second language processing. Language Learning, pp. 199–230 (2006)
4. Hickok, G., Poeppel, D.: The cortical organization of speech processing. Nat. Rev. Neurosci. 8, 393–402 (2007)
5. Rodd, J.M., Davis, M.H., Johnsrude, I.S.: The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. Cereb. Cortex 15(8), 1261–1269 (2005)

6. Ferree, T.C., Luu, P., Russell, G.S., Tucker, D.M.: Scalp electrode impedance, infection risk, and EEG data quality. Clin. Neurophysiol. 112, 536–544 (2001)
7. Salman, A., Turovets, S., Malony, A., Poolman, P., Davey, C., Eriksen, J., et al.: Noninvasive conductivity extraction for high-resolution EEG source localization. Advances in Clinical Neuroscience and Rehabilitation 26, 27–28 (2005)
8. Turovets, S., Salman, A., Malony, A., Poolman, P., Davey, C., Tucker, D.: Anatomically constrained conductivity estimation of the human head tissues in vivo: computational procedure and preliminary experiments. In: Paper presented at the 7th Conference on Biomedical Applications of Electrical Impedance Tomography, Seoul, Korea (2006)
9. Pascual-Marqui, R.D.: Standardized low resolution brain electromagnetic tomography (sLORETA): technical details. Methods & Findings in Experimental & Clinical Pharmacology 24D, 5–12 (2002)

Analyzing Neural Correlates of Attentional Changes during the Exposure to Virtual Environments: Application of Transcranial Doppler Monitoring

Beatriz Rey[1], Vera Parkhutik[2], José Tembl[2], and Mariano Alcañiz[1,3]

[1] Instituto en Bioingeniería y Tecnología Orientada al Ser Humano, Universidad Politécnica de Valencia, Camino de Vera s/n, 46022 Valencia, España
[2] Neurology Service, Hospital Universitari La Fe, Valencia, Spain
[3] CIBER Fisiopatologia Obesidad y nutrición (CB06/03), Instittuto Carlos III, Spain
{brey,malcaniz}@labhuman.i3bh.es, tembl_jos@gva.es, parkhutic_ver@yahoo.com

Abstract. Transcranial Doppler monitoring (TCD) has been proposed as a tool to be used in Augmented Cognition (AugCog) systems to monitor brain activation during the performance of different cognitive tasks. In the present study, the main goal is to analyze variations in blood flow velocity (BFV) measured by TCD during the exposure to a virtual reality environment when there are changes in the focus of attention of the participants. Two abrupt events are forced during the navigation in a virtual environment in order to change their focus of attention to the real world. In one of them, the screen goes completely blue, and in the other one, a mesh appears in front of the virtual environment making it difficult to visualize. Results show that BFV values in both middle cerebral arteries remain similar when the first event occurs, but there is an increase during the second event. The origin of this increment may probably be found in the higher difficulty of having a mesh in front of the virtual environment, requiring more attention than before. These results show that changes in the stimuli can generate modifications in BFV that can be monitored by TCD, and can be useful for AugCog applications.

Keywords: Augmented Cognition, Virtual Reality, Transcranial Doppler, Neurophysiological Data, Cognitive State Assessment.

1 Introduction

Transcranial Doppler (TCD) is an ultrasound diagnosis technique used to monitor the blood flow velocity (BFV) variations in major cerebral arteries with high temporal resolution [1]. If the neurovascular coupling is adequate [2], these variations reflect changes in cerebral blood flow, which increases during the performance of mental tasks, such as reading, arithmetic operations, visual stimulation, attention, verbal tasks, motor tasks, visuospatial tasks and memory [3-6].

In the field of Augmented Cognition (AugCog), TCD has been applied to analyze brain activity during vigilance tasks [7-8]. It has been found that the vigilance

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 212–220, 2011.

decrement in detection rate over time is accompanied by a decrease in BFV in both MCAs [9]. This reduction only happens when the observers are asked to actively monitor the stimuli, and not when they are asked just to look at the vigilance displays with no task to be performed. Other studies [10] have focused on abbreviated vigilance tasks. Although they found a significant decline in performance, there was no significant change in BFV measures over time, which does not coincide with earlier findings from long-duration tasks.

TCD has also been applied during simulated air traffic control tasks [11], in order to monitor the influence of automation cues of varying reliability on vigilance performance. Performance effects for cueing found in the experiment were closely followed by changes in BFV just in the right MCA in conjunction with low salience signals.

Recently, TCD has been proposed as a tool to measure brain activity during the exposure to virtual environments (VE) that can be used in Augmented Cognition (AugCog) systems [12]. It has already been used to monitor BFV during the exposure to VEs in different immersive and navigation conditions [13, 14], in order to analyze neural correlates of the sense of presence, which is the feeling of the participant of being in the VE. Results from these studies have shown that there is an increase in BFV of middle and anterior cerebral arteries (MCAs and ACAs) when the participant starts the exposure to the VE. Furthermore, significant differences have been found between the increases in BFV observed in various virtual reality (VR) configurations associated with different levels of presence in the participants.

In those previous studies [13-14], no abrupt changes in the presentation of the stimuli occurred during the automatic or guided navigations through the VE. The focus of attention of the participant was directed upon the VE during the whole experience. However, during the navigation in a virtual environment, there are times when the participant may switch the interpretation of the sensory inputs as coming from the VE or as coming from the real world [15]. These changes in the focus of attention from the virtual to the real world have often been described as breaks in presence [16], because when they occur, the participants become aware that they are participating in a computer-mediated experience in a research laboratory and they stop feeling present in the virtual world. Some studies have been conducted to analyze both the subjective interpretation and the physiological correlates of those changes [17-18]. The moments when the participants changed their focus of attention were forced by the experimental design, in order to control the time when those switches occur. Specifically, in these studies, there were moments when the projections were abruptly changed to white, generating a sudden anomaly in the visual perception. An averaged galvanic skin response during the anomalies was presented in one of the studies [18]. Heart rate was also analyzed, and a decrease was observed during the experimental abrupt modifications in the visual projections.

In the present study, our goal is to analyze which are the effects on BFV of changes in the focus of attention from the virtual to the real world in participants of virtual reality experiences, as a complement to previous studies that have been based on peripheral physiological measures [18]. Our hypothesis is that there will be changes in BFV during those sudden modifications in the projection stream, either as an increase in mean BFV (occasioned by the surprise and rise in the level of concentration due to an unexpected event) or as a decrease in mean BFV (generated by a diminution in brain activation because the visual stimuli have disappeared and navigation in the VE is no longer possible).

2 Method

In the following subsections, a description of the methodology used for the experimental sessions and BFV data analysis can be found.

2.1 Subjects

Seventeen right-handed volunteers aged between 21 and 64 years (mean age, 29.47; standard deviation, 12.19) participated in the study. All the participants gave their informed consent prior to their inclusion in the study. Handedness was established during the previous interview.

2.2 Transcranial Doppler Monitoring

A commercially available 2-MHz pulsed-wave TCD unit (Doppler-Box™ Compumedics Germany GmbH) was used to obtain a bilateral continuous measurement of the Doppler signal. This unit allowed the on-line calculation of BFV during the experiment. It was connected to a PC in which DWL Doppler software (QL software) was used to receive data from the Doppler Box. The apparatus was connected to a PC in which DWL® Doppler software (QL software) was installed. The monitored signals were stored on the PC hard disk during the experiments for off-line analysis.

Details about the insonation technique can be found in other studies (i.e., [19]). With this technique, both hemispheres can be simultaneously monitored through the temporal windows using two probes. Probes were attached to the user's head using the probe holder provided with the device. Each probe was located to monitor vessels between 50 and 55 mm depth, which allowed the registration of left and right middle cerebral arteries (MCA-L and MCA-R) BFV.

2.3 Virtual Reality Settings

The environment was retro-projected in a 4 x 1.5 m metacrilate screen. Users navigated using a Logitech Rumblepad Joypad (Logitech, Fremont, CA, USA).

The VE displayed on the screen was a park with different elements such as trees, benches and swings. Some images of the environment are shown in Fig. 1.

The environment was programmed using Brainstorm eStudio software (Brainstorm Multimedia, Madrid, Spain). Participants could navigate freely through the park while they heard to ambient sounds and background music through stereo loudspeakers.

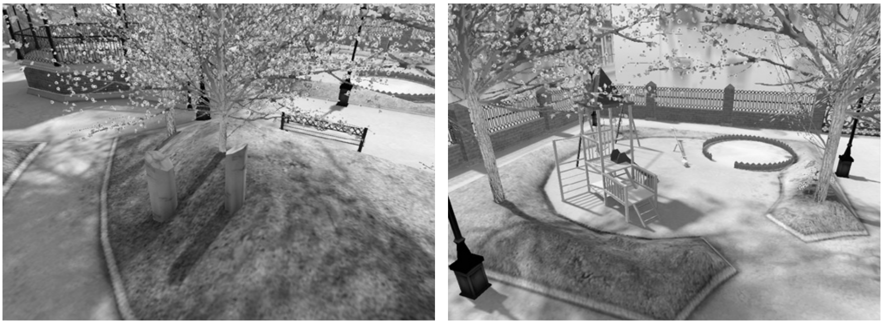


Fig. 1. Images of the virtual environment (park) used during the experimental sessions

2.4 Experimental Procedure

Previously to the beginning of the virtual reality experience, a neurosonologist adjusted the angle and the insonation depth of each probe once the subject was sitting in front of the screen to monitor the MCAs. When the probes had been adjusted, there was a training stage where the participants learnt to navigate in a simple environment using the joypad.

Then, the exposure to the virtual environments (parks) started. Globally, this phase lasted 210 s. This period was divided in several stages:

1. Repose (40 s). The screen remained black. The participant had to be relaxed and wait until the next stage started.
2. Free navigation A (70 s). The subject navigated freely through the park.
3. Rupture A (20 s). Abruptly, the screen went completely blue. The participants did not know that this change was going to happen, so they suddenly became aware that they were participating in a virtual experience and changed their focus of attention to the real world.
4. Free navigation B (40 s). The environment was shown again and the participant navigated freely.
5. Rupture B (20 s). A mesh appeared in front of the VE making it difficult to visualize it. Subjects could temporarily change the focus of attention to the real world. However, in this case, it was possible for them to continue the navigation but visualizing the environment partially occluded by the mesh.
6. Free navigation C (20 s). The VE appeared again and participants navigated through it.

2.5 BFV Signal

MCA-L and MCA-R BFV signals were captured and stored during the experiment on the PC hard disk with a sampling rate of 100 Hz.

The registries for the different vessels were validated by the neurosonologist. Only those cases in which measurements from both vessels were available (14 subjects) have been included in the analysis to allow the comparison between vessels.

Fig. 2. Photograph of one of the participants navigating with the joypad through the virtual environment of the park

2.6 Data Analysis

Prior to other processing steps, the BFV signal is filtered using a low-pass filter with a cut-off frequency of 10 Hz. Then, the filtered BFV signal is transformed to percentage relative units, simply dividing each sample by the arithmetic mean of the full set of data during the whole examination time and multiplying by 100 [20] (eq.1). Mean BFV in the different monitored vessels during each of the different periods was calculated.

$$X[n] = 100 \cdot \frac{x[n]}{\frac{1}{N} \cdot \sum_{n=0}^{N-1} x[n]} \tag{1}$$

A two-way analysis of variance (ANOVAs) with repeated measures was applied to analyze the effects in the dependent variables (BFV) of the within-subjects factors:

- The vessel under study (MCA-R, MCA-L).
- The period (Repose, Free navigation A, Rupture A, Free Navigation B, Rupture B, Free Navigation C).

If Mauchly's test indicated that the assumption of sphericity had been violated, Greenhouse Geisser corrections were applied. Paired comparisons were calculated to allow the comparison between mean BFV in consecutive periods.

3 Results

Mean BFV values (and standard deviation) in the different periods are shown in percentage relative units in Table 1 (MCA-R) and Table 2 (MCA-L).

Table 1. Normalized MCA-R BFV values in the different periods (number of subjects = 14)

Period	Mean MCA-R BFV	MCA-R BFV Standard Deviation
Repose	91.04	4.64
Free Navigation A	103.26	1.64
Rupture A	102.14	3.66
Free Navigation B	100.09	1.27
Rupture B	102.66	3.20
Free Navigation C	101.64	2.67

Table 2. Normalized MCA-L BFV values in the different periods (number of subjects = 14)

Period	Mean MCA-L BFV	MCA-L BFV Standard Deviation
Repose	91.96	6.49
Free Navigation A	103.34	2.01
Rupture A	102.99	4.42
Free Navigation B	99.50	1.93
Rupture B	101.77	3.64
Free Navigation C	100.56	3.87

Results from the ANOVA applied to BFV mean show a significant effect for the period factor ($F(2.120,27.561)=19.928$, $p<0.001$). No significant effect was found for the vessel under study factor or for the interaction factor. Paired comparisons applied to consecutive periods revealed the following significant differences:

- A significant increment in BFV between the initial repose and the Free Navigation A ($p<0.001$), in both MCA-L and MCA-R.
- A significant increment in BFV between the Free Navigation B and the Rupture B, in MCA-L ($p=0.035$) and MCA-R ($p=0.005$).
- A significant decrement in BFV between the Rupture B and the Free Navigation C, only in MCA-L ($p=0.016$)

4 Discussion

The present work has analyzed the BFV responses of participants in a VR experience during forced changes in the focus of attention from the virtual to the real world. The interpretation of the obtained results and the main contributions of the work will be discussed in the following paragraphs.

Although there are previous studies that have used TCD to analyze brain activation during the exposure to VR settings [13, 14], they have only considered global differences between periods of repose and periods of exposure to the VE. The results of those studies have shown that there is an increment in the mean BFV during the exposure period in comparison with the mean BFV during the previous repose period.

In the present study, this increase in BFV during the navigation is also observed. The results show a significant difference between the mean BFV during the initial repose period and the mean BFV during the first navigation period, both in MCA-L and MCA-R, which supply mainly the lateral parts of the brain [21]. As has already been discussed in the previous studies [13-14], the origin of this increment can be found in different factors, such as the complex interaction between visuospatial interaction tasks, attentional tasks, and the creation and execution of a motor plan. The users have an active role in the navigation because they are creating a motor plan [22], but this role cannot be observed during the baseline, so it may be one of the factors that justify the increase in BFV that is observed when the navigation starts. All these results are in accordance with studies about navigation in videogames using TCD [23-24].

However, in the present study, we are interested in the evolution of the BFV during the exposure, with the objective of analyzing which are the effects on BFV that can be generated by any attentional changes that occur during this period. In order to have controlled experimental conditions, and following previous experimental approaches [17-18], the influence of the two sudden ruptures in the visual stream that were forced during the experience has been analyzed.

Results show that BFV remains similar during the Free Navigation A, during the Rupture A and during the Free Navigation B (no significant difference in mean BFV between those periods is observed). There is only a trend to a decrease in mean BFV during the rupture. A possible explanation for this trend can be found in the fact that, when the rupture occurs, participants of the experience stop focusing on the VE and become aware that they are in a laboratory participating in an experimental procedure. That generates an interruption in all the tasks that were happening during the navigation (such as visuospatial interaction tasks, attentional tasks, and the creation and execution of a motor plan). Furthermore, participants stop making movements with their arms to control the joypad. All these factors may justify the decreasing trend in BFV that is observed in both vessels.

On the other hand, when the Rupture B occurs (and a mesh partially occludes the VE), there is a new significant increment in BFV in both hemispheres. After that, BFV remains similar in the rest of the experiment in the case of MCA-R, with no significant differences. In the case of MCA-L, there is still a significant decrement when the normal navigation is restored.

Let us discuss possible explanations for the increment in BFV that has been observed during the Rupture B. Although maybe initially the participant may change his/her focus of attention from the virtual to real world, there is not a complete separation from the VE. In fact, the VE remains visible although partially occluded by the mesh. The participant keeps on trying to advance with the joypad, so movements with the arms and hands to control it continue, but with a higher difficulty because the environment is not visualized correctly. More attention than in the previous period is required during the Rupture B, and that may be having an influence on the increase in BFV during this period.

Further studies would be required to obtain more detailed conclusions about the effects of the change of the focus of attention from the virtual to the real world. Globally, with the results of the present study, the preliminary conclusions are that an abrupt interruption of the visual stream (that completely eliminates the visualization of the VE) does not have a high impact in BFV, and only a trend to a decrease is observed. However, if the changes in the visual projections still allow that the participant visualizes part of the VE, significant increases in mean BFV are observed, maybe because a higher level of attention is required to continue the navigation.

It would be interesting to continue analyzing in future studies the influence on BFV that any changes of attention during the exposure to a VE may have. Using TCD, it is possible to obtain information closely related to the processes that occur in the brain in response to those changes. This kind of information will help us to understand any changes in BFV that occur during the exposure to VR environments in the AugCog field.

Acknowledgments. The authors would like to thank José Lores, from Almevan, S.L., for his support with the transcranial Doppler equipment used in this work.

References

1. Aaslid, R., Markwalder, T.M., Nornes, H.: Noninvasive Transcranial Doppler Ultrasound Recording of Flow Velocity in Basal Cerebral Arteries. J. Neurosurg. 57, 76–774 (1982); Smith, T.F., Waterman, M.S.: Identification of Common Molecular Subsequences. J. Mol. Biol. 147, 195–197 (1981)
2. Iadecola, C.: Regulation of the Cerebral Microcirculation during Neural Activity: Is Nitric Oxide the Missing Link? Trends Neurosci. 16, 206–214 (1993)
3. Risberg, J.: Regional cerebral blood flow in neuropsychology. Neuropsychologia 24, 135–140 (1986)
4. Daffertshofer, M.: Functional Doppler testing. In: Hennerici, M., Meairs, S. (eds.) Cerebrovascular ultrasound, pp. 341–359. Cambridge University Press, Cambridge (2001)
5. Stroobant, N., Vingerhoets, G.: Transcranial Doppler Ultrasonography Monitoring of Cerebral Hemodynamics during Performance of Cognitive Tasks: A Review. Neuropsychol. Rev. 10, 213–231 (2000)
6. Duschek, S., Schandry, R.: Functional Transcranial Doppler sonography as a Tool in Psychophysiological Research. Psychophysiology 40, 436–454 (2003)
7. Warm, J.S., Matthews, G., Tripp, L., Hancock, P.A.: Cerebral Hemodynamics and Brain Systems in Vigilance. In: Schmorrow, D.D. (ed.) Foundations of Augmented Cognition, pp. 707–708. Lawrence Erlbaum Associates, Mahwah (2005)
8. Warm, J.S., Parasuraman, R.: Cerebral Hemodynamics and Vigilance. In: Parasuraman, R., Rizzo, M. (eds.) Neuroergonomics. The Brain at Work, pp. 146–158. Oxford University Press, New York (2007)
9. Schnittger, C., Sönke, J., Anouschen, A., Münte, T.F.: Relation of Cerebral Blood Flow Velocity and Level of Vigilance in Humans. Cognitive Neurosci. and Neuropsychol. 8, 1637–1639 (1997)
10. Helton, W.S., Hollander, T.D., Warm, J.S., Tripp, L.D., Parsons, K., Matthews, G., Dember, W.N., Parasuraman, R., Hancock, P.A.: The abbreviated vigilance task and cerebral hemodynamics. J. Clin. and Exp. Neuropsychol. 29, 545–552 (2007)
11. Hitchcock, E.M., Warm, J.S., Matthews, G., Dember, W.N., Shear, P.K., Tripp, L.D., Mayleben, D.W., Parasuraman, R.: Automation Cueing Modulates Cerebral Blood Flow and Vigilance in a Simulated Air Traffic Control Task. Theor. Issues in Ergon. Sci. 4, 89–112 (2003)
12. Rey, B., Alcañiz, M., Naranjo, V., Tembl, J., Parkhutik, V.: Transcranial Doppler: A Tool for Augmented Cognition in Virtual Environments. In: Schmorrow, D.D., Estabrooke, I.V., Grootjen, M. (eds.) FAC 2009. LNCS, vol. 5638, pp. 427–436. Springer, Heidelberg (2009)

13. Alcañiz, M., Rey, B., Tembl, J., Parkhutik, V.: A Neuroscience Approach to Virtual Reality Experience Using Transcranial Doppler Monitoring. Presence, Teleoperators & Virtual Environments 18(2), 97–111 (2009)
14. Rey, B., Alcañiz, M., Tembl, J., Parkhutik, V.: Brain Activity and Presence: a Preliminary Study in Different Immersive Conditions using Transcranial Doppler Monitoring. Virtual Reality 14(1), 55–65 (2010)
15. Slater, M., Brogni, A., Steed, A.: Physiological responses to breaks in presence: A pilot study. In: Proceedings of the 6th Annual International Workshop on Presence (2003)
16. Slater, M., Steed, A.: A virtual presence counter. Presence: Teleoperators & Virtual Environments 9, 413–434 (2000)
17. Garau, M., Friedman, D., Widenfeld, H.R., Antley, A., Brogni, A., Slater, M.: Temporal and spatial variations in presence: Qualitative analysis of interviews from an experiment on breaks in presence. Presence: Teleoperators & Virtual Environments 17, 293–309 (2008)
18. Slater, M., Guger, C., Edlinger, G., Leeb, R., Pfurtscheller, G., Antley, A., Garau, M., Brogni, A., Friedman, D.: Analysis of physiological responses to a social situation in an immersive virtual environment. Presence: Teleoperators & Virtual Environments 15, 553–569
19. Ringelstein, E.B., Kahlscheuer, B., Niggemeyer, E., Otis, S.M.: Transcranial Doppler sonography: Anatomical landmarks and normal velocity values. Ultrasound in Medicine and Biology 16, 745–761 (1990)
20. Sitzer, M., Diehl, R.R., Hennrici, M.: Visually evoked cerebral blood flow responses: Normal and pathological conditions. J Neuroimaging 2, 65–70 (1992)
21. Angevine, J.B., Cotman, C.W.: Principles of neuroanatomy. Oxford University Press, New York (1999)
22. Holden, M.K., Todorov, E.: Use of virtual environments in motor learning and rehabilitation. In: Stanney, K.M. (ed.) Handbook of Virtual Environments: Design, Implementation, and Applications, pp. 999–1026. Lawrence Erlbaum Assocaites, Mahwah (2002)
23. Kelley, R.E., Chang, J.Y., Scheinman, N.J., Levin, B.E., Duncan, R.C., Lee, S.C.: Transcranial Doppler assessment of cerebral flow velocity during cognitive tasks. Stroke 23, 9–14 (1992)
24. Vingerhoets, G., Stroobant, N.: Lateralization of cerebral blood flow velocity changes during cognitive tasks: A simultaneous bilateral transcranial Doppler study. Stroke 30, 2152–2158

Neuroergonomic Assessment of Simulator Fidelity in an Aviation Centric Live Virtual Constructive (LVC) Application

Tom Schnell[1], Alex Postnikov[2], and Nancy Hamel[3]

[1] Operator Performance Laboratory (OPL), University of Iowa, Iowa City, Iowa, 52242, USA
[2] Advanced Technology Center, Rockwell Collins, 400 Collins Rd., Cedar Rapids, IA 52498, USA
[3] Advanced Infoneering, Inc., 433 Hwy 1 W, Iowa City, IA 52246, USA
thomas-schnell@uiowa.edu, apostnik@rockwellcollins.com, nhamel@advancedinfoneering.com

Abstract. This paper describes a recent human factors study that was performed on a flight simulator and in a fighter trainer jet aircraft to quantify the cognitive effects of simulator fidelity. There are many parameters that could be manipulated to affect physical fidelity in a simulator and we want to point out that in this study we make no claims of having covered a large portion of the possible fidelity design space. Rather, this study provides a comparison of trainee performance in a low to mid-level simulator with the performance obtained in a real fighter jet training aircraft using state-of-the-art operator state characterization equipment. As this study is ongoing, only partial data is shown in this paper.

Keywords: Neurocognitive measures, operator state characterization, flight training.

1 Introduction

This paper describes a human factors study, which was funded by the Office of Naval Research (ONR) as part of a Small Business Technology Transfer Research (STTR) program at Advanced Infoneering, Inc. [1]. The study involved the measurement of neuroergonomic parameters including eye gaze behavior, electroencephalogram (EEG), heart rate variability (HRV), and mission specific measures of performance in pilots performing a close air support (CAS) task using a pop-up bombing maneuver in a simulator and a real fighter jet trainer aircraft.

The technology that was developed under this STTR will find application in civilian and military flight training and technology testing applications. One application that is especially well-suited for neuroergonomic performance assessment technology is the emerging area of Live Virtual Constructive (LVC) training. LVC is a relatively new paradigm in aviation training that has considerable potential to revolutionize the way aviators are trained and prepared for their missions. LVC incorporates live aircraft, virtual simulators, and constructive entities into a single

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 221–230, 2011.

environment that provides training participants with an opportunity to interact the same way how they would interact when performing their real missions in theater. Live aircraft are connected to the network of ground-based simulators using high-bandwidth digital datalinks and dedicated robust data protocols [2]. In this fashion, LVC not only supports the training of the pilots in the live aircraft and flight simulators, but also the training of other participants, such as airborne and ground-based controllers and their support teams including Joint Terminal Attack Controllers (JTACs) and Joint Forward Observers (JFOs). Rockwell Collins in collaboration with the Operator Performance Laboratory (OPL) recently demonstrated the huge training potential of LVC by enabling a real JTAC to receive LVC training to regain night currency during a demonstration at the 2010 Interservice Industry Training Simulation Education Conference (I/ITSEC) in Orlando [3]. During this training for credit, the JTAC in training controlled the OPL jet aircraft flying in Iowa. From this JTAC training station in Orlando he prosecuted a simulated close air support mission against virtual targets that were overlaid in the real world.

Naval aviation flight training is performed using a combination of procedure trainers, flight simulators and live aircraft [4]. While flight simulators and procedure trainers have become very capable and flexible, there are still many skills that naval aviators need to acquire in live training and fleet aircraft. However live flight training can be very costly and logistically difficult to accomplish and the current fleet of aircraft is fairly thinly stretched across training and war fighting operations. LVC is an integration concept that incorporates live, virtual, and constructive elements into a single environment, in an attempt to leverage the best of each world to minimize logistics and maximize training effectiveness [2, 5]. What makes LVC so attractive is its ability to connect airborne and ground-based assets in a net centric training exercise that can be geographically distributed [6].

The overarching objective of LVC training is to improve the effectiveness of the delivery of content while at the same time achieving a reduction in operational costs and enhanced flight safety. Cost-effective delivery of instruction will be enabled by the inherently embedded and net centric capability of LVC [2, 5], where the reduction is demonstrated by a smaller number of training flights required to complete the tactical tasks and component skills called for in the training syllabus and also through a reduction in the required number of flights to provide live opposing force necessary for readiness training of another pilot [7]. LVC also requires the development and indoctrination of new concepts of operation (CONOPS) including methods of exercise planning, briefing, air traffic and range control, rules of engagement (ROE), handling of emergencies, performance evaluation, and debriefing. Our team has performed LVC research and demonstrations for the past two years [2, 6], and we feel that LVC has great potential to reduce cost by reducing or eliminating logistical complexities and enhancing training effectiveness by enabling early immersion of trainees in complex net centric distributed exercises that draw on many dimensions of the cognitive-perceptual stimulation that is necessary to prepare our warfighters for effective operation in theater.

Designers of virtual environments such as flight simulators are faced with difficult cost-benefit trade-offs that may affect its fidelity and its training effectiveness or transfer of training. The construct of fidelity has several dimensions, including physical fidelity, functional fidelity, and cognitive fidelity. Interaction of different fidelity dimensions have an impact on trainee immersion, presence, and buy-in [8]. In flight simulators, physical fidelity relates to the accuracy of the physical layout of the crew station and how closely the visual, auditory, haptic, vestibular, and flight dynamic stimuli mimic those that will be experienced in the real aircraft. Functional fidelity primarily relates to how accurately the simulated crew station equipment acts like the operational equipment and cognitive fidelity is a quantification of how closely the human factors effects of the virtual environment track with those that will be found when training in the real aircraft.

This paper describes the synergistic combination of our recent developments in aviation LVC technology in conjunction with a human factors study to specifically investigate the cognitive effects of simulator fidelity. The Operator Performance Laboratory (OPL) has two L-29 jet training aircraft [6], each modified with an evaluation cockpit in the rear seat, integrated instrumentation pods, a ground support infrastructure, and a neuroergonomic operator monitoring and evaluation system [9, 10]. Additionally, the OPL has developed a matching ground simulator with a functionally identical simulated avionics set up and the same operator monitoring system. The avionics, datalink, and LVC concept of operations work at OPL was funded over the past two years by Rockwell Collins.

2 LVC Research Apparatus

Our current LVC infrastructure consists of two L-29 jet training aircraft (Fig. 1) and two flight simulators, one being of a fast jet factor and the other one being of a transport aircraft form factor. Each aircraft is instrumented with an evaluation cockpit in the rear seat, integrated range instrumentation pods, a ground support infrastructure, and an operator monitoring and evaluation system. A third, piston powered, aircraft is available for use as data link relay and/or as an airborne command-and-control platform. The flight test assets are interconnected to a ground station using a range instrumentation, datalink that can transmit in several formats, including the Advanced Range Data System (ARDS) protocol. A command-and-control ground station with two high gain pan-tilt rotator systems is located at the OPL flight operations center at the Iowa City municipal airport. This ground station provides the interconnection between the airborne and ground based assets. All ground assets communicate with each other using the HLA protocol. By using this constellation of airborne and ground-based assets, we can test the performance of multiple crews in an LVC exercise.

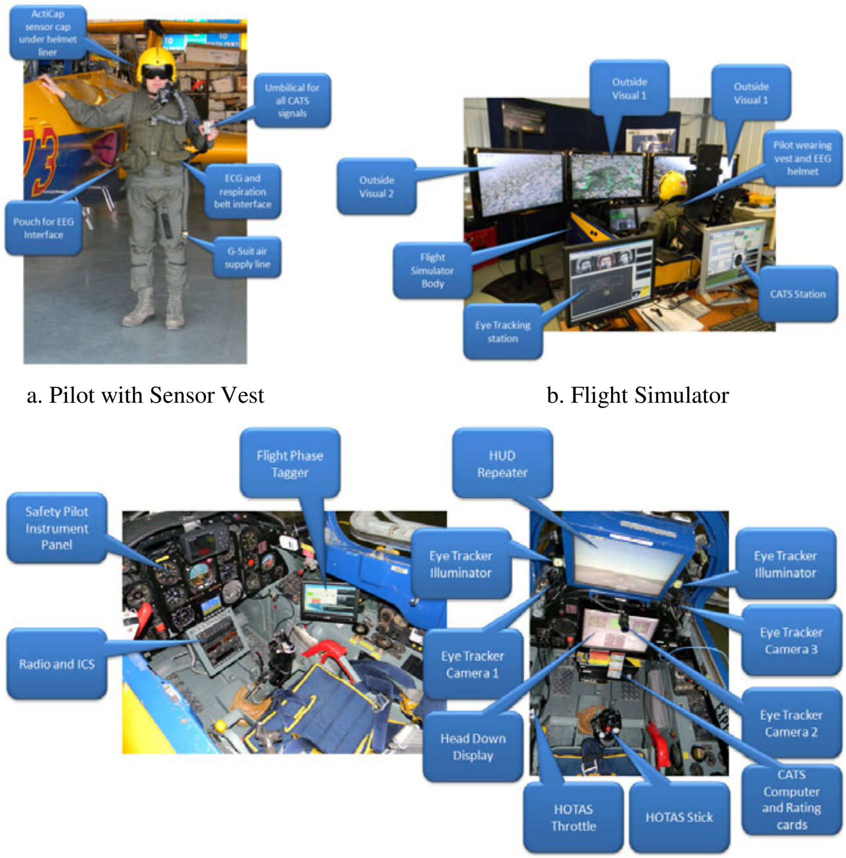
To simplify the deployment of the neurocognitive and physiological sensors on the pilot we have integrated the EEG electrodes in the liner of a flight helmet. The respiration belt and ECG electrodes are worn under the flight suit connecting to the peripheral electronics that are integrated in a pilot survival vest as shown in Fig. 2a. This level of integration provides for a ruggedized instrumentation package with a single point umbilical connection to the aircraft or flight simulator. Fig. 2b shows a rear quartering view of the fixed base flight simulator that was used in this study. The flight simulator features three channels of outside visuals, subtending a total of 135° lateral visual field of view (FOV) or around 45° per channel and a vertical field of view of 25 degrees. The outside visual (OSV) channels 2 and 3 were used to manipulate the fidelity of the flight simulator with low fidelity corresponding to the condition where OSV 2 and 3 were off and medium fidelity when OSV 2 and 3 were on. In the present study, high fidelity corresponded to runs in the L-29 jet.

Standard F/A-18 head up display (HUD) symbology was overlaid on OSV channel 1, providing the participant with symbology to fly a bombing run with a Continuous Computed Impact Point (CCIP) for Mark-82 dumb bombs. The head-down display (HDD) showed the layout of an F/A-18 instrument panel with Stores Management System (SMS), map page, Up Front Controls (UFC), and a Horizontal Situation Indicator (HSI) page. The HDD is a touch screen so that all SMS and UFC functions can be activated by touch.

Hawk 11 Hawk 12 Range Integration Pod

Fig. 1. OPL's Instrumented L-29 Fighter Jet Trainer Aircraft

Fig. 2c shows the front cockpit of the L-29 jet where the safety pilot (SP) operates. The SP performs all maneuvering on the ground, take-off, landing, and repositioning of the aircraft between runs. The SP uses standard aircraft instruments to navigate in US airspace under FAR part 91 flight rules. Two VHF radios are available to allow the SP to simultaneously communicate with air traffic control (ATC) and the command and control ground station on separate frequencies. A side display touch screen called the Phase Tagger (Fig. 2c) is available to the SP to start and stop the recorder, tag events to check the video data link integrity, and to check CATS and the integrity of the eye tracker. The rear cockpit is the crew station that the evaluation pilot (EP, experiment participant) occupies. A daylight readable 15 inch touch screen display installed in the head-down position that allows presentation of any avionics symbology as per program requirements. The symbologies can be driven either with PC board dedicated avionics graphics processors. In this experiment, the symbologies were identical to the ones used in the simulator and represented an F/A-18 instrument panel. A daylight readable 15 inch touch screen in the head-up display (HUD) position provided the same outside visuals and F/A-18 HUD symbology as in the simulator. The lateral FOV of the HUD display was 45° which made the imagery displayed on it conformal with the real world. Therefore, a pilot in the rear crew station had an essentially unrestricted view of the surroundings, with the central 45° being a computer generated photorealistic inset and the remaining view being the real world.



a. Pilot with Sensor Vest b. Flight Simulator

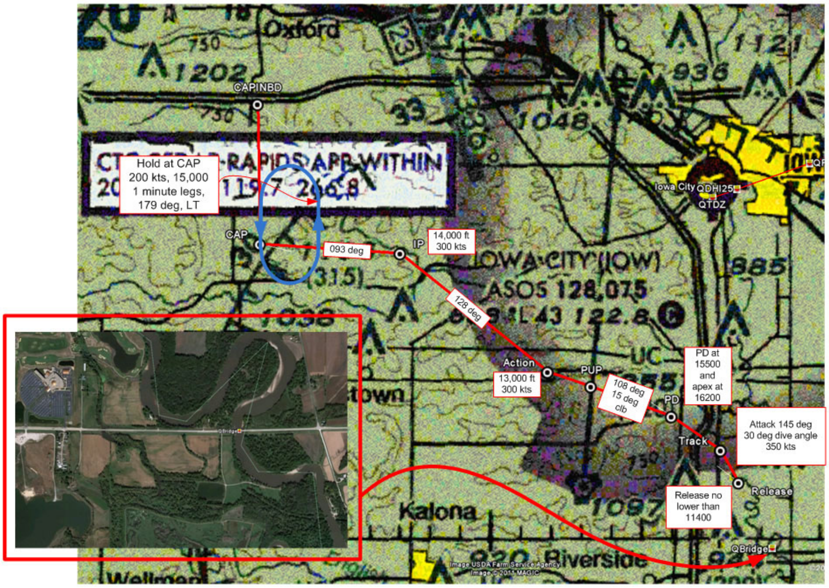
c. L-29 Front Cockpit d. Rear Cockpit

Fig. 2. Instrumented Flight Simulator and Matching L-29 Flight Test Aircraft

3 Experimental Design and Procedure

The authors of this paper fully understand that simulator fidelity is a very complex concept and that our simple experiment does not color the entire gamut of fidelity. We tested three levels of fidelity as a between subjects factor. Participants in the low fidelity group flew their mission in the flight simulator using only one channel of the outside visuals. Participants in the medium fidelity group flew their mission in the simulator using all three channels of visuals. Participants in the high fidelity group flew their missions in the jet. Each group consisted of five pilots who had no military tactical flight experience.

The mission consisted of a holding pattern at a combat air patrol (CAP) point followed by a series of waypoints leading to an offset pop-up bombing pattern with a 15° climb and a 30° dive angle to deliver a Mk-82 low drag general-purpose bomb onto a target represented by the middle of the bridge deck across a river (Fig. 3). We chose this mission profile because it has a wide range of perceptual, motor, and cognitive demands ranging from a simple holding pattern, a dynamic high-speed route with precise turns and crossing altitudes, culminating in a relatively complex pop-up bomb delivery pattern that requires precise management of pitch, bank, speed, and heading in a very short amount of time.



Mission Number	: 76Y	Aircraft: F/A-18 Loadout: 6xMk 82, 500 rds 20 MM	1. IP Alpha 2. 108, 2.2 Left 3. 9.6 4. 610
Sortie Number	:		
Pilot Call Sign	:____________	Target: Hwy 22 Bridge over Iowa River, East of Creekside Casino, 41° 29.769'N, 91° 30.658'W, 610	5. Bridge 6. 41° 29.769'N, 91° 30.658'W 7. None 8. 750 Meters W 9. Egress West to CAP
Date	:____________		

Fig. 3. Holding Pattern, Route, and Pop-Up Attack Pattern

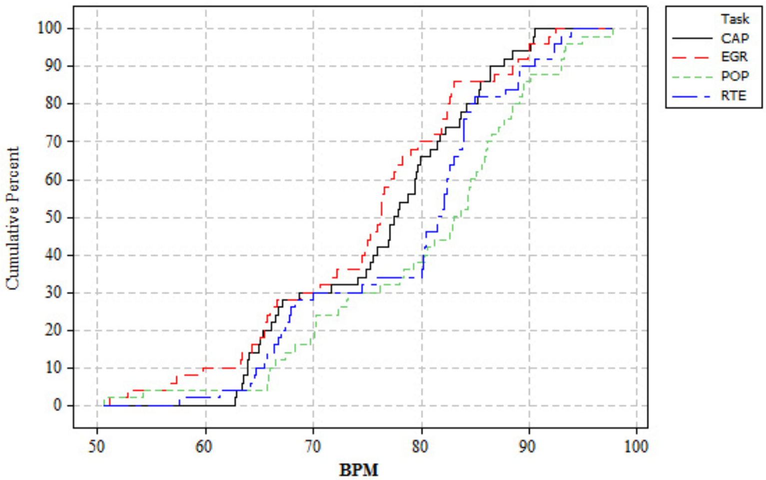
Normal pop-up weapons delivery is performed following a low-level route at 500 feet AGL. The pop-up itself actually serves to increase altitude to several thousand feet to allow for a reasonable amount of tracking time in a dive during which the bomb guidance symbology can be tracked against the desired target. In this experiment, we did not fly the route at a low level for reasons of flight safety and compliance with the required speed limit of 250 kts below 10,000 feet. Rather, we started the rout at 15,000 feet gradually descending to 13,000 feet just prior to the pop, with apex altitude of 16,200 feet, 6 seconds of tracking time and the release altitude at or above 11,400 feet. Fig. 3 shows the mission profile that was flown identically in the simulator and the real aircraft. Each pilot was given simulator training to acclimate to the flight symbology and to learn the mechanics of the route and pop. A total of 10 minutes of simulator training was provided to allow the pilot to acclimate the flight symbology. Following that, a total of 30 min. of simulator training was provided to teach the participants the basics of the route following and pop-up bombing pattern. This training was always administered with only the center channel available in the flight simulator. After training was complete, data collection was performed for score. The flight simulator groups performed their runs using either one or three channels of OSV, depending on which fidelity group they belong to. The jet group always performed their runs in the real world in the L 29 jet.

The data was analyzed with CATS that has provisions to access the tags that are placed by the experimenters throughout the runs to be able to separate the holding pattern from the route and the pop-up bombing pattern. CATS generated over 60 mission technical, neurocognitive, and physiological performance metrics. Mission technical performance was quantified in terms of the flight path accuracy (speed, offset angle, pull-up-point, climb angle, pull-down-point, apex altitude, dive angle, tracking time, release altitude, and accuracy of the weapon delivery). The physiological assessment consisted of six metrics of heart rate and short-term heart rate variability. Additionally, eye gaze metrics of performance included fixation duration, fixation count within areas of interest, lateral and vertical fixation dispersion, and distance between fixations. CATS generated over 65 neurocognitive metrics based on the average, RMS, and standard deviation of EEG power in the frontal, midline, occipital, and sensorimotor areas. Subjective workload data was collected after each pop using the Bedford workload scale and situation awareness data was collected using the SART scale.

4 Preliminary Results

Data analysis is still ongoing at the time of writing of this paper (March 1, 2011). Data for all pilots in the flight simulator has been collected at this time. Flight test data collection has been slow due to adverse winter weather, but we are making good progress with data for three subjects in the jet group already being collected. As the data analysis is continuing we are going through dozens of combinations of the dependent measures to determine which ones are statistically significantly able to predict the level of pilot workload. Preliminary results are shown in this paper for the flight simulation groups. One exciting finding is that heart rate (Fig. 4) is highly predictive of workload for the task used in this experiment.

The cumulative histograms in Fig. 4 shows that heart rate increases for increasing task demand. This is indicated by the right shift of the curves towards hire beats per minute numbers. Holding at CAP demands the least workload from the pilot. Flying the route requires considerable concentration and demands a significant amount of workload from the pilot to precisely cross the waypoints at the assigned altitudes and speeds. Flying the pop-up bombing maneuver requires very precise pull-up timing, accurate flight path angle control in the climb with simultaneous tracking of the offset heading, monitoring of the approaching pull-down altitude, proper selection of the bank angle (about 135°) at the pull-down point, a sufficient pull to achieve the correct heading change to the final attack heading and dive angle on the pitch ladder, and precise dive angle and final attack heading tracking of the CCIP bomb fall line (BFL) onto the target with a bomb release at wings level and at or above, release altitude. This entire sequence takes around 40 seconds to complete, and failure to accomplish any of the sub tasks is likely to make it impossible to achieve proper tracking and bomb release.



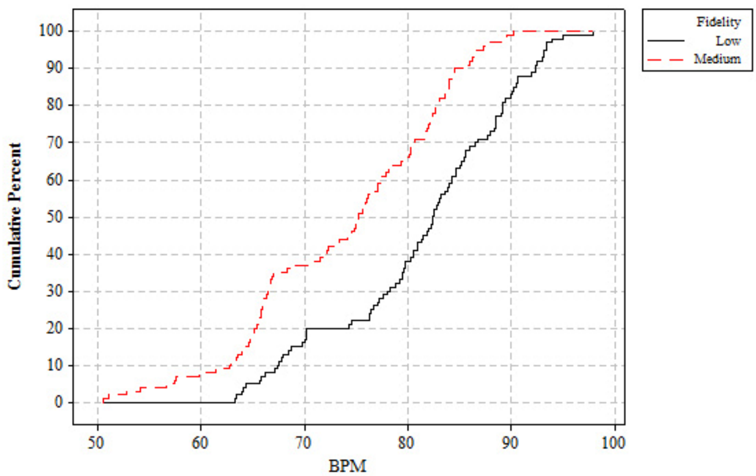
Note:N=10 participants in Flight Simulator. CAP=low workload, RTE=Medium workload, POP=High workload, EGR=Low workload. Repeated measures Anova for Task $F_{3,27}$=17.36, p<0.0001

Fig. 4. Heart Rate as a Function of Mission Task Difficulty

The egress following the bomb release consists of a simple pull to a 10° flight path angle while tracking towards the egress waypoint. This demand is well characterized by the simple heart rate (beats per minute) as indicated in Fig. 4. What is rather astounding is how quickly heart rate responds to increases and reductions of workload demand as indicated by the egress (EGR), heart rate curve. The egress follows only seconds after the pop-up bombing segment has been completed, yet the heart rate responds correctly. A repeated measures analysis of variance (ANOVA) of heart rate against task indicates statistical significance with $F_{3,27}$=17.36, p<0.0001.

Fig. 5 shows the heart rate of the pilots as a function of the level of fidelity, with low fidelity being representative of the five pilots who flew the flight simulator with only one channel of the outside visuals and medium fidelity being representative of the five pilots who flew the simulator with all three visual channels. The higher heart rate for the lower fidelity simulation clearly indicates that the pilots had a higher workload when only one channel of outside visuals was present.

Using the very large data set that we have amassed in this study we are going to continue to use statistical, data mining, and neural network methodologies to find the best combination of physiological, neurocognitive, and flight technical performance metrics in an effort to create a robust model to predict workload. Going forward, we propose to use the existing LVC framework consisting of flight simulators (one at OPL, two at Rockwell Collins), two instrumented fighter trainer jet aircraft, and a JTAC training station to quantify neuroergonomic measures of effectiveness in pilots performing in multi-participant LVC exercises.



Note:N=10 participants in Flight Simulator. CAP=low workload, RTE=Medium workload, Low Fidelity = 1 Channel OSV, Medium Fidelity = 3 channels, Anova for Task $F_{1,198}=33.71$, $p<=0.0001$

Fig. 5. Heart Rate as a Function of Fidelity Level

5 Conclusions

The CATS neurocognitive, physiological workload measurement package described in this paper has performed very well in our flight simulator and instrumented fighter jet trainer. State-of-the-art active shielding electrodes have helped us to mitigate the effects of adverse noise and signal acquisition. In our experiment we have demonstrated that this package can be rapidly deployed on the pilot was performing I dynamics tactical maneuvering in the real fighter jet training aircraft. Perhaps the most significant conclusion of this paper is that heart rate appears to be a reliable, yet simple method to characterize pilot workload demand.

References

1. Schnell, T.: Tools for Virtual Environment Fidelity Design Guidance: Quality of Training Effectiveness Assessment (QTEA) Tool, in Office of Naval Research, ONR Code 00, Office of Naval Research, ONR Code 00: Arlington, VA (2008)
2. Hoke, J.: Live Virtual Constructive Training. In: Proceedings of MODSIM World 2010 Conference, Human Dimension Track (2010)
3. Walker, K.: An exercise in realism (2011), `http://www.tsjonline.com/story.php?F=5503217`
4. CNATRA. CNATRA 21 Strategic Vision (2010) `https://www.cnatra.navy.mil/docs/cnatra_21_vision.pdf`

5. DoD, Department of Defense Directive on Military Training, Department of Defense, Washington, DC (2009)
6. Drake, D.L., et al.: Improving the Immersive Environment in the Virtualized Cockpit. In: Proceedings of the Fall Simulation Interoperability Workshop (SIW), Florida Mall Conference Center, Orlando, Florida (2009)
7. Sheehan, J., et al.: Human System Capabilities-Based Training System Acquisition in Naval Aviation. In: Proceedings of the Human Systems Integration Symposium 2009. American Society of Naval Engineers (2009)
8. Alexander, A., Brunye, L., Weil, T.: From Gaming to Training: A Review of Studies on Fidelity, Immersion, Presence, and Buy-in and Their Effects on Transfer in PC-Based Simulations and Games (2005); `http://www.darwars.org/downloads/DARWARS%20Paper%2012205.pdf`
9. Schnell, T., Melzer, J.E., Robbins, S.J.: The cognitive pilot helmet: enabling pilot-aware smart avionics. In: Head- and Helmet-Mounted Displays XIV: Design and Applications, SPIE, Orlando, FL, United states (2009)
10. Schnell, T., Cornwall, R., Walwanis, M., Grubb, J.: The quality of training effectiveness assessment (QTEA) tool applied to the naval aviation training context. In: Schmorrow, D.D., Estabrooke, I.V., Grootjen, M., et al. (eds.) FAC 2009. LNCS, vol. 5638, pp. 640–649. Springer, Heidelberg (2009)

Brain Activity of Young and Adult Hebrew Speakers during Lexical Decision Task: fNIR Application to Language

Itamar Sela[1,2], Tzipi Horowitz–Kraus[2], Meltem Izzetoglu[1], Patricia A. Shewokis[1], Kurtulus Izzetoglu[1], Banu Onaral[1], and Zvia Breznitz[2]

[1] School of Biomedical Engineering, Science & Health Systems, Drexel University, Philadelphia, PA

[2] The Edmond J. Safra Brain Research Center for the Study of Learning Disabilities, University of Haifa, Haifa, Israel

Abstract. The process of reading activates a large-scale neural network which includes different cortical brain regions. This network is thought to be age-dependent and changes throughout the process of reading acquisition. The frontal lobe is considered to be related to higher, executive, functions. We conducted a functional Near InfraRed Spectroscopy (fNIR) study in order to compare frontal lobe performance during a Lexical Decision Task (LDT) among two different age-groups: children and adults. Data indicated significant differences with age in LDT behavioral performance, and brain activity in the upper left frontal lobe. The young group exhibited slower reaction times and lower accuracy in addition to differences both in the level of blood oxygenation as well in the blood oxygenation timeline. The current study's results suggest 1) the involvement of the frontal lobe during the process of reading and that 2) frontal lobe activity is modified with the age of maturity.

Keywords: Neuroimaging, fNIR, Lexical Decision Task, Developmental language.

1 Introduction

One of the common methods of investigating neural networks related to language is the Lexical Decision Task (LDT) [1-7], which involves the identification of words and pseudowords. It has been suggested that word identification occurs through orthographic processing and pseudoword identification through phonological processing [8]. As such, this task is often used in an attempt to investigate the developmental and impaired aspects of word decoding processes. Previous neuroimaging research suggested that several brain regions are involved in LDT performance, including the superior and the inferior frontal lobe [1, 5, 6, 9]. According to these studies, the superior frontal lobe is involved in the process of decision making [3], where an input from more posterior brain regions involved in semantic information processing (mainly the inferior parietal lobe, specifically the angular gyrus) evokes a positive intra-lexical decision response following a word

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 231–239, 2011.

stimulus, or a negative response which is triggered by an extra-lexical temporal threshold [5, 10]. The inferior frontal lobe is thought to be more involved in differentiating between frequent and non-frequent words [6, 9].

fNIR is an emerging, non-invasive brain-imaging technology that allows for the measurement of hemodynamic changes within the brain. It is a portable, affordable, and easy-to-use device that is considered to be more tolerant to movement artifacts as compared to other brain imaging modalities such as electroencephalography (EEG) and functional Magnetic Resonance Imaging (fMRI). Due to these many attractive attributes, fNIR has become commonly used in various areas of cognitive research [5, 11-14]. Specifically, several studies focused on the involvement of the frontal lobe in different aspects of language. For example, Sakatani et al. [15] used near infrared spectroscopy to show the effect of aging on the left prefrontal cortex activity during a series of lingual and memory tasks. Quaresima et al. [14] reported on the involvement of the left Broca in the process of language translation task. Watanabe et al. [16] correlated between language dominance and handedness. In addition, Hofmann et al. [5] used near infrared spectroscopy to demonstrate the involvement of the left superior and inferior frontal lobe in the performance of LDT. Although these studies support the notion that the left frontal lobe is involved in language, the role of the frontal lobe in the process of reading acquisition still remains unclear.

The purpose of the current study was to find whether, and to what extent, the frontal lobe is involved in the performance of the LDT. Specifically, we investigated whether there are age-related differences, in the frontal lobe activation during the performance of the task.

2 Method

Participants: Twenty-two adults (age 25.1±2.48, 9 females and 13 males) and 25 7th grade children (age 12.65±0.467, 13 females and 12 males), participated in the study. All subjects fall into the criteria of a regular reading definition based on Standard Hebrew Reading Test. The decoding score was 92.98±36.52 for word and 47.33±23.40 for pseudowords per minute for the adults and 76.96±21.05 for word and 31.27±10.30 for pseudowords per minute for the young group ($F_{(1,46)}=4.012$, $p<0.05$; $F_{(1,46)}=10.57$, $p<0.01$, for words and pseudowords, respectively). The adults were paid volunteers and the teenagers were compensated with a gift at school. All participants had nonverbal IQs in the normal range (100 and above) as measured by the Raven Standard Progressive Matrices [17]. All participants were native Hebrew speakers from a middle-class background. All subjects were right-handed, displayed normal or corrected-to-normal vision in both eyes, and were screened for normal hearing. None of the participants reported chronic use of medications. Informed consent approved by the University of Haifa ethics committee was obtained prior to each participant's participation in the study.

Apparatus: Two computers were employed. The first computer presented the LDT stimuli via ePrime software (Psychology Software Tools, Inc. http://www.pstnet.com) and collected the participants' reaction times. The second computer hosted the fNIR system (fNIR Devices LLC; http://www.fnirdevieces.com). The fNIR device used in this study was composed of two main parts- a head piece holding the light sources and detectors, and a control box for data acquisition with a sampling rate of 2 Hz. The flexible fNIR sensor consists of four light sources and ten detectors designed to image cortical areas underlying the forehead (dorsolateral and inferior frontal cortices). With a fixed source-detector separation of 2.5 cm, this configuration results in a total of 16 voxels. The control box was connected to the computer for data collection and storage which were utilized by the COBI studio software (Drexel University). In order to synchronize the two computers, a COM cable was used to send online event triggers from the ePrime software to the COBI studio software. Matlab software (Version 2010a, The Mathworks, Natick, MA) was used for the signal processing and to prepare data for statistical analysis which was performed using IBM SPSS (Version 18, IBM SPSS Inc., Chicago, IL).

Task: The Lexical Decision Task [18] included 96 trials, of which 48 trials included high frequency words in the Hebrew language [19] and the remaining 48 trials included pseudowords created from the same letters as the real words. The stimuli were presented for 400 ms horizontally in the center of the screen in white on a gray background. Each stimulus was comprised of 4-5 Hebrew letters, each letter one-quarter of an inch (0.6 cm) in diameter. The participants were seated approximately 80 cm from the computer screen and were asked to press with their right hand '1' for word and '2' for pseudowords. The between-trials time interval was set to 10 seconds with a jitter of ± 4 seconds to allow sufficient time for the hemodynamic response to fully evolve [20, 21]. An fNIR resting baseline of 10 seconds was recorded prior to the performing of the LDT, which was used as a reference in the computation of the relative blood oxygenation changes [18].

Behavioral Reaction Extraction: The reaction time of the trial was defined as the time starting from the stimulus onset until participant’s reaction is received. Reaction time and accuracy for each trial were first obtained from the LDT log files. Then, for each participant, mean reaction times for words and pseudowords and for correct and incorrect reactions were calculated. Due to the ceiling effect obtained in the LDT, only correct reaction trials were used in the statistical analysis.

fNIR Data Processing and Feature Extraction: Once the heart pulsation, respiration and movement artifacts were removed, fNIR intensity measurements were first converted to relative changes in hemodynamic responses in terms of oxygenated (OxyHb) and deoxygenated hemoglobin (DeoxyHb) using the modified Beer-Lambert law (MBLL) [22]. Note that, since there is an age difference between the two study groups, an age-dependent correction to the path length factor was integrated in MBLL to accurately extract the hemodynamic signals [14, 23]. Then, Oxygenation, which was defined as the subtraction of the DeoxyHb from the OxyHb, was computed. Finally, once oxygenation data epochs were segmented from the stimuli onset to 15 seconds later for each trial, fNIR features such as minimum, maximum and mean value, and time to reach minimum and maximum value, were extracted for each Oxygenation trial epoch, voxel and participant. For the statistical analysis, each parameter was averaged over trials per participant. Noisy segments, which mainly occurred due to movement artifacts, were excluded from the statistical analysis.

Statistical analysis: A series of 2 X 2 mixed model Analysis Of Variance (rmANOVA's) tests were conducted in order to verify age (children X adults) and stimulus-type (word X pseudoword) differences in the research parameters. In addition, in cases where significant age by stimulus-type interaction was found, appropriate t-tests were applied. The rmANOVA was applied separately to different variables including mean reaction time, accuracy, and fNIR features which were obtained from the Oxygenation data of 16 channels. Since fNIR is considered to have a low signal to noise ratio, in order to verify data integrity, in each of the analyses, the fNIR parameters were first tested for normal distribution using the Kolmogorov-Sminrnov test of normality. In cases where the test of normality failed, and outliers were found, the outliers were screened out from the analysis, and then the test of normality was run for the second time.

3 Results

The comparison of behavioral responses and fNIR features between the two age groups (children and adults) and stimuli (word and pseudoword) revealed both group differences as well as stimulus-dependent differences.

3.1 Behavioral Reaction Results

Reaction time (Fig. 1a): Results revealed a significant group effect ($F(1,37)=21.35$, $p<0.001$). The younger group exhibited a slower reaction as compared to the adults. A significant stimulus-type effect ($F_{(1,37)}=22.21$, $p<0.001$) was also obtained. Both groups exhibited a slower reaction time to the pseudoword stimulus as compared to the word. No group and stimulus-type interaction was found ($F_{(1,37)}=2.80$, $p=0.103$).

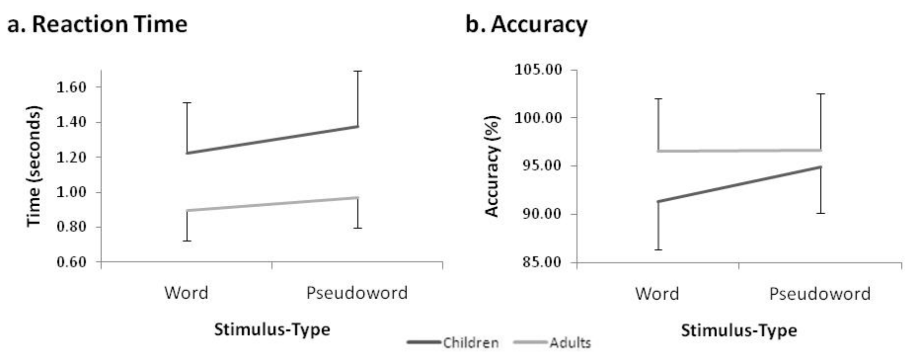


Fig. 1. The children (dark line) and adults (light line) mean (a) reaction time and (b) accuracy performance in LDT. Error bars represent group's standard deviation.

Accuracy (Fig. 1b): Data indicated a significant group effect ($F_{(1,35)}=29.143$, $p<0.001$). The adult group obtained higher accurate reactions as compared to the younger group. A significant stimulus-type effect ($F_{(1,37)}=8.686$, $p<0.01$) was also found, for both groups accuracy was higher for pseudowords as compared to words. Furthermore a significant group by stimulus-type interaction ($F_{(1,37)}=7.38$, $p=0.01$) was also found. The interaction stems from lower accuracy rate in words ($t_{(35)}=-5.24$, $p<0.001$) and not for pseudowords ($t_{(35)}=-1.925$, $p=0.062$) among the younger group.

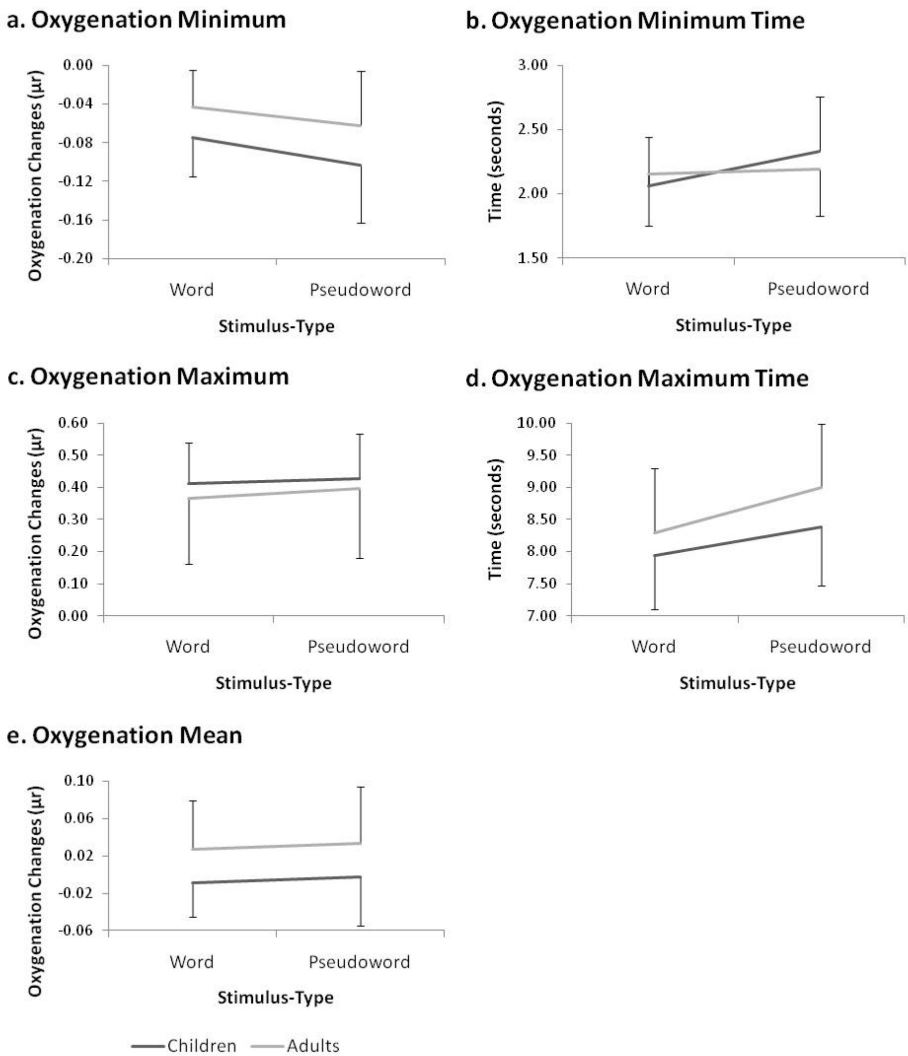


Fig. 2. The children (dark line) and adult (light line) Oxygenation results measured at Channel 3. Error bars represent group's standard deviation.

fNIR Results: The statistical analysis on the features extracted from oxygenation data epochs within the time interval of 15 seconds following the stimulus onset revealed that stimulus-type and age effects were found mainly in Channel 3 which is located in the upper side of the mid-left frontal lobe. Fig. 2 presents the results for the fNIR analysis.

Minimum value of Oxygenation (Fig. 2a): Data revealed significant group effect ($F_{(1,33)}$=9.522, p<0.01). The children exhibited a larger decline in the value of the Oxygenation as compared to the adults for both words and pseudowords. A significant stimulus-type effect was also found ($F_{(1,33)}$=4.139, p=0.050). For both groups a lower minimum value was found under the pseudowords condition as compared to words. No group and stimulus-type interaction was found ($F_{(1,33)}$=0.181, p=0.673).

Minimum-time in which the Oxygenation signal reached its lower value (Fig. 2b): revealed a non-significant group effect ($F_{(1,36)}$=0.094, p=0.760), non-significant stimulus-type effect ($F_{(1,36)}$=2.278, p<0.140), and no group by stimulus-type interaction ($F_{(1,36)}$=1.394, p=0.245).

Maximum value of the Oxygenation signal (Fig. 2c): revealed no significant group effect ($F_{(1,37)}$=0.552, p=0.462) and stimulus-type effect ($F_{(1,37)}$=1.800, p=0.188) as well as no group by stimulus-type interaction ($F_{(1,37)}$=0.231, p=0.634).

Maximum-time in which the Oxygenation signal reached its maximum value (Fig. 2d): indicated a significant group effect ($F_{(1,37)}$=4.650, p<0.05), the children reached the maximum value of Oxygenation faster as compared to the adults. In addition a stimulus-type effect was also obtained ($F_{(1,37)}$=8.370, p<0.01) where both groups exhibited a longer time to reach the maximum Oxygenation value under the pseudowords condition as compared to the word. No group by stimulus-type interaction ($F_{(1,37)}$=0.432, p=0.515) was found.

Mean Oxygenation value (Fig. 2e): Data revealed a significant group effect ($F_{(1,32)}$=10.077, p<0.01) where the adults group showed a higher mean value of Oxygenation. No stimulus-type effect ($F_{(1,32)}$=0.253, p=0.619) or group by stimulus-type interaction ($F_{(1,32)}$=0.000, p=0.992) were found.

4 Discussion

Overall, the behavioral results of the current study demonstrate an advantage of the adults in their performance of the LDT in terms of accuracy and reaction time as compared to the 7th graders group. Moreover, fNIR results reveal evidence for both the involvement of the left frontal lobe in the performance of the LDT as well as age related differences in terms of cortical oxygenation.

By using fNIR and behavioral measures the current data indicated a clear developmental trend in accuracy and reaction time as well as brain activity in the left frontal lobe during performance of LDT tasks. Compared to the adults, the young population exhibited significantly slower reaction time (Fig. 1a), and lower accuracy (Fig. 1b), for both words and pseudowords. Furthermore when performing the LDT the young population showed a higher decline in the upper left frontal lobe Oxygenation value (Fig. 2a), a faster Maximum-Time (Fig. 2d) and an overall lower mean Oxygenation value (Fig. 2e). The Oxygenation minimum value obtained shortly after the stimulus onset represents a fast reduction in the amount of oxygen in the sampled voxel [20, 21]. This was previously suggested to be related to an initial consumption of oxygen reserves within the voxel by its local neurons. Thus, it can be suggested that the lower minimum Oxygenation value during LDT performance that was exhibited by the younger group may represent higher neural activity in the upper left frontal lobe. Support for this notion can be also seen in the fact that the maximum value of Oxygenation emerge faster among the young participants and may represent a faster inflow of oxygenated blood into the sampled voxel. In sum, the longer processing time, higher rate of errors and the upper left frontal lobe activation seen in the fNIR parameters during the LDT among the younger readers as compared to the adults may suggest that the young group may need to invest more effort in an attempt to process the LDT. Although the young readers who took part in this study were all non-disabled readers at the beginning of secondary school and their reading performance was within the normal range it seems that for this population the process of distinguishing between words and pseudowords has not yet fully automatized and requires more mental effort than for mature readers.

Our data indicated that the young group of readers exhibited significantly more errors when processing words as compared to the adults and no significant differences were found between the two groups in pseudowords accuracy. However, for both groups reaction time for pseudowords was slower than for words. Mounting evidence suggests that LDT represents processing of the orthographic (words) and the phonological (pseudowords) routes in reading [8]. Based on the dual-route theory [8], it can be suggested that pseudowords identification is based on the slower sequential phonological route [2]. However, it is conceivable that by the time a reader reaches secondary school, after more than six years of print exposure, the identification of pseudoword patterns becomes more precise and almost similar to a mature reader. However, recognition of real word relies on the identification of its orthographic pattern and on the retrieval of its exact meaning from the mental lexicon. It seems that more than six years of print exposure and reading practice is needed in order to bring the brain circuitry to automatic activation. This notion might be more pronounced in reading Hebrew, as the Hebrew script has two forms, i,e., pointed Hebrew (in shallow orthography) for 1^{st}-5^{th} graders and un-pointed (deep orthography) scripts from 5^{th} grade onwards. It is conceivable that the members of the younger group in the current study, who were in 7^{th} grade, did not yet fully mastered reading in deep orthography and, as a result, exhibited a higher number of errors when identifying words.

Finally, the results of the current study support the notion that the upper left frontal lobe takes part in the process of lexical decision [5]. It was previously shown that the upper left frontal lobe is involved in decision making [3, 24]. It has neuronal connections with more posterior brain regions [25], which were suggested to be involved in semantic information processing [5, 10]. According to the Multiple Read-Out model [10, 26], a lexical decision can be made when the semantic information process leads to a positive intra-lexical trigger or a negative extra-lexical trigger. The relatively lower level of minimum Oxygenation value under the pseudowords condition suggests a higher level of oxygen consumption at the beginning of the information processing, that is, higher effort was made in the pseudoword condition. In addition, the relatively slower fresh blood inflow under the pseudoword condition, also suggests that the pseudoword information processing required more time than the word information processing. Overall, the fNIR results were in line with the behavioral results where a significant stimulus-type effect was found in both reaction time and accuracy, indicating longer and more complex information processing for pseudowords as compared to words.

References

1. Binder, J.R., et al.: Neural Correlates of Lexical Access during Visual Word Recognition. Journal of Cognitive Neuroscience 15, 372–393 (2003)
2. Carreiras, M., Mechelli, A., Estevez, A., Price, C.J.: Brain Activation for Lexical Decision and Reading Aloud: Two Sides of the Same Coin? Journal of Cognitive Neuroscience 19, 433–444 (2007)
3. Fiebach, C., Ricker, J., Friederici, B., Jacobs, A.D., Inhibition, A.M.: Facilitation in Visual Word Recognition: Prefrontal Contribution to the Orthographic Neighborhood Size Effect. NeuroImage 36, 901–911 (2007)
4. Frishkoff, G.A., Perfetti, C.A., Westbury, C.: ERP Measures of Partial Semantic Knowledge: Left Temporal Indices of Skill Differences and Lexical Quality. Biological Psychology 80, 130–147 (2009)
5. Hofmann, J., Markus, et al.: Differential Activation of Frontal and Parietal Regions during Visual Word Recognition: An Optical Topography Study. NeuroImage 40, 1340–1349 (2008)
6. Ischebeck, A., Indefrey, P., Usui, N., Nose, I., Hellwig, F., Taira, M.: Reading in a Regular Orthography: An fMRI Study Investigating the Role of Visual Familiarity. Journal of Cognitive Neuroscience 16, 727–741 (2004)
7. Meyer, D.E., Schvaneveldt, R.W.: Facilitation in Recognizing Pairs of Words: Evidence of a Dependence between Retrieval Operations. Journal of Experimental Psychology 90, 227–234 (1971)
8. Coltheart, M., Curtis, B., Atkins, P., Haller, M.: Models of Reading Aloud: Dual-Route and Parallel-Distributed-Processing Approaches. Psychological Review 100, 589–608 (1993)
9. Carreiras, M., Mechelli, A., Price, C.J.: Effect of Word and Syllable Frequency on Activation during Lexical Decision and Reading Aloud. Human Brain Mapping 27, 963–972 (2006)
10. Grainger, J., Jacobs, M., Arthur: Orthographic Processing in Visual Word Recognition: A Multiple Read-Out Model. Psychological Review 103, 518–565 (1996)
11. Ayaz, H., et al.: Cognitive Workload Assessment of Air Traffic Controllers using Optical Brain Imaging Sensors. In: Marek, T., Karwowski, W., Rice, V. (eds.) Neuroergonomics, Human Factors Design, and Special Populations, pp. 21–31. CRC Press Taylor & Francis Group, Boca Raton (2010)
12. Izzetoglu, K., Bunce, S., Onaral, B., Pourrezaei, K., Chance, B.: Functional Optical Brain Imaging using Near-Infrared during Cognitive Tasks. International Journal of Human-Computer Interaction 17, 211–227 (2004)
13. Menda, J., et al.: Optical Brain Imaging to Enhance UAV Operator Training, Evaluation, and Interface Development. Journal of Intelligent Robotic Systems 61, 423–443 (2011)
14. Quaresima, V., Ferrari, M., van der Sluijs, Marco, C.P., Menssen, J., Coiler, W.N.J.M.: Lateral Frontal Cortex Oxygenation Changes during Translation and Language Switching Revealed by Non-Invasive Near-Infrared Multi-Point Measurements. Brain Research Bulletin 59, 235–243 (2002)
15. Sakatani, K., Lichty, W., Xie, Y., Li, S., Zuo, H.: Effects of Aging on Language-Activated Cerebral Blood Oxygenation Changes of the Left Prefrontal Cortex: Near Infrared Spectroscopy Study. Journal of Stroke and Cerebrovascular Diseases 8, 398–403 (1999)
16. Watanabe, E., et al.: Non-Invasive Assessment of Language Dominance with Near-Infrared Spectroscopic Mapping. Neuroscience Letters 256, 49–52 (1998)
17. Raven, J.C.: Guide to the Standard Progressive Matrices. H. K. Lewis, London (1960)

18. Breznitz, Z., Misra, M.: Speed of Processing of the Visual-Orthographic and Auditory-Phonological Systems in Adult Dyslexics: The Contribution of "Asynchrony" to Word Recognition Deficits. Brain and Language 85, 486–502 (2003)
19. Frost, R.: The Word-Frequency Database For Printed Hebrew, `http://atar.mscc.huji.ac.il/~frost`
20. Izzetoglu, M., Nioka, S., Chance, B., Onaral, B.: Single trial hemodynamic response estimation in a block anagram solution study using fNIR spectroscopy. In: Proc. of ICASSP, vol. 5, pp. 633–636 (2005)
21. Miezin, F.M., Maccotta, L., Ollinger, J.M., Petersen, S.E., Buckner, R.L.: Characterizing the Hemodynamic Response: Effects of Presentation Rate, Sampling Procedure, and the Possibility of Ordering Brain Activity Based on Relative Timing. NeuroImage 11, 735–759 (2000)
22. Izzetoglu, M., Bunce, S.C., Izzetoglu, K., Onaral, B., Pourrezaei, K.: Functional Brain Imaging Using Near-Infrared Technology for Cognitive Activity Assessment. IEEE Engineering in Medicine and Biology Magazine, Special Issue on the Role of Optical Imaging in Augmented Cognition 26, 38–46 (2007)
23. Duncan, A., et al.: Optical Pathlength Measurements on Adult Head, Calf and Forearm and the Head of the Newborn Infant using Phase Resolved Optical Spectroscopy. Physics in Medicine and Biology 40, 295–304 (1999)
24. Goswami, U.: Neuroscience and Education: From Research to Practice? Nature Reviews Neuroscience 7, 406–411 (2006)
25. Ligges, C., Blanz, B.: Survey of fMRI Results regarding a Phonological Deficit in Children and Adults with Dyslexia: Fundamental Deficit Or Indication of Compensation? Z Kinder Zugendpsychiatr Psychother 35, 107–115 (2007)
26. Jacobs, A.M., Graf, R., Kinder, A.: Receiver Operating Characteristics in the Lexical Decision Task: Evidence for a Simple Signal-Detection Process Simulated by the Multiple Read-Out Model. Journal of Experimental Psychology, Learning, Memory, and Cognition 29, 481–488 (2003)

Brain in the Loop: Assessing Learning Using fNIR in Cognitive and Motor Tasks

Patricia A. Shewokis[1,2], Hasan Ayaz[1], Meltem Izzetoglu[1],
Scott Bunce[3], Rodolphe J. Gentili[4], Itamar Sela[1,5],
Kurtulus Izzetoglu[1], and Banu Onaral[1]

[1] School of Biomedical Engineering, Science & Health Systems,
Drexel University, Philadelphia, PA
[2] College of Nursing and Health Professions, Drexel University, Philadelphia, PA
[3] Penn State Hershey Neuroscience Institute, Penn State University, Hershey, PA
[4] Department of Kinesiolgy, Graduate Program in Neuroscience & Cognitive Science,
University of Maryland, College Park, MD
[5] The Edmond J. Safra Brain Research Center for the Study of Learning Disabilities,
University of Haifa, Haifa, Israel
{shewokis,ayaz,meltem,ki25,banu.onaral}@drexel.edu,
sbunce@hmc.psu.edu, rodolphe@umd.edu, itamar.sela01@gmail.com

Abstract. The skill acquisition process and learning assessments are dependent upon the quality and extent of practice of the tasks. Typically, learning is inferred from behavioral and cognitive results without taking into account the role of the brain in the learning loop. In this paper we discuss the neural mechanisms of learning and skill acquisition using fNIR with 3D spatial navigation tasks (e.g., MazeSuite), a center-out reaching movement task during which adaptation to new tool use was performed and mathematical problem solving tasks. Further, this research study compared and contrasted multiple analysis methods, which include general linear models of repeated measures during acquisition, retention and transfer phases of learning, learning curve analyses, the testing of fit of various learning models (i.e., power, exponential or other non-linear functions) and relationships between neural activation and behavioral measures.

Keywords: Practice, Learning, Optical Brain Imaging, Analysis Methods, Functional Near Infrared Spectroscopy, fNIR, Prefrontal Cortex.

1 Introduction

The role of practice is crucial in the skill acquisition process and for assessments of learning. By examining the cognitive and behavioral output during the performance and learning of selected cognitive and motor tasks, along with a detailed examination of the neural activity obtained from functional near infrared (fNIR) spectroscopy, it may be possible to gain insight into the impact that practice has on learning, transfer and the skill acquisition processes. This paper discusses the neural mechanisms of learning and skill acquisition using fNIR with 3D spatial navigation tasks (e.g., MazeSuite [1]),

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 240–249, 2011.

center-out reaching movement task during which adaptation to a new tool use was performed (e.g., [2]), and mathematical problem solving tasks. Prior to the examination of different methods of analyses, we provide a brief review of the literature on fNIR, skill acquisition, neural aspects of cognitive- motor control and the impact of practice on neural plasticity.

Functional near infrared spectroscopy (fNIR) has been used as a noninvasive tool to monitor changes in concentration of oxygenated hemoglobin and deoxygenated hemoglobin at the cortex [3-5]. Moreover, fNIR technology allows the design of portable, safe, affordable and accessible brain activity monitoring systems that can be used in both laboratory and ecologically valid natural settings. The fNIR sensor, scans primarily the prefrontal cortex to monitor executive functions [6-8] while being able to process out movement artifact post-acquisition (e.g., [9]) or during real-time data acquisition [10, 11].

During skill acquisition, there is a relationship between cognition and motor function [12]. Georgopoulous (2000) posited that the goal for the neural aspects of motor control is to discover which facet of the cognitive–motor function are processed by given areas of the cortex during performance of a specific task [12]. This neural plasticity and flexibility is critical during the skill acquisition process and illustrates the need for multiple cortical sites to be measured during learning. Importantly, the prefrontal cortex (PFC) serves as the highest cortical area responsible for motor planning, organization and regulation. In addition, PFC plays an important role in the integration of sensory and mnemonic information, the regulation of cognitive function and action, and works with other cortical circuits with executive functions including working memory and inhibitory control [13, 14].

There is considerable evidence that the sensory and motor areas of the brain are dynamically maintained in both normal and brain-injured animals and humans, and are continuously modulated in response to activity, behavior, and skill acquisition [15, 16]. Repeated motor performance or practice as part of motor learning is crucial to promoting the cortical changes that result in functional improvement [17]. Motor learning, which is inferred from external observations of improvements in performance, occurs in various internal processes that are associated with practice or experience that drives the acquisition of motor skills [17-19]. Karni et al. [19] hypothesized that fast learning involves processes that identify and develop an optimal routine for the execution of the task while slow learning reflects the on-going long-term changes of the movement pattern that may occur at a structural level [19]. These processes establish a relationship between the improvements associated with motor learning and changes in cortical pathways that facilitate the improvements.

Practice and its influence on skill learning have been studied across a range of motor, visuomotor, perceptual and cognitive tasks, and from disparate research perspectives. To briefly summarize this literature, four main patterns of practice-related activation change can be distinguished [20]. Practice can lead to an increase or a decrease in activation at the brain areas that are involved in task performance. The differences in brain activation as a result of practice can be due to (1) a functional redistribution of brain activity, in which some initial areas of activation decrease, whereas other initial areas of activation increase, and (2) a functional reorganization of brain activity, i.e., the pattern of activation increases and decreases occur in distinct brain areas as well as the initial areas.

Our aim is to present multiple methods of analysis which include general linear models of repeated measures during acquisition, retention and transfer phases, testing of learning curves and the testing of fit of various learning models (i.e., power, exponential or other non-linear functions) during selected tasks. Discussion focuses on the use of different methods of analysis, interpretation of the data and implications for training and learning.

2 Method - fNIR Device, MazeSuite and Spatial Navigation Tasks

2.1 fNIR Device

For all tasks described in this paper, Drexel's continuous wave fNIR system was used as the neuroimaging device. The fNIR device is connected to a flexible sensor pad that contains 4 light sources with wavelength peaks at 730 nm and 850 nm and 10 detectors which are designed to sample cortical areas underlying the forehead at 2Hz. The fNIR device has a fixed source-detector separation of 2.5 cm, resulting in a total of 16 measurement locations (voxels) [21]. For data acquisition and visualization, COBI Studio software (© Drexel University, 2010) was used.

2.2 MazeSuite and Spatial Navigation Tasks

For the design and presentation of interactive stimulus during spatial navigation tasks,

MazeSuite (Drexel University; www.mazesuite.com) has been used [1]. MazeSuite is a set of software tools to help researchers prepare, present and analyze navigational and spatial experiments.

The study involved the monitoring of the PFC area to assess changes in cognitive activity during the acquisition and learning of computer maze tasks for blocked (BLK) and random (RAN) orders. The PFC is thought to be involved in the maintenance of working memory and response selection [22]. Each subject performed 315 acquisition trials (i.e., 105 trials for each of the 3 mazes) across three days and 72 hours following acquisition, 30 retention and 20 transfer trials (using 2 different mazes) were performed in a random order. PFC activity was monitored during all phases for 16 optode sites using fNIR. Dependent measures included relative changes in the mean oxygenated hemoglobin (oxy-hb) and behavioral measures of total time and path length for the mazes.

2.3 Participants and Specific Aim for Spatial Navigation Task

Seven healthy adults consented to participate in the study. All were right-handed and they were randomly assigned to either a blocked (n=4) or random practice order.

The specific aim was to identify and characterize the neuroplasticity changes that occur in the practice schedules at the cortical, cognitive and behaviors levels across the acquisition, memory and transfer phases of computer maze tasks.

2.4 Spatial Navigation Task Results for General Linear Model (GLM)Analysis

For acquisition, two 2 X 3 X 3 (Practice Schedule X Task X Day) mixed model ANOVAs with repeated measures on the last two factors were calculated on mean path length (arbitrary units (a.u.)) and mean oxygenation change (µmolar). Significant interactions for the behavioral mean path length were Practice Schedule X Task X Day [$F_{(4,10)}$ =4.25, p = 0.002]; Practice Schedule X Task [$F_{(2,10)}$= 4.61, p 0.01]; Practice Schedule X Day [$F_{(2,10)}$ = 3.78, p =0.023] with a Task main effect $F_{(2,10)}$ = 67.90, $p < 0.001$. The most important oxygenation changes were in the left PFC at channel 5 with a significant interaction for Practice Schedule X Day [$F_{(2,10)}$ = 5.36, p = 0.005] and a Day main effect [$F_{(2,10)}$ = 31.69, $p < 0.001$].

During retention, the only significant behavioral effect was for task [$F_{(2,10)}$ = 94.07, $p < 0.001$]. Maze 2 had the shortest path travelled (M $\pm$ SD; 30.77 $\pm$ 0.88 a.u.) compared to 41.38 $\pm$ 3.29 and 38.60 $\pm$ 1.09 for mazes 1 and 3, respectively. Figure 1 illustrates the Oxy Practice Schedule X Task interaction effect for channel 5, located on the left PFC [$F_{(2,10)}$ = 8.53, p = 0.003 (H-F correction)]. No significant effects were found for mean path length and mean oxygenation for transfer.

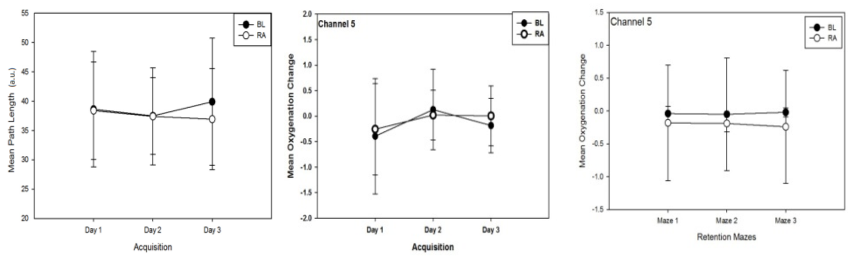


Fig. 1. Mean Path Length (left panel) and Mean Oxygenation Changes (middle panel) for Significant Acquisition Practice Schedule by Day and Mean Oxygenation Changes for (right panel) Retention Phase Practice Schedule by Task Interactions for Representative Behavioral and Left Prefrontal Cortical Area Activation. Error bars are standard deviations.

2.5 Spatial Navigation Task Results for Learning Curve Analysis

Given that the group results indicated that maze 2 had the shortest distance traveled for retention (see Sec. 2.4) and that there are different predictions for learning based on the practice order of the tasks during acquisition, we calculated learning curves for selected behavioral and neural measures during the acquisition for an individual learning under a BLK order (Subject 1) and for an individual learning in a RAN order (Subject 2). Depicted in Figure 2 are learning curves for the first trials 1-10. For the behavioral measures of total maze time (sec) and average velocity in maze (a.u./sec), a power law is the best fit for the data for both the BLK ($y(x) = 11.247x^{-0.217}$; $y(x) = 3.084x^{0.1565}$) and RAN ($y(x) = 11.895x^{-0.178}$; $y(x) = 2.836x^{0.1147}$) practice schedules, respectively. Maze 2 metabolic results showed linear models of best fit for acquisition trials 1-10 for BLK ($y(x) = 0.0463x + 0.0366$) and RAN ($y(x) = 0.0338x + 1.6828$) practice orders.

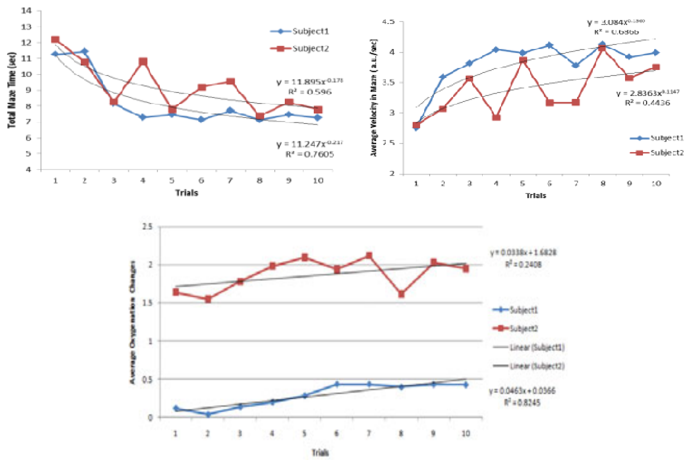


Fig. 2. Fitted learning curves for selected behavioral (upper left and upper right panels) and hemodynamic measures (lower panel) for Maze 2 acquisition for trials 1-10 for a BLK (subject 1) and RAN (subject 2) practice order

3 Method – Center Out Task

3.1 Participants and Apparatus

Five healthy right-handed adults gave informed consent prior to participation and had normal vision or corrected to normal vision. Participants were seated facing a computer screen and performed a 'center-out' drawing movement which linked a center target with one of four peripheral targets. A 16-channel fNIR system was positioned on the forehead. Limb movements were not observed by participants with the outcome available on the screen. Specific details of the device and experiment are reported in [2]. The experiment was divided into three segments: i) pre-training, ii) training (180 trials – 20 trials X 9 blocks; task was to adapt to a 60^0 counterclockwise screen cursor rotation), and iii) post-training. The instructions were to draw a line as fast and as straight as possible from the center to the target.

3.2 Data Processing and Statistical Analyses

The 2D pen position was low-pass filtered using a 5Hz eighth order Butterworth filter. Motor performance was computed as Movement Time (MT) and defined as the amount of time that elapsed between the pen leaving the home circle until it entered the peripheral target. After standard fNIR signal preprocessing, blood oxygenation changes within the PFC were calculated using the modified Beer Lambert Law [6-9] and oxygenation (Oxy – difference between oxygenated hemoglobin and deoxygenated hemoglobin) was derived. For both MT and Oxy, the values were standardized to the pre-training stage to account for individual differences in the baseline and to focus on changes due to training and adaptation.

$$StdP_i = \frac{P_i - Mean(P)_{pre-training}}{SD(P)_{pre-training}} \quad (1)$$

With P_i (Parameter) is the value of the MT or Oxy for the i^{th} trial during the training period. $Mean(P)_{pre\text{-}training}$ and $SD(P)_{pre\text{-}training}$ are the pre-training parameters computed during the pre-training session. After computation, the StdP (standardized parameter) values were averaged across blocks and subjects. The average MT and Oxy standardized values were analyzed for the best fitting curves to assess learning.

3.3 Results of the Center Out Task for Tool Use

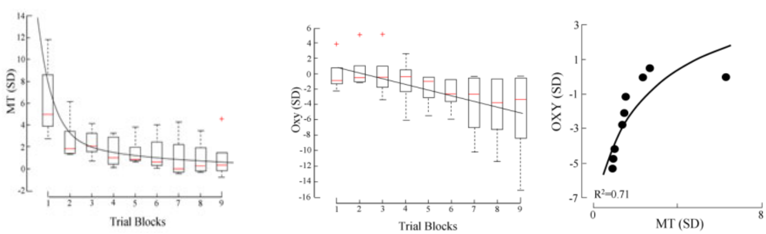


Fig. 3. Changes in movement time (MT) (left panel) and oxygenation (Oxy) (middle panel) throughout adaption to the task with corresponding cortical hemodynamic changes. Using the median values, for the MT an exponential learning curve best fit the adaptation changes [$y(x)= 22.87e^{-1.41x} + 2.67e^{-0.16x}$] while the OXY learning curve had a linear curve of best fit for the adaptation changes [$y(x) = -0.435x +0.543$ with $r^2 = 0.785$]. The relationship between neural activation and behavior on the center out adaptation task is illustrated with the mean standardized oxygenation values and the standardized behavioral measure of movement time. (right panel)

4 Method – Mathematical Problem Solving Task

4.1 Participants and Tasks

A subsample of two right-handed adults provided informed consent and performed a series of mathematical tangram puzzles. Figures 4 and 5 provide a schematic of the experimental procedure and an example of the control and an animal tangram puzzle, respectively.

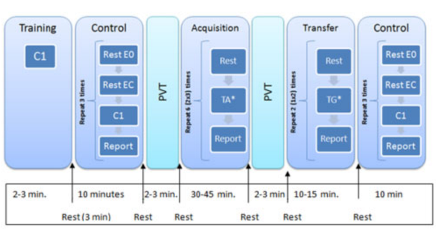


Fig. 4. Schematic Diagram of Phases in the Tangram Protocol with Approximate Durations**.** C1-control without timeout; Rest EO – Rest Eyes Open; Rest EC – Rest Eyes Closed, Report – effort reporting; PVT –Psychomotor Vigilance Test; TA*-Tangrams – Animals; TG* - Tangrams – Geometric.

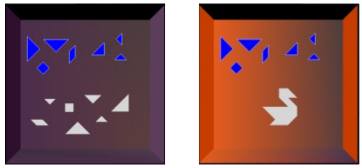


Fig. 5. Screen shots of the control task (left panel), and the animal-shaped swan task with a timeout warning (right-panel) in which the background changes color so that the subject knows the time limit for completion of the task is approaching

4.2 Processing and Statistical Analyses

A two stage analysis model is the basis to analyze the hemodynamic measures of a statistical parametric mapping (SPM) model that has been used for fMRI analyses [23]. SPM models employ a univariate approach for determining estimates of parameters of a GLM at each voxel measure in the head. Our focus is to illustrate the utility of a stage one analysis prior to proceeding to Stage two. To accomplish a stage one – a subsample of two subjects were selected. Stage one is a subject-level specific GLM (paired t-test) where the mean oxygenation is assessed and parameter estimates derived. GLM tests were conducted on the control values paired with the early learning (trials 1-3), late learning (trials 4-6) and transfer (trials 1-2).

4.3 Results of Mathematical Problem Solving Task

The results of the paired t-tests for the stage one analysis are reported below in Table 1 using the oxygenated hemoglobin metabolic measure for neural activation.

Table 1. Mean oxygenated hemoglogin (HBO2) values comparing the control (C), early learning (EL), late learning (LL) and transfer (T) phases for voxel 2 (directional, one-tailed tests). Significant effects are in bold

Subject	Z-value C vs.EL	Z-value C vs. LL	Z-value C vs. T	p-value C vs. EL	p-value C vs.LL	p-value C vs. T
5	-1.716	-1.912	-0.215	**0.039**	**0.027**	0.525
6	-2.140	-0.801	-0.599	**0.029**	0.212	0.283

5 Discussion

Although assessments of learning area best exemplified by retention (i.e., memory) and transfer (i.e., generalizability) tests, it is important to determine differences across acquisition trial blocks as a first indicator of learning. In our study, the three mazes that were practiced during acquisition with feedback were then tested without feedback for a few trials (e.g., 10 of each maze) in retention. For transfer, two novel mazes were created that had additional corridors and different starting and ending points than the mazes originally practiced. Ten trials each of the two new mazes created a transfer phase which was used to determine how well each subject generalized their learning to the new mazes. We found acquisition and retention differences for both behavioral and hemodynamic measures. These findings are consistent with the expected changes in behavior and neural activation as assessments measures of learning.

Important for our understanding of learning processes, Speelman and Kirsner [24] note that learning curves are individualized based on the performers previous experiences and learning a new task is practice of previously acquired skills within a new context. This point is precisely the findings of our behavioral and neural learning curves when the subjects learn spatial navigation of the mazes under either a BLK or RAN practice order. In Fig. 3, as expected, subject 1 (BLK practice) reaches an asymptote faster than subject 2 (RAN practice). Subject 2 is slower in reaching an asymptote because there are no more than two consecutive trials of the same task practiced, consequently more information and comparisons between tasks are made resulting in slower performance time and velocity. This finding is corroborated by the similar oxygenation patterns for the BLK and RAN practice, however, RAN practice requires more effort (see Fig. 3) and indicated higher average oxygenation per trial. The learning curve analyses of a spatial navigation task (maze 2), revealed a power function for the behavioral measures and a linear function for oxygenation. In general, if the learning of a task follows a power law, then slowing down of learning is based on a decreasing percentage of the amount of the task to be learned [25]. For the spatial navigation task behavioral measures there is a physiological and mechanical limit as to how fast any individual can traverse the maze. These constraints include how proficient the individual is in manipulating the input devices, the speed of the computer processor as well as how quickly the individual remembers the maze pattern.

For the center out task, exponential models best fit the MT data (see Fig. 2), indicating that the learning is based on a fixed percentage of what remains to learned in the task [24, 25]. However, the Oxy data was best fit by a linear function which represents a minimum level of activation to perform the task (0.543 units of Oxy) while the slope reduces -0.435 units per trial block increase. Therefore, there is a reduction in neural activation as adaptation occurs and the task is acquired. In addition, we assessed the relationship between the MT and Oxy (see Fig. 3 right panel) which showed that 71% of the variance in Oxy can be explained by MT. The non-linear relationship is primarily influenced by the longer movement time and higher neural activation of the first trial block noting the earliest stage of learning. If the first trial block of learning is removed the relationship between oxygenation and MT improves – accounting for 88 % variation which is a 17% improvement.

The learning of mathematical tangram problem solving tasks using a stage one subject level analysis [23], revealed that both subjects had increased activation of the left PFC during the early stages of learning compared to the control task (Table 1). This finding is comparable to the work of Lee and colleagues [26] where they noted that the 'left frontal gyrii are implicated in arithmetic problem solving tasks. Although our subsample is in the direction expected, additional subjects and a stage two group analysis is needed.

In summary, incorporating the brain in the learning loop using a portable, robust, safe and non-invasive fNIR optical imaging device was demonstrated using several different tasks and learning paradigms. We used a variety of methods to assess learning at the individual and group levels of analyses. Typically, inferences about learning have been based on behavioral measures. We contend that incorporating neural activation measures during performance and learning provides important insights into the learning processes that have implications for designing new training, instructional and rehabilitation paradigms.

Acknowledgments. The spatial navigation task was funded in part by the Commonwealth of Pennsylvania # 4100037709 subcontract #240468 and Drexel University subcontract #280773.

References

1. Ayaz, H., Allen, S.L., Platek, S.M., Onaral, B.: Maze Suite 1.0: a complete set of tools to prepare, present, and analyze navigational and spatial cognitive neuroscience experiments. Behav. Res. Methods 40, 353–359 (2008)
2. Gentili, R.J., Hadavi, C., Ayaz, H., Shewokis, P.A., Contreras-Vidal, J.L.: Hemodynamic Correlates of Visuomotor Motor Adaptation by Functional Near Infrared Spectroscopy. In: Conf. Proc. IEEE Eng. Med. Biol. Soc., pp. 2918–2921 (2010)
3. Chance, B., Zhuang, Z., UnAh, C., Alter, C., Lipton, L.: Cognition-activated low-frequency modulation of light absorption in human brain. Proceedings of the National Academy of Sciences of the United States of America 90, 3770–3774 (1993)
4. Strangman, G., Boas, D.A., Sutton, J.P.: Non-invasive neuroimaging using near-infrared light. Biological psychiatry 52, 679–693 (2002)
5. Villringer, A., Planck, J., Hock, C., Schleinkofer, L., Dirnagl, U.: Near infrared spectroscopy (NIRS): a new tool to study hemodynamic changes during activation of brain function in human adults. Neuroscience letters 154, 101–104 (1993)
6. Izzetoglu, M., Izzetoglu, K., Bunce, S., Ayaz, H., Devaraj, A., Onaral, B., Pourrezaei, K.: Functional near-infrared neuroimaging. IEEE Trans. Neural Syst. Rehabil. Eng. 13, 153–159 (2005)
7. Izzetoglu, K., Bunce, S., Onaral, B., Pourrezaei, K., Chance, B.: Functional Optical Brain Imaging Using Near-Infrared During Cognitive Tasks. International Journal of Human-Computer Interaction 17, 211–227 (2004)
8. Ayaz, H., Willems, B., Bunce, B., Shewokis, P.A., Izzetoglu, K., Hah, S., Deshmukh, A., Onaral, B.: Cognitive Workload Assessment of Air Traffic Controllers Using Optical Brain Imaging Sensors. In: Marek, T., Karwowski, W., Rice, V. (eds.) Advances in Understanding Human Performance: Neuroergonomics, Human Factors Design, and Special Populations, pp. 21–31. CRC Press Taylor & Francis Group (2010)
9. Izzetoglu, M., Chitrapu, P., Bunce, S., Onaral, B.: Motion artifact cancellation in NIR spectroscopy using discrete Kalman filtering. Biomedical engineering online 9, 16 (2010)
10. Ayaz, H., Izzetoglu, M., Shewokis, P.A., Onaral, B.: Sliding-window Motion Artifact Rejection for Functional Near-Infrared Spectroscopy. In: Conf. Proc. IEEE Eng. Med. Biol. Soc., pp. 6567–6570 (2010)
11. Ayaz, H., Shewokis, P.A., Bunce, S., Schultheis, M., Onaral, B.: Assessment of Cognitive Neural Correlates for a Functional Near Infrared-Based Brain Computer Interface System. In: Schmorrow, D.D., Estabrooke, I.V., Grootjen, M. (eds.) FAC 2009. LNCS, vol. 5638, pp. 699–708. Springer, Heidelberg (2009)
12. Georgopoulos, A.P.: Neural aspects of cognitive motor control. Current Opinion in Neurobiology 10, 238–241 (2000)

13. Procyk, E., Goldman-Rakic, P.S.: Modulation of dorsolateral prefrontal delay activity during self-organized behavior. Journal of Neuroscience 26, 11313–11323 (2006)
14. Miller, E.K., Cohen, J.D.: An integrative theory of prefrontal cortex function. Neuroscience 24, 167–202 (2001)
15. Bütefisch, C.M., Davis, B.C., Wise, S.P., Sawaki, L., Kopylev, L., Classen, J., Cohen, L.G.: Mechanisms of use-dependent plasticity in the human motor cortex. Proceedings of the National Academy of Sciences of the United States of America 97, 3661 (2000)
16. Nudo, R.J.: Functional and structural plasticity in motor cortex: implications for stroke recovery. Physical medicine and rehabilitation clinics of North America 14, S57–S76 (2003)
17. Hummelsheim, H.: Rationales for improving motor function. Current Opinion in Neurology 12, 697–701 (1999)
18. Shepherd, R.B.: Exercise and training to optimize functional motor performance in stroke: Driving neural reorganization? Neural Plasticity 8, 121–130 (2001)
19. Karni, A., Meyer, G., Rey-Hipolito, C., Jezzard, P., Adams, M.M., Turner, R., Ungerleider, L.G.: The acquisition of skilled motor performance: fast and slow experience-driven changes in primary motor cortex. Proceedings of the National Academy of Sciences of the United States of America 95, 861 (1998)
20. Kelly, A., Garavan, H.: Human functional neuroimaging of brain changes associated with practice. Cerebral Cortex 15, 1089 (2005)
21. Ayaz, H., Izzetoglu, M., Platek, S.M., Bunce, S., Izzetoglu, K., Pourrezaei, K., Onaral, B.: Registering fNIR data to brain surface image using MRI templates. In: Conf. Proc. IEEE Eng. Med. Biol. Soc., pp. 2671–2674 (2006)
22. Wood, J.N., Grafman, J.: Human prefrontal cortex: processing and representational perspectives. Nat. Rev. Neurosci. 4, 139–147 (2003)
23. Gläscher, J.: Visualization of group inference data in functional neuroimaging. Neuroinformatics 7, 73–82 (2009)
24. Speelman, C., Kirsner, K.: Beyond the Learning Curve: The Construction of Mind. Oxford University Press, Oxford (2005)
25. Ritter, F.E., Schooler, L.J.: The learning curve. In: Kintch, W., Smelser, P.N., Baltes, P. (eds.) International Encyclopedia of the Social and Behavioral Sciences, Pergamon/Oxford, UK, pp. 8602–8605 (2001)
26. Lee, K., Lim, Z.Y., Yeong, S.H.M., Ng, S.F., Venkatraman, V., Chee, M.W.L.: Strategic differences in algebraic problem solving: Neuroanatomical correlates. Brain Res. 1155, 163–171 (2007)

Neurocognitive Patterns: Using Brain, Behavior, and Context to Infer User Intent

Webb Stacy

Aptima, Inc.,
12 Gill Street, Suite 1400, Woburn, MA 01801
wstacy@aptima.com

Abstract. Neurocognitive Patterns is a system that will offer execution options to users as soon as they form an intention to act. It will accomplish this by combining neural signals, user behavior, and contextual knowledge to determine when a user has a goal, and what that goal is. Because it will leverage the user's neural signals and behavioral history, the options it will provide to the user will be available quickly. Because it will leverage real-time contextual and background knowledge, its estimates concerning the user's goal will be accurate. Our initial target domain is UAV operators, but we expect it will be of use to other military decision-makers in Command and Control settings. We also expect that Neurocognitive Patterns will be a useful tool in Cognitive Neuroscience in general for interpreting neural signals in the presence of salient contextual information.

Keywords: Neural Signals, Behavioral Measures, User Intent, Using Contextual Information.

1 Introduction

Modern military operations are increasingly complex and increasingly electronic, and this often has the effect of increasing operator workload. System developers have incorporated many technological developments, but they have not kept up with our increasing understanding of the cognitive capabilities of the operators themselves. Though there is a recently-renewed, increased interest in establishing a system's usability and utility during development in order to mitigate the increasing technology-workload mismatch, there are more dramatic possibilities for improvement.

One such possibility is to use neural signals from the operator as the ultimate unobtrusive form of input: the operator merely has to think about something to make it so. A solution like this uses neurocognitive signals from operators to determine their intended actions so that computers can automatically develop plans to execute those actions. Unfortunately, traditional brain-computer interface approaches are not suitable. They use stereotypic and non-specific neural signals, do not scale with task complexity, and cannot handle changing task environments and contexts. What is needed is a way to leverage the neural signals of the operators' natural cognitive processes that can adapt to the dynamics of modern complex military environments.

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 250–256, 2011.

An immediate difficulty is that, with Cognitive Neuroscience's current understanding of the meaning of neural signals from the brain, it is extremely difficult to provide general interpretations of underlying cognitive activities (i.e., of specific "thoughts"). There has been progress in this regard, of course. For example, Luu and his colleagues [1] identified the neural signature of visual intuition from dense-array electroencephalography (dEEG) data, but it is currently not possible to know the content or correctness of that intuition. As another example, Researchers at CMU identified functional Magnetic Resonance Imaging (fMRI) correlates of semantic components of simple nouns such as "apple." [2] Both studies represent impressive advances in Neuroscience, but they are a far cry from identifying user intent in a complex task environment. It follows that a solution that provides useful neurocognitive control of human-machine systems will need to supplement neural data from the operator with additional contextual knowledge and situational information.

2 Towards a Solution

A system currently under development called Neurocognitive Patterns will provide such a solution. It is a system that will combine neural data, user behavior, and contextual knowledge to determine when a user has a goal, and what that goal is. It will plan ways to accomplish the goal, present them to the user, and execute the plan that the user chooses.

As illustrated in Figure 1, Neurocognitive Patterns will collect the neural signals and user behavior logs and will combine these with mission context, situational updates, and background knowledge in long term memory (LTM) to infer the user's goal. Neurocognitive Patterns will then automatically plan one or more ways to accomplish the user's inferred goal, and will present them to the user for approval. Once approved, it will interact with the workstation to execute the plan. Users will wear equipment to collect neural signals such as dEEG but will otherwise engage with their workstations in the normal manner.

Emerging research in Neuroscience has been able to identify neural correlates associated with a user intending to do something, especially in the motor domain, though it has proved difficult to determine exactly what that intention might be. The key idea in Neurocognitive Patterns is to leverage those neural correlates of intent to identify *when* the user intends to take action, and to use high-level contextual knowledge from sources like the mission, operator behavior, and previously-established context to determine *what* the user intends to do.

The identification of an "intent" signature in the operator's neural signals will indicate that the operator would like to accomplish a newly-formed goal. This will prompt Neurocognitive Patterns to identify the nature of that goal.

A naïve approach to identifying user goals will not solve the problem. For example, simple aggregation of user behaviors like keystrokes or mouse clicks is rarely productive as the meaning of those interaction behaviors is obscure without some understanding of context, such as the computer applications that are receiving those keystrokes and mouse clicks, and the overall mission of the user. Clearly, context must be part of the equation.

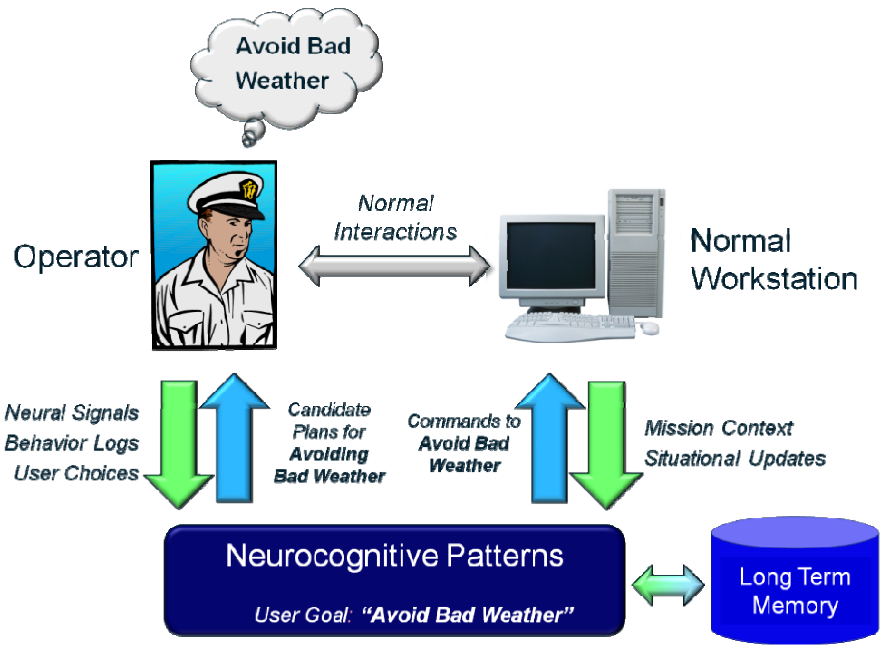


Fig. 1. Neurocognitive Patterns will speed decisions and allow users to make them easily and unobtrusively

Neurocognitive Patterns will incorporate context via an approach called Top-Down/Bottom-Up (TDBU) processing. In TDBU processing, hypotheses ("top down") about the meaning of sensor signals ("bottom up")—neural signals, keystrokes, mouse clicks, and possibly other sensors—are generated and then tested against actual sensor data. The hypotheses originate from contextual knowledge and from previously-perceived aspects of the situation.

The TDBU approach is often effective for machine perception tasks. For example, a machine vision task might be to identify visual objects from a camera image. The TDBU approach might use the fact that the setting is indoors in a kitchen, and that a counter, a sink, and a floor have already been identified, to generate and test the hypothesis that a certain set of pixels on the counter represents a microwave oven. In Neurocognitive Patterns, the task is to identify operator intent, and the TDBU approach might use facts like this:

- The mission is for an unmanned aerial vehicle (UAV) operator to identify possible weapons caches without being spotted.
- A group of people has been identified on the ground.
- The operator's previous behavior indicates that he has been trying to get a closer look by flying at a lower altitude.

to generate the hypothesis that the operator intends to take action to make the vehicle less visible. This hypothesis will then be tested against the operator "sensor signals." Neurocognitive Patterns will consider the operator's neural signals and their recent interaction with their computer system.

Neurocognitive patterns will thus treat the interpretation of neural signals as a top-down/bottom-up machine perception problem. The effect will be a system that provides the opportunity for automated planning assistance to operators without the operators having to explicitly indicate their intent. This will provide an unprecedented level of system efficiency and effectiveness. Further, the general approach of treating the interpretation of neural signals as a machine perception problem can find wide applicability in the field of Cognitive Neuroscience in general.

3 Overview of the System

Neurocognitive Patterns will have four main modules, shown in Figure 2, which will enable the system to detect the nature and timing of user goals, to automatically develop plans to accomplish them, to present those plans to the user, and to execute the user's choice. The process will start with the Goal Identification Module, which will use neural signals to detect the timing of when a user intends to do something, will draw on a variety of contextual and background information to determine the set of goals the user is likely to be pursuing, and will examine this set of goals in relation to the user's recent behavior to deduce their current most likely intent. Given those candidate goals, the Goal Planning Module will automatically construct a plan to accomplish them, and will present the plans to the user via the Graphical User Interface (GUI). If there is a plan with which the user concurs, they will indicate it using the GUI, and that plan will be executed by the Goal Execution Module. If there is no plan with which the user agrees, the user will have the option to indicate whether the goal was wrong or whether the plan was not acceptable, and Neurocognitive Patterns will learn from this feedback.

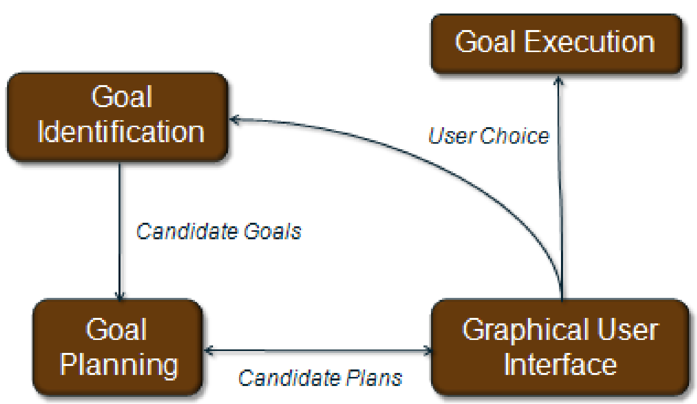


Fig. 2. Top-level view of Neurocognitive Patterns

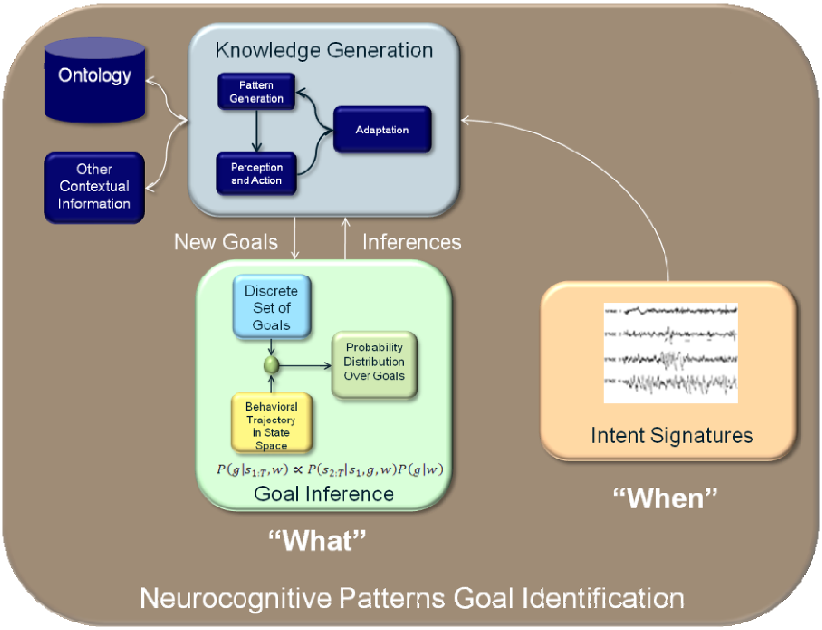


Fig. 3. Components of the Goal Identification Module

As Figure 3 shows, the Goal Identification Module for Neurocognitive Patterns is heart of the system that interprets neural signals. It will accomplish the timely detection of the user's intent to act by finding signatures of intent (the Intent Signatures Component) in the neural patterns, and will determine the user's goals using two components, the Knowledge Generation Component and the Goal Inference Component.

The Knowledge Generation Component, based on a system called Cognitive Patterns [3], will reason over available context and background information to narrow down the set of probable user goals, together with associated user behaviors (the state space) and their probabilities. When confronted with a novel situation, as represented in the contextual and background information, it will have the capability of generating a potential new goal and probabilistic state space, giving it the capability of behaving sensibly in the face of what would otherwise be puzzling and uncertain information.

Cognitive Patterns is an adaptive system based on theoretical concepts from Cognitive Science, such as Conceptual Blending [4], dynamically interpreted perceptual symbol systems [5], and a strong theoretically and empirically justified model of the human eye-brain visual system [6].

The Knowledge Generation (KG) module adds a layer of flexible, temporarily-generated knowledge to the reasoning system. Because it can generate knowledge easily, it can accommodate anomalies and novel events when necessary. The KG component comprises three major internal subcomponents: Pattern Generation, Perception and Action, and Adaptation, which together enable it to create situationally-relevant abstract patterns, to match sensory input to a suitable abstract pattern in a multilayered TDBU fashion (similar to the mechanisms used for visual perception in the brain), and generate new abstract patterns corresponding to "intuitive leaps."

Armed with this set of goals and associated information, the Goal Inference (GI) Component, based on a system called MIMIC [7], will use a mathematical model of user intent to determine the probability distribution over that set of goals, given the user's recent (bottom-up) behavior, thus completing the TDBU strategy.

The GI Component will take as input a model of priorities and goals which will be provided by the KG component will provide. It will attempt to infer (1) the *goals* that are most likely to have generated the human's actions and (2) the most likely *priorities* for the person, of which each goal is a concrete realization. It will do this by combining the model with the user's behavioral history using Bayesian inference to invert this causal relation to infer the most probable goals and priorities that gave rise to the action sequence [8].

To perform this inference, we compare all possible goal hypotheses against each other in terms of how well they explain observed action sequences. Specifically, for each goal hypothesis, it will compute the posterior probability, given observed Actions and the Environment:

$$\begin{aligned} &P(\text{ Goal} \mid \text{Actions}, \text{Environment}) \\ &\propto\ P(\text{ Actions} \mid \text{Goal}, \text{Environment})P(\text{ Goal} \mid \text{Environment}) \end{aligned} \tag{1}$$

Goals with a sufficiently high probability will be considered to have been identified, and these will be passed to the Planning Module to determine the options to be presented to the user.

4 Conclusion

The field of Cognitive Neuroscience is making progress at an unprecedented rate. This new knowledge will lead to dramatic improvements both in the kinds of automated support we can provide our warfighters and in the technology for interpreting neural signals in meaningful and useful ways. We hope that this effort represents a contribution in both directions.

Acknowledgements. I would like to acknowledge the support of the Office of the Secretary of Defense and the Office of Naval Research, especially LCDR Joseph Cohn, Ph.D., for strong support of this effort. I would also like to acknowledge strong collegial encouragement of Alexandra Geyer, Ph.D., and Brian Riordan, Ph.D., both of Aptima. The opinions expressed here are those of the author and do not necessarily reflect the official position of the sponsors or the Department of Defense.

References

[1] Luu, P., Geyer, A., Fidopiastis, C., Campbell, G., Wheeler, T., et al.: Reentrant Processing in Intuitive Perception. PLoS ONE 5(3), e9523 (2010), doi:10.1371/journal.pone.0009523

[2] Just, M., Cherkassky, V., Aryal, S., Mitchell, T.: A Neurosemantic Theory of Concrete Noun Representation Based on the Underlying Brain Codes. PLoS ONE 5(1) (2010), DOI: 10.1371/journal.pone.0008622

[3] Stacy, W., Cohn, J.V., Geyer, A., Wheeler, T.: A Cognition-Based Control System for Autonomous Robots. In: Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting, vol. 54, pp. 1473–1477 (2010)
[4] Fauconnier, G., Turner, M.: The Way We Think: Conceptual Blending and the Mind's Hidden Complexities. Basic Books, New York (2002)
[5] Barsalou, L.W.: Abstraction as dynamic interpretation in perceptual symbol systems. In: Gershkoff-Stowe, L., Rakison, D. (eds.) Building object categories. Carnegie Symposium Series, pp. 389–431. Erlbaum, Majwah (2005)
[6] Kveraga, K., Ghuman, A.S., Bar, M.: Top-down predictions in the human brain. Brain and Cognition 65, 145–168 (2007)
[7] Riordan, B., Bruni, S., Schurr, N., Freeman, J., Ganberg, G., Cooke, N.J., Rima, N.: Inferring user intent with Bayesian inverse planning: Making sense of multi-UAS mission management. In: Proceedings of the 20th Behavior Representation in Modeling and Simulation Conference (BRIMS), Sundance, Utah (2011)
[8] Baker, C.L., Saxe, R., Tenenbaum, J.B.: Action understanding as inverse planning. Cognition 113(3), 29–349 (2009)

Behavioral and Brain Dynamics of Team Coordination Part I: Task Design

E. Tognoli[1], A.J. Kovacs[1], B. Suutari[1], D. Afergan[2,3], J. Coyne[2], G. Gibson[2], R. Stripling[4], and J.A.S. Kelso[1,5]

[1] Center for Complex Systems and Brain Sciences, Florida Atlantic University, Boca Raton, FL
[2] Naval Research Laboratory, Washington, DC
[3] Strategic Analysis Inc., Arlington, VA
[4] Office of Naval Research, Arlington, VA
[5] Intelligent Systems Research Center, University of Ulster, Derry, N. Ireland
{tognoli,kovacs,suutari,kelso}@ccs.fau.edu, dafergan@sainc.com, {coyne,gibson}@itd.nrl.navy.mil, roy.stripling@navy.mil

Abstract. In this study, pairs of subjects performed a team-intensive task with the shared goal of clearing a virtual room from threats. Our goal was to identify signatures of efficient team work from a dynamic analysis of both subjects' brain signals and behavioral performance. An ecologically valid task of room clearing was designed and a novel analysis framework was developed to address the challenge of understanding complex, continuous social processes at both behavioral and brain levels. In the present paper, we detail the design of the task, and present validation techniques undertaken to acquire and analyze high-quality and accurately timed neurobehavioral information. A companion paper will discuss the neurobehavioral findings and their implications.

Keywords: Neuromarkers - EEG - neurobehavioral dynamics - social behavior - complexity.

1 Introduction

One of the most extreme forms of team coordination is observed when members' survival and safety depend upon efficient team interactions, such as when Marines neutralize dangers in a confined urban environment. During such tasks, like clearing a room in a hostile environment, a host of behavioral, cognitive and social processes have to be coordinated in space and in time in a context-dependent fashion. The right process at the wrong time may be deleterious to performance. The goal of this study was to create a framework to quantify the dynamics of neurobehavioral processes unfolding during such ecologically valid tasks that place a high demand on both individual and team coordination.

Our framework aims at quantifying inter-individual variability in team performance, team compatibility, and intra-individual skill learning characteristics of novices trained to perform team-intensive tasks. To combine neural and behavioral dynamics, we developed new tools aimed at revealing the link between brain and

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 257–264, 2011.

behavior in real time [1]. These tools are geared to circumvent a limitation of conventional neuroscientific studies in which only a handful of processes can be assessed at once and whose organization is typically determined by the experimenter.

The underlying tenet of our work is that identifying dynamical neuromarkers and linking them causally to truly complex behaviors that evolve adaptively over time adds critical information. For example, in other contexts such as recovering from head injury, it has been shown that even though behavioral indices have returned to normal, the underlying neural circuitry has certainly not [2]. Also, in typical studies performance error and deficiency are revealed only for a subset of environmental circumstances, whereas deficient neural processes are more frequent, and precede the onset of observable behavioral errors. We argue that a dynamic neurobehavioral framework is all the more important for high-level tasks such as team coordination, because of the complex and intricate architecture of the behavioral, cognitive and social processes that must be recruited for successful performance. In the following, we present some preliminary findings from a very rich data base as well as the methodological framework which is based on the concepts and principles of Coordination Dynamics [3].

2 Materials and Methods

2.1 Subjects

Nine pairs of subjects participated in the experiment (n=18, 1 female, 17 male) with an age range of 20 to 45 years (mean = 28.2). All subjects had normal or corrected to normal vision, and no motor dysfunction. All but one subject was free of psychoactive medication. The results from the medicated subject were not different from others and his data were included in the group analysis. Upon successful completion of the experiment, subjects received a $20 gift card. The protocol was approved by the Florida Atlantic University Internal Review Board and was in accordance with the Declaration of Helsinki. Informed consent was obtained from all subjects.

2.2 Behavioral Task

The task was designed to retain the essence of key behavioral, perceptual, cognitive, social and attentional processes that participate in successful team work (Figure 1A), and followed the main lines of an instructional video of room clearing by the VIRTE program at Clemson University. The processes were integrated into a videogame, in which participants shared the same top down perspective of their virtual environment (Figure 1C). The task was designed and run under XNA (Microsoft Co). Pairs of subjects sat at a table facing a computer screen while holding an Xbox controller (Nyko, China, see Figure 1B). They performed coordinated room clearing, stacking and cueing one another to entry, moving to and covering their areas of responsibility and deciding upon firing at occasionally present enemy (and sometimes friendly) avatars. Subjects controlled their avatar's position and direction of gaze, as they navigated through a series of 32 buildings each composed of 5 successive rooms, with their virtual environment becoming visible upon the avatars' spatial exploration.

Trials started with both avatars stacking along a closed door (Figure 1D-E). At a ready signal by the partner conveyed through touch 'tap' in the real situation and provided through a vibration of the Xbox controller; (verbal communication was not allowed due to EEG artifacts), the avatar closest to the door initiated motion, opened the door with the press of a button on the Xbox controller (Figure 1B) and chose as destination either the left or right corner adjacent to the door. The partner's task was to follow immediately and orient to the opposite corner. Subjects were instructed to follow one of two entry techniques. One entry pattern named "crossover" required the avatar to move diagonally to the opposite corner alongside the wall on which the door is located (movement path for the green avatar in Figure 1D, blue avatar in Figure 1E). The other entry pattern named "buttonhook" required performing a 180^{o} turn (blue avatar in Figure 1D, green avatar in Figure 1E). Given that the leader did not communicate to the follower the type of entry he/she was about to perform, the decision of the follower had to be based on perceptual information regarding the leader's entry pattern. After crossing the door threshold, subjects were asked to move the avatars uninterruptedly until they reached the corners of domination (corners 1 and 2 in Figure 1D-E).

Information about the room was not provided at once, but revealed itself as the avatars explored their environment. A "fog-of-war" was initially present, and avatars' cone of vision revealed the details of the room (walls, enemies and friendly inhabitants). While moving toward the corners of domination, subjects were instructed to initiate scanning of their environment--a behavior called pieing—aimed at optimizing exploratory gaze behavior. That is, they had to divert the cones of vision from their heading direction and rotate it toward the center of the room. This behavior speeds up the discovery of threats in the environment. Upon reaching the corners of domination, the avatars had to adopt a pattern called "interlocking sector of fire", in which their cones of vision was intersecting on the median position of the opposite wall.

Rooms were either empty or had a friend (circular shaped avatar displayed in magenta color) or enemy (orange color) located at pseudorandom spatial locations. The complete session included a total of 52 friends and 52 enemies. The task was designed such that there was never more than one inhabitant per room. When an inhabitant was discovered, subjects were required to make a shoot/no-shoot decision, while following instructions not to slow down their path through the room.

Each action sequence (clearing a building) was composed of 5 successive rooms, all of rectangular shape, with varying dimensions in order to maintain a level of uncertainty on the part of the avatars. After reaching the corners of domination and having secured the current room, avatars were moved to restack along the next door: the sequence of room clearing was then repeated until the entire building was cleared.

In a practice/familiarization session that took place one day preceding the experiment proper, subjects were instructed on correct room clearing techniques and practiced in the virtual environment for about an hour. Warm-up practice was repeated the day of the main experiment while subjects' scalps were prepared for EEG recording. To improve their learning of the task, following each behavioral sequence subjects were asked to provide self-rating on key performance variables such as proximity to the partner upon entry, absence of slowing down during entry, shooting and early pieing readiness. Ratings were provided on a scale from 1 to 10, with 10 being rated as a very good performance, and 1 as a poorly executed performance.

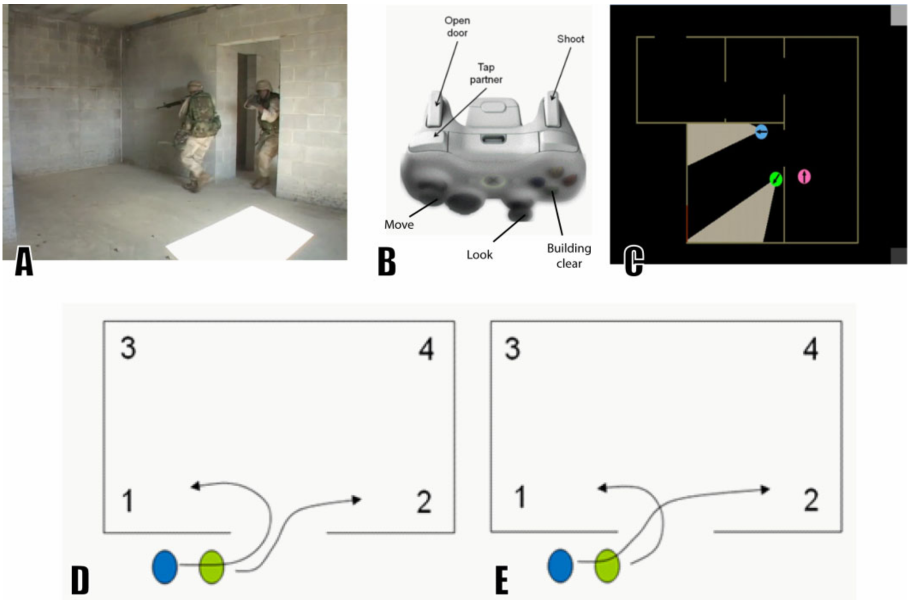


Fig. 1. (A) A live room clearing upon which the task is based (reproduced from Muth and Hoover). (B and C) show the Xbox controllers used in the experiment and a snapshot of the task respectively. (D-E) depicts the two entry patterns to corners of domination 1 and 2. Subjects identified their avatar as an isoluminant colored circle (blue and green). Leader performed either a crossover (D) or a buttonhook entry pattern (E), follower performed the complementary path.

2.3 Behavioral Recording

Behavioral data collection was handled by the computer running the room clear software. A log entry line was written at each graphical refresh for each avatar, which included position and direction of gaze, room location, and binary state variables for a number of behavioral events including tap, door opening, detection of enemy and/or friendly avatars, shooting, and friend or enemy elimination.

2.4 EEG Recording

The experiment took place in a sound-attenuated Faraday chamber. Dual-EEG was recorded by using two 60 channel EEG caps with Ag-AgCl electrodes arranged according to the 10% system [4] including midline and rows 1-8. Signals from both subjects were fed to a single amplifier (Synamp2; Neuroscan, TX) equipped with two distinct referential montages. This specially designed dual-EEG system ensured no delay between the EEGs acquired from each subject [5]. EEG signals were measured with the respective grounds located at FPz sites and the references at the corresponding linked mastoids. Impedances were maintained below 10 KΩ. Vertical and horizontal electro-ocular recordings were also collected to measure saccades and eye blinks. Additionally, noise prevention strategies (shielding and guarding of noise-emitting equipment) were applied to ensure that the behavioral apparatus did not corrupt EEG recordings.

2.5 Timing Validation and Syncing

To optimize the room clear software and to ensure that records of the avatar behavior and EEG data were accurately synchronized during experimental data collection, we developed an analog timing verification apparatus that served to synchronize the virtual room clear environment and the EEG system. The apparatus included time-telling (attached to virtual environment) and time-sensing elements (attached to EEG system). Time-telling elements were reserved patches on the corner of the screens that cycled through luminance increments at each refresh of the computer graphics (refresh rate, 60Hz), indicating the temporal progression of the room clear software. Time-sensing elements were photodiodes placed above the time-telling patches that transduced luminance values and transmitted them directly as an analogue signal to the EEG amplifier. Because EEG amplifiers are designed for accurate time registration, we took the analog trace of task progression provided by the photodiode as an objective measure of time-passing, and compared the temporal events of the room clear software against this standard.

During design of the software, measurement of time by the analog timing verification apparatus allowed us to identify and refine graphical processes and computations that slowed or perturbed execution of the videogame. For instance, we spared computer resources by identifying optimal programming of bullet trajectories. We also identified the Graphical Processing Unit (GPU) clock as the most reliable clock for use in the logging of events within the XNA framework. During experimental data collection, analog timers were used (offline) to resynchronize behavioral data collected in the room clear computer and the EEG measurements acquired and stored in its computer. Sections of the screens occupied with analog timers were masked to the subjects during the actual experiments.

2.6 Behavioral Analysis

Individual and collective performance measures were computed and analyzed from the log data (see section 2.3). Key performance measures were:

- leader entry readiness: delay between tap by follower to motion initiation by leader
- coordinated entry: delay between leader and follower room entry
- time to pie: delay between room entry and the initiation of pieing behavior
- time to shoot: delay between enemy detection and successful shooting
- time to reach corner: delay between room entry and corner reaching, normalized to distance
- entry slow-down: presence of motion deceleration in fatal funnel
- shooting slow-down: presence of motion deceleration during firing
- trajectory error: diversion of trajectory in fatal funnel
- gaze overlap at room clearing completion; percentage redundancy in room scanning
- time to clear room: delay between leader movement and room cleared

To determine coordinated as well as uncoordinated behavior, trials were divided into split-halves on the basis of the above performance measures, for coordinated entry, pieing individual and collectively, and corner of domination coordinated behavior. These variables also formed a behavioral context of broader temporal scale (room and building). Further, each time instant was characterized by a number of event-specific descriptors (see 2.3) which contribute to the behavioral context on shorter time scales. Finally, descriptors of role were added, such as leader (first avatar to enter) and follower; agent or observer of a given behavior; and executed entry pattern (buttonhook or crossover).

2.7 EEG Analysis

There is reason to believe that brain oscillations are the language of the brain and provide important dynamic "neuromarkers" (and neuromarker dynamics). By recording simultaneously between two brains, this study seeks to understand how neuromarkers are modulated in a team setting of some consequence.

Oscillations come in a variety of frequencies, themselves the intertwined result of temporal properties of neural circuitry (faster frequencies for shorter circuits, e.g. [6]) and functionally relevant time scales by virtue of interactions with the body and the environment (faster frequency for briefer functional processes). Within each of these frequency bands, transient dynamics is observed that has a typical duration of one-to-two cycles during waking EEG [1]. The present EEG analysis aims to identify how the self-organizing activity of the brain supports the many complex and temporally overlapping behavioral processes described in sections 2.2 and 2.6.

In order to analyze free-flowing behaviors in the ecologically valid task studied here, we developed a framework for *continuous* EEG analysis. Unlike typical methods, continuous EEG does not rest on iterative protocols that average neural activity over multiple realizations of the same sensorimotor events [7,8]. Instead, the goal is to uncover the link between spontaneously occurring behavioral variables and neural events, and further, to describe the optimal temporal distribution of neural activity for task performance (section 2.6). To do so, each subject's EEG is parsed into its constituting patterns of oscillatory activity, using specially designed spatiotemporal (segmentation of continuous, bandpass filtered EEG), spatiospectral (spatial patterns of Fourier power in short windows) and spatiotemporo-spectral techniques (wavelet spatial patterns). Co-occurrence between brain patterns and behavioral descriptors is then assessed with the goal of establishing causal relations between them.

Spatiotemporal analysis aims to identify synergistic patterns of brain activity as significant units of brain and behavior [9], or in other words, to read the correspondence between brain patterns and corresponding behavioral processes continuously, as in a musical score. Our analysis proceeds through four steps: frequency band identification, bandpass filtering, segmentation and classification. The resulting spatiotemporal patterns are compared between behavioral conditions. Frequency bands are selected on the basis of data-driven and conceptual constraints. When data indicates a specific oscillation of interest, for instance a neuromarker that differentiates between coordinated and uncoordinated social behavior (e.g. [5]), then that band is retained for subsequent analysis. To accommodate for inter-individual variability, this type of frequency selection is best accomplished on a subject-by-subject basis. If no prior information exists for specific oscillations, bands are chosen that meet the time scale of behaviors under investigation (e.g. faster frequencies for more transient behavioral processes). Broadband EEG is then filtered within this band with care (1) to choose a filter cutoff that prevents spatiotemporal pattern clipping and (2) to avert phase distortion with the use of zero-phase shift filtering techniques (phase information is essential to infer cortical self-organization; [1,3]). Results of these signal processing methods reveal a succession of transient spatial patterns that are segmented with visualization techniques [10] or algorithms based on the rotating wave approximation [11]. This parsing is followed by the classification of constitutive patterns and further analysis of their relation to behavioral descriptors.

Spatio-spectral analysis is a complementary method for the detection of dynamic neuromarkers, which is less precise temporally, but computationally faster. It addresses the spatial distribution of high-density EEG spectra. Because of the typical time scale of EEG patterns (1-2 cycles), Fourier analysis is performed in short segments (as short as relevant for the frequency band of interest). Short segments are generally conducive to poor spectral resolution, and high spectral resolution is critically required to distinguish closely overlapping neuromarkers [5]. To circumvent this limitation, "optimized zero-padding" is applied [12]: the signal is first split into epochs, the mean removed, multiplied with a Tukey window to minimize spectral leakage and padded with a suitable number of zeros to increase spectral resolution to the desired value. This technique preserves peak location of neuromarkers at the cost of spreading their power to a broader band (which is controlled by the optimized zero-padding technique). Resulting neuromarkers are examined for their correlation with behavioral descriptors.

Finally, a spatiotemporo-spectral approach was developed to explore more comprehensively the different frequency bands, their dynamic task-dependent modulation and their mutual interactions. Each EEG channel was subjected to a continuous wavelet transform to obtain time-frequency-power distributions. Complex Morlet wavelet was used as the mother function. The spatial distribution of high amplitude time-frequency energy was examined in relation to behavioral events: this technique does not assume frequency bands *a priori*, and is well-suited to reveal the natural time scales at which brain oscillations live during the course of room clearing task performance.

3 Summary

In the previous, we have presented a behavioral task that retains essential components of team work in the team-intensive task of room clearing. In addition to its obvious relevance for training in simulated environments, the task was designed to illuminate a novel dynamical framework for the analysis of brain and behavior intricacies. In a companion paper [13], we will present candidate neuromarkers for efficient room clearing and discuss key theoretical issues relating to successful team coordination.

Acknowledgments. The technical support of William McLean is acknowledged. This work is supported by the US Office of Naval Research Contract N000140510117. JASK and ET are also supported by NIMH Grant MH080838, NSF Grant BCS0826897and the Davimos Family Endowment for Excellence in Science.

References

1. Tognoli, E., Kelso, J.A.S.: Brain coordination dynamics: True and false faces of phase synchrony and metastability. Progress in Neurobiology 87, 31–40 (2009)
2. Jantzen, K.J., Anderson, B., Steinberg, F.L., Kelso, J.A.S.: A prospective functional MR imaging study of mild traumatic brain injury in college football players. American Journal of Neuroradiology 25, 738–745 (2004)
3. Kelso, J.A.S.: Dynamic patterns: the self-organization of brain and behavior. MIT Press, Cambridge (1995)
4. Chatrian, G.E., Wirch, A.L., Edwards, K.H., Turella, G.S., Kaufman, M.A., Snyder, J.M.: Electrophysiological techniques in audiology and otology - cochlear summating potential to broad-band clicks detected from the human external auditory meatus - a study of subjects with normal hearing for age. Ear and Hearing 6, 130–138 (1985)
5. Tognoli, E., Lagarde, J., DeGuzman, G.C., Kelso, J.A.S.: The phi complex as a neuromarker of human social coordination. Proc. Natl. Acad. Sci. USA 104, 8190–8195 (2007)
6. Bressler, S.E., Tognoli, E.: Operational principles of neurocognitive networks. Int. J. Psychophysiology 60, 139–148 (2006)
7. Tognoli, E.: EEG Coordination Dynamics: Neuromarkers of social coordination. In: Fuchs, A., Jirsa, V.K. (eds.) Coordination: Neural, Behavioral and Social Dynamics, pp. 309–323. Springer, Heidelberg (2008)
8. Tognoli, E., DeGuzman, G.C., Kelso, J.A.S.: Interacting humans and the dynamics of their social brains. In: Wang, R., Gu, F. (eds.) Advances in Cognitive Neurodynamics (II), pp. 139–143. Springer, Heidelberg (2010)
9. Kelso, J.A.S.: Synergies: Atoms of brain and behavior. Adv. Exp. Med. Biol. 629, 83–91 (2009)
10. Benites, D., Tognoli, E., DeGuzman, G.C., Kelso, J.A.S.: Brain coordination dynamics: Continuous EEG tracking of the neural functional organization in a social task. Psychophysiology 47, S75–S75 (2010)
11. Fuchs, A., Tognoli, E., Benites, D., Kelso, J.A.S.: Neural correlates of social coordination: Spatiotemporal analysis of brain and behavioral measures. In: Society for Neuroscience, 40th meeting, vol. 293, p. 17 (2010)
12. Suutari, B., Weisberg, S., Tognoli, E., Kelso, J.A.S.: Neuromarkers of Individual and Social Behavior (submitted)
13. Tognoli, E., Kovacs, A.J., Suutari, B., Afergan, D., Coyne, J., Gibson, G., Stripling, R., Kelso, J.A.S.: Behavioral and brain dynamics of team coordination part II: neurobehavioral performance (this issue)

Using Neurophysiological Data to Inform Feedback Timing: A Pilot Study

Jennifer Vogel-Walcutt and Julian Abich

University of Central Florida, 3100 Technology Parkway,
Orlando, FL 32826
jvogel@ist.ucf.edu

Abstract. In an effort to achieve a level of knowledge comparable to that which typically results from individual tutoring, innovative models of adaptive computer-based training are continually being tested and refined. Despite these efforts, adaptive computerized training programs still fall significantly short of the gold standard of one-on-one instruction. In response, this study used a previously developed model defining when to apply instructional feedback during learning in order to improve efficiency. Specifically, we compared the combination of performance and neuro-physiological indices to performance alone as indicators for when to adapt training. Contrary to our hypotheses, this study failed to demonstrate positive impact on knowledge acquisition, knowledge application, perceived cognitive load, or training efficiency. However, based on observational data, it is suspected that participants in neither group possessed enough available working memory capacity to attend to the supporting material. Consequently, this may account for the lack of differential findings.

Keywords: Feedback, EEG, physiological measures, simulation based training, adaptive intelligent systems.

1 Introduction

In an effort to achieve a level of knowledge comparable to that which typically results from individual tutoring, innovative models of adaptive computer-based training are continually being tested and refined [1, 2, 3]. Despite these efforts, adaptive computerized training programs, though superior to traditional classroom-based settings [4], still fall significantly short of the gold standard of one-on-one instruction [5, 6, 7]. In response, previous research has investigated the use of electroencephalogram (EEG) inputs to better inform when to provide adaptive training interventions, finding that workload data, when combined with performance data, can significantly better predict future performance compared to using performance data alone [8]. Based on these data, this paper describes a preliminary investigation using workload measures from the EEG to inform adaptation choices within a Simulation Based Training (SBT) environment.

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 265–274, 2011.

1.1 Adaptive Trainers

Current adaptive trainers predominately alter instructional content or strategies for providing trainee support based on performance data alone [9]. It is hypothesized that one of the major reasons human tutors are more effective than these training systems is because a tutor has a richer source of data on which to base adaptations in instructional strategy. Specifically, in addition to performance data, an individual tutor observes a learner's behaviors, reactions, and emotional responses in real-time. Such observational data is subsequently used to inform the prescription of more optimal interventions.

Given the effectiveness of individualized human instruction, significant research has been devoted to identifying how specific components of tutoring impact learning. For example, the use of emotional recognition to inform instructional adjustments results in an improvement of 55% over traditional classroom environments [10]. Another parameter of consideration, and the focus of this paper, is the amount of working memory being utilized, or cognitive load. Working memory has a finite capacity that when exceeded, results in information loss [11]. However, too little cognitive load suggests that the learner is not fully engaged in the activity. Therefore, it has been posited that optimizing cognitive load may result in improved knowledge acquisition, assimilation in long-term memory, and eventual retrieval and application [12] and [3]. Thus, it is expected that consideration of cognitive load during learning, in addition to performance data, may provide additional information about the learner, allowing for more targeted and appropriate instructional intervention.

1.2 Cognitive Load

Cognitive load refers to the amount of working memory capacity utilized to complete a task [13]. It is broken down into three categories: Intrinsic, extraneous, and germane [13] and [14]. Intrinsic load is that which is inherent in the learning material itself. Extraneous load is the amount of working memory capacity expended on information that does not pertain to the learning material and germane load is the mental effort devoted to acquiring and developing schemas [15]. Thus, the goal of instructional design is to maximize the amount of working memory capacity devoted to germane load, minimize the amount of load devoted to extraneous information, and optimize the amount of effort devoted to intrinsic load [14]. To accomplish this goal, two pieces of information are necessary. First, we must understand how the instructional intervention will affect the cognitive load of the learner so that it can be adjusted (provided/removed/altered) as needed. Second, it is necessary to measure the cognitive load of the learner in real-time so that these adjustments can be made appropriately for each individual. In this study, instructional feedback, is the strategy utilized to improve training.

1.3 Feedback and Timing

The extensive reviews of the feedback literature find only a small, positive overall learning impact [16], [17], and [18]. However, the more adaptive the feedback is to the learner's needs, the better impact it has on the learning process [19], [20], and [21]. Therefore, if it is possible to better predict when an individual requires instructional support, it is more likely that feedback will be provided only when needed. Consequently, cognitive overload or distractions during learning can be reduced. Unfortunately, despite their expected positive impact, strategies such as feedback during a lesson or an activity can have the unwanted side effect of overwhelming working memory capacity. As such, the most efficient application of this strategy is needed. One way to improve the efficiency is to utilize real-time indices that can help improve the predictions of when learners require intervention.

1.4 EEG

The EEG can provide real-time cognitive workload measurements during learning tasks. It can be synced with the task timing and provide further insight into the amount of cognitive workload used to complete the task. To date, the EEG has been supported in the literature for its balance of usability and accuracy for measuring cognitive workload [22], [23] and [24]. Several studies have been conducted to validate its use in this context. Berka, et al., [23] was able to use Advance Brain Monitoring's (ABM) wireless headset EEG to discern varying cognitive workload levels during cognitive and assessment tasks using a combination of alpha, beta, and theta wave outputs. After conducting a similar study [25], they utilized the data to develop a generalizable model of cognitive workload that reflects changes in working memory load during learning. Subsequently, using these models, Vogel-Walcutt, et al. [8] were able to combine the output with performance data and better predict future performance when compared to predicting using past performance data alone.

1.5 Current Study

Therefore, based on previous research [8], [23], and [25], this study used the predictive data to develop a model for when to apply instructional feedback during learning in order to improve efficiency. Specifically, we compared the combination of performance and neuro-physiological indices to performance alone as indicators for when to adapt training. Objective, real-time measures of cognitive load were assessed throughout performance using an EEG [Advanced Brain Monitoring (ABM)]. Responses were then categorized into one of four classifications: hits (correct decision, high workload), misses (incorrect decision, high workload), guesses (correct decision, low workload) or slips (incorrect decision, low workload).
or slips (incorrect decision, high workload). During the performance only group, all incorrect decisions received feedback prompts (see fig. 1). In the performance plus workload group, misses and guesses received feedback (see fig. 2).

Our hypotheses were as follows:
Providing instructional feedback prompts during training based on performance plus workload data will result in:

> *H1: more effective knowledge acquisition.*
> *H2: more effective knowledge application.*
> *H3: lower perceived cognitive load during training and assessment.*
> *H4: more efficient training.*

2 Method

2.1 Participants

29 undergraduate students (12 male, 17 female) participated in the current study with a mean age of 18.9. Due to the protected nature of the assessment material, individuals were required to be United States citizens to participate. Further, they were required to have no prior knowledge of the subject matter in order to examine the impact of this training cycle on novice learners. The participants were recruited through a web-based human subject pool management software and were compensated with class credit.

2.2 Materials

Apparatus

Advanced Brain Monitoring's (ABM) wireless EEG sensor. The ABM EEG [23] sensor consists of a wireless headset containing equally distributed sensors throughout the cap and fitting over the head like a small hat. The headset is combined with the B-Alert® Software which is used to extract the user's data in real-time from the EEG so that it can be compared to the current learning experience [23] and [25]. The software output yields the probability of workload levels (the likelihood a person is experiencing high or low workload). Before analysis, all values are standardized to account for individual baseline differences.

Simulation Tasks

Threat-Assessment Training System (ThreATS). ThreATS [26] is a training tutorial used to familiarize participants to the task context. This program consists of an introductory component and two additional levels of training focused on specific decisions participants must make while using the USMC's Deployable Virtual Training Environment (DVTE) simulator. Participants viewed a series of training videos depicting the job of a Fire Support Team (FiST), and specifically, the role of the Forward Observer – Artillery (during a Call For Fire (CFF) task.

Decision-Making Assessment Scenarios (DMAS). The DMAS [26] requires participants to engage in computer simulated "Call for Fire" (CFF) scenarios. Each scenario presents participants with a battlefield that contain friend and enemy targets that are either stationary or moving. Participants must determine the threat level of the targets and use that information to decide the correct warning order (using multiple shots for moving units with imprecise coordinates or use a single shot for static units with precise coordinates), the correct sequence in which to destroy the targets, and the correct method of engagement (determine the proper type of ammunition for each target). Scenario Reference Sheets were provided during the DMAS to help participants distinguish between friend and foe targets, assess the layout of the scenario, and provide information about how to complete the DVTE radio sheet during the simulation.

Measures

Biographical Questionnaire (BQ). This 14-item questionnaire addresses personal identifiers such as age, race, gender, military experience, and degree of comfort with and frequency of use of computers.

Prior Knowledge Questionnaire (PKQ). Developed by the current authors, the PKQ consists of 4 free-response questions assessing their prior knowledge of or experience with the elements of the simulated task.

Cognitive Load Questionnaire (CLQ). The CLQ [15] is a single-item measure of perceived cognitive load during a task or set of tasks. Participants rate subjective impressions of cognitive load on a 9-point Likert-type scale, with higher scores indicating greater perceived cognitive load.

Declarative Knowledge Test (DKT). Developed by the current authors, the DKT consists of 12 factually-based multiple choice items designed to assess the extent to which the participant understands the proper terms used during the CFF task.

Procedural Knowledge Test (PKT). Developed by the current authors, the PKT consists of 7 factually-based multiple choice items designed to assess the extent to which the participant understands the proper execution of the CFF task.

Conceptual Knowledge Test (CKT). Developed by the current authors, the CKT is comprised of 9 factually-based multiple-choice items designed to assess conceptual knowledge regarding the understanding of the components involved in the task.

Integrated Knowledge Test (IKT). Developed for use in the current study, the IKT is comprised of 9 free-response items designed to assess inferences about and deeper knowledge of the FiST.

2.3 Procedure

Participants were split into two groups: Performance Only (PO) and Performance plus Workload (PW). After providing informed consent, participants completed the BQ. For those participants in the PW group, a baseline was established to account for individual differences in resting workload levels.

Introductory Phase - All participants watched a ThreATS video and completed a CLQ. They then completed the DKT, PKT, CKT and a CLQ regarding the experience of filling out the three knowledge tests. Next, a practice scenario in the DMAS was completed to familiarize participants with the simulator. Participants then completed a CLQ regarding the practice mission.

Training Phase - The two training phases of the experiment followed the same pattern. Participants were first asked to watch a training video and then answer the CLQ, after which they conducted a simulated mission in the DMAS. Following the mission, they were again asked to complete the CLQ. After each decision in the

DMAS training missions, participants received a 15-second instructional prompt in the upper right corner of their screen. Prompts were provided based on the heuristics developed for their group (see figs. 1 & 2).

Assessment Phase - Participants completed the DKT, PKT, CKT, IKT, and the CLQ regarding the knowledge tests. Next, participants completed a simulated mission, similar in complexity to that of the second training phase, however, no feedback prompts were provided to either group during this phase. The total time required to complete the entire study was approximately 2.5 hours.

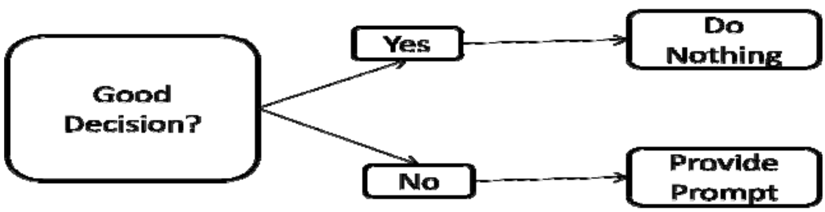


Fig. 1. Performance Group Flowchart

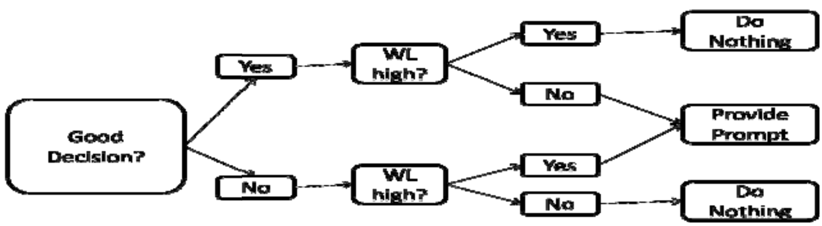


Fig. 2. Neuro-physiological + Performance Group Flowchart

3 Results

H1: Knowledge acquisition - A one-way between subjects analysis of variance (ANOVA) was used to test knowledge acquisition among both groups. Contrary to the hypothesis, the knowledge acquisition scores of the group that received feedback based on performance and neuro-physiological measures did not differ significantly from those of the performance alone group [Declarative $F(1, 27) = .194$, Procedural $F(1, 27) = 1.53$, $p = .227p = .663$, Conceptual, $F(1, 27) = .110$, $p = .742$, and Integrated $F(1, 27) = .083$, $p = .775$; see table 1]. These data suggest that the utilization of neuro-physiological data to indicate optimal intervention timing did not improve knowledge acquisition.

H2: Knowledge application - A one-way between subjects ANOVA was used to test decision making (DM) scores among both groups. Contrary to our hypothesis, DM scores did not differ significantly between groups [$F(1, 27) = .240$, $p = .628$; see table 1]. These data suggest that additional information provided by the neuro-physiological sensors to aid intervention timing did not improve DM skills.

H3: Cognitive load - A one-way between subjects ANOVA was used to compare subjective measures of cognitive load (CL) among both groups. Contrary to our hypothesis, CL scores did not differ significantly between groups during the training (F(1,27) < .001, p = .992) or assessment phases [F(1,27) = 3.51, p = .072 (after scenario); F(1,27) = .134, p = .718 (after tests); see table 2]. These data suggest that perceived cognitive load was not lowered when using neuro-physiological sensors to assist intervention timing.

H4: Efficient training - A one-way between subjects ANOVA was used to test training efficiency among both groups. Contrary to our hypothesis, no significant differences in training efficiency between groups were found [F(1,27) = 2.52, p = .124]. These data suggest that learning efficiency was not impacted by the additional specificity of the intervention timing.

Table 1. Means and Standard Deviations of Knowledge and Decision-making Scores

		Perf.		Perf. + Neuro	
Variables		M	SD	M	SD
Declarative	Pre	10.67	1.35	10.57	1.40
	Post	10.80	1.37	11.00	1.04
Procedural	Pre	15.73	2.09	14.79	3.97
	Post	17.80	3.36	15.71	5.54
Conceptual	Pre	4.93	1.67	3.93	1.49
	Post	7.53	1.36	7.36	1.50
Integrated	Post	6.70	1.59	6.92	2.42
Decision Making*	Post	-2.23	1.83	-2.57	1.86

Note. N = 29. Pre = Completed during introductory phase. Post = Completed during assessment phase. *DM was scored using penalty points; 0 was a perfect score

Table 2. Means and Standard Deviations of Cognitive Load Scores

	Perf.		Perf. + Neuro	
Variables	M	SD	M	SD
Training	5.93	1.34	5.93	1.14
Assessment	4.33	1.72	5.43	1.40
Knowledge Tests	5.60	1.24	5.43	1.28

Note. N = 29.

4 Discussion

The goal of the study was two-fold. First, we aimed to utilize neuro-physiological measures of cognitive load in combination with performance data to better classify learners' errors. Second we aimed to use these data to help inform an adaptive training system to provide tailored feedback at an optimal time to those individuals requiring intervention. Contrary to our hypotheses, however, this study failed to demonstrate positive impact on knowledge acquisition, knowledge application, perceived cognitive load, or training efficiency when incorporating these measures. The current results indicate that the proposed solution failed to positively impact training effectiveness and efficiency in its current form. However, based on observational data some inferences can be made regarding the lack of impact. For example, it was observed that few participants looked at the provided instructional feedback prompts. This may be due to the highly visual-strain of the task. It is possible that participants in neither group possessed enough available working memory capacity to attend to the supporting material. Consequently, this may account for the lack of differential findings.

4.1 Limitations of Current Study

Several limitations may have led to the lack of differences between groups including, task shedding, task overload, and small sample size. Task shedding may result as a consequence of task overload. Based on current Cognitive Load Theory (CLT) [11] and [13], it is believed that working memory has a limited capacity. It is possible that participants were already utilizing their full working memory capacity during the activity and were therefore unable to additionally attend to the feedback that was provided during this highly visual task. Since the feedback provided was visual, it may have overloaded the visual channel of working memory, leading to the inattention or inability to process the information in the feedback prompts.

Additionally, as in most studies, sample size may be another factor to account for the lack of significant differences. It is difficult to make inferential conclusions based on a weak sample size, since larger sampling error tends to be present in small samples. Accordingly, the data should be considered with caution.

4.2 Future Research

Future efforts may benefit from altering feedback placement sizing (i.e. full screen prompts). If it was the case that participants were already cognitively overloaded by the intrinsic load of the task, then it may be necessary to close the simulation in order to allow participants sufficient time to read and digest the instructional material. This could be accomplished by providing full screen prompts. In doing so, participants may experience a temporary reduction in task load which in turn may reduce task shedding. Otherwise, seeking an alternative modality all together, such as utilizing auditory prompting, may be more effective. Lastly, increasing the sample size based on a power analysis would improve the statistical power and ultimately lead to more confident conclusions.

Acknowledgement. This work is supported in part by the Office of Naval Research Grant N0000141010113. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the ONR or the US Government. The US Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

References

1. Paquette, G.: Designing Virtual Learning Centers. In: Adelsberger, H.H., Pawlowski, J.M., Collis, B. (eds.) Handbook of Information Technologies for Education and Training, pp. 249–272 (2002)
2. McLaughlin, Luca, J.: A learner-centered approach to developing team skills through web-based learning and assessment. British Journal of Educational Technology 33(5), 571–582 (2002)
3. Kalyuga, S., Sweller, J.: Rapid Dynamic Assessment of Expertise to Improve the Efficiency of Adaptive E-learning. Educational Technology Research and Development 55(3), 83–93 (2005)
4. Bayraktar, S.: A meta-analysis of the effectiveness of computer-assisted instruction in science education. Journal of Research on Technology in Education 34(2), 173–188 (2001-2002)
5. Bloom, B.S.: The 2 sigma problem: the search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher 13(6), 4–16 (1984)
6. Shute, V.J., Psotka, J.: Intelligent tutoring systems: Past, Present and Future. In: Jonassen, D. (ed.) Handbook of Research on Educational Communications and Technology, Scholastic Publications (1994)
7. Iglesias, A., Martinez, P., Aler, R., Fernandez, F.: Learning teaching strategies in an adaptive and intelligent educational system through reinforcement learning. Applied Intelligence 31, 89–106 (2009)
8. Vogel-Walcutt, J.J., Marino-Carper, T., Bowers, C., Nicholson, D.: Utilizing Learners' Internal States to Drive Feedback Decisions: A Preliminary Investigation (Manuscript under review) (2010)
9. Bolton, A., Campbell, G., Schmorrow, D.D.: Towards a closed-loop training system: Using a physiological-based diagnosis of the trainee's state to drive feedback delivery choices. In: Schmorrow, D.D., Reeves, L.M. (eds.) HCII 2007 and FAC 2007. LNCS (LNAI), vol. 4565, pp. 409–414. Springer, Heidelberg (2007)
10. Porayska-Pomsta, K., Mavrikis, M., Pain, H.: Diagnosing and acting on student affect: the tutor's perspective. UMUAI 18(1-2), 125–173 (2008)
11. Sweller, J., Van Merrienboer, J.J.G., Paas, F.G.W.C.: Cognitive architecture and instructional between feedback timing, content and modality under high cognitive workload. Educ. Psych. Rev. 10, 251–296 (1998)
12. Mousavi, S.Y., Low, R., Sweller, J.: Reducing cognitive load by mixing auditory and visual presentation modes. J. Educ. Psych. 87(2), 319–334 (1995)
13. van Merrienboer, J.J.G., Sweller, J.: Cognitive load theory and complex learning: Recent developments and future directions. Educ. Psych. Rev. 17(2), 147–177 (2005)
14. Paas, F., Renkl, A., Sweller, J.: Cognitive Load Theory: instructional implications of the interaction between information structures and cognitive architecture. Instr. Sci. 32(1-2), 1–8 (2003)
15. Paas, F., Tuovinen, J.E., Tabbers, H., Van Gerven, P.W.M.: Cognitive load measurement as a means to advance cognitive load theory. Educ. Psych 38(1), 63–71 (2003)
16. Kluger, A.N., DeNisi, A.: Effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. Psych. Bull. 119(2), 254–284 (1996)
17. Bangert-Drowns, R.L., Kulik, C.-L.C., Kulik, J.A., Morgan, M.T.: The instructional effect of feedback in test-like events. Rev. Educ. Res. 61(2), 213–238 (1991)

18. Mory, E.H.: Feedback research revisited. In: Jonassen, D.H. (ed.) Handbook of Research for Educational Communications and Technology, Simon & Schuster Macmillan, New York (2004)
19. Shute, V.J.: Focus on formative feedback. Rev.Educ. Res. 78(1), 153–189 (2008)
20. Dieterle, E., Murray, J.: Realizing adaptive instruction (Ad-In): The convergence of learning, instruction, and assessment. In: Presented at the 13th annual conference on HCII, San Diego, CA, July 19-24 (2009)
21. Wulfeck, W.: Adapting instruction. In: Presented at the 13th annual conference on human computer interaction international, San Diego, CA, July 19-24 (2009)
22. Shute, V.J., Zapata-Rivera, D.: Using an evidence-based approach to assess mental models. In: Ifenthaler, D., Pirnay-Dummer, P., Spector, J.M. (eds.) Understanding models for learning and instruction: Essays in Honor of Norbert M. Seel, pp. 23–41. Springer, Heidelberg (2008)
23. Luu, P., Poulson, C., Tucker, D.: Neurophysiological measures of brain activity: Going from the scalp to the brain. In: Presented at the 13th annual conference on HCII, San Diego, CA, July 19-24 (2009)
24. Berka, C., et al.: EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. Aviat Space Environ. Med. 78(5), 231–244 (2004)
25. St. John, M., Kobus, D.A., Morrison, J.G., Schmorrow, D.: Overview of the DARPA augmented cognition technical integration experiment. IJHCI 17(2), 131–149 (2004)
26. Berka, C., et al.: Real-time analysis of EEG indexes of alertness, cognition, and memory acquired with a wireless EEG headset. IJHCI 17(2), 151–170 (2007)
27. Vogel-Walcutt, J.J., Nicholson, D.: Applied Learning Science Team Update. In: Paper presented at the ONR HPT&E: AITE program review, San Diego, CA, January 22 (2009)

Part III

Augmented Cognition, Social Computing and Collaboration

Modelling User Behaviour and Interactions: Augmented Cognition on the Social Web

Ching-man Au Yeung and Tomoharu Iwata

NTT Communication Science Laboratories
2-4 Hikaridai, Seika-cho, Soraku-gun
Kyoto, 619-0237, Japan
{auyeung.chingman,iwata.tomoharu}@lab.ntt.co.jp

Abstract. Social sharing on the Web has become very popular in recent years. However, as the amount of information grows rapidly it becomes difficult for a user to discover relevant information. The principle of augmented cognition can be applied to help users on the Social Web. This can be done by modelling the behaviours and interactions of the users in a system in order to discover implicit relations among the users. We describe two related approaches to model user behaviours for different types of social sharing sites. We show that the methods can be used to help users identify social relations that are more important to them, as well as items that are more relevant to their interests.

1 Introduction

In recent years, Social Web applications have become very popular. Users establish social networks online and share their favourite items on the Web. However, as the amount of information and the size of one's social network grow rapidly, it becomes very difficult for a user to discover relevant information, or to know which acquaintance is more reliable as the source of relevant information. Augmented cognition aims at extending our ability to process information with computers. This principle can be applied to the Social Web by modelling the behaviours and interactions of the users in a system, discovering implicit relations among users and supporting users by recommendations.

We describe two related approaches to model user behaviours and interactions. Firstly, we consider systems where users may not have established any explicit social relation. We describe a probabilistic model [2] of how users choose to collect different items when influenced by different factors. Secondly, we consider product rating systems in which users have established explicit trust relations among themselves. We describe an extension to the standard matrix factorisation technique to estimate the strength of trust relations among users [1]. We also demonstrate that our methods give more accurate predictions of the actions and preferences of the users.

By analysing user behaviours and interactions, our methods can be used to reveal implicit social relations among users in Social Web applications. The results can help users identify social relations that are more important to them, and to retrieve information and items that are more relevant to their interests.

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 277–287, 2011.

In the next section, we briefly review works related to our research. In Section 3 and 4, we describe in detail our two proposed methods. We present some experimentation results in Section 5. Finally, we conclude the paper and mention some future research directions in Section 6.

2 Related Works

Online social networks have recently attracted the attention of researchers from different disciplines. While in the past it was difficult to collect data of social interactions among a large number of people, recently online social networking sites have provided rich data for studying real world social networks [4,12].

Several methods have been proposed to model social influences and information diffusion. In the common threshold model, a user would take an action if the number of his/her neighbours who have taken the same action reaches a certain threshold [5,7]. Song et al. [18,19], propose to use a Markov chain generated from the activity histories of the users to model information flow in a network. Tang et al. [20] introduce the notion of Topic Affinity Propagation to model social influence in a network with respect to different topics.

Our work is also related to research on recommendation systems based on social sharing data. For example, Shepitsen et al. [17] generate personalised recommendation by matching user profiles to clusters of tags obtained from a hierarchical clustering process. Bogers et al. [3] and Parra et al [16] present comparative studies of different collaborative filtering techniques on CiteULike. There are also attempts to make use of explicit social ties in social tagging systems to improve performance of collaborative filtering [8].

On the other hand, trust networks as a special form of social networks have also received much attention. They are usually implemented on product review or rating sites, on which making recommendations to users is an important application. For example, Ma et al. propose several different methods for incorporating trust relations in the standard matrix factorisation technique for collaborative filtering [14,13]. On the other hand, Jamali et al. [6] proposes TrustWalker, a random walk model that combines both trust-based recommendation and collaborative filtering based on item similarity.

Finally, there are a few works that focus on estimating the strength of trust and its effect on the opinions and ratings given by the users. For example, Matsuo and Yamamoto [15] present a hypothesis on the bi-directional effects between trust relations and item ratings. Focusing on more general social networks, Xiang et al. [22] proposes a generative model to estimate relationship strength in online social networks based on observed user interactions.

Overall, there is a substantial amount of works that investigate how user behaviours in social networks can be modelled. However, we believe that two issues have largely been overlooked. Firstly, when no explicit social ties are present, a method that can be used to model the implicit interactions among the users is not available. Secondly, when explicit relations have been established among the users, it is generally assumed that these relations truly reflect the similarity between the users. However, our study [1] shows that trust relations do not necessarily imply similarity. Hence a effective method for estimating the strength of trust relations is needed.

3 Implicit Influence in Social Sharing Sites

In this section, we focus on social sharing sites where users may not establish explicit ties among themselves. Examples of this kind include Delicious[1] and LibraryThing[2]. A social sharing system $\mathcal{S}$ of this kind can be defined as follows:

Definition 1. *A social sharing system $\mathcal{S}$ is a tuple $\mathcal{S} = \langle U, I, Y \rangle$. U is a set of users, I is a set of shared items among the users, and Y is a set of posts. A post $(u, i, t) \in Y$ represents the fact that $u \in U$ posts/adopts $i \in I$ at time t.*

To model user behaviours, we consider how users come to adopt certain items. On the one hand, users are free to introduce any new item to a social sharing site. On the other hand, users are also likely to discover something interesting in other users' collection. In many cases, social relations cannot be treated as the only means through with users influence one another. As a user's collection is publicly available in most cases, influence exists even when a social tie does not exist. In the following, we describe a model that explains how users adopt different items in a social sharing site under the influence of different factors.

3.1 A Model of Social Sharing

In our model, we assume that when a user decides to adopt an item by going through a two-step process: the user first selects a factor that would reveal to him/her a set of items, and then he/she chooses an item from that set. Here, the factor can be another user, the list of recent items or popular items, or even the user him/herself. Mathematically, the probability that user u would adopt item i at time t is defined as:

$$P(i|u,t) = \sum_{u' \in U} P(u'|u) G(i|u',t), \tag{1}$$

where $P(u'|u)$ represents the probability that u is influenced by the factor u' when he/she attempts to adopt something. $G(i|u',t)$ represents the probability that item i is chosen when factor u' is selected at time t.

The above definition is flexible and can be used to model a wide range of behaviours in different social sharing systems. When u' is a real user, $P(u'|u)$ represents the influence of u' on u. When u' is the list of popular items, $P(u'|u)$ represents how likely u would adopt something popular. Different factors can be modelled by defining a different $G(i|u',t)$. Below are some factors that are likely

---

[1] Delicious: `http://www.delicious.com/`

[2] LibraryThing: `http://www.librarything.com/`

to be found in a common social sharing system: (1) influence from other users, (2) list of recent/new items, (3) list of popular items, (4) random browsing of the items, and (5) 'Self-influence'.

For example, for the list of popular items, we can define $G(i|u', t)$ to be proportional to the number of users who have adopted i so far. $G(i|u', t)$ for other factors can be defined in a similar fashion. In addition, 'self-influence' refers to cases in which the user simply discovers an item external to the system and is the first one to adopt the item. In this case, we can assume that the user is influenced by him/herself. Let u_n be the factor of self-influence. In this case, instead of defining a particular $G(i|u_n, t)$, we can simply estimate $P(u_n|u)$ by the proportion of items of which u is the first user to adopt it in the system.

3.2 Parameter Estimation

The parameters of the model are the probabilities $P(u'|u)$. Given a history of user activities, we can estimate the parameters by maximising the log-likelihood of observed data under the constraint that $\sum_{u'} P(u'|u) = 1$. Note that if we model 'self-influence' as described above, we can estimate $P(u_n|u)$ in advance. Hence, our constraint becomes $\sum_{u'} P(u'|u) = 1 - P(u_n|u)$. Let U_A be the set of factors except u_n. The log-likelihood of the observed data is given by:

$$\log L = \sum_{(u,i,t)\in Y} \log \sum_{u'\in U_A} P(u'|u)G(i|u', t). \tag{2}$$

To estimate the parameters, we can employ the EM algorithm. In the E-step, we compute the posterior probability using the Bayes rule:

$$P(u'|u, i, t) = \frac{P(u'|u)G(i|u', t)}{\sum_{u''\in U_A} P(u''|u)G(i|u'', t)}. \tag{3}$$

In the M-step, we obtain the next estimate of the probabilities $P(u'|u)$ as follows:

$$P(u'|u) \propto \sum_{(u,i,t)\in Y} \sum_{i\in I} P(u'|u, i, t). \tag{4}$$

By iterating the above two steps until convergence, we obtain estimates for the probabilities $P(u'|u)$. As a result, we can generate recommendations to users by using $P(u'|u)$ (influential users) and $P(i|u, t)$ (items that the users may find interesting).

4 Strength of Trust Relations in Product Rating Sites

In this section, we turn our attention to social sharing sites in which users establish trust relations among themselves. In many proposals of using trust relations to generating recommendations, trust relations are usually taken at face value, i.e. it is usually assume that users who trust each other tend to have similar interests and opinions. However, our study [1] of a popular product review sites, Epinions, reveals that this is not usually the case. Hence, we believe it is necessary to estimate the true strength of the trust relations before utilising them in generating recommendations.

We propose an extension to matrix factorisation, which is commonly used to analyse user preferences in rating systems, to estimate the strength of trust relations among users. The basic matrix factorisation technique aims at revealing the latent factors that determine the ratings given by the users. Our extension allows us to explain user ratings by both latent factors as well as the different degree of influence they receive from their trusted neighbours. In other words, we consider matrix factorisation as a tool for *predicting ratings* as well as for *studying social relations among users*.

4.1 Estimating Trust by Matrix Factorisation

In general, a product rating system consists of a set U of users and a set I of items. Users express their interests or preferences in different items by rating the items with scores from a specific range. The interactions between the M users and the N items can be represented by an $M \times N$ matrix $\mathbf{R}$, where $M = |U|$ and $N = |I|$. An element r_{ui} in $\mathbf{R}$ indicates the rating of user u on item i. We represent the set of observed ratings as O.

Matrix factorisation aims at finding out the latent factors that can be used to explain the ratings given by the users. This is done by decomposing $\mathbf{R}$ into a $M \times K$ matrix $\mathbf{P}$ and a $N \times K$ matrix $\mathbf{Q}$, where K is the number of latent factors. Here, we extend standard matrix factorisation by considering trust relations among the users.

Let $\mathbf{G}$ be an $M \times M$ matrix encoding the trust network that is established *explicitly* by the users themselves. g_{uv}, an element in $\mathbf{G}$, equals to 1 if u trusts v, and 0 otherwise. In addition, we let $\mathbf{S}$ be an $M \times M$ matrix that holds the *estimated* trust relations: $s_{uv} = 0$ if $g_{uv} = 0$, and $s_{uv} \geq 0$ if $n_{uv} = 1$. The values of s_{uv} will be estimated in the matrix factorisation process, and they represent the strengths of the trust relations among the users.

Thus, for a particular tuple (u, i, r), two factors are at play in determining the rating r. Firstly, the rating is determined by the latent factor model. Secondly, u gives i a particular rating because he/she is influenced by some users he/she trusts. If the users trusted by u give high ratings, u should also tend to give high ratings.[3] Based on the above idea, a rating r_{ui} in the $\mathbf{R}$ can be estimated by:

$$\widehat{r}_{ui} = \alpha \sum_{k=1}^{K} p_{uk} q_{ik} + (1 - \alpha) \frac{\sum_{\forall v, g_{uv} > 0} s_{uv} r_{vi}}{\sum_{\forall v, g_{uv} > 0} s_{uv}} \tag{5}$$

where p_{uk} and q_{ik} are elements of the matrices $\mathbf{P}$ and $\mathbf{Q}$ respectively, and α is a parameter than controls the contributions of the two factors. In this model, the

[3] When a user trusts no other users, we assume that he/she trusts a virtual user who has rated all items by their respective mean ratings.

values of s_{uv} are to be estimated based on the differences between ratings given by pairs of users. If two users give very different ratings to the same products, s_{uv} will be small even if a trust relation exists between them.

4.2 Parameter Estimation

In this model, the parameters are p_{uk}, q_{ik}, and s_{uv}. To estimate the values of these parameters, we solve an optimisation problem that involves minimising the following regularised sum-of-squared error:

$$\min \quad \frac{1}{2}\sum_{(u,i,r)\in O}(r_{ui} - \alpha\sum_{k=1}^{K} p_{uk}q_{ik} - (1-\alpha)\frac{\sum_{\forall v, g_{uv}>0} s_{uv}r_{ui}}{\sum_{\forall v, g_{uv}>0} s_{uv}})^2 + \frac{\beta}{2}(\sum_{u,k} p_{uk}^2 + \sum_{i,k} q_{ik}^2), \tag{6}$$

where the last component is a regularisation term to avoid the parameters from taking on large values that might result in overfitting.

In our implementation, we use gradient descent to solve this optimisation problem. Parameters are initialised with random values. Let $e_{ui} = r_{ui} - \widehat{r}_{ui}$ be the error of estimation, γ be a learning meta-parameter, and there be a constraint that $\sum_v s_{uv} = 1$, the followings are the update rules for the different parameters:

$$p_{uk} \leftarrow p_{uk} + \gamma(\alpha \cdot e_{ui} \cdot q_{ik} - \beta \cdot p_{uk}) \tag{7}$$

$$q_{ik} \leftarrow q_{ik} + \gamma(\alpha \cdot e_{ui} \cdot p_{uk} - \beta \cdot q_{ik}) \tag{8}$$

$$s_{uv} \leftarrow s_{uv} + \gamma((1-\alpha) \cdot e_{ui} \cdot \frac{r_{vi}\sum_{\forall v, g_{uv}>0} s_{uv} - \sum_{\forall v, g_{uv}>0} s_{uv}r_{vi}}{(\sum_{\forall v, g_{uv}>0} s_{uv})^2}). \tag{9}$$

At the end of the training period, we should obtain a weight s_{uv} for each trust relations established between some users u and v. These can be considered as the true strengths of trust relations among the users. Together with the learnt values of the matrices $\mathbf{P}$ and $\mathbf{Q}$, we can also generate predictions for unknown ratings.

5 Experiments

To study the effectiveness of our proposed methods, we carry out experiments by using datasets collected from two popular social sharing sites, namely Delicious and Epinions.

5.1 Capturing Implicit Influence in Delicious

Delicious is a popular social bookmarking site. We use a dataset that is publicly available for research purpose [21].[4] It contains bookmarking histories of

[4] http://www.dai-labor.de/index.php?id=1726

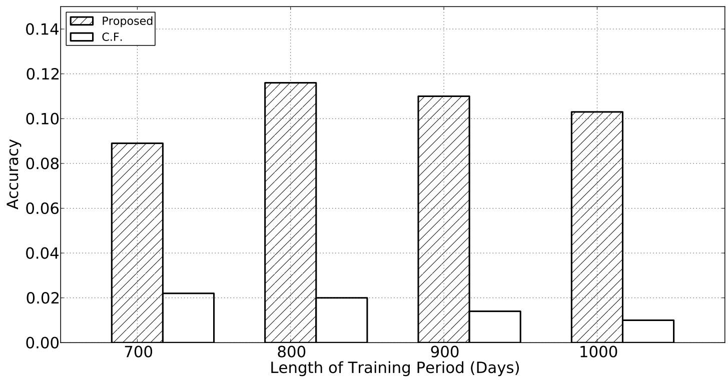


Fig. 1. Average accuracy of the model under different conditions across all 50 tags, compared to simple collaborative filtering

over 950,000 users and about 50 million bookmarks, spanning the period from September 2003 to December 2007. To avoid data sparsity, we remove users who have only adopted one item. In addition, we choose to train our model on the datasets of the 50 most frequently used tags in the dataset.

In this experiment, we study whether the probabilities $P(u'|u)$ can be used to generate accurate recommendations to the users. Firstly, we split the dataset into two parts, respecting the chronological order of the observations. We train the model using adoption histories of users in the first part, and test our model on those in the second part. For a particular user u, using the probabilities $P(u'|u)$ obtained from the training process, we predict the next items that will be adopted by u after time t.

Recommendation is done by first calculating the probability $P(d|u,t)$ using Equation 1 and then ranking the items in descending order of their probability of being adopted by the user u. We consider it successful if the next items adopted by the user appear in the top m results of the ranking. We define *accuracy*, our performance measure, as the average proportion of items that are adopted by the user and at the same time appear in our top results.

The model under evaluation incorporates all the five factors mentioned in Section 3.1. We train a separate model for each of the 50 datasets. We test our model by using different amount of training data (i.e. data from the first 700, 800, 900 and 1,000 days). For each dataset, we randomly sample 1,000 users and collect the next 10 items they adopt in the testing period. Users and items that did not appear in the training period are ignored. We then use the probabilities $P(u'|u)$ obtained in the training period and Equation 1 to come up with a ranked list of items. In our experiment we set $m = 50$. For the purpose of comparison, we implement a simple k-nearest-neighbour collaborative filtering (C.F.) method.

Figure 1 show the result of our experiment. We observe our model can be used to predict item adoption at a much higher accuracy when compared with the simple collaborative filtering algorithm. In other words, for a particular user, other users who have similar adoption histories do not necessarily possess items that are interesting to him/her. Instead, users who are found to be influential to a particular user are useful for predicting item adoption.

Our results reveal an interesting fact about the importance of considering the temporal order of adoption in making recommendations. Collaborative filtering does not consider this order and therefore is not able to distinguish between *followers* and *influencers*. However, this distinction is important because followers are more likely to adopt items from influencers but not vice versa. The probabilities $P(u'|u)$, which is asymmetric for a pair of users, is able to model this distinction, and as a result is able to generate more accurate recommendations.

5.2 Estimating Strength of Trust Relations in Epinions

Next, we conduct an experiment using data collected from Epinions, in order to evaluate our proposed matrix factorisation method. Epinions is a popular product rating sites. Users write reviews and give ratings to a wide rane of products. From the Web site of Epinions, we collect over 900,000 ratings given by about 60,000 users. We also collect all the existing trust relations among these users.

Regarding the experiment, it should be noted that there are many matrix factorisation techniques that are shown to achieve very high accuracy in predicting ratings. Our objective is not to compete with the state-of-the-art algorithms. Instead, we want to demonstrate that our method can be used estimate strength of trust relations among users, which helps us to generate more accurate rating predictions.

For comparison, we consider predictions made by the standard matrix factorisation method, which only considers the latent factors but not the social relations among the users. The metric used to measure performance is the standard root-mean-squared-error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{(u,i,r)\in T}(r_{ui} - \widehat{r}_{ui})^2}{|T|}} \tag{10}$$

where T is the set of testing data (a set of ratings). A lower RMSE means that the predictions are more accurate.

An important parameter in our model is α, which controls the ratio of contributions from the latent factor model and the trust relations. We test different value of α to investigate its effect on performance. We set the other parameters as follows: $\gamma = 0.001$ (learning meta-parameter), $\beta = 0.01$ (regularisation) and $K = 20$ (number of latent factors). Results of using 80% of the datasets as training data and the rest as testing data are presented in Figure 2.

We can observe that our proposed method constantly achieve lower RMSE than the standard matrix factorisation method. This shows that the estimated strengths of trust relations contribute to more accurate predictions. Our proposed method performs better when α is larger (in the range of [0.5,0.8]). A larger α means that a higher weight would be put on the latent factors.

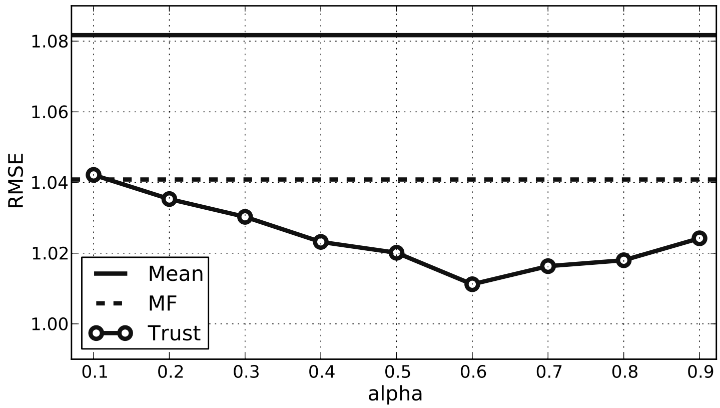


Fig. 2. RMSE of rating predictions in Epinions. 'Mean' refers to predicting a rating using the mean of existing ratings of an item. 'MF' refers to the standard matrix factorisation method. 'Trust' refers to our proposed method.

Therefore, it suggests that the latent factors are still important in generating accurate predictions. However, it also shows that an appropriate combination of the two components is crucial to achieving higher accuracy.

The optimal value of α probably depends on the specific characteristics of the system under study. In addition of setting α manually, it is possible to estimate α in the training process, or even assign different α values to different users. We plan to investigate these extensions in the future.

6 Conclusions

In this paper, we presented two methods for modelling user behaviours in social sharing sites. The first one can be used to model probabilistically user behaviours in social sharing sites in which explicit social ties are not available. It can be used to discover influential users and generate recommendations. The second one targets review and rating sites in which users maintain a trust network among themselves. The method based on matrix factorisation can be used to estimate the true strength of their trust relations by analysing their common ratings. It can also be used to predict product ratings given by the users more accurately, compared to methods that do not consider the strength of trust relations.

Our objective in this paper is to apply the principle of augmented cognition to assist users in processing huge amount of information and social ties in online social networking systems. In many cases, users can only maintain a flat list of social relations, and cannot really distinguish between acquaintances who share information that is more relevant to themselves and those who are less likely to do so. The methods presented in this paper thus augment the ability of the users to handle online relations and to discover information that is more relevant to their interests.

There are several directions for extending this work. The first direction, common to both methods, is to consider variations of the strength of (implicit) relations across different topics and contexts. For example, user A may be influenced by user B in terms of electronic goods, but by user C instead in terms of movies. In other words, there is a need to consider different contexts when modelling user behaviours. In addition, in both cases it would be interesting to rank the users by how influential they are. Hence, we are planning to develop user ranking algorithms based on the results produced by these two methods.

References

1. Au Yeung, C., Iwata, T.: Strength of social influence in trust networks in product review sites. In: WSDM 2011, ACM Press, New York (2011)
2. Au Yeung, C., Iwata, T.: Capturing Implicit Influence in Online Social Sharing. In: HT 2010, ACM Press, New York (2010)
3. Bogers, T., van den Bosch, A.: Recommending scientific articles using citeulike. In: RecSys 2008, pp. 287–290. ACM Press, New York (2008)
4. Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J., Suri, S.: Feedback effects between similarity and social influence in online communities. In: KDD 2008, pp. 160–168. ACM Press, New York (2008)
5. Goyal, A., Bonchi, F., Lakshmanan, L.V.S.: Learning influence probabilities in social networks. In: WSDM 2010, ACM Press, New York (2010)
6. Jamali, M., Ester, M.: Trustwalker: a random walk model for combining trust-based and item-based recommendation. In: KDD 2009, pp. 397–406. ACM Press, New York (2009)
7. Kempe, D., Kleinberg, J., Tardos, E.: Maximizing the spread of influence through a social network. In: KDD 2003, pp. 137–146. ACM Press, New York (2003)
8. Konstas, I., Stathopoulos, V., Jose, J.M.: On social networks and collaborative recommendation. In: SIGIR 2009, pp. 195–202. ACM Press, New York (2009)
9. Koren, Y.: Factorization meets the neighborhood: a multifaceted collaborative filtering model. In: KDD 2008, pp. 426–434. ACM Press, New York (2008)
10. Koren, Y.: Collaborative filtering with temporal dynamics. In: KDD 2009, pp. 447–456. ACM Press, New York (2009)
11. Leskovec, J., Adamic, L.A., Huberman, B.A.: The dynamics of viral marketing. ACM Trans. Web 1(1), 5 (2007)
12. Leskovec, J., Singh, A., Kleinberg, J.M.: Patterns of influence in a recommendation network. In: Ng, W.-K., Kitsuregawa, M., Li, J., Chang, K. (eds.) PAKDD 2006. LNCS (LNAI), vol. 3918, pp. 380–389. Springer, Heidelberg (2006)
13. Ma, H., Lyu, M.R., King, I.: Learning to recommend with trust and distrust relationships. In: RecSys 2009, pp. 189–196. ACM Press, New York (2009)
14. Ma, H., Yang, H., Lyu, M.R., King, I.: Sorec: social recommendation using probabilistic matrix factorization. In: CIKM 2008, pp. 931–940. ACM Press, New York (2008)
15. Matsuo, Y., Yamamoto, H.: Community gravity: measuring bidirectional effects by trust and rating on online social networks. In: WWW 2009, pp. 751–760. ACM Press, New York (2009)
16. Parra, D., Brusilovsky, P.: Collaborative filtering for social tagging systems: an experiment with citeulike. In: RecSys 2009, pp. 237–240. ACM Press, New York (2009)

17. Shepitsen, A., Gemmell, J., Mobasher, B., Burke, R.: Personalized recommendation in social tagging systems using hierarchical clustering. In: RecSys 2008, pp. 259–266. ACM Press, New York (2008)
18. Song, X., Tseng, B.L., Lin, C.-Y., Sun, M.-T.: Personalized recommendation driven by information flow. In: SIGIR 2006, pp. 509–516. ACM Press, New York (2006)
19. Song, X., Chi, Y., Hino, K., Tseng, B.L.: Information flow modeling based on diffusion rate for prediction and ranking. In: WWW 2007, pp. 191–200. ACM Press, New York (2007)
20. Tang, J., Sun, J., Wang, C., Yang, Z.: Social influence analysis in large-scale networks. In: KDD 2009, pp. 807–816. ACM Press, New York (2009)
21. Wetzker, R., Zimmermann, C., Bauckhage, C.: Analyzing social bookmarking systems: A del.icio.us cookbook. In: Proc. of Mining Social Data Workshop, collocated with ECAI 2008, pp. 26–30 (2008)
22. Xiang, R., Neville, J., Rogati, M.: Modeling relationship strength in online social networks. In: WWW 2010, pp. 981–990. ACM Press, New York (2010)

Brain Signatures of Team Performance

Silke Dodel[1,⋆], Joseph Cohn[2], Jochen Mersmann[3], Phan Luu[4], Chris Forsythe[5], and Viktor Jirsa[1,6]

[1] Center for Complex Systems and Brain Sciences, Florida Atlantic University, Boca Raton, FL, USA
dodel@ccs.fau.edu
[2] Defense Advanced Research Projects Agency[⋆⋆], USA
[3] CodeBox Computerdienste GmbH, Stuttgart, Germany
[4] Electrical Geodesics, Inc., Eugene, OR, USA
[5] Cognitive Science and Applications, Sandia National Laboratories, Albuquerque, NM, USA
[6] Theoretical Neuroscience Group, CNRS UMR 6233, Université de la Méditerranée, Institute of Movement Sciences, Faculté des Sciences du Sport, Marseille, France

Abstract. We report results from a dual electroencephalography (EEG) study, in which two-member teams performed a simulated combat scenario. Our aim was to distinguish expert from novice teams by their brain dynamics. Our findings suggest that dimensionality increases in the joint brain dynamics of the team members is a signature of increased task demand, both objective, e.g. increased task difficulty, and subjective, e.g. lack of experience in performing the task. Furthermore in each team we identified a subspace of joint brain dynamics related to team coordination. Our approach identifies signatures specific to team coordination by introducing surrogate team data as a baseline for joint brain dynamics without team coordination. This revealed that team coordination affects the subspace itself in which the joint brain dynamics of the team members are evolving, but not its dimensionality. Our results confirm the possibility to identify signatures of team coordination from the team members' brain dynamics.

Keywords: team, coordination, manifold, dimension, brain, dynamics, subspace, EEG.

1 Introduction

Real-world tasks and military missions often require the coordinated efforts of many team members for successful completion. Success within a team format is

⋆ Corresponding author.
⋆⋆ Per 5 C.F.R. 3601.108 and DoD JER 5500.7-R, 2-207: "The views, opinions, and/or findings contained in this article/presentation are those of the author/presenter and should not be interpreted as representing the official views or policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the Department of Defense."

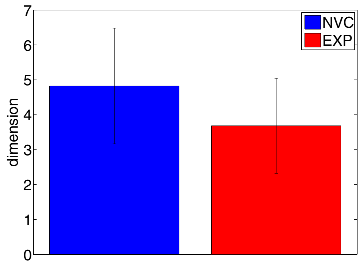
D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 288–297, 2011.

critically dependent on the ability of members of the team to work effectively together. A team whose members can work effectively together to accomplish team objectives may be considered as an expert team. Understanding how an expert team functions has been a topic of study in diverse fields, such as sports psychology and business management. The objective is simple: if we can understand how expertise develops within a team and the relevant variables that determine how an expert team operates, then training can be directed. Recently, Dodel et al. have developed a new approach for the study of team dynamics that avoids the need to understand team dynamics in detail and yet can still capture complex behaviors often expressed by expert teams [1]. This approach starts with the measurement of behaviors performed by members of a team. These behaviors evolve over time as the team advances towards the goal. The unfolding of behavior over the course of goal achievement defines a trajectory. If the same task is repeatedly performed and measured, multiple trajectories are defined and span a manifold which is shaped by constraints imposed by the task itself as well as by the interaction between the team members. Expert manifolds can be defined and used as the criterion standard for understanding novice manifolds, thus removing the need for a coach or teacher to evaluate team performers. In this study, we extend the concept of manifolds to the neural domain to help us further understand the complexity of expert team dynamics. We used dual electroencephalography (EEG) to simultaneously record the brain activity of the members of two expert teams and two novice teams, respectively. Each team consisted of two subjects which performed a test scenario in a simulated realistic and challenging combat situation. The scenario has been intentionally structured so that it necessitates extensive coordination and communication between the team members. Each trial comprises a time point ("turning point") after which simulated hostilities occur. The three specific goals of our analyzes were to (1) characterize differences between novice and expert teams based on the brain dynamics of the team members, (2) characterize differences in brain dynamics before and after the onset of simulated hostilities (3) find signatures of team coordination in the brain.

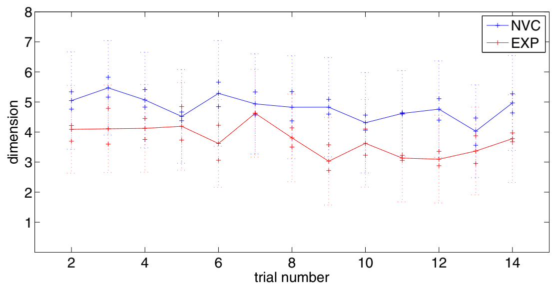
2 Results

2.1 Dimensionality of Brain Dynamics in Experts and Novices

One of our key hypotheses on team coordination states that coordinated team dynamics evolves along a particular manifold, the geometry of which reflects task related constraints as well as effects of team coordination. Such a manifold has been successfully constructed for behavioral team data and is hypothesized to exist for neural team data as well. There are multiple ways to define such a manifold. Here we computed the local subspaces of the joint brain dynamics of the team members to approximate the team manifold (see Materials and Methods). We found that the local subspaces changed rapidly over time, thereby reflecting the highly dynamic nature of the brain signals. As an approximation of the local dimensionality of the manifold, we first assessed the intrinsic dimensionality of the local subspaces in which the joint brain dynamics evolves using a sliding window of 320 ms (see Materials and Methods). Taking the average over the whole time interval of the trials, we found that novices had a higher mean intrinsic dimensionality of their joint brain dynamics than experts (Fig. 1(a)). This result was highly consistent over trials (Fig. 1(b)).



(a)



(b)

Fig. 1. Dimensionality of the joint brain dynamics (sliding window length: 320 ms) of novices (blue) and experts (red). (a) Mean over time and trials. Error bars: standard deviation. (b) Mean over time for each trial averaged over both teams of the same team level, respectively. Dotted lines: standard deviation.

2.2 Brain Dynamics before and after Onset of Simulated Hostilities

One of our hypotheses was that the degree of team coordination will be different before and after the turning point (the onset of simulated hostilities) and that this difference will be stronger for the expert team than for the novice team. Comparing the time-averaged intrinsic dimensionality of the joint brain dynamics before and after the turning point, we found that there is a tendency towards higher dimensionality after the turning point, in particular for the experts (Fig. 2).

The overall effect is small and can be considered only as a trend, but the effect was highly consistent over trials as assessed by computing significance values from the binomial distribution. In the expert team the effect of dimensionality increase was highly consistent with a significance of $p < 0.001$. The effect was less consistent for the novice team ($p < 0.03$). In addition we assessed the effect also in the single subject data, where it was significantly consistent in both subjects of the expert team ($p < 0.001$ and $p < 0.011$, respectively), but only in subject 2 ($p < 0.005$) of the novice team. Furthermore in the experts the increase in mean dimensionality after the turning point occurred congruently in both joint and single subject brain dynamics for most of the trials. This was not the case in the novice team.

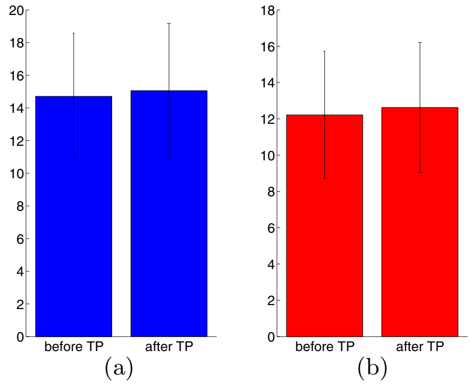


Fig. 2. Mean dimensionality of the joint brain dynamics before and after the turning point (TP), which marks the onset of simulated hostilities. Sliding window length: 10s. (a) novices (blue), (b) experts (red), error bars: standard deviation.

2.3 Manifold Spanned by the Joint Brain Dynamics of the Team

To get insight about the overall subspace in which the joint brain dynamics on the manifold evolves, we computed a team subspace from the local subspaces (see Materials and Methods). The team subspace for one expert team in a 20 s interval around the turning point is shown in Figure 3. Three of the four dominant spatial modes show localized activity over right prefrontal electrodes in one or both team members, indicating that joint activity in this area in both team members could play a role in team cognition. At each point in time the team subspace accounts for about 10-40% of the data as shown by the reconstruction quality of the data with respect to the subspace (Fig. 3(b)) with the highest reconstruction qualities occurring after the turning point.

2.4 Signatures of Team Coordination

So far we have analyzed signatures of team performance by approximating the hypothesized team manifold dynamically by identifying subspaces of the joint brain dynamics of the team members. Here we extend this approach to identify the aspect of coordination in a team. To determine signatures of team coordination we created surrogate teams with the same performance level as the original teams but without team coordination. This was achieved by pairing the data from two subjects of the same teams, but from different trials (e.g. data from trial 3 in subject 1 combined with data from trial 4 in subject 2). The surrogate data hence serves as a baseline to isolate effects of team coordination. When surrogate data are constructed the alignment of coordinating dynamic elements is lost and scrambled, which is equivalent to a loss of team coordination. As a consequence, the team manifold should change for the surrogate data set.

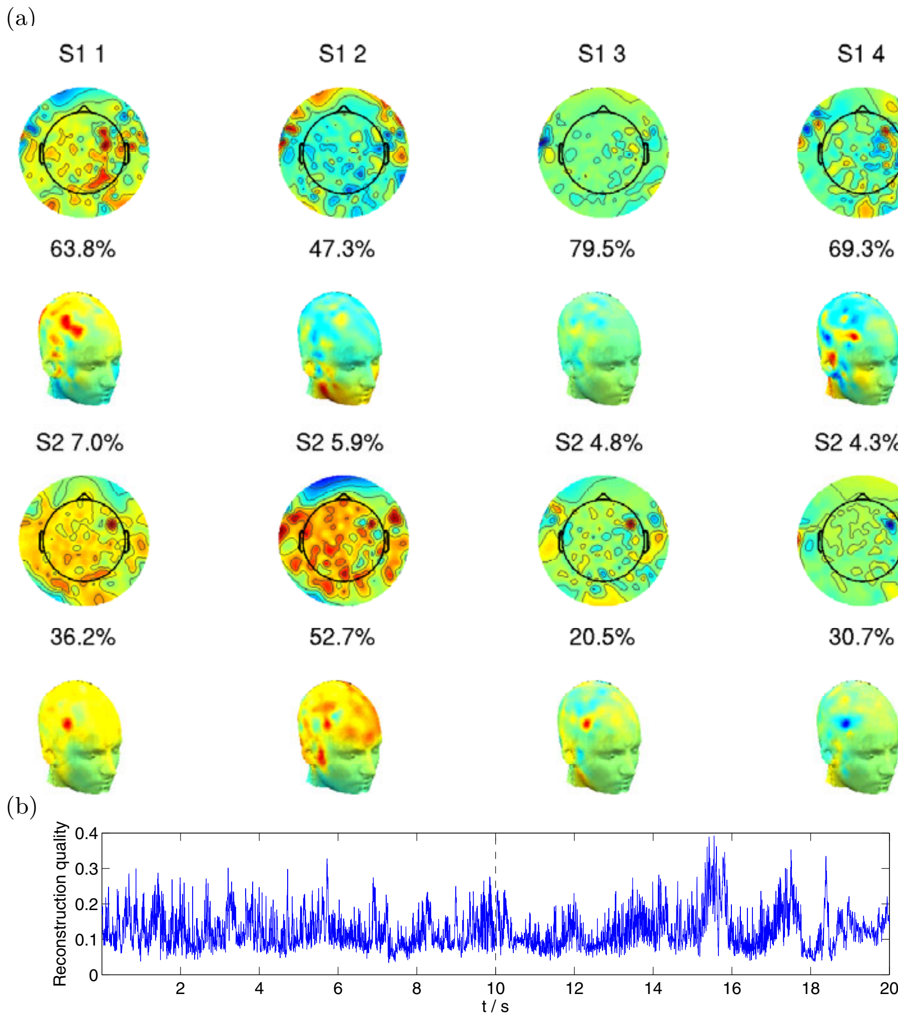


Fig. 3. Subspace in which the joint brain dynamics of the expert team evolves in an interval of 20 s around the turning point (dashed line). (a) Topographic maps and head plots of the first four spatial modes of the 16-dim. team subspace of expert team 1. For visualization purposes each subject has its own color map centered at zero (green). Percentages on top of the head plots: contribution of each subject to the total spatial mode. Percentages on top of the topographic maps of subject 2: contribution of the spatial mode to the total variance of the data in the team subspace. (b) Mean reconstruction quality of the expert team data with respect to the team subspace.

We first tested this hypothesis by comparing the dimensionality of the team subspaces of true and surrogate teams and second we computed the similarity of the two types of subspaces (see Materials and Methods). While we found that the dimensionality of the two types of subspaces was essentially the same, the team subspaces of the real teams differed from the team subspaces of the surrogate teams in about three dimensions. We could hence identify a three dimensional subspace of joint brain dynamics of the team members which is specifically related to team coordination. The dominant spatial modes of this subspace had similar features as the dominant spatial modes of the team subspace (cf. Fig. 3), in particular it also showed localized activity over right prefrontal electrodes in both team members.

3 Materials and Methods

3.1 Data Acquisition and Preprocessing

High density EEG data (256 channels) with a sampling rate of 250 Hz was acquired simultaneously from the two team members of four teams (two novice teams, two expert teams) in a simulated combat scenario where the subjects were coordinating to accomplish a common goal. Data from 14 trials were acquired from each team. Each trial lasted about 20 minutes and comprised a time point ("turning point") after which simulated hostilities occurred. We aligned all data sets with respect to the turning point and equalized the time intervals before and after the turning point, resulting in trial lengths of about 16 and 26 minutes for experts and novices, respectively. To account for transiently faulty electrodes, in every data set of each subject we discarded the 30 electrodes with the highest variances over time, still leaving almost 90% of the electrodes available for analysis. The data was cleaned from eye blink artifacts by an in-house program developed by EGI.

3.2 Team Manifold and Dimensionality

The measured brain dynamics of the teams evolve in a high-dimensional state space (here: 2×256 channels $= 512$ dimensions). A team manifold was approximated by computing subspaces of the joint brain dynamics of the two team members using a sliding window. For each window the subspace was computed by performing a singular value decomposition (SVD) of the joint data of subjects 1 and 2, created by concatenating the channel data of the two subjects at each point in time. Prior to concatenating the total spatio-temporal variance of the data of the two subjects was equalized to eliminate effects of inter-subject variability in signal strength. Dimensionality of the manifold was assessed by the dimensionality d of the subspaces for each time window from the SVD of the data matrix according to

$$d = N + 1 - \sum_{i=1}^{N} \frac{\sum_{j=1}^{i} \sigma_j^2}{\sum_{l=1}^{N} \sigma_l^2} \tag{1}$$

where N is the total number of singular values σ_j, $j \in \{1, \ldots, N\}$.

3.3 Team Subspace and Reconstruction Quality

The team subspace was determined by performing an SVD over the concatenation of the basis vectors of all subspaces of the joint brain dynamics of the team members, weighted by their singular values and using their respective dimensionality. Reconstruction quality $r(t)$ of the data with respect to the team subspace was assessed at each point in time t by

$$r(t) = \frac{\|\mathbf{P}\mathbf{v}(t)\|^2}{\|\mathbf{v}(t)\|^2} \tag{2}$$

where $\mathbf{P}$ is the projection matrix onto the team subspace and $\mathbf{v}(t)$ is the joint data at time t. Furthermore $r(t) \in [0, 1]$ with 1 indicating perfect reconstruction.

3.4 Similarity of Team and Surrogate Subspace

We assessed the similarity between the team subspaces of the true team and the surrogate team by determining whether they had a common subspace. A common subspace can be determined by solving an eigenvalue problem as follows. Let $\mathbf{X}$ be an $m \times n$ matrix, $n \leq m$, the column vectors of which span the team subspace, and $\mathbf{Y}$ an $m \times k$ matrix, $k \leq m$, the column vectors of which span the surrogate subspace. The vectors in the common subspace may be written in both bases $\mathbf{X}$ and $\mathbf{Y}$ using an $n \times 1$ vector $\mathbf{a}$ and a $k \times 1$ vector $\mathbf{b}$ with the relation

$$\mathbf{X}\mathbf{a} = \mathbf{Y}\mathbf{b} \tag{3}$$

We can solve eq. (3) for $\mathbf{a}$ or $\mathbf{b}$ using the pseudo-inverses $\mathbf{X}^+ := (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ and $\mathbf{Y}^+ := (\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T$ of $\mathbf{X}$ and $\mathbf{Y}$, respectively. This yields

$$\mathbf{a} = \mathbf{X}^+\mathbf{Y}\mathbf{b} \tag{4}$$

$$\mathbf{b} = \mathbf{Y}^+\mathbf{X}\mathbf{a} \tag{5}$$

Inserting eq. (5) into eq. (4) yields

$$\mathbf{a} = \mathbf{M}\mathbf{a} \tag{6}$$

with the $n \times n$ matrix $\mathbf{M} := \mathbf{X}^+\mathbf{Y}\mathbf{Y}^+\mathbf{X}$. Eq. (6) is an eigenvalue problem which is solved by the eigenvectors to the eigenvalue 1 of the matrix $\mathbf{M}$. The multiplicity of the eigenvalue 1 equals the dimensionality of the common subspace, and the eigenvectors $\mathbf{a}$ can be used to obtain a basis of the common subspace in terms of the basis $\mathbf{X}$ of the team subspace. The subspace reflecting team coordination consists in the complement of the common subspace and can be constructed from the eigenvectors to the eigenvalues $\neq 1$ of $\mathbf{M}$.

4 Discussion

Behavioral signatures of coordination typically are based on the variability of measures that characterize behavioral dynamics. The underlying assumption is that coordination induces coupling between individuals and hence decreases the variability related to independent behavior. Typical examples include the variance of the relative phase (see [2] for a review) and variability across manifolds (Uncontrolled Manifold (UCM)) [3]. More recent approaches to characterize coordination include the description of behavioral processes [4–8] using mathematical representations of flows on manifolds in the state space of a given system. The beauty of this approach is its generality and its ability to account for complex behaviors which may be represented by curved manifolds in state space. Dodel et al. [1] applied this approach to team coordination. They reconstructed the flow on the manifold in the shared behavioral state space of all team members and demonstrated that this manifold may serve as a tool to compare different levels of team performance against a gold standard (the expert team performance). Here we have taken this effort one step further, and considered the existence of a shared brain state space in analogy to the behavioral state space of [1]. The brain state space is spanned by all electrodes of all team members. Our hypothesis was that, if team coordination is indeed reflected in neuroimaging data, a common manifold exists in the brain state space along which the joint brain dynamics of the team members evolves. Such manifolds have been recently postulated to dominate the resting state brain dynamics of individuals [9–12], and may reflect temporal or spatial modulation in the brain [13].

In the current study we have attempted to identify manifolds of brain dynamics which reflect processes of team performance and coordination in the members of a team. Our specific goals were to (1) characterize differences between novice and expert teams based on the brain dynamics of the team members, (2) characterize differences in brain dynamics before and after the onset of simulated hostilities (3) find signatures of team coordination in the brain. Our results indicate that novice and expert teams exhibit different characteristics in their brain dynamics as measured by dual EEG when performing a highly nontrivial ongoing task. In particular brain dynamics in expert teams were lower dimensional than in novice teams. Increased task demand was associated with a consistently higher dimensionality in the expert team, whereas this effect was less consistent in the novices.

Investigation of team cognition from behavioral measures has a long tradition [14–16]. More recently, there is also increasing interest in analyzing how interactions between humans are reflected in the individuals' brain activity [17–23]. Our approach is different from the latter approaches in that it allows identifying team coordination by directly comparing signatures of joint brain activity in team members with and without team coordination. This is achieved by producing surrogate data in which team coordination is selectively removed while retaining the performance level of the individual team members. The introduction of a surrogate team provides a baseline and could be used to develop significance measures. Ideally surrogate data should consist of trials with the exact same behavior of the subjects as in the original trial, but without team coordination. To a first approximation of this ideal we used data where the individual subject data was taken from different trials of the same task. Using surrogate data we identified a subspace which was related specifically to team coordination. The spatial modes spanning this subspace showed localized joint activity over the right prefrontal electrodes in both team members, which indicates that this area could play a role in team coordination.

Given the complex nature of brain imaging signals (spiking neuron networks generate oscillatory dynamics that is only partially picked up by non-invasive brain imaging), we did not pursue a detailed reconstruction of the shared manifold in brain state space, but nevertheless found some evidence of its existence. The phenomenological characterization of the manifold in terms of non-invasive brain imaging is not unique. Therefore the reconstruction of the manifold could be improved by using various derived properties of the data, such as separate frequency bands, instead of the raw data. Furthermore, here we have used only team level (novice or experts) and task demand (before and after onset of simulated hostilities) as task descriptors. More task descriptors such as communication status between the team members or information about situational content could be used to refine the analysis and identify brain activity patterns related to specific team situations.

5 Conclusion

This study shows proof of concept that even in a highly uncontrolled real-world task setting it is possible to identify signatures of team performance and team coordination from the brain dynamics of the team members. In particular, our results indicate that dimensionality increases in brain activity is a signature of increased task demand, both objective, e.g. increased task difficulty, and subjective, e.g. lack of experience in performing the task. An integral part of our approach is the identification of brain signatures of team coordination by means of surrogate team data. While our results do not support the use of dimensionality as a signature of team coordination, we were able to identify a subspace of brain dynamics which is related to team coordination. This is the first evidence that a manifold of team coordination may exist in the brain state space across all team members. If that is the case, the manifold is a prime candidate for a neural biomarker of team coordination.

References

1. Dodel, S., Pillai, A., Fink, P., Muth, E., Stripling, R., Schmorrow, D., Cohn, J., Jirsa, V.: Observer-independent dynamical measures of team coordination and performance. In: Danion, F., Latash, M.L. (eds.) Motor Control, pp. 72–103 (2010)
2. Kelso, J.A.S.: Dynamic Patterns: The Self-Organization of Brain and Behavior. The MIT Press, Cambridge (1995)
3. Scholz, J.P., Schöner, G.: The uncontrolled manifold concept: identifying control variables for a functional task. Exp. Brain. Res. 126(3), 289–306 (1999)
4. Jirsa, V.K., Kelso, J.A.S.: The Excitator as a Minimal Model for discrete and rhythmic movement Coordination. Journ. Motor. Behav. 37(1), 35–51 (2005)
5. Huys, R., Studenka, B.E., Zelaznik, H.N., Jirsa, V.K.: Distinct timing mechanisms are implicated in distinct circle drawing tasks. Neuroscience Letters 472(1), 24–28 (2010)

6. Huys, R., Fernandez, L., Bootsma, R.J., Jirsa, V.K.: Fitts law is not continuous in reciprocal aiming. Proc. R. Soc. B 277(1685), 1179–1184 (2009)
7. Huys, R., Studenka, B.E., Rheaume, N.L., Zelaznik, H.N., Jirsa, V.K.: Distinct Timing Mechanisms Produce Discrete and Continuous Movements. PLoS Comput. Biol. 4(4), e1000061 (2008), doi:10.1371/journal.pcbi.1000061
8. Calvin, S., Jirsa, V.K.: Perspectives on the Dynamic Nature of Coupling in Human Coordination. In: Huys, R., Jirsa, V.K. (eds.) Nonlinear Dynamics in Human Behavior. SCI, vol. 328, pp. 91–114. Springer, Heidelberg (2010)
9. Deco, G., Jirsa, V.K., McIntosh, A.R.: Emerging concepts for the dynamical organization of resting state activity in the brain. Nature Reviews Neuroscience 12, 43–56 (2011)
10. Mcintosh, A.R., Kovacevic, N., Lippe, S., Garrett, D., Grady, C., Jirsa, V.K.: The development of a noisy brain. Archives Italiennes de Biologie 148, 323–337 (2010)
11. Ghosh, A., Rho, Y., McIntosh, A.R., Kötter, R., Jirsa, V.K.: Noise during rest enables the exploration of the brain's dynamic repertoire. Plos Comp. Biol. 4(10), e1000196 (2008), doi: 10.1371/journal.pcbi.1000196
12. Deco, G., Jirsa, V.K., Sporns, O., McIntosh, A.R., Kötter, R.: The Key Role of Coupling, Delay and Noise in Resting Brain Fluctuations. PNAS 106, 10302–10307 (2009)
13. Banerjee, A., Tognoli, E., Assisi, C., Scott, J., Jirsa, V.: Mode Level Cognitive Subtraction (MLCS) quantifies spatiotemporal reorganization in large-scale brain topographies. NeuroImage 42(2), 663–674 (2008)
14. Salas, E., Cooke, N.J., Rosen, M.A.: On Teams, Teamwork, and Team Performance: Discoveries and Developments. Human Factors 50(3), 540–547 (2008)
15. Cooke, N.J., Gorman, J.C., Duran, J.L., Taylor, A.R.: Team cognition in experienced command-and-control teams. Journal of Experimental Psychology: Applied 13(3), 146–157 (2007)
16. DeChurch, L.A., Mesmer-Magnus, J.R.: The Cognitive Underpinnings of Effective Teamwork: A Meta-Analysis. Journal of Applied Psychology 95(1), 32–53 (2010)
17. Tognoli, E., Lagarde, J., DeGuzman, G.C., Kelso, J.A.S.: The phi complex as a neuromarker of human social coordination. PNAS 104(19), 8190–8195 (2007)
18. Lindenberger, U., Li, S.-C., Gruber, W., Müller, V.: Brains swinging in concert: cortical phase synchronization while playing guitar. BMC Neuroscience 10, 22–34 (2009)
19. Stevens, R.H., Galloway, T., Berka, C., Sprang, M.: Can Neurophysiologic Synchronies Provide a Platform for Adapting Team Performance? In: Schmorrow, D.D., Estabrooke, I.V., Grootjen, M. (eds.) FAC 2009. LNCS, vol. 5638, pp. 658–667. Springer, Heidelberg (2009)
20. Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., Garnero, L.: Inter-Brain Synchronization during Social Interaction. PLoS ONE 5(8), e12166 (2010), doi:10.1371/journal.pone.0012166
21. Schippers, M.B., Roebroeck, A., Renkena, R., Nanettia, L., Keysers, C.: Mapping the information flow from one brain to another during gestural communication. PNAS 107(20), 9388–9393 (2010)
22. Anders, S., Heinzleb, J., Weiskopf, N., Ethoferd, T., Haynes, J.-D.: Flow of affective information between communicating brains. NeuroImage 54(1), 439–446 (2011)
23. Astolfi, L., Toppi, J., Fallani, F.V., Vecciato, G., Salinari, S., Mattia, D., Cincotti, F., Babiloni, F.: Neuroelectrical Hyperscanning Measures Simultaneous Brain Activity in Humans. Brain Topogr 23, 243–256 (2010)

Team Coordination Dynamics and the Interactive Approach: Emerging Evidence and Future Work

Jamie C. Gorman

Texas Tech University, Lubbock, TX

Abstract. In the study of coordination and teamwork, the primacy of team interaction is emphasized in an interactive approach. The interactive approach lies in stark contrast to the traditional, shared cognition approach to understanding team cognition. An overview of team coordination dynamics, an interactive approach rooted in nonlinear dynamics, is provided. Results from a series of experiments on team coordination dynamics are summarized. Finally, future research directions, inspired by those results, are considered.

Keywords: Nonlinear dynamics, Teams, Team coordination, Teamwork.

1 Introduction

Teamwork enables groups of individuals to accomplish complex tasks in a variety of work settings, including business, military, medical, and educational settings. Not surprisingly, interest in team cognition and coordination, and their relation to team effectiveness, is on the rise. Ideally, during the execution of a complex cognitive task, teams allow cognitive work to be efficiently distributed across a heterogeneous division of knowledge, skills, and abilities. Certainly, a good theoretical understanding of this phenomenon—*team cognition*—and its real-time coordination would have the potential to benefit work in sociotechnical environments.

1.1 Shared Cognition and the Interactive Approach

The search for mechanisms of effective teamwork and its development has traditionally centered on shared knowledge structures across team members. This approach is known as the shared cognition approach [1, 2, 3]. In the shared cognition approach, team cognition—the ability to think and react as a team—is conceptualized by two central questions concerning shared knowledge structures: What overlaps? and What is complementary? According to this approach, the development of overlapping and complementary knowledge is tantamount to shared cognition and, therefore, team cognition.

Shared knowledge structures can be quite complex. A mental model is a knowledge structure that allows an individual to describe, explain, and predict states of a system [4]. A *shared mental model* (SMM) is generally defined as the degree of overlap of team members' mental models; hence, the SMM construct addresses the question of what overlaps. SMMs extend the mental model construct to teamwork [5, 6]: here, the team is the system and a SMM allows team members to describe,

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 298–307, 2011.

explain, and predict team interaction. At the same time, a *transactive memory system* (TMS) [7, 8, 9] is structured knowledge about whom to ask for the correct piece of information [10, 11, 12]. By viewing shared cognition as relational knowledge across team members, TMSs address the question of what is complimentary.

The paradigm for linking shared cognition to team effectiveness is the Input-Process-Output (IPO) causal framework [13, 14, 15], wherein shared knowledge is the input, team interaction the process, and team effectiveness the output [16, 17, 18]. In the IPO framework shared cognition is indirectly linked to team effectiveness via the moderating variable, team interaction processes. Multiple studies have empirically linked shared cognition to team effectiveness using the IPO causal framework. For instance, the development of SMMs through pre-task planning has been empirically linked to the anticipation of team member interaction during task performance, thereby allowing for increased efficiency of communications [19]. In this way, SMMs may provide an adaptive coordination mechanism—so-called *implicit coordination*—during times of high workload [20, 21]. Research has similarly tied longevity of team membership to the development of TMSs, which are linked to adaptive, "backing-up" team interactions [22]. That is, long-tenured teams that have developed such a knowledge structure have members that tend to know when, and from whom, to request and accept help, ultimately leading to increased team effectiveness, particularly in high-stress environments.

Though the characterization of effective teamwork as the shared contents of team member knowledge has been (to some degree) empirically validated in some studies, it has been rejected in others. Over a variety of studies the development of shared knowledge (either overlapping or complimentary) is not concomitant with the acquisition of team effectiveness [23, 24], but the development of effective team interaction processes is consistently linked with enhanced team effectiveness [25, 26]. Therefore, the central construct of team interaction processes may provide a better foundation than shared knowledge for a theory of team cognition, especially as team cognition relates to team effectiveness [27].

Perhaps most limiting is the view that something as dynamic as effective teamwork could be based on a structure. Structures are (relatively speaking) static entities and, as it stands, the substrates upon which they may be modified during task performance are neither explicitly nor implicitly specified in the shared cognition approach. However, team interaction processes are inherently dynamic and may provide the appropriate substrate. Of course, shared cognition theorists have become aware of the inconsistency between static knowledge and dynamic task performance and have adjusted the theory accordingly (e.g., "complilational emergence" of shared cognition through team-member interaction; [28]. Indeed, it has recently been acknowledged that a feedback arrow from O to I may be required to account for new results [29]. On the other hand, if team knowledge is viewed not as antecedent to, but as incidental to, team interaction processes, then team interaction becomes the substrate upon which aspects of teamwork may be modified.

Static constructs may not satisfactorily explain something as dynamic as effective teamwork. Team interaction, on the other hand, seems to fit the bill because it is an inherently dynamic process: During teamwork, an appropriate level of awareness and agency to act as a team must be dynamically assembled through team-member interaction. Further, team-level awareness and agency to act is partially determined as the situation unfolds. Therefore, the content of team interactions may rely just as heavily upon the exigencies of an unfolding situation, as upon static knowledge. Such a viewpoint is the thesis of an interactive approach to team cognition (as opposed to the shared cognition approach). The next section describes team coordination dynamics, an interactive approach rooted in nonlinear dynamics, for studying the development of coordination as it relates to team cognition and effectiveness.

1.2 Team Coordination Dynamics

Team coordination is a process wherein team members adjust their interactions in accordance to a changing environment. As a skill, it develops when teams interact in a dynamic environment. Team interaction patterns unfold in discontinuous bursts and lulls of activity that can be described using nonlinear dynamics. Inspired by concepts from ecological psychology [30, 31], coordination dynamics [32], and Haken's [33] synergetics, team coordination dynamics employs nonlinear dynamics as a paradigm for understanding how team coordination skill develops. Just as nonlinear dynamics pervade coordination in motor and molecular systems; team coordination dynamics is the paradigm used for understanding the emergence of coordinated behavior in teams. Of course, the concept of team coordination considered from this viewpoint is quite different from shared cognition.

To say that coordination is *achieved* by team members would be to say that coordinated behavior was somehow stored within the members themselves. And that is very much how team coordination has been viewed traditionally, from a shared cognition perspective. Certainly teams have members with requisite knowledge; teams would fail otherwise. Teams must communicate this knowledge cooperatively. However, it is not enough to just communicate cooperatively. The purpose of assembling a team is to set up a system that can handle a complex, changing task environment that no one individual, working alone, could comprehend. If that is the case then it is clear that there will be changing environmental conditions that a team can experience but an individual cannot. Now, can coordination be stored in these as-yet unexperienced conditions of the task environment? It might be better to say that coordination is not stored at all. Rather, it is dynamically assembled across team interactions, such that the team remains focused on achieving its goals in the context of a dynamic environment.

2 Emerging Evidence

In this section I describe research on three-person teams conducted in an uninhabited air vehicle (UAV) simulator located in Mesa, AZ [34]. In the UAV task, the three team members—pilot, navigator, and photographer—must interact over headsets to photograph reconnaissance targets over a series of 40 min missions. In each mission, teams have to photograph 11-12 reconnaissance targets. Each team member has their own computer workstation that displays information specific to that team-member role as well as general flight information (e.g., current heading, altitude, and speed). Team members were seated in the same room with their backs to each other, such that they only communicated verbally over the headsets.

2.1 Team Mixing

Gorman, Amazeen, and Cooke [35] had three-person (UAV) teams (who had already been trained to photograph reconnaissance targets) return after a retention interval with either the same team members ("Intact") or different team members ("Mixed"). An order parameter—a collective measure of team interaction at photograph points—was developed to measure team coordination dynamics.

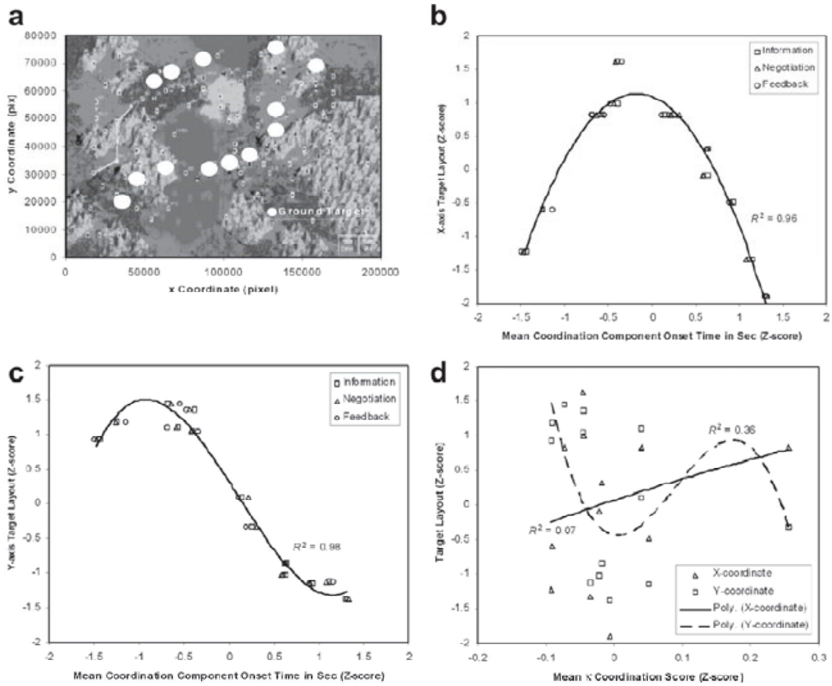


Fig. 1. The environmental layout of the UAV targets in two dimensions [x, y] (a); the relationship between target x coordinates and κ components (b); the relationship between target y coordinates and κ components (c); the relationship between [x, y] and κ. (Reprinted from [36].)

The team interactions that were the components of the order parameter were the three primary team member functions for photographing each target: (1) the navigator sends target *information* (I) to the pilot; (2) the pilot and photographer *negotiate* (N) an appropriate airspeed and altitude; and (3) the photographer provides *feedback* (F) on the status of the target photograph. Timestamps corresponding to the exact time each of these interaction functions occurred for each UAV target were collected during the experiment. These functions must be dynamically combined in a specific order to photograph each target, such that I→N→F. The order parameter, which captures these relationships, is called κ and is given in Equation 1. κ was computed for each target and is a dimensionless (unit-free) quantity because time cancels in the numerator and denominator. κ < 1 indicates a lack of coordination relative to the

I→N→F relationship, and $\kappa = 1$ is indeterminate. As long as the I→N→F relationship is satisfied, $\kappa > 1$; however, fluctuations of $\kappa > 1$ are indicative of variance in team coordination from target to target. Now, if team coordination is the dynamic assembly of team interactions in accordance to changes in the task environment, then κ should fluctuate with the changing task environment. Indeed, fluctuations ("adjustments") in κ components correspond precisely to the dynamic layout of targets in the UAV environment [36] (Figure 1).

$$\kappa_i = \frac{\text{time}(\text{F}_\text{i}) - \text{time}(\text{I}_i)}{\text{time}(\text{F}_\text{i}) - \text{time}(\text{N}_\text{i})} \qquad (i\ =\ 1, 2, \ldots, \#\text{targets}). \tag{1}$$

Nonlinear dynamical indices, calculated from teams' κ time series, revealed that Mixed teams had more flexible coordination, as indexed by the Hurst exponent, and greater coordination stability, as indexed by the Lyapunov exponent, than Intact teams. Counter to expectations from a shared cognition approach, Mixed teams were also more adaptive as indexed by the correlation between stability and overcoming roadblocks, which are novel events that perturb coordination dynamics. These results further suggest that Mixed teams were better able to adjust their interaction dynamics to the changing demands of the task environment. Both the coordination dynamics and subjective process ratings were correlated with team effectiveness [37], but the coordination dynamics provided insight into the Mixed team advantage: By allowing teams to experience more of the possible relations that could occur, mixing team members may have allowed teams to spontaneously self-organize coordination attractors that remained stable under novel task requirements.

2.2 Perturbation Training

In a second experiment, also conducted in the UAV task context, Gorman, Cooke, & Amazeen [38] trained teams by perturbing elements of the κ order parameter during task acquisition. This was accomplished by actively interrupting either the I, N, or F functions, forcing teams to find new ways to work around those perturbations. Different coordination links in the I→N→F sequence were perturbed multiple times during training. We found that perturbation-trained teams performed as well as teams trained using traditional methods (e.g., cross-training) under routine conditions and outperformed those teams under novel conditions, although cross-trained teams had higher levels of shared knowledge. Similar to the effect of team mixing, perturbing coordination during task acquisition may allow teams to self-organize coordination attractors that remain stable across a range of possible task conditions. This interpretation is also consistent with recent research in the motor literature that suggests unpredictable practice elements can lead to acquisition of adaptive skill [39].

3 Future Work

Much of the next round of research on team coordination dynamics will focus on three intertwined areas: real-time dynamics; perturbation training; and cross-level team coordination dynamics.

3.1 Real-Time Dynamics of Team Interaction

Given that adaptive team synergies are created, ideally we would like to detect threats to those synergies as they occur. Real-time dynamical analysis of team communication [40] is a method for detecting team coordination anomalies as they occur. The primary challenge of dynamics in real time is a methodological one. Nonlinear dynamics are usually analyzed using long time series (e.g., a minimum of 1,024 observations). However, to detect threats to team synergy, dynamical parameters have to be calculated on incoming data streams of unknown length. The approach we have employed is to calculate the dynamical pattern of interest (e.g., the Lyapunov exponent) for the incoming data stream using k different windows of size 2^k. In principle, this is a moving window analysis, except that there are multiple windows of different sizes, and the data are moving through the window.

In theory, a parameter estimate at smaller window sizes should contain a great deal of error due to the small sample size. Conversely, moving to larger window sizes should yield better estimates. The variability (*SD*) of the estimate for each window size, updated in real time, is then viewed across window size using linear regression: $\log_2(SD) = v*k$. The slope v is always negative and has a maximum of zero (i.e., if the variability scaled perfectly with window size). Steeper (more negative) values of v indicate that the dynamical parameter is relatively stable as window size increases: the team communication dynamics are not changing. However, if v begins to fluctuate toward zero, then variance persists across increasing window sizes: the dynamical parameter estimate is becoming unstable across window sizes. This is characteristic of team communication dynamics undergoing change.

When team communication dynamics do not change, the team is operating within the bounds of their own *intrinsic* dynamics. However, when there is a critical change in the environment—the *extrinsic* dynamics—the team's communication dynamics must also change to accommodate the perturbation. We have validated this approach using team communication data in which an outside confederate intelligence agent briefly interrupts team interaction. In that study [40], the intelligence agent perturbation caused a significant upward shift in v, but the team recovered their intrinsic dynamics (i.e., a negative shift in v) soon after the intelligence agent was withdrawn.

3.2 Perturbation Training

Similar to the notion of differential learning [39], perturbation training involves randomly inserting off-task elements into the coordination process during task acquisition. Ideally, training should exercise coordination variability to match the variability of the post-training environment. The general idea is that exposure to noisy interactions during training may allow teams to transfer training to even noisier post-training environments. The mechanism though which teams acquire this skill, perturbation training, may create opportunities for bottom-up organization of new coordination links in response to random changes in the training environment.

Because, in perturbation training, the organization of new coordination links is not knowledge-driven, team members are compelled to unconsciously adjust interactions in unplanned ways to maintain a balance between team dynamics and environmental dynamics. If task acquisition takes place under these conditions, then learning may continue to occur when teams transfer their coordination skill to novel post-training situations. In other words, because novel (unexpected) events in the post-training environment are similar to the conditions of initial task acquisition, perturbation training may transform novel post-training events into opportunities for continued skill development.

3.3 Cross-Level Team Coordination Dynamics

Cross-level coordination dynamics is a program of research that seeks to measure team coordination across motor (e.g., postural kinematics; [41]), cognitive (e.g., communication content; [42]), and physiological (e.g., neurophysiological synchronies; [43]) team subsystems. Efforts toward understanding how these subsystems interact to dissipate perturbations may facilitate interventions designed to enhance team resiliency under novel task conditions.

4 Conclusion

Sociotechnical work environments demand distributions of specialty knowledge across team members. To be sure, shared knowledge plays some role in a variety of sociotechnical tasks. It takes more than knowledge, however, to perform effectively in highly-dynamic, high-risk task environments. Teams must continually adapt their interactions to the changing dynamics of the task environment. Rooted in adaptive interaction, team coordination dynamics takes a nonlinear dynamics perspective on what means to learn and achieve coordinated behavior. Results from taking such a perspective have been promising thus far. However, as outlined in the last section of this paper, more work is needed to better understand dynamic nature of team synergies and how to detect threats to them. A better understanding of how coordination mechanisms, operating at many different levels of analysis, are dynamically linked may facilitate future training and learning interventions to enhance team resiliency under novel task conditions.

Acknowledgments. Potions of this research are funded by Office of Naval Research Grants N00014-07-M-0352 and N00014-11-M-0129. The UAV experiments were funded by Air Force Office of Scientific Research Grant FA9550-04-1-0234 and Air Force Research Laboratory Grant FA8650-04-6442 awarded to Dr. Nancy J. Cooke. The findings, views, and opinions expressed in this paper are the author's and do not necessarily represent the views of the funding agencies. I would like to acknowledge my colleagues, Nancy J. Cooke, Polemnia G. Amazeen, Steven M. Shope, and Eric E. Hessler, who contributed empirically and theoretically to the ideas expressed in this paper.

References

[1] Banks, A.P., Milward, L.J.: Distributed mental models: Mental models in distributed cognitive systems. Journal of Mind and Behavior 30, 249–266 (2009)

[2] Cooke, N.J., Salas, E., Cannon-Bowers, J.A., Stout, R.: Measuring team knowledge. Human Factors 42, 151–173 (2000)

[3] DeChurch, L.A., Mesmer-Magnus, J.R.: The cognitive underpinnings of effective teamwork: A meta-analysis. Journal of Applied Psychology 95, 32–53 (2010)

[4] Rouse, W.B., Morris, N.M.: On looking into the black box: Prospects and limits in the search for mental models. Psychological Bulletin 100, 349–363 (1986)

[5] Cannon-Bowers, J.A., Salas, E., Converse, S.: Shared mental models in expert team decision making. In: Castellan, N.J. (ed.) Individual and group decision making, pp. 221–246. Lawrence Erlbaum Associates, Hillsdale (1993)

[6] Langan-Fox, J., Code, S., Langfield-Smith, K.: Team mental models: Techniques, Methods, and Analytic Approaches. Human Factors 42, 242–271 (2000)

[7] Hollingshead, A.B.: Distributed knowledge and transactive processes in decision-making groups. In: Neale, M.A., Mannix, E.A. (eds.) Research on managing groups and teams, pp. 102–123. JAI, London (1998)

[8] Moreland, R.L.: Transactive memory: Learning who knows what in work groups and organizations. In: Thompson, L.L., Levine, J.M., Messick, D.M. (eds.) Shared cognition in organizations: The management of knowledge, pp. 3–32. Erlbaum, Mahwah (1999)

[9] Wegner, D.M.: Transactive memory: A contemporary analysis of the group mind. In: Mullen, B., Goethals, G.R. (eds.) Theories of group behavior, pp. 185–208. Springer, New York (1986)

[10] DeChurch, L.A., Mesmer-Magnus, J.R.: The cognitive underpinnings of effective teamwork: A meta-analysis. Journal of Applied Psychology 95, 32–53 (2010)

[11] Mohammed, S., Klimoski, R., Rentsch, J.R.: The measurement of team mental models: We have no shared schema. Organizational Research Methods 3, 123–165 (2000)

[12] Smith-Jentsch, K.A., Kraiger, K., Cannon-Bowers, J.A., Salas, E.: Do familiar teammates request and accept more backup? Transactive memory in air traffic control. Human Factors 51, 181–192 (2009)

[13] Hackman, J.R.: The design of work teams. In: Lorsch, J. (ed.) Handbook of organizational behavior, pp. 315–342. Prentice-Hall, Englewood Cliffs (1987)

[14] McGrath, J.E.: Groups: Interaction and Performance. Prentice-Hall, Englewood Cliffs (1984)

[15] Steiner, I.D.: Group process and productivity. Academic Press, New York (1972)

[16] DeChurch, L.A., Mesmer-Magnus, J.R.: Measuring shared team mental models: A meta-analysis. Group Dynamics: Theory, Research, and Practice 14, 1–14 (2010)

[17] Marks, M.A., Mathieu, J.E., Zaccaro, S.J.: A temporally based framework and taxonomy of team processes. Academy of Management Review 26, 356–376 (2001)

[18] Marks, M.A., Zaccaro, S.J., Mathieu, J.E.: Performance implications of leader briefings and team-interaction training for team adaptation to novel environments. Journal of Applied Psychology 85, 971–986 (2000)

[19] Stout, R.J., Cannon-Bowers, J.A., Salas, E., Milanovich, D.M.: Planning, shared mental models, and coordinated performance: An empirical link is established. Human Factors 41, 61–71 (1999)

[20] Entin, E.E., Serfaty, D.: Adaptive team coordination. Human Factors 41, 312–325 (1999)

[21] MacMillan, J., Entin, E.E., Serfaty, D.: Communication overhead: The hidden cost of team cognition. In: Salas, E., Fiore, S.M. (eds.) Team Cognition: Process and Performance at the Inter and Intra-individual Level. American Psychological Association, Washington (2004)

[22] Smith-Jentsch, K.A., Kraiger, K., Cannon-Bowers, J.A., Salas, E.: Do familiar teammates request and accept more backup? Transactive memory in air traffic control. Human Factors 51, 181–192 (2009)
[23] Mathieu, J.E., Heffner, T.S., Goodwin, G.F., Salas, E., Cannon-Bowers, J.A.: The influence of shared mental models on team process and performance. Journal of Applied Psychology 85, 273–283 (2000)
[24] Levesque, L.L., Wilson, J.M., Wholey, D.R.: Cognitive divergence and shared mental models in software development project teams. Journal of Organization Behavior 22, 135–144 (2001)
[25] Cooke, N.J., Kiekel, P.A., Helm, E.: Measuring team knowledge during skill acquisition of a complex task. International Journal of Cognitive Ergonomics 5, 297–315 (2001)
[26] Cooke, N.J., Salas, E., Kiekel, P.A., Bell, B.: Advances in measuring team cognition. In: Salas, E., Fiore, S. (eds.) Team cognition: Understanding the factors that drive process and performance, pp. 83–106. American Psychological Association, Washington (2004)
[27] Cooke, N.J., Gorman, J.C., Myers, C.W., Duran, J.L.: Theoretical underpinnings of interactive team cognition. In: Salas, E., Fiore, S., Letsky, M. (eds.) Theories of team cognition: Cross-disciplinary perspectives (in press)
[28] Kozlowski, S.W.J., Klein, K.J.: A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In: Klein, K.J., Kozlowski, S.W.J. (eds.) Multilevel theory, research, and methods in organizations, pp. 3–90. Jossey-Bass, San Francisco (2000)
[29] Ilgen, D.R., Hollenbeck, J.R., Johnson, M., Jundt, D.: Teams in organizations: From input-process-output models to IMOI models. Annual Review of Psychology 56, 517–543 (2005)
[30] Gibson, J.J.: The ecological approach to visual perception. Houghton Mifflin, Boston (1979)
[31] Cooke, N., Gorman, J., Rowe, L.J.: An ecological perspective on team cognition. In: Salas, E., Goodwin, J., Burke, C.S. (eds.) Team Effectiveness in Complex Organizations: Cross-disciplinary Perspectives and Approaches. SIOP Organizational Frontiers Series, pp. 157–182. Taylor & Francis, Abington (2009)
[32] Kelso, J.A.S.: Dynamic patterns: The self-organization of brain and behavior. MIT Press, Cambridge (1995)
[33] Haken, H.: Synergetics: An introduction. Springer, Berlin (1977)
[34] Cooke, N.J., Shope, S.M.: Synthetic task environments for teams: CERTT's UAV-STE. In: Handbook on human factors and ergonomics methods, pp. 46-1-46-6. CLC Press, LLC, Boca Raton, FL (2005)
[35] Gorman, J.C., Amazeen, P.G., Cooke, N.J.: Team coordination dynamics. Nonlinear Dynamics, Psychology, and Life Sciences 14, 265–289 (2010)
[36] Cooke, N.J., Gorman, J.C.: Interaction-based measures of cognitive systems. Journal of Cognitive Engineering and Decision Making 3, 27–46 (2009)
[37] Gorman, J.C., Cooke, N.J.: Changes in team cognition after a retention interval: The benefits of mixing it up (under review)
[38] Gorman, J.C., Cooke, N.J., Amazeen, P.G.: Training adaptive teams. Human Factors 52, 295–307 (2010)
[39] Schollhorn, W.I., Beckmann, H., Michelbrink, M., Sechelmann, M., Trockel, M., Davids, K.: Does noise provide a basis for the unification of motor learning theories? International Journal of Sport Psychology 37, 186–206 (2006)
[40] Gorman, J.C., Hessler, E.E., Amazeen, P.G., Cooke, N.J., Shope, S.M.: Dynamical analysis in real time: Detecting perturbations to team communication (submitted)

[41] Shockley, K., Santana, M.-V., Fowler, C.A.: Mutual interpersonal postural constraints are involved in cooperative conversation. Journal of Experimental Psychology: Human Perception & Performance 29, 326–332 (2003)
[42] Gorman, J.C., Foltz, P.W., Kiekel, P.A., Martin, M.J., Cooke, N.J.: Evaluation of latent-semantic analysis-based measures of team communications. In: Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting, pp. 424–428 (2003)
[43] Stevens, R.H., Galloway, T., Berka, C., Behneman, A.: Temporal sequences of neurophysiologic synchronies can identify changes in team cognition. In: Proceedings: Human Factors and Ergonomics Society 54th Annual Meeting, San Francisco, CA, September 27-October 1, pp. 190–194 (2010)

Performance-Based Metrics for Evaluating Submarine Command Team Decision-Making

Eric Jones[1], Ronald Steed[2], Frederick Diedrich[1], Robert Armbruster[3], and Cullen Jackson[1]

[1] Aptima, Inc., Woburn, MA, USA
[2] UpScope Consulting, Mystic, CT, USA
[3] CGI, Norfolk, VA, USA
{ejones,diedrich,cjackson}@aptima.com, ronaldsteed@gmail.com
robert.armbruster@stanleyassociates.com

Abstract. Successful submarine operations—those that accomplish the mission while maintaining security and safety—depend on numerous factors, including the capabilities of various sensor systems, the reliability of algorithms, and the proficiency of the crew. Among the most critical elements is Command Team decision-making and the underlying processes that create a cohesive and effective team. As a team, submarine commanders must successfully contend with complexities associated with safety and security as they build an understanding of the operational environment in order to accomplish their mission. Hence, opportunities to enhance training to support Command Team decision-making are essential. This paper describes a framework used to develop performance measures to support formative assessment of the submarine Command Team. Results are reported here from a study at the Naval Submarine School concerning the validity and utility of the measures in relation to capturing essential aspects of performance.

Keywords: performance measures, formative assessment, decision-making, teamwork, submarine.

1 Introduction

1.1 The Challenge

Successful submarine operations—those that accomplish the mission while maintaining security and safety—are exceedingly difficult to achieve. These operations depend on numerous factors including items such as the capabilities of various sensor systems, the reliability of algorithms, and the proficiency of the crew. Among the most critical elements is Command Team decision-making and the underlying processes that create a cohesive and effective team. As a team, submarine commanders must successfully contend with complexities associated with safety and security as they build an understanding of the operational environment in order to accomplish their mission. This cannot be accomplished by individuals acting independently while executing their assigned tasks. Rather, the objective for the Command Team is to

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 308–317, 2011.

leverage their combined resources. Opportunities to enhance training or system performance to support Command Team decision making are essential, given that the decisions are complex with severe consequences for error. A critical question, therefore, is how can learning systems develop essential team processes?

To address this challenge, we believe that the key is to create learning systems that are learner-centered, knowledge-centered, assessment-centered, and community-centered. More specifically, we follow a framework developed by the National Research Council, Commission on Behavioral and Social Sciences and Education, which outlines the requirements for effective learning systems [1]: First, to be *learner-centered*, a learning system must address what learners bring to the table—what they know, what they do not know, how they learn, and what they are motivated to learn. Second, to be *knowledge-centered*, effective learning systems must stress "sense-making" rather than mere memorization or execution of procedures without understanding. Third, to be *assessment-centered*, learning systems must provide formative feedback and not only summative evaluation. Fourth, to be *community-centered*, learning systems must be relevant to, and reflect, the communities in which they are embedded such that the learning is view as being meaningful.

Within the context of this overall approach, our focus is on the creation of assessment tools that support effective Command Team decision making. It is well-established that team performance depends not only on individual skills, but also on teamwork processes [2]. For instance, one framework employed by the Submarine Force is known as Team Dimensional Training (TDT) [3], [4]. TDT catalogues important aspects of teamwork that are associated with high-performing teams. Key teamwork skills identified in this framework include items such as information transfer (e.g., providing information prior to requests), communication (e.g., providing complete standard reports), supporting behavior (e.g., providing back-up to team members), and team leadership (e.g., stating clear and appropriate priorities). Accordingly, with a focus on team processes, in the work reported here we concentrate on the creation of tools to support formative assessment. Our goal is to create assessment tools that do not merely measure a Command Team against a standard, but that serve to guide learners and instructors to an understanding of learner state along with an appreciation of the path toward success (i.e., what "right" looks like). We strive to ensure that the assessment tools are community-centered in that they focus learning on meaningful, critical teamwork-oriented foundations for the Submarine Force.

1.2 The Development of Formative Assessment Tools

The measures we developed are designed to facilitate formative feedback regarding decision making and teamwork in the context of Intelligence, Surveillance, and Reconnaissance (ISR) missions. They focus on the Command Team, including the Commanding Officer (CO), Executive Officer (XO), Officer of the Deck (OOD), Junior Officer of the Deck (JOOD), and Junior Officer of the Watch (JOOW). To construct the measures, we used an approach named the Competency-based Measures for Performance Assessment Systems (COMPASS) method [5]. COMPASS employs an intensely interactive process in which operators work directly with scientists to identify behaviors that need to be measured. Here, the work relied on operational expertise (retired individuals with experience as CO and XO, as well as active-duty Tactical Readiness Evaluation (TRE) team members) combined with theories of measurement and teamwork, drawing heavily on the TDT framework. Overall, the approach explicitly strived to leverage theoretical foundations while deliberately expressing measures within the context of the issues, terminology, and focus of the submarine community.

The outcome was a set of approximately 100 measures, referred to in this paper as Exceptional Expertise for Submarine Command Team Decision Making (E2SCDM) measures [6]. As a sample in Figure 1 illustrates, the E2SCDM measures have the potential to support formative assessment in several ways. First, they capture current behavior in clear terms, guiding both the instructor and leaner in describing this behavior. Second, by providing indications of ideal performance, the measures serve to illustrate what the learners need to do in the future (e.g., proactively transfer information to the team). The measures therefore help to guide attention to essential items and illustrate desired behaviors. They have the potential to serve as the backbone of assessments tools that move beyond scores (or pass/fail) and into assessments that describe meaningful challenges facing the learner (e.g., see [7], for descriptions of alternative "report cards" that focus on descriptions of current and future abilities).

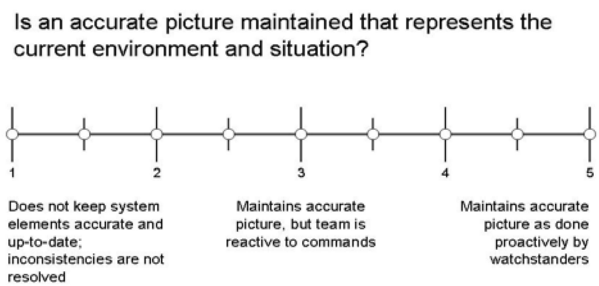


Fig. 1. Sample measure reflecting building of operational picture, highlighting proactive transfer of information

The purpose of the effort reported here was to explore the utility of the E2SCDM measures in relation to capturing essential aspects of team performance and providing feedback. To do so, we conducted a study at the Naval Submarine School (NSS) in Groton, CT, in which the measures were employed during Intermediate and Advanced Pre-Deployment Training events (IPDT and APDT respectively). While this approach limited our data collection to a small sample of three Watch Sections over four days, it allowed us to focus on an intact crew preparing for deployment, thereby enabling exploration of measure utility in examining a real Command Team in a real training setting. Our specific objectives were as follows: first,

- To capture data to illustrate how the measures might be used in the future to examine teams and provide feedback to guide learning.

- To examine the measures in relation to a current Fleet standard training assessment tool, the Continuing Training Support System (CTSS). CTSS is a comprehensive tool covering a wide range of items grouped within a few hundred categories of behavior (attributes). Our objective was to explore whether we could add value to this tool by focusing explicitly on Command Team decision making. Moreover, we sought to understand whether the measures could provide additional sensitivity with respect to team processes.
- To examine whether multiple observers tended to rate behavior similarly.
- To examine whether the measures of team processes produced overall assessments similar to those produced by current standard tools.

Below we begin by presenting the methods of our study, and then the results and potential implications. Given the small sample size that resulted from our focus on an intact Command Team, we did not concentrate on statistical analyses. Rather, we focus on trends gleaned from our examination of actual operational teams, and conclude that the approach does indeed show promise for providing formative assessments of team processes that support effective decision making.

2 Method

2.1 Participants and Apparatus

Participants included a crew conducting IPDT and APDT exercises at NSS, Groton, CT, in 2010. The focus of the study was on the CO, XO, OOD, JOOD, and JOOW as they interacted with each other and the crew as a whole. Three different Watch Sections were observed, two per day both in the IPDT and APDT events, for a total of four days of data collection. The crewmembers that were observed ranged in rank from O1 to O5. The average time spent at sea was 18 months, with a standard deviation of 27 months.

The study utilized existing training scenarios at NSS, with a primary focus on ISR exercises, in the Submarine Multi-Mission Team Trainer (SMMTT). No modifications were made to existing training scenarios since our objective was to observe a typical training cycle.

Data collection consisted primarily of observations performed by three retired submarine Officers, with prior experience as CO and XO, who acted as expert observers. These subject matter experts observed and collected data using E2SCDM measures which were implemented in Aptima's Scenario-based Performance Observation Tool for Learning In Team Environments (SPOTLITE) data collection tool, which allows for electronic collection of observer-based data. These same observers also collected observations using a subset of CTSS attribute sheets applicable to the ISR missions that were observed. When using CTSS to guide their observations, the subject matter experts took notes throughout the exercise and recorded the CTSS scores post-exercise. A comparison between these observer-based collection methods, with a focus on opportunities in which to add value, is included in *Results and Discussion*. Additional self-report data (e.g., workload, situation awareness) and system-based data (e.g., solution quality) were collected as well, but are not reported here.

2.2 Procedure

Informed consent was obtained during an initial brief prior to the PDT events, along with general information regarding the experience of the participants. Collection took place during the morning and afternoon over four days (two each during the IPDT and APDT events). The only exception to this schedule was the Monday session of the APDT in which we observed the afternoon and evening sessions because of technical difficulties with the SMMTT trainer. Each Watch Section lasted approximately four hours. Each expert observer used either the E2SCDM measures (as implemented within SPOTLITE) or the CTSS attribute sheets to rate performance for the duration of the Watch Section. The order in which these tools were used was varied, meaning that for a single session the raters were using E2SCDM measures, CTSS, or a combination of both (one rater assigned to each). One observer participated in the Monday session of the IPDT only. These data are included in the analysis where noted, but omitted from discussions of inter-rater reliability due to the limited data set. The experimental design permits several analyses both within and between Watch Sections that reasonably balances the treatments given the small number of opportunities to observe the PDT events.

While the E2SCDM measures assume a variety of forms, the predominant structures are that of Yes/No questions and 5-point Likert scales (with half-point increments) that are behaviorally anchored at levels 1, 3, and 5 (less than acceptable, acceptable, and exceptional performance respectively expressed as behavioral anchors; see Figure 1). Our analysis therefore focused on these two question types. Also, the observers were provided flexibility in deciding which measures to rate, based on the behavior they were observing. There was not a fixed set of measures that was rated for each session, but rather a library from which to select based on the developing scenario. The measures could be summoned multiple times within a session as well, as the observers decided to capture certain behaviors at various points throughout the scenario evolution. Similarly, the observers applied the same scrutiny to the CTSS attributes, grading only those sheets that were applicable and not a consistent set throughout the study. CTSS scores ranged from 0 to 3 with anchors reflecting the frequency of occurrence of an attribute (none of the time to all of the time, respectively). In the analyses reported below, results are reported based on the measures taken in each watch section, with a focus on comparison of ratings of like measures where possible.

3 Results and Discussion

The research team compiled and examined the data from various perspectives to understand the extent to which the E2SCDM measures support the design objectives. We begin by illustrating use of the measures for formative purposes through an example, and then move on to provide analyses with respect to trends concerning reliability, sensitivity, and validity (with respect to current Fleet tools).

3.1 Use of Measures for Formative Feedback

To illustrate use of the E2SCDM measures for formative feedback, the “top ten” and “bottom ten” Likert score items were compiled and averaged to provide an overview of team performance during the IPDT. Note that our objective was neither an extensive analysis nor a complete evaluation of the crew (that would require data over multiple days across a variety of events), but rather, we sought to illustrate how an instructor can use the measures to understand a team through emergent themes. These themes could serve as the basis for providing feedback during debriefing sessions. In this study, the crew was rated highly in several categories that are associated with desirable performance and favorable mission outcomes: pre-watch information management, command-level feedback, and risk/gain evaluations. These top ten measures ranged from averages of 5.0 to 4.1, with standard errors ranging from 0.0 to 0.3.

The bottom ten lowest-rated Likert measures describe areas in which the crew could improve given the challenging scenario events: accurate picture of safety/stealth, flexibility/adaptability, and self-assessment. Scores ranged from 2.8 to 1.0, with more variability than the top ten scores (standard errors ranging from 0.2 to 0.8). This range of data permits finer distinctions in performance: for example, the measures pertaining to self-assessment exhibited large standard errors, indicating that the crew demonstrated a wide range of proficiency in this area, possibly impacted by specific scenario events. On the other hand, the item regarding maintaining an accurate tactical picture had a small standard of error, indicating consistent behavior throughout the scenario.

Collectively, these findings provide clear evidence that could be used to support team assessment. For instance, while the Command Team did well to provide feedback to the team, self-assessment by watch-standers was not as effective as what might have been ideal. In addition, while the team prepared well for their watch sections by sharing information, they had more difficulty maintaining an accurate picture that enabled flexibility during the watch. Moreover, in addition to diagnosing those areas that challenged the crew given the difficult scenario events, the E2SCDM measures also provided a mechanism through which to guide improvements to performance (e.g., proactively transfer information to build the picture). Indeed, the anchors that support each question describe what "right" behavior looks like and can be incorporated by an instructor or team member during a training exercise to provide actionable, formative feedback to the crew.

3.2 Measures Sensitivity

The E2SCDM measures were generally designed with the use of a 5-point rating scale, with half-point increments, thus enabling nine possible rating levels. From a perspective of sensitivity, a critical question was the extent to which these possible ratings were actually employed. For instance, if a limited range of ratings were used, these would data suggest that the measures provide limited insight into the range of performance seen in PDT-like training events. In addition, given our interest in determining possible ways in which to add value to the current standard tool, it is useful to compare ratings employing CTSS and the E2SCDM measures on similar events. A wider distribution of meaningful ratings suggests possible ways to enhance current assessment instruments.

Figure 2 demonstrates the frequency and distribution of E2SCDM Likert scores and CTSS grades for the APDT event (only APDT data were examined in this analysis to reflect use of the measures following the greatest possible familiarization with the instruments within the confines of the study). Almost two thirds of all CTSS grades were a 2, which is equivalent to a passing grade ("performs behavior most of the time"). Ratings at the extremes of the scales were relatively limited. The E2SCDM Likert scores were relatively evenly distributed across the nine values to capture behavior, demonstrating not only that the observers had more options to rate performance, but that these options were actually used. No score rating was given more than 25% of the time, as opposed to the approximate 66% of CTSS grades that were a 2. These data suggest that overall, the E2SCDM measures provide a range of rating values, with the potential to capture more subtle variation in performance.

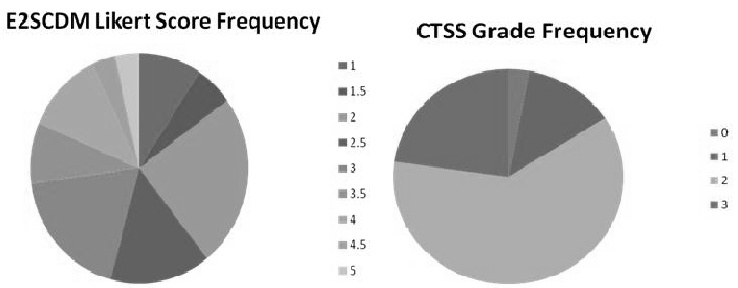


Fig. 2. Frequency and distribution of scoring using the E2SCDM measures (5-point scale with half-point increments; left) and the CTSS attributes (4-point scale; right) for the APDT event

3.3 Measures Reliability

The behavioral anchors associated with the E2SCDM measures not only facilitate formative feedback, but also have the potential to promote consistency with which the observers use the tool. To the greatest extent possible, subjective language was avoided when constructing the anchors, and care was taken to ensure that the descriptions were firmly rooted in behavior that is observable expert [6]. In this section, we therefore explore the extent to which the measures enabled consistency in ratings. A comparison was made between those measures that were rated by both observers within the one session in which both raters were using the E2SCDM measures during the APDT event. Measures that were rated multiple times during this session were averaged to achieve an overall rating.

No two Likert scores differed by more than 1.5 points (on the 5 point scale) and over 85% of the ratings were less than 1 point apart or identical. This is promising, and shows (e.g.) that at no time during the session did one observer determine the crew was performing exceptionally (5) while the other considered the performance only acceptable (3). The Yes/No (Y/N) measures were also examined, which showed high percent agreement (approximately 90% identical ratings). Though not behaviorally anchored, the Y/N questions were constructed to enhance objectivity through clear indications of the behaviors to be observed. A more principled statistical analysis demonstrates that there is high agreement in the manner in which each observer used the tool (covariance = 0.64, alpha= 0.78). Kappa, however, indicates a fair to moderate agreement (kappa = 0.34). While these trends of high inter-rater reliability are promising, the sacrifice in sample size that was made to observe a real PDT exercise does not permit strong conclusions to be made.

Similarly, the inter-rater reliability between observers using CTSS was high: 47% of the ratings were identical, and 100% of the ratings were within 1 point of each other. This shows that CTSS is a reliable tool from which multiple raters can reach similar conclusions regarding performance. When compared to the E2SCDM measures, trends suggest that these measures have the potential to add additional sensitivity without destroying reliability among multiple observers.

3.4 Measures Validity

Given that CTSS is a current standard Fleet tool, it is important to determine whether the E2SCDM measures draw, to an extent possible given the varying and specific topics of focus, similar conclusions regarding overall crew proficiency. Accordingly, Figure 3 displays the average E2SCDM Likert scores and CTSS average grades for all sessions across IPDT and APDT data collection events (standard error displayed as well). For those days that exhibit only one tool, this indicates that both observers used the same observation recording method. Data from the third observer is included in the Monday AM and PM IPDT sessions. For the purpose of this analysis, the 4-point CTSS scale was normalized to the same 5-point scale used with the E2SCDM measures (calculated using: normalized score = (CTSS+1)*(5/4), such that a "0" in CTSS was equivalent to approximately "1" and a "3" was equivalent to a "5").

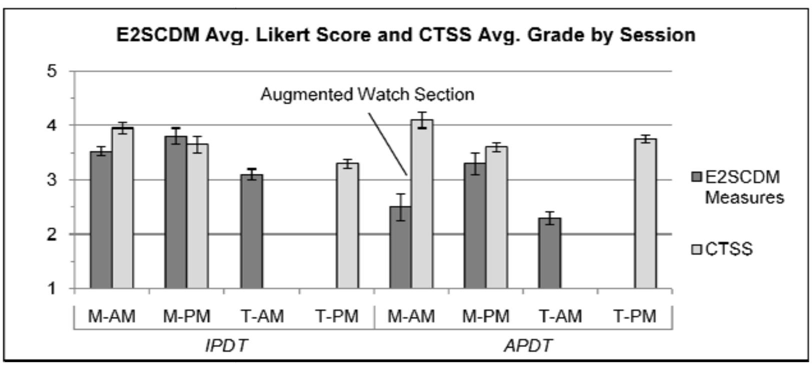


Fig. 3. Average E2SCDM Likert measure scores and CTSS grades by session. CTSS grades are normalized to a 5-point scale.

When both tools were used, there was generally good agreement between the ratings, such that for 3 of 4 observations that used multiple tools, mean ratings were within 0.5 normalized points, with CTSS slightly higher for two observations and E2SCDM slightly higher for one observation. The exception to this trend was a large difference seen for the M-AM session of the APDT event. This day was unique for a number of reasons including: technical difficulty with the SMMTT delayed the PDT start, the OOD was required to replace an absent crewmember and thus participated in both the AM and PM Watch Sections, and the length of the AM section was shortened. These confounding factors may point to the conclusion that the E2SCDM measures are more sensitive to the amount of time spent observing the crew, whereas the CTSS measures are less so. This difference may also reflect the fact that the E2SCDM measures are intended to be taken multiple times and averaged, whereas the CTSS ratings were made once at the end of each Watch Section. E2SCDM scores were significantly lower that day, and as suggested by the subjective feedback of the observers, more time may be required for them to make accurate judgments of team performance. However, given the peculiar nature of this particular collection event, it is difficult to be certain why the scores diverged for that particular Watch Section. Overall the results suggest some consistency with CTSS, with additional data required to reach definitive conclusions.

4 Summary and Conclusions

The findings of this study serve to illustrate the potential utility of developing measures for provision of formative feedback. In the context of pre-deployment training events, the measures were utilized to create themes of crew performance that could be used to guide learning. Given the construction of the measures, the anchors provide substantial guidance with respect to desired behaviors that could serve as reminders to crews who are learning and instructors who are watching for key events. Moreover, our findings also suggested that the measures investigated here exhibited acceptable levels of reliability between the raters while enabling a greater range of ratings to detect more subtle distinctions in performance. Hence, relative to the current Fleet standard tool, these data suggest that the E2SCDM measures have the potential to provide additional value with respect to formative feedback with increased sensitivity without a loss of reliability.

The findings also indicated overall agreement between ratings using the CTSS tool and the E2SCDM measures (despite disagreement in one test session). Given that CTSS is the current Fleet standard assessment tool, this finding suggests convergent validity with current tools, which are assumed to predict actual operational performance. Collectively, then, these trends suggest that the measures have the potential to provide a valid assessment tool for Command Team decision making, thus potentially adding value in addition to that of current tools. However, it should be noted that due to the small sample size, and poor agreement associated with one observation session, these results must be interpreted as suggestive but not conclusive.

Given these findings, future work should focus on refinement of the current measures, as well as further expansion of the measurement concept. Development should focus on testing with a range of crews, levels of proficiency, raters, and instructional events so as to further explore utility and validity. The result will be a revised set of measures suitable for use in assessment applications with respect to Command Team decision making. Overall, the preliminary findings reported here indicate that the methods promise to promote the creation of a comprehensive assessment toolset. Through an intense focus on team processes that support cohesion, decision making, and overall effectiveness, these measures provide a mechanism to support development of submarine Command Team decision making.

Acknowledgements. This work was completed under a contract with the Office of Naval Research (ONR). Any opinions, findings and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of ONR. We gratefully acknowledge the support and assistance of the crew that participated in the study, Kip Krebs, Ann Silva, Joseph Gabriel, Peter St. Jacques, Jack O'Sullivan, Advanced Solutions for Tomorrow, the Naval Submarine School, Groton, CT, and at the Naval Undersea Warfare Center Division Newport.

References

1. Bransford, J.D., Brown, A.L., Cocking, R.R.: How people learn: Brain, mind, experience, & school. National Academy Press, Washington (2000)
2. Salas, E., Rosen, M.A., Held, J.D., Weissmuller, J.J.: Performance measurement in simulation-based training: A review and best practices. Simulation & Gaming 40(3), 328–376 (2009)
3. Smith-Jentsch, K.A., Zeisig, R.L., Acton, B., McPherson, J.A.: Team dimensional training: A strategy for guided team self-correction. In: Cannon-Bowers, J., Salas, E. (eds.) Making decisions under stress: Implications for individual and team training. American Psychological Association, Washington (1998)
4. Smith-Jentsch, K.A., Johnston, J.H., Payne, S.C.: Measuring team-related expertise in complex environments. In: Cannon-Bowers, J., Salas, E. (eds.) Making decisions under stress: Implications for individual and team training. American Psychological Association, Washington (1998)
5. MacMillan, J., Entin, E.B., Morley, R.M., Bennett Jr., W.R.J.: Measuring team performance and complex and dynamic military environments: The SPOTLITE method. Military Psychology (in press)
6. Jones, E., Jackson, C., Diedrich, F., Durkee, K., Geldhauser, H.: Measures of Command Team decision making in submarine ISR missions. Technical Report, Aptima, Inc., Woburn, MA (2009)
7. Pelligrino, J.W., Chudowsky, N., Glaser, R.: Knowing what students know. National Academy Press, Washington (2001)

Multi-Modal Measurement Approach to Team Cohesion

Camilla C. Knott, Alexandra Geyer, Jason Sidman, and Emily Wiese

Aptima, Inc., 12 Gill Street, Suite 1400, Woburn, MA 01801, USA
{ccnott,ageyer,sidman,wiese}@aptima.com

Abstract. Team performance is a function, in part, of team cohesion: a dynamic process that is reflected in the tendency of a group to remain united in the pursuit of its goals and objectives (Carron 1982). We propose that a multi-modal measurement approach that integrates data from a variety of sources is critical to forming a comprehensive understanding of the relationship between team cohesion and performance, and can afford measurement of the hard-to-assess social component of team cohesion. Moreover, the use of a multi-modal measurement technique can afford flexibility in measuring across a variety of environments and selecting the most relevant measurement tools to minimize the technical footprint required for the assessment of teams and individuals in an operational environment.

Keywords: Team cohesion, multi-modal measurement, team performance.

1 Introduction

In military environments, where consequences of poor team performance can be catastrophic, it is critical for commanders and leaders to be able to quickly assess team performance to ensure safety and mission success. That performance is a function, in part, of **team cohesion**: a dynamic process that is reflected in the tendency of a group to remain united in the pursuit of its goals and objectives (Carron 1982). Hence, team cohesion has been related to team performance and its associated components but not reliably (e.g., Bowers, Urban, & Morgan, 1992; also see Salas, Bowers, & Canon-Bowers, 1995 for a review). Meta-analyses of the team cohesion and performance literature points to a generally positive relationship between team cohesion and performance, but one that is complex and highly variable across tasks (Evans & Dion, 1991; Gully, Devine, & Whitney, 1995; Oliver, Harman, Hoover, Hayes & Pandhi, 1999; Salo & Siebold, 2005).

Part of the problem lies in the complexity of team cohesion itself. Team cohesion is commonly broken down into task and social cohesion components. Task cohesion is a dimension of team cohesion that reflects team's ability to work together as a group to complete tasks required to support a common goal, and social cohesion, reflects affinity among team members. However, even these components are still complex and dynamic. Another part of the problem may be attributable to the methods used to measure these components of team cohesion. Though conceptually very different – Task cohesion should reflect performance and cognitive processing and Social cohesion should reflect affective and social processes – both components

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 318–324, 2011.

are commonly measured using self-report questionnaires or surveys. Likewise, in some cases even performance variables (which vary across studies) have been measured through questionnaires or surveys (either self-report or observer-based). However, surveys can only offer a limited viewpoint about team cohesion and performance.

We argue that a single measurement method will be inadequate for an accurate and comprehensive assessment of the relationship between team cohesion and performance. In this paper, we will focus primarily on essential task work components and teamwork functions that must be present for effective performance (Salas, Dickinson, Converse & Tannenbaum, 1992), specifically, back-up behaviors, coordination, feedback, and, communication (Dickinson & McIntyre, 1997) as well as workload, which has been shown to reliably affect performance and may, likewise, have an impact on team cohesion. We propose that a multi-modal measurement approach integrating data from a variety of sources is not only critical for developing a comprehensive and accurate assessment of team cohesion, but it can also afford a potentially effective method for measuring the hard-to-assess social component of team cohesion. Moreover, the use of a multi-modal measurement technique can provide flexibility in measurement methods across a variety of environments and in selection of the most relevant measurement tools to minimize the technical footprint required for the assessment of teams and individuals in an operational environment.

2 Use Case

To frame the measurement problem, we have developed a scenario anti-piracy training scenario in which a Combined Task Force (CTF) is monitoring international waters for pirate activity. The CTF is composed of 2 US Navy (USN) ships, a Royal Navy (RN) ship, and a French Navy ship. Each ship has a selection of airborne assets (unmanned aerial vehicles (UAVs) and helicopters) it can employ to assist in the mission. Data link compatibility issues between the USN, RN, and French ships preclude real-time data exchange between the CTF assets, with the exception of UHF line-of-sight voice communications. Because the CTF is composed of ships from a variety of countries, each ship must be considered with multi-level security issues and must therefore monitor the content of their communications to others. The CTF's mission is to prevent acts of piracy against any commercial or private vessel operating in international waters. To that end, the rules of engagement (ROEs) state that CTF assets should be employed to deter and/or prevent pirate intercepts by show of force, primarily by positioning assets. Weapons engagement is permitted only in self-defense or to negate a hostile attack. The team of trainees has been training together for a period of two weeks.

In this simplified training scenario, the trainees are the commanders of each ship. They are responsible for communicating with the other CTF members, coordinating activities with the other CTF members, and directing their own assets in appropriate intelligence, surveillance, reconnaissance, and targeting activities. The training objectives for this scenario include multiple aspects of Social and Task cohesion:

- Support inter-team collaboration/communication while adhering to multi-level security considerations.
- Manage the team members' task/workload.
- Identify and track all pirate vessels.
- Intercept/suppress pirate attacks while adhering to stated rules of engagement.
- Achieve mission objectives within fuel, weapons, etc. allotment.

The trainee performance in each of these areas is being assessed throughout the exercise using various measure types. Observer-based assessments are captured via an instructor, system-based measures are captured directly from the simulation data stream, trainees provide self-assessments during the scenario execution, and neurophysiological/physiological data is collected using various measurement tools which could include eye tracker, heart rate monitor, electroencephalography (EEG), respiration monitors, etc., depending on the availability and practicality. Table 1 below shows a mapping of each performance measurement source to each training objective.

Table 1. Summary of performance measurement source relevant for each training objective

Training Objectives	System-based	Observer-based	Self-Report	Neuro/physiological
Inter-team collaboration/communication	x	x	x	x
Task/Workload Management	x	x	x	x
Coordinate target tracking and identification	x	x		x
Coordinate intercept/suppress attacks	x	x		
Achieve mission objectives within fuel, weapons, etc. allotment	x	x	x	

As the training scenario unfolds, an unidentified track is detected and begins to close in on a known commercial vessel, the MV Sirius Star. The CTF must communicate with each other to determine which ship's assets should be used to investigate the track. Using their limited communication mechanisms (chat messages and voice communication), the CTF must determine which assets are in the vicinity of the track and have the resources (e.g., fuel, weapons, etc.) available to investigate the track. It is determined that a helicopter from a French Navy ship is the best choice to investigate the track, even though it is low on fuel and available weapons. From an assessment perspective, the inter-team collaboration/communication training objective is most relevant in this situation. The quantity and quality of communication can becaptured using system-based measurement. Number of chat messages and voice communications that occur between the CTF members can be used to analyze communication quantity. Observer-based measurements add context to those automated measures by assessing the CTF's ability to coordinate effectively given the communication delays and multi-level security issues imposed. Quality of communication and coordination may be assessed by querying the trainees directly, through self-report measures. Likewise, analysis of key words or even tone can also provide some information about social cohesion. In addition, coordination could be assessed by querying the trainees directly, through self-report measures as well as by collecting neurophysiological and physiological data.

Task/workload management is another objective that would benefit from neurophysiological and physiological assessment because system-based, observer-based, and self-report measures often lack the sensitivity to detect increases in workload prior to the point at which it affects performance. Ideally, increases in workload should be detected before it can lead to performance degradations to allow interventions to be introduced at the right time to mitigate those performance degradations. Neurophysiological and physiological measures of workload allow for such early detection. Moreover, on-line measures of workload circumvent the problem of when to administer self-report questionnaires.

While investigating the unidentified track, the French helicopter determines that the track is a pirate vessel heading towards the MV Sirius Star. During its investigation, the helicopter comes under attack and defends itself. However, given the French helicopter's fuel and weapons status, a second helicopter must be deployed to defend the Sirius Star. The CTF attempts to coordinate this activity, but the communication delays, combined with the added stress of dealing with an armed pirate vessel, results in confusion. The result is that two replacement helicopters are deployed: one from the French ship and one from the US ship. They arrive at the pirate vessel before it reaches the Sirius Star, thereby preventing a pirate attack. Again, the quantity and quality of communications between the CTF ships can be captured via automated performance measures. In addition, the ability of the CTF to manage their assets' fuel and weapons load and determine the appropriate assets to deploy can be assessed automatically through system-based measures. Finally, the helicopters' ability to effectively engage the pirate vessel can also be assessed automatically (system-based) through timeliness and accuracy measures. Observer-based measurements can capture additional detail about the coordination issues experienced by the team, high-level assessments about the tactics employed to intercept and suppress the pirate attack, and the ability of the CTF to effectively manage its assets to achieve mission objectives. Self-report measures can be used to obtain an understanding of the trainee's view of workload during this situation and the coordination amongst the CTF members. As in the previous situation, for on-line measurement of workload, neurophysiological and physiological data can be collected. Additionally, eye tracking data can be used to provide more detailed information about the CTF communication by analyzing whether or not the ship commanders are "seeing" relevant chat messages and scan patterns can be analyzed to assess coordination of tracking activities. Throughout the exercise, neurophysiological and physiological measures can be collected to assess individual preferences/likeness (e.g., Doherty, et al., 2006) as well as frustrations (e.g., Abler, et al., 2005) without requiring explicit acknowledgement from team members. While possibly controversial, these measures could provide some insight into the social and affective components of team cohesion.

After the training exercise, the data collected can be used to gain a comprehensive understanding of team performance - not one type of data can provide a complete and accurate picture of team performance or cohesion. The system-based data frequently lack the context required for complete interpretation; observer-data do not capture detailed data (e.g., timing and accuracy); self-report data provide only the perspective of the trainees and could be subject to trainee biases; and similar to system-based data, neurophysiological/physiological data require additional context to put the fine grained individual analyses in the context of a team. By viewing these data in combination with one another, however, the benefits of each measurement source are realized and the costs associated with each can be minimized. Most importantly, each measurement type can assist in the interpretation of the team's performance at specific points throughout the exercise. For example, the lack of communication between CTF members (as assessed using system-based measures) can help make sense of observer and self-report ratings of poor performance that occurred shortly thereafter. Just like completing a puzzle, obtaining an objective view of team member workload during that time period through neurophysiological/physiological sources can provide the final piece in completing the picture of team performance (what went right/wrong and why) during that time period.

3 Future Directions

The value of the multi-modal approach to measurement we propose is the ability to construct a more comprehensive assessment of team cohesion, and indeed it requires a considerable effort to determine which constructs, measures, and combinations of measures should contribute to that assessment. We believe that there are additional questions to explore based on that foundational effort. In particular, we foresee a need for the ability to deduce meaningful assessments in impoverished research and operational conditions when the full set of measures is simply not available. While the use case described above assumes a full array of measurement equipment capabilities, in reality, and particularly as we progress to more operational settings rather than research ones, such an array is unlikely to be available. Following a program to develop a comprehensive multi-modal approach to team measurement, we believe an equally valuable exercise for this reason would be the decomposition of team cohesion measures to determine the minimum measures, or combinations of measures, required to assess team cohesion constructs.

A critical relationship to establish among the measures would be redundancies; different measures or combinations of measures that lead to the same conclusions about team cohesion. Redundancies can serve multiple purposes. First, they can be used to validate measures. For example, certain self-report measures about performance could be validated by physiological measures that detect the indicators of that same construct. Second, they could be used to identify "surrogates" or "proxies" when certain measures are unavailable. The key challenge when identifying surrogates and proxies is distinguishing between measures that complement each other and measures that are redundant. Measures that complement each other produce a more comprehensive picture of a construct by addressing it from multiple angles (as described above). Measures that are redundant address the same part of the construct; they are interchangeable.

Mistaking complementarity for redundancy would produce major errors in assessing team cohesion. For example, knowing that people can feel uncomfortable discussing other team member's performance through self-report should cue researchers to explore the use of other potential sources of data that can provide complementary information. However, when measuring a construct that can be assessed in multiple ways, such as communication (see Table 1 above), one might not need physiological measures given other data sources that can provide redundant information that are less intrusive and cumbersome to collect. Or, it may be that three sources of data are more than what is needed to assess communication, and adequate assessments can be made based on only observer and system-based measures instead, thereby relieving participants of the burden of responding to questions.

As such, the challenge for multi-modal assessment is not simply to be all-encompassing, but to derive "just enough" assessment to make useful conclusions about team cohesion. Such a research program would involve not only building up the comprehensive set of measures and combinations of measures for assessing team cohesion, but pruning the measures down to the fewest that will yield the greatest assessment value.

References

1. The Abler, B., Walter, H., Erk, S.: Neural correlates of frustration. Neuroreport 16, 669–672 (2005)
2. Bowers, C.A., Urban, J.M., Morgan Jr., B.B.: The study of crew coordination and performance in hierarchical team decision making. Team Performance Laboratory Technical Report No. 92-1. University of Central Florida, Orlando, FL (1992)
3. Carron, A.V.: Cohesiveness in sport group: Interpretations and consideration. Journal of Sport Psychology 4, 123–138 (1982)
4. Dickinson, T.L., McIntyre, R.M.: A conceptual framework for teamwork measurement. In: Brannick, M.T., Salas, E., Prince, C. (eds.) Team performance assessment and measurement. Theory, methods, and applications, pp. 19–43. Lawrence Erlbaum, Mahwah (1997)
5. Evans, C.R., Dion, K.L.: Group cohesion and performance: A meta-analysis. Small Group Research 22(7), 175–186 (1991)
6. Gully, S.M., Devine, D.J., Whitney, D.J.: A meta-analysis of cohesiveness and performance: Effects of level of analysis and task interdependence. Small Group Research 26(4), 497–520 (1995)
7. Mathieu, J.E., Heffner, T.S., Goodwin, G.F., Salas, E., Cannon-Bowers, J.A.: The influence of shared mental models on team process and performance. Journal of Applied Psychology 85(2), 273–283 (2000)
8. Oliver, L.W., Harman, J., Hoover, E., Hayes, S.M., Pandhi, N.A.: A qualitative integration of the military cohesion literature. Military Psychology 11(1), 57–83 (1999)

9. Salas, E., Dickinson, T.L., Converse, S., Tannenbaum, S.I.: Toward an understanding of team performance and training. In: Swezey, R.W., Salas, E., Bowers, C.A., Cannon-Bowers, J.A. (eds.) Teams: Their Training and Performance. Military Psychology: Special Issue on Team Processes,Training, and Performance, vol. 7 (2), Ablex, Norwood (1995)
10. Salo, M., Siebold, G.L.: Cohesion components as predictors of performance and attitudinal criteria. In: The Annual Meeting of the International Military Testing Association, Singapore, November 7-10 (2005)
11. O'Doherty, J.P., Buchanan, T.W., Seymour, B., Dolan, R.J.: Predictive Neural Coding of Reward Preference Involves Dissociable Responses in Human Ventral Midbrain and Ventral Striatum. Neuron 49, 157–166 (2006)

Communications-Based Automated Assessment of Team Cognitive Performance

Kiran Lakkaraju, Susan Stevens-Adams, Robert G. Abbott, and Chris Forsythe

Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185
{klakkar,smsteve,rgabbot,jcforsy}@sandia.gov

Abstract. In this paper we performed analysis of speech communications in order to determine if we can differentiate between expert and novice teams based on communication patterns. Two pairs of experts and novices performed numerous test sessions on the E-2 Enhanced Deployable Readiness Trainer (EDRT) which is a medium-fidelity simulator of the Naval Flight Officer (NFO) stations positioned at bank end of the E-2 Hawkeye. Results indicate that experts and novices can be differentiated based on communication patterns. First, experts and novices differ significantly with regard to the frequency of utterances, with both expert teams making many fewer radio calls than both novice teams. Next, the semantic content of utterances was considered. Using both manual and automated speech-to-text conversion, the resulting text documents were compared. For 7 of 8 subjects, the two most similar subjects (using cosine-similarity of term vectors) were in the same category of expertise (novice/expert). This means that the semantic content of utterances by experts was more similar to other experts, than novices, and vice versa. Finally, using machine learning techniques we constructed a classifier that, given as input the text of the speech of a subject, could identify whether the individual was an expert or novice with a very low error rate. By looking at the parameters of the machine learning algorithm we were also able to identify terms that are strongly associated with novices and experts.

1 Introduction

Situation awareness is a key factor in determining the effectiveness of team performance. This paper describes work undertaken to develop automated measures of team performance. There is particular interest in automated performance assessment in complex domains that involve a large and heterogeneous set of entities. The task of the E-2 Hawkeye, a U.S. Airborne Early Warning (AEW) aircraft, is a prime example in that three Naval Flight Officers (NFOs) must co-ordinate their activities to classify air and sea entities, communicate with air and sea commanders, and give direction to a variety of friendly assets. It is essential that E-2 NFOs sustain situation awareness with respect to overall battlespace management and communicate among themselves with respect

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 325–334, 2011.

to ongoing status, developing situations and handling of specific entities. This communication occurs through a variety of mechanisms including voice communications, hand gestures and written notes. Within military and other training domains, the need exists for technologies to assist instructors, enabling them to accomplish more with available resources. Automated performance assessment represents one technique proposed to accomplish this objective by allowing computers to focus on assessing mundane facets of performance, while instructors focus their attention on higher-level cognitive processes. Previously reported research has provided experimental evidence for the training efficacy of automated performance assessment [4,5]. The current paper discusses the extension of these capabilities to team communications, which represents a vital component in establishing and sustaining team situation awareness.

1.1 Simulation Training

A significant cost in simulation-based training is the workload on human instructors to monitor student actions and provide corrective feedback. For example, the U.S. Navy trains Naval Flight Officers for the E-2-Hawkeye aircraft using a high-fidelity Weapons Systems Trainer (E-2 WST). Currently this requires a separate instructor to observe each student within the context of team performance and provide instruction based on observed misunderstandings, inefficient task execution, ineffective or inappropriate actions, etc. Individualized instruction contributes to high training costs. Intelligent tutoring systems target this need, but they are often associated with high costs for knowledge engineering and implementation. New technologies are required that assist instructors in providing individually-relevant instruction. AEMASE, a tool developed at Sandia National Laboratories, is one such technology.

1.2 AEMASE

Sandia National Laboratories has shown the feasibility of automated performance assessment tools such as the Sandia-developed Automated Expert Modeling and Student Evaluation (AEMASE) software. One technique employed by AEMASE is the grading of student performance by comparing their actions to a model of expert behavior. Models of expert behavior are derived by collecting sample data from simulator exercises or other means and then employing machine learning techniques to capture patterns of expert performance. During training, the student behavior is compared to the expert model to identify and target training to individual deficiencies. Another technique utilized by AEMASE is the grading of student performance by comparing their actions to models of good and/or poor student performance. Students with good and bad performance are identified and machine learning techniques are employed to construct models of these two types of performance in the same manner as expert performance. Student performance from other training sessions is then compared to these models to identify and target training to individual deficiencies. Both techniques avoid the costly and time-intensive process of manual knowledge elicitation and expert system implementation [1].

In a pilot study, AEMASE achieved a high degree of agreement with a human grader (89%) in assessing tactical air engagement scenarios [1]. However, the 68 trials assessed utilized only four subjects under three different initial training scenarios and the range of correct behaviors was quite limited. The current study provides a more rigorous empirical evaluation of the accuracy of these assessments. User modeling, based on behavioral and/or physiological measures, will be a key component of technologies implementing augmented cognition tools for training.

2 Purpose of Study

In our study, two-person teams of novices and experts engaged in a simulation-based scenario numerous times. We were particularly interested in differences between the expert and novice teams on facets of verbal communications. We expected that the expert teams would be more similar to each other than to the novice teams in terms of communication and that the novice teams to be more similar to each other than to the expert teams.

3 Methods

3.1 Participants

Eight participants took part in the experiment. Four participants were expert E-2 NFOs who had extensive in-flight experience and served as our Subject Matter Experts (SMEs). These four experts comprised two, two-person expert teams. The other four participants were novices and employees of Sandia National Laboratories. These novices met the demographics of an entry-level E-2 NFO and had undergone prior E-2 training, which enabled them to successfully complete the current mission on the EDRT. The four novices comprised two, two-person novice teams.

3.2 Materials

Materials included an E-2 Deployment Readiness Trainer (EDRT) simulator that was obtained from the Naval Air Systems Commands Manned Flight Simulator organization. The Joint Semi-Automated Forces (JSAF) simulation software was used to create and drive the scenario. The scenario was written by an expert E-2 NFO and consisted of a complex mission involving hostilities and an all-out air and sea engagement. In addition, the Sandia-developed Automated Expert Modeling and Student Evaluation (AEMASE) software was used in the analyses of the data.

3.3 Procedure

Participants were recruited via email or phone. Each two-person team was run separately over the course of one week. Both participants in the team were asked

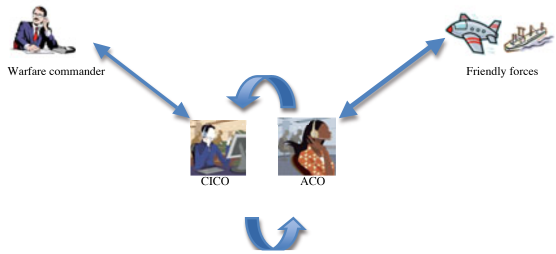


Fig. 1. Communication between the team members, warfare commanders and friendly forces

to sign an informed consent at the beginning of their first session. They were then introduced to the scenario and were informed of their roles for performing the scenario. There were two roles, the Air Control Officer (ACO) who only communicates only with the friendly forces and the Combat Information Center Officers (CICO) who only communicate with the warfare commanders (Figure 1). Each team performed the scenario 14 times (over the course of a week) and each person played the same role for all iterations of the scenario. The scenario required the team to keep in constant communication with each other in order to successfully complete the mission. The scenario was written such that one team member could only communicate with the warfare commanders and the other team member could only communicate with the friendly forces. Thus, the team members had to communicate in order to keep each other up-to-date on the happenings and orders given in the scenario. The participants behavioral responses, voice communications and biometric (i.e., EEG) data were collected for all iterations of the scenario. At the end of the last iteration of the scenario, the participants were debriefed and thanked for their participation.

4 Available Data

We use the following shorthand for the 8 subjects.

EX1ACO Expert team 1, ACO NFO.
EX1CICO Expert team 1, CICO NFO.
EX2ACO Expert team 2, ACO NFO.
EX2CICO Expert team 2, CICO NFO.
No1ACO Novice team 1, ACO NFO.
No1CICO Novice team 1, CICO NFO.
No2ACO Novice team 2, ACO NFO
No2CICO Novice team 2, CICO NFO.

4.1 Metrics

Initial analysis identified metrics that differentiated expert from novice teams with respect to voice communications, yet may be reliably measured using current voice recognition technology (i.e. assumes a literal transcription may not be reliably achieved). Based on interaction with subject matter experts, reservist E-2 NFOs, three aspects of team communication were identified: (1) when a team communicates, (2) what they communicate, and (3) how they communicate. By studying when NFOs communicate, the responsiveness of the team to external events and information flow within the team may be assessed. However, what NFOs communicate is just as important each utterance should transmit important information communicated in a clear and understandable manner. Finally, the phonetic characteristics of the communication (e.g., tone and rhythm) play an important role in conveying cues such as urgency or importance. We focus on the first two aspects in this paper.

4.2 Pedal Presses

In order to operate the radio, subjects had to press and hold a pedal that would open a communication channel. Releasing the pedal closed the radio channel. The pedal presses provide information about when and how long subjects communicated.

4.3 Communications

Each subject performed 14 scenarios. Scenario 3 was manually transcribed and included filler words (um, ah, er, m, etc). The text of each subject was contained in a text file. Several of the other scenarios were automatically transcribed using the Sphinx speech-to-text engine. For each subject, per scenario, a document was created that contained all of their speech.

The Text package of the Cognitive Foundry [2] was used to calculate term based representations of the speech of each of the subjects. The basic idea is to transform each document into a vector that indicates the terms in a document and then calculate the similarity by comparing the vectors. Weighting factors are applied to terms in order to emphasize rare terms. Figure 2 describes the process. We used the *tfidf* weighting factor which is described below.

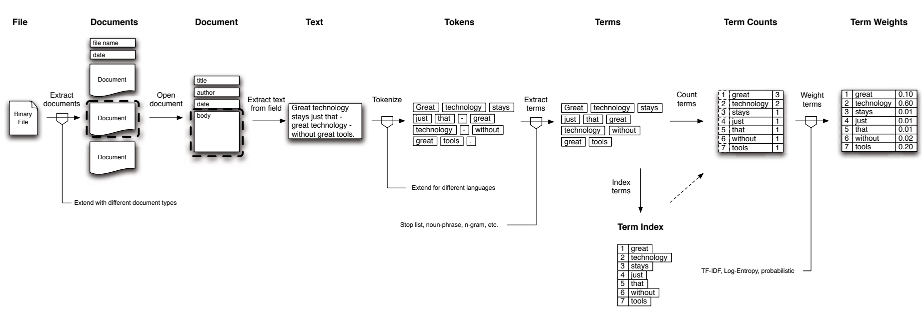


Fig. 2. Overview of the generation of a document vector

There are numerous ways of comparing two vectors. One common measure is the Euclidean distance. However, following work in this area, we use cosine similarity – which calculates the cosine of the angle between the two vectors. This quantity ranges from 0 to 1.

Term Frequency Inverse Document Frequency (TFIDF) is a commonly used term weighting system that assigns a weight to each term in a document that is a function of its frequency (how often it occurs within the document, abbreviated tf) and its document frequency (in how many of the other documents the term appears as well, abbreviated df). The exact function is:

$$\text{Tfidf} = \text{tf}\ log(n/\text{df}). \tag{1}$$

where n is the total number of documents. Note that if a term appears in all documents the tfidf weight will be 0 (log(1)=0). Thus, a zero-document is often used that contains no terms so that the df will always range from 1 to n-1. In this work we always used a zero-document.

The intuition behind tfidf is to heavily weight terms that are high frequency and occur in only a few documents. Those terms should help to differentiate between the documents.

5 When Individuals Communicate

We first looked at whether there are differences in the frequency of communications between the novice and expert teams. The pedal presses of the subjects provides information on when a verbal communication occured and its length.

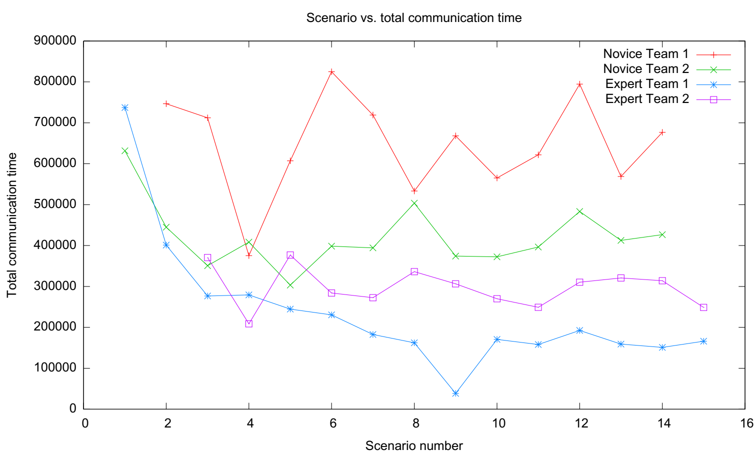


Fig. 3. Scenario vs. total time in communication

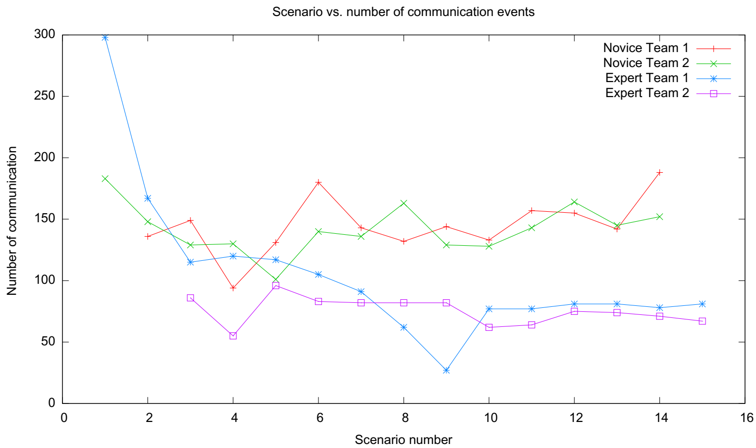


Fig. 4. Scenario vs. total number of communication events

Figure 4 shows the number of communication events per team over all scenarios. We see some very interesting differences between experts and novices. During the early scenarios expert and novice communication occurred at roughly the same rate. However by scenario 6 both expert teams have a significantly lower number of communication events. Figure 3 shows the duration of communication – the length of time the pedal was depressed. Novices had significantly longer communications, which corresponds to the higher number of communication events.

6 What Teams Communicate

We also looked at the language of the NFOs in the experiments. We will describe two analyses, one based on comparing the similarity of term vectors, and the other based on using machine learning techniques to learn a classifier of expert and novice language.

Figure 5 shows the similarity between each NFO and all the other experts and novices on the manually transcribed scenario (self similarity was not counted). We can see that for 7 of 8 subjects, the two most similar subjects (using cosine-similarity of term vectors) were in the same category of expertise (novice vs expert). This indicates that all the experts were using similar terms to each other, but not with novices.

Results for the Sphinx transcribed data showed similar patterns. Figure 6 shows the similarity for one of the automatically transcribed scenarios. In this scenario, the same pattern as before holds, experts are more similar to other experts than themselves, whereas novices are more similar to each other than to other experts. Only one novice NFO (N1ACO) was actually more similar to the experts than novices.

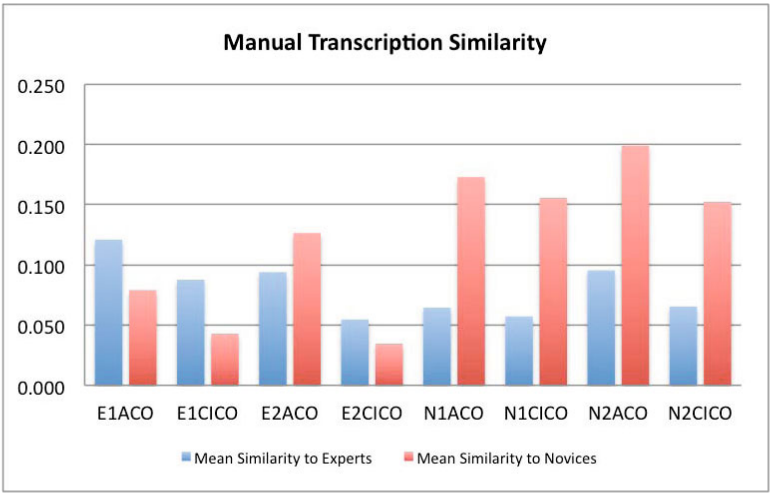


Fig. 5. Similarity between experts and novices on manual transcription

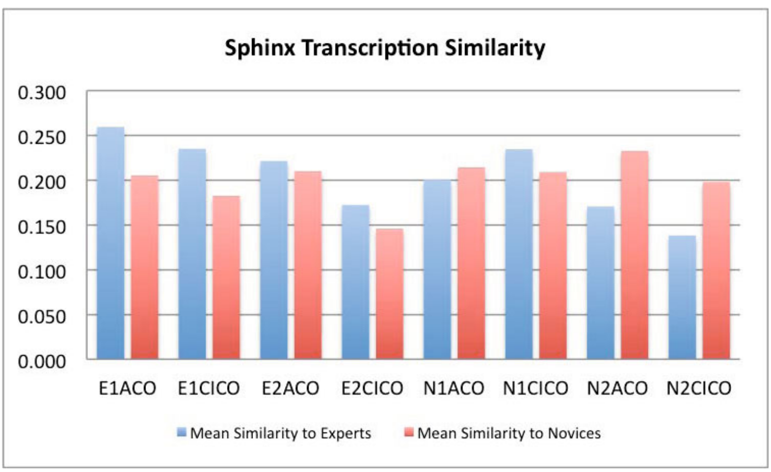


Fig. 6. Similarity between experts and novices on an automatically transcribed scenario

Our next question was to understand if particular terms of the subjects could indicate their expertise. To explore this, we utilized the perceptron learning algorithm to classify utterances of the subjects. This experiment uses the learning package of the Cognitive Foundry [3].The idea is to use the perceptron learning algorithm to train a classifier that can distinguish between expert and novice team members.

The input to the perceptron algorithm is the term frequency vector – a vector where each element represents a term and the value of the element is the frequency of the term (this is exactly the *tf* value mentioned before).

The classifier learned to distinguish between the experts and novices with an error rate of .05 (5% of the utterances of a subject were misclassified). Through studying the weights of the classifier, we are able to determine terms that differentiate novices from experts. Table 1 shows some of the top terms. Interesting differences can be seen; for instance novices use the word "feet" to indicate altitude of entities, whereas experts did not use that word at all.

Table 1. Terms that strongly differentiated experts from novices

Novices	Experts
charlie	tango
oh	alpha
cool	whiskey
this	zulu
feet	roger
thanks	advised
successfully	kill

7 Discussion and Future Work

The results above, while preliminary, indicate that language based metrics can help in differentiating between expert and novices. In addition, we have shown that we find the same patterns when using automated speech-to-text technology. This opens the door for large scale automated assessment. We found that experts communicate with lower frequency than novices, and that the overall time for communication is less with experts than novices. We also find that experts and novices differ in their terms, with experts being more similar to other experts and novices to other novices. Finally, through using the perceptron algorithm to classify the text of experts and novices, we have identified some of the words that differ between novices and experts.

There are many avenues for future work. Given the small number of participants, we did not perform any statistical analyses on our results. We would like to see if we can find the same differences among a larger set of experts and novices.

Preliminary analysis of phonetic patterns did not distinguish between experts and novices. However, our subject matter expert did indicate that phonetic differences were present in the data. Investigation into this aspect is one of the future goals for this project.

Acknowledgements

We would like to thank Justin Basilico and John Battista for their help in this project.

Sandia is a multi-program laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United Sates Department of Energy's national Nuclear Security Administration under contract DE-AC04-94AL85000.

References

1. Abbott, R.G.: Automated expert modeling for automated student evaluation. In: Ikeda, M., Ashley, K.D., Chan, T.-W. (eds.) ITS 2006. LNCS, vol. 4053, pp. 1–10. Springer, Heidelberg (2006)
2. Basilico, J., Benz, Z., Dixon, K.R.: The cognitive foundry: A flexible platform for intelligent agent modeling. In: Proceedings of the 2008 Behavior Representation in Modeling and Simulation (BRIMS) Conference (2008)
3. Basilico, J., Diegert, C.F., Heath, Z., Ko, T.H., Linebarger, J.M., Pancerella, C.M., Parker, E.P., Shneider, M.S., Williams, P.A.: Large scale social simulation. Tech. Rep. 2008-7831, Sandia National Laboratories (2008)
4. Stevens-Adams, S.M., Basilico, J., Abbott, R.G., Giesler, C., Forsythe, J.C.: Performance assessment to enhance training effectivness. In: Proceedings of the Interservice/Industry Training, Simulation, and Education Conference, No. 10228, pp. 1–9(2010)
5. Stevens-Adams, S.M., Basilico, J., Abbott, R.G., Giesler, C., Forsythe, J.C.: Using after-action review based on automated performance assessment to enhance training effectiveness. In: Proceedings of the Human Factors and Ergonomics Society 54th Meeting, pp. 2309–2315 (2010)

Visual Analytics of Social Networks: Mining and Visualizing Co-authorship Networks

Carson Kai-Sang Leung, Christopher L. Carmichael, and Eu Wern Teh

The University of Manitoba, Winnipeg, MB, Canada
kleung@cs.umanitoba.ca

Abstract. Co-authorship networks are examples of social networks, in which researchers are linked by their joint publications. Like many other instances of social networks, co-authorship networks contain rich sets of valuable data. In this paper, we propose a visual analytic tool, called SocialVis, to analyze and visualize these networks. In particular, SocialVis first applies frequent pattern mining to discover implicit, previously unknown and potential useful social information such as teams of multiple frequently collaborating researchers, their composition, and their collaboration frequency. SocialVis then uses a visual representation to present the mined social information so as to help users get a better understanding of the networks.

Keywords: Human-computer interaction, data mining, frequent patterns, social network analysis and mining, social computing, social information, data visualization, information and knowledge visualization, visualizing social interaction, augmented cognition.

1 Introduction and Related Work

Over the past few years, the rapid growth and exponential use of social digital media has led to an increase in popularity of social networks and the emergence of social computing. In general, *social networks* [7,25] are structures made of social entities (e.g., individuals, corporations, collective social units, or organizations) that are linked by some specific types of interdependency (e.g., kinship, friendship, common interest, beliefs, or financial exchange). A social entity is connected to another entity as his next-of-kin, friend, collaborator, co-author, classmate, co-worker, team member, or business partner. *Social computing* [15,23,24] aims to computationally facilitate social studies and human-social dynamics in these networks as well as to design and use information and communication technologies for dealing with social context. It includes the development of human-computer interaction technologies for augmenting cognition [1,8]—i.e., naturally extending the minds of social entities so that they could effectively perform conscious mental activities such as solving problems, making decisions, acquiring new knowledge, and connecting with others—by social information and collective intelligence. Intuitively, *collective intelligence* [19] is a shared or group intelligence that emerges from the collaboration of some social entities. Joint publications are examples of solid outcomes of such collaboration.

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 335–345, 2011.

To facilitate augmented cognition, it is better to mine useful social information from the social networks.

Social network mining discovers implicit, previously unknown and potentially useful social information. Examples of mining tasks include predicting links [3], learning influence probabilities [9], and discovering suspicious groups [21]. In this paper, we apply another mining task to an important type of social networks. Specifically, we apply *frequent pattern mining* [13,14,18] (which was introduced [2] to analyze shopping market basket data for revealing shopper behaviour) to *co-authorship networks* for discovering important social information such as teams of frequently collaborating researchers, their composition, and their collaboration frequency. The mined information is helpful in applications like academic author ranking and expert recommendation. Related works on mining co-authorship networks mainly focused on different mining tasks (than finding frequent patterns about collaboration teams) such as classifying origins of researcher names [4] and analyzing supportiveness between pairs of researchers [11].

As "a picture is worth a thousand words", having a visual representation is generally more comprehensive to users than its textual representation. This explains why several visualizers have been proposed to visualize results (e.g., association rules [5], shopper patterns [6,16,17], clusters [20]) of various traditional data mining tasks. Similarly, while it is important to discover useful frequent social patterns from co-authorship networks, it is equally important to be able to visualize these patterns. Common visual representations of these networks include *node-link diagrams* [12], in which each node represents a social entity (researcher) and each edge connecting two nodes represents a linkage (co-authorship) between the two entities. This social information can also be represented in a *socio-matrix* (i.e., an adjacency matrix) [7]. However, node-link diagrams or socio-matrices do not necessarily capture frequency information associated with researchers and their co-authorship (e.g., number of papers authored by a researcher or the number of joint publications between two researchers). To capture multi-researcher co-authorship, one may use other representations such as *hypergraphs* [11] or *bipartite graphs* [22]. However, as frequency information is captured implicitly by these representations, users may encounter difficulties in counting frequency (due to overlapping clusters in hypergraphs or crossing-over lines in bipartite graphs). In this paper, we use an alternative representation in our proposed visual analytic tool called ***SocialVis***, which visualizes co-authorship networks so that it not only shows collaborators of user-selected researchers but also all the linkages among them. It clearly and explicitly presents frequency information for individual researchers and for pairs of researchers. Moreover, it shows the composition of teams of multiple researchers and their frequency information even for large co-authorship networks. Our ***key contribution*** of this invited paper is our proposal of SocialVis, which analyzes and visualizes social networks like co-authorship networks. SocialVis discovers useful social information and results of collective intelligence (e.g., publications) from the networks and allows users to visualize this information so that it helps them understand the networks and augment cognition. In general, SocialVis can serve as a standalone tool for mining and visualizing the networks and as a complement to existing tools (especially those that have features such as spotting and displaying interesting patterns but do *not* provide the frequency information of the patterns).

This paper is organized as follows. Next section discusses different visual representations of co-authorship networks. We propose our SocialVis in Section 3 and present evaluation results in Section 4. Conclusions are given in Section 5.

2 Representing Co-authorship Networks

Co-authorship networks are commonly represented as *node-link diagrams* [12], in which each node represents a researcher and each edge represents co-authorship between the two researchers. Fig. 1(a) presents a node-link diagram for a 1-degree egocentric network showing some selected collaborators of researcher Ng; Fig. 1(b) presents a node-link diagram for a 1.5-degree egocentric network, in which each dashed edge represents co-authorship between selected collaborators of Ng. (For simplicity of illustration, only some but not all collaborators of Ng are shown in the figures.) The social information depicted by the node-link diagram can be equivalently represented in a *socio-matrix* (i.e., an adjacency matrix) [7], in which every row and column is indexed by a researcher and each non-diagonal cell (x,y) keeps a Boolean value indicating the presence or absence of co-authorship between the two corresponding researchers x and y. See Fig. 1(c). However, node-link diagrams or socio-matrices do not necessarily capture quantitative information such as publication counts of researchers.

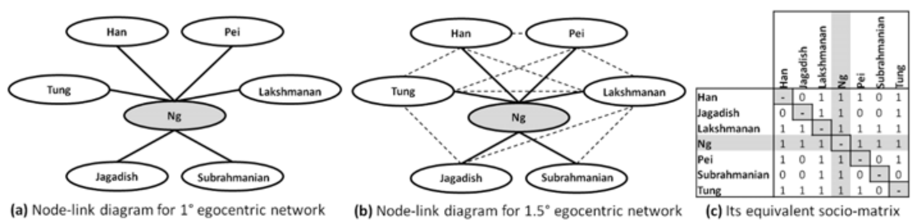


Fig. 1. Node-link diagrams & a socio-matrix

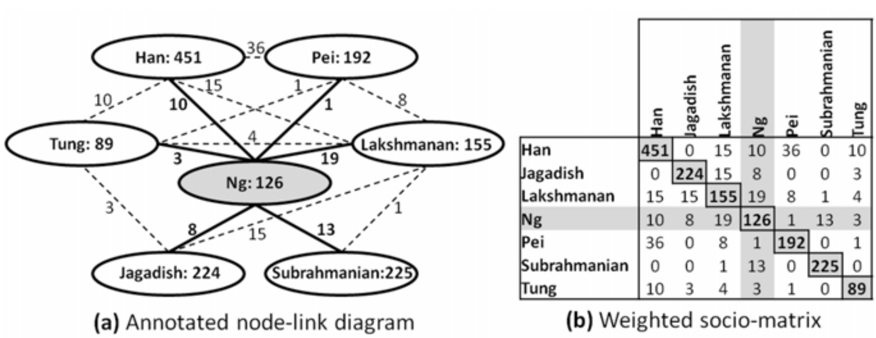


Fig. 2. An annotated node-link diagram & weighted socio-matrix for 1.5° egocentric network

To show quantitative information, one could use different node sizes and edge thickness in a node-link diagram. However, it is not easy to compare the size of two nodes or the thickness of two edges. Alternatively, one could annotate each node or edge with the quantitative information. Fig. 2(a) presents a node-link diagram for a 1.5-degree egocentric network showing some selected collaborators of researcher Ng, in which node x is annotated with the number of papers authored by researcher x and edge xy is annotated with the number of joint publications by researchers x & y. Similarly, one could replace the Boolean value in each non-diagonal cell (x,y) in a socio-matrix (indicating the presence or absence of some publications between the two researchers x & y) by an integer indicating the number of joint publications between x & y. The number of papers authored by a researcher could either (i) be augmented to the row or column label or (ii) be captured by the diagonal cell. See Fig. 2(b). While these representations show quantitative information for individual social entities and their pairwise relationships, they do not show the relationships among multiple social entities.

In many situations, relationships simultaneously involve more than two individuals in the network (e.g., papers co-authored by more than two researchers). For these situations, one could use (i) a *hypergraph* to group multiple researchers into the same cluster if they co-authored the same paper, (ii) a *dual hypergraph* to group multiple papers into the same cluster if they are co-authored by the same researcher, or (iii) a *bipartite graph* to link researchers to their corresponding joint publications. While the use of these three types of graphs depicts the composition of multi-entity relationships, these graphs do not clearly and explicitly provide users with frequency information. Moreover, multiple researchers may coauthor the same paper, and multiple papers may be coauthored by the same researcher. As such, clusters of entities (researchers) in hypergraphs often overlap with each other, and clusters of relationships (joint publications) in dual hypergraphs often overlap with each other. Linkages between researchers and papers in bipartite graphs often cross over each other. Hence, the use of these graphs can be quite unwieldy when depicting large social networks. An alternative visual representation is needed.

3 Analyzing and Visualizing Social Information with SocialVis

Given co-authorship network data (such as DBLP Bibliography records), our proposed visual analytic tool—called SocialVis—analyzes social networks and visualize social information. First, it applies frequent pattern mining algorithms [2,10] to find teams of frequently collaborating researchers and their collaboration frequency. Then, SocialVis represents the discovered frequent patterns in a two-dimensional space where the x-axis lists the researchers and the y-axis shows the number of their (solo or joint) publications.

To facilitate quick lookup researchers of user interest, SocialVis arranges researchers in *alphabetical order* on the x-axis. Besides this default ordering, SocialVis can also arrange researchers in *descending order of the number of their publications* (which gives users a quick insight about the frequency distribution of research publications because researchers with more publications appear on the left-hand-side and those with fewer publications appear on the right-hand-side).

Moreover, users do not need to select *all* researchers. Users can select one or more researchers based on their interest (e.g., select researcher Ng and some of his collaborators) for further analysis and visualization.

To clearly show the number of publications, SocialVis explicitly lists only the existing frequency values on the *y*-axis. This avoids large gaps between existing frequency values. Besides this default listing, SocialVis can also show the frequency values in linear scale, which allows users to get insight about the density or distributions of frequencies.

3.1 Visualizing Individual Researchers

When given a co-authorship network, a commonly asked question is as follow:

> **Q1.** How many papers published by this researcher?

The answer to Q1 may indicate how active this researcher is, which helps in ranking researchers in the network. To visualize the answer, SocialVis represents the number of papers authored by each individual researcher using a diamond-shaped icon ◀▶ (composed of a left-pointing triangle and a right-pointing triangle) in a two-dimensional space. The *x*-position of the icon indicates the researcher name, and the *y*-position of the icon indicates the number of his publications. See Fig. 3(a) for a screenshot of SocialVis, which explicitly shows the number of publications authored by each of the above seven selected researchers. From this figure, we can easily look up the number of Ng's publications (i.e., 126 papers). When researchers are arranged in descending order of the number of publications as shown in Fig. 3(b), we can easily observe that—among the seven selected researchers—Han published the most (with 451 papers) and followed by Subrahmanian (who published 225 papers).

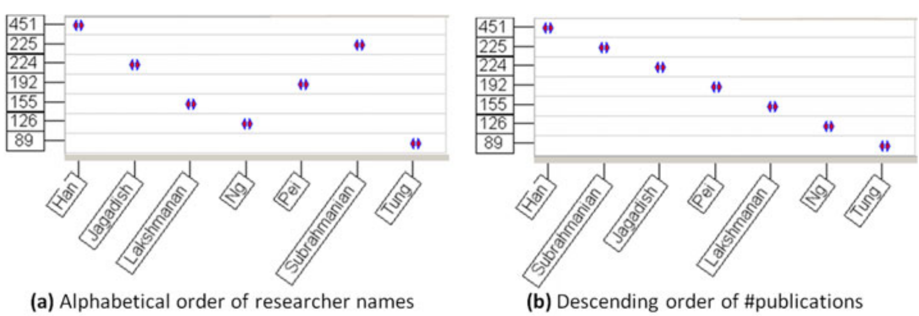


Fig. 3. SocialVis shows the numbers of papers authored by individual researchers

3.2 Visualizing Pairs of Researchers

Besides Q1, the following is the next commonly asked questions:

> **Q2.** Did this researcher collaborate with another researcher? If so, how many papers co-authored by them?

Answers to Q2 help users understand pairwise connections—in the form of joint publications—between pairs of researchers. SocialVis represents each pairwise connection using a horizontal line linking the left-pointing and right-pointing triangles (representing the two researchers) in the form of a bi-direction arrow ◂—▸. The use of horizontal lines avoids crossing over of lines (as in bipartite graphs). The *y*-position of the line explicitly indicates the number of co-authored papers. For example, Fig. 4(a) shows that Han & Pei co-authored 36 papers. It also shows that Han co-authored 10 papers with Ng and 10 papers with Tung.

When combined with the information depicted by Fig. 3, we can infer that, among 451 papers published by Han, 10 of them were co-authored with Ng (which means the remaining 451–10 = 441 papers were either solo publications of Han or the results of his other collaboration in which Ng did not participate).

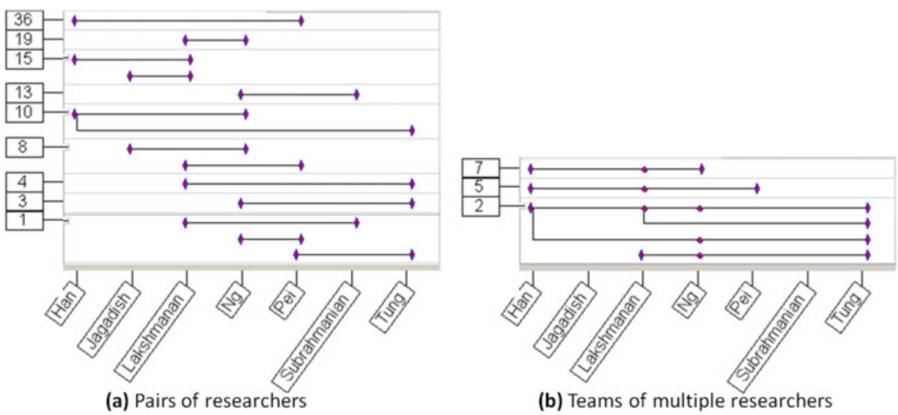


Fig. 4. SocialVis shows the numbers of papers co-authored by pairs or teams of researchers

3.3 Visualizing Collaborating Researchers in Teams

With the above two features of SocialVis (Sections 3.1 & 3.2), we can visualize the publication counts of each individual researcher and pair of researchers. For instance, we observed from Fig. 4(a) that there are pairwise connections between Han, Lakshmanan and Ng (e.g., Lakshmanan & Ng co-authored 19 papers, Han & Lakshmanan co-authored 15 papers, and Han & Ng co-authored 10 papers), from which we can infer that they together co-authored at most 10 papers (i.e., an *upper bound* for the number of their joint publications). Since we cannot infer the *exact* number of joint publications, it is unclear whether or not these three researchers collaborated together in a team. Hence, a logical question is:

> **Q3.** Did these k researchers (where $k \geq 3$) collaborate in a team? If so, exactly how many joint papers co-authored by them?

Answers to Q3 help users understand social linkage not only between two researchers but among multiple researchers. SocialVis uses a horizontal line to connect two triangles and k–2 circles representing all k researchers in a team (e.g., ◂–●–●–▸ represents a team of 4 researchers). The *y*-position of the line explicitly indicates the number of their joint publications. For example, Fig. 4(b) clearly shows that Han, Lakshmanan & Ng together co-authored 7 papers (cf. the upper bound of 10 papers inferred without using this feature of visualizing multi-researcher teams). As Ng participated in only 7 of 15 papers co-authored by Han & Lakshmanan, the remaining 15–7 = 8 papers were either written only by both Han & Lakshmanan or written together with their other collaborators.

Fig. 4(b) also shows that Han, Lakshmanan, Ng & Tung together co-authored 2 papers. This means that, among the 7 papers co-authored by the first three researchers, Tung co-authored only 2 of them, but he did not participate in the other 5 publications. Moreover, observing that the number of joint publications for Han, Lakshmanan & Tung is also 2, we can conclude that Ng participated in *all* the 2 papers jointly written by Han, Lakshmanan & Tung.

When combining the information depicted by Fig. 4(b) with that by Fig. 4(a), we make the following interesting observation: Although Han, Ng & Pei collaborated in pairs (with Han & Pei co-authored 36 papers, Han & Ng co-authored 10 papers, and Ng & Pei co-authored 1 paper), they did not write a joint paper together as indicated by the absence of any horizontal line connecting all three of them. This is different from the aforementioned ⟨Han, Lakshmanan, Ng⟩ team, in which the three researchers collaborated in pairs and all together.

Based on the numbers of joint publications of the ⟨Han, Lakshmanan⟩ and ⟨Han, Lakshmanan, Ng⟩ teams shown in Fig. 4, we conclude that Ng participated in only 7 out of the 15 papers co-authored by both Han & Lakshmanan. Furthermore, when observing the number of papers authored by Han is 451, we conclude Lakshmanan & Ng participated in only 7 out of these 451 papers (which implies that the remaining 451–7 = 444 papers were either solo publications of Han or the results of his other collaboration in which Lakshmanan & Ng did not participate together).

3.4 Visualizing the Entire or Partial Collaborating Teams

Based on the above observations, users can visualize the frequency information for teams of k researchers (for any $k \geq 1$) using the above three features of SocialVis (Sections 3.1-3.3). For example, users can conclude that Lakshmanan & Ng participated in 7 out of the 451 papers published by Han, but they may have difficulties in determining how many of the remaining 444 papers were written solely by Han and how many involved other collaborator. So, the following question is not uncommon:

> **Q4.** Did we have the complete list of co-authors for this paper? If so, how many co-authors are there? How many papers were jointly co-authored by all and only those researchers in this team?

Answers to Q4 help users understand (i) whether they found the *entire* frequently collaborating team or just a subset of it, (ii) the composition of the entire team, and (iii) the collaboration frequency of the entire team. To distinguish a complete team from a partial team, SocialVis replaces the right-pointing triangle with a bar ▮ for the complete team. See Fig. 5(a), which shows that Han is a sole author of 32 publications and Ng & Subrahmanian jointly published 8 papers (without any other co-authors). The figure also shows that the ⟨Han, Lakshmanan, Ng, Tung⟩ team co-authored 2 papers and its subset—the ⟨Han, Lakshmanan⟩ team—co-authored another 2 papers.

When we combine all the information depicted by Figs. 3-5, we get a better understanding of the networks. Recall from Section 3.2, we knew that Ng participated in 10 of 451 Han's papers, but we were uncertain about the remaining 441 papers. Now, with Fig. 5(a), we know Han wrote 32 papers alone, which means he co-authored the remaining 441–32 = 409 papers with researchers other than Ng. Fig. 5(a) also clears up the uncertainty in Section 3.3: (i) Among the 8 papers co-authored by Han & Lakshmanan but not Ng, 2 were written only by both Han & Lakshmanan (which means the remaining 6 were with Han & Lakshmanan's other collaborators besides Ng). (ii) Among the 444 Han's papers not co-authored with both Lakshmanan & Ng, 32 were solo publications of Han (which means the remaining 412 were the results of his other collaboration in which Lakshmanan & Ng did not participate together).

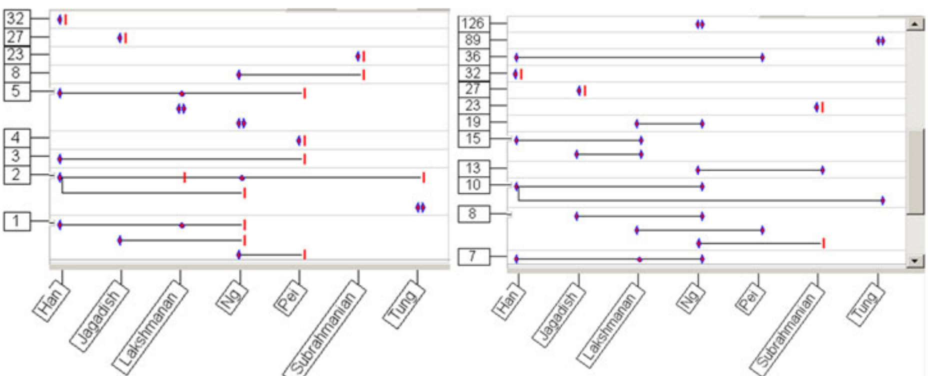


Fig. 5. SocialVis shows entire teams of researchers and their publication counts

3.5 Visualizing Large Co-authorship Networks

For clarity of our illustration, we only show features of SocialVis one at a time in each of the above figures (e.g., teams of multiple researchers in Fig. 4(b)). In general, users could select one or more features so that SocialVis displays the corresponding results on the same screen. See Fig. 5(b).

To visualize large co-authorship networks, SocialVis gives users an overview of all social information mined from the networks. To avoid over-crowdedness, SocialVis only displays some labels on the axes. As researcher names are (by default) arranged in alphabetical order on the *x*-axis, users can easily determine the hidden names. Moreover, SocialVis provides users with interactive features for selection/filtering so that they can focus on the area of user interest (e.g., some specific researchers and/or collaboration frequencies). Users can then zoom into, or out of, that area. SocialVis also provides users with scrollbars on both *x*- and *y*-directions so that users can easily scroll & explore different areas of the mined results and effectively access the useful social information about the networks.

4 Evaluation

To assess the effectiveness of SocialVis (our visual analytic tool) in conveying important social relationships (e.g., co-authorship information) and their frequency information mined from the social networks, we conducted a user evaluation. The evaluation was primarily case-based, within which users were required to answer different questions based on the information depicted by SocialVis. Sample questions include the following: How many papers are co-authored by Ng? Among them, how many were his sole publication? Did Ng collaborate with Han & Lakshmanan together? What is the number of their joint publications?

We recruited 18 participants, and none of them was exposed to our proposed SocialVis before. We began the evaluation by presenting our SocialVis and asking them to explore it at their own will. We did not give them any information regarding what the icons and representations meant in the visualization. We first questioned them on what they were able to identify. Due to our intuitive representation, the results showed that 78% of the participants were able to identify the basic meaning behind the representations (e.g., teams of *k* researchers and numbers of their joint publications). Only 22% of the participants had slight problems in distinguishing entire teams from partial teams. Afterwards, we gave the participants information on how to read the graphs (especially, the differences between right-pointing triangles and bars). Then, all participants were able to correctly answer all the given questions.

Moreover, these participants were asked to answer the same set of questions using other graphical representations of the network such as annotated node-link diagrams, weighted socio-matrices, hypergraphs, dual hypergraphs, and bipartite graphs. As expected, participants were only able to answer questions about individual or pairs of researchers but not teams of multiple researchers using the first two types of graphs. Most participants found it difficult to answer questions about multi-researcher teams using the latter three types of graphs as answers were not explicitly shown as by SocialVis. Participants need to manually dig out the information from overlapping clusters (in hypergraphs or dual hypergraphs) or crossing-over lines (in bipartite graphs) and to carefully count the numbers. As SocialVis clearly and explicitly provides the frequency information, participants can easier read this information.

5 Conclusions

A co-authorship network is one type of social networks, in which researchers are connected by their joint publications. In this paper, we proposed SocialVis to analyze and visualize these networks. Specifically, it applies frequent pattern mining to find useful social information such as the entire or partial teams of *k* frequently collaborating researchers and numbers of their joint or solo publications. It also presents these mining results graphically to users so that they can easily visualize the valuable social information about outcomes of collective intelligence and get a good understanding of the networks, which in turn helps users augment cognition. In general, SocialVis can serve either as a standalone visual analytic tool for revealing interesting social relationships among multiple entities in the networks or as a complement to existing visualizers by providing users with additional quantitative information such as publication counts. As ongoing work, we are extending SocialVis to analysis and visualize higher dimensional relationships such as where (venue) did the research papers published and/or when (year) did they published.

Acknowledgments. This project is partially supported by Natural Sciences and Engineering Research Council of Canada (NSERC) in the form of research grants.

References

1. Adams, R., Gill, S.P.: Augmented cognition, universal access and social intelligence in the information society. In: Schmorrow, D.D., Reeves, L.M. (eds.) FAC 2007, HCII 2007. LNCS (LNAI), vol. 4565, pp. 231–240. Springer, Heidelberg (2007)
2. Agrawal, R., Imielinski, T., Swami, A.N.: Mining association rules between sets of items in large databases. In: ACM SIGMOD 1993, pp. 207–216 (1993)
3. Benchettara, N., Kanawati, R., Rouveirol, C.: Supervised machine learning applied to link prediction in bipartite social networks. In: ASONAM 2010, pp. 326–330 (2010)
4. Biryukov, M.: Co-author network analysis in DBLP: classifying personal names. In: MCO 2008. CCIS, vol. 14, pp. 399–408. Springer, Heidelberg (2008)
5. Blanchard, J., Guillet, F., Briand, H.: Interactive visual exploration of association rules with rule-focusing methodology. KAIS 13(1), 43–75 (2007)
6. Carmichael, C.L., Leung, C.K.-S.: CloseViz: visualizing useful patterns. In: ACM UP 2010, pp. 17–26 (2010)
7. Carrington, P.J., Scott, J., Wasserman, S. (eds.): Models and Methods in Social Network Analysis. Cambridge University Press, Cambridge (2005)
8. Chi, E.H.: Augmented social cognition: using social web technology to enhance the ability of groups to remember, think, and reason. In: ACM SIGMOD 2009, pp. 973–984 (2009)
9. Goyal, A., Bonchi, F., Lakshmanan, L.V.S.: Learning influence probabilities in social networks. In: ACM WSDM 2010, pp. 241–250 (2010)
10. Han, J., Pei, J., Yin, Y., Mao, R.: Mining frequent patterns without candidate generation: a frequent-pattern tree approach. Data Mining and Knowledge Discovery 8(1), 53–87 (2004)
11. Han, Y., Zhou, B., Pei, J., Jia, Y.: Understanding importance of collaborations in co-authorship networks: a supportiveness analysis approach. In: SDM 2009, pp. 1111-1122 (2009)
12. Hansen, D.L., Shneiderman, B., Smith, M.A.: Analyzing Social Media Networks with NodeXL. Morgan Kaufmann, Burlington, MA (2011)
13. Lakshmanan, L.V.S., Leung, C.K.-S., Ng, R.T.: Efficient dynamic mining of constrained frequent sets. ACM TODS 28(4), 337–389 (2003)
14. Leung, C.K.-S., Carmichael, C.L.: FpVAT: a visual analytic tool for supporting frequent pattern mining. SIGKDD Explorations 11(2), 39–48 (2009)
15. Leung, C.K.-S., Carmichael, C.L.: Exploring social networks: a frequent pattern visualization approach. In: IEEE SocialCom 2010, pp. 419–424 (2010)

16. Leung, C.K.-S., Irani, P.P., Carmichael, C.L.: FIsViz: a frequent itemset visualizer. In: Washio, T., Suzuki, E., Ting, K.M., Inokuchi, A. (eds.) PAKDD 2008. LNCS (LNAI), vol. 5012, pp. 644–652. Springer, Heidelberg (2008)
17. Leung, C.K.-S., Irani, P.P., Carmichael, C.L.: WiFIsViz: effective visualization of frequent itemsets. In: IEEE ICDM 2008, pp. 875–880 (2008)
18. Leung, C.K.-S., Khan, Q.I., Li, Z., Hoque, T.: CanTree: a canonical-order tree for incremental frequent-pattern mining. KAIS 11(3), 287–311 (2007)
19. Lévy, P.: Toward a self-referential collective intelligence some philosophical background of the IEML research program. In: Nguyen, N.T., Kowalczyk, R., Chen, S.-M. (eds.) ICCCI 2009. LNCS (LNAI), vol. 5796, pp. 22–35. Springer, Heidelberg (2009)
20. Makanju, A., Brooks, S., Zincir-Heywood, A.N., Milios, E.E.: LogView: visualizing event log clusters. In: PST 2008, pp. 99–108 (2008)
21. Milani Fard, A., Ester, M.: Collaborative mining in multiple social networks data for criminal group discovery. In: IEEE SocialCom 2009, pp. 582–587 (2009)
22. Misue, K.: Visual analysis tool for bipartite networks. In: Lovrek, I., Howlett, R.J., Jain, L.C. (eds.) KES 2008, Part II. LNCS (LNAI), vol. 5178, pp. 871–878. Springer, Heidelberg (2008)
23. Ugai, T., Aoyama, K.: Organization diagnosis tools based on social network analysis. In: Smith, M.J., Salvendy, G. (eds.) Human Interface 2009, Part I, HCII 2009. LNCS, vol. 5617, pp. 181–189. Springer, Heidelberg (2009)
24. van Ham, F., Schulz, H.-J., DiMicco, J.M.: Honeycomb: visual analysis of large scale social networks. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009, Part II . LNCS, vol. 5727, pp. 429–442. Springer, Heidelberg (2009)
25. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge (1994)

The Crowdsourcing Design Space

Yasuaki Sakamoto, Yuko Tanaka, Lixiu Yu, and Jeffrey V. Nickerson

Center for Decision Technologies, Stevens Institute of Technology
{ysakamot,yuko.tanaka,lyu3,jnickerson}@stevens.edu

Abstract. Crowdsourcing is a new kind of organizational structure, one that is conducive to large amounts of short parallel work: thousands of individuals may work for several minutes on tasks, their outputs aggregated into a useful product or service. The dimensions of this new organizational form are described. Areas for future research are identified, focusing on open-ended tasks and the coordination structures that might foster collective creativity.

Keywords: Crowdsourcing, distributed cognition, organizational design, peer production, collective creativity, human computation.

1 Introduction

Crowdsourcing, the assembling of strangers to accomplish a task [1-4], has the potential to transform the nature of work [5]. Many companies are sensing this, and adapting it to solve problems and provide services [6-8]. Crowdsourcing is becoming popular because the Internet-based infrastructure to support the management of crowds has grown and strengthened. This infrastructure provides opportunities not just for businesspeople, but also for researchers who want to study distributed cognition [9]. Specifically, crowds can participate in experiments: crowds can be simultaneously assembled in both control and experimental conditions, so the effects of social networks, organizational structure, and information flow can be studied (e.g., [10, 11]). Because of its potential for both business change and fundamental research, the crowdsourcing phenomenon deserves systematic study through analysis and experimentation.

There are several ways crowdsourcing differs from other forms of organization. First, the assembly of a crowd can happen quickly, and need only persist for short amounts of time – minutes, hours, or days. In contrast, social institutions tend to persist for many years, and take time to build. For example, the hiring of one thousand individuals into a company is a large undertaking, but a crowd of 1000 can be pulled together and asked to perform a simple task, all in the space of less than one hour. Moreover, a crowd can be assembled from around the world, constituted from different nationalities, cultures, and professions.

What motivates crowds to participate? Some sites provide money, others provide reputational incentives, and still others provide neither of these. But crowds will perform tasks they find interesting, whether or not they receive monetary incentives. In particular, games have proved popular: game players have helped label images [12] and discover ways to fold proteins [13]. These successes suggest that crowds might be

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 346–355, 2011.

assembled to solve a wide variety of problems, if such problems can be broken down into fun-to-play activities.

In fact, for many tasks, a crowd might be more efficient and effective than an expert. For example, Galton showed that a crowd could guess the weight of an ox better than a farmer [14]. Other work showed that groups of people performed better than individuals on tasks such as traversing a maze if their decisions were aggregated [15, 16]. Such techniques are effective probably because aggregation cancels out individual errors and reinforces correct solutions [17].

As a way of synthesizing crowdsourcing research, we will describe the dimensions that define a design space for crowdsourcing systems. We then discuss four dimensions in particular: the nature of the task, the communication methods, the levels of hierarchy, and the workflow. Along the way we point out possibilities for structuring the crowd in new ways, and opportunities for understanding more about collective cognition.

2 Dimensions of Crowdsourcing

A design space identifies the range of possible solutions for a design problem [18]. The space is delineated by set of dimensions that structure and constrain the set of decisions made in creating a design.

Malone et al. [19] analyzed a related space – collective intelligence – in terms of four broad questions: *Who, What, How*, and *Why*: applying this to crowdsourcing leads to a breakdown of the space into the users of the system, the tasks to be performed, the method for crowdsourcing, and the motivation of the crowd. The authors also identify more specific dimensions they call *genes* that can be combined into *genomes* in order to constitute a new application.

In contrast, Quinn and Bederson [2] use a different set of dimensions to describe another related space they call human computation: *Motivation, Human Skill, Aggregation, Quality Control, Process Order, and Task-request Cardinality.* These can be reconciled with the four broad questions of Malone et al.: *Aggregation* and *Process Order* address a big part of *How*; *Task-request Cardinality* and *Human Skill* relate to *What. Motivation* focuses on how incentives are used to encourage participation. Table 1 shows the four broad dimensions, and several specific dimensions from Quinn and Bederson.

We add several dimensions to the mix. In the *Who* category, we include demographics and level of expertise. Participants vary along demographic dimensions, such as age and nationality, as well as along the level of expertise they have in the task domain. These additional dimensions may become important when allocating tasks and assessing crowd performance. In the *What* category, we add the task domain (e.g., image labeling) the nature of the task (e.g., generation) and the output (e.g., list of terms). These dimensions highlight the context for crowdsourcing. The *How* category includes an evaluation method. Evaluation processes can increase the quality of responses. *How* also includes incentives as distinct from workers' motivations, an aspect of *Why*. That is, one can offer cash but still find that workers perform the task for different reasons: there is a rich literature on the relation between motivation and incentives that can be brought to bear on these two dimensions [20, 21]. We further include visibility of outputs as a dimension of *How* – for example, one might restrict crowd members from seeing each other's work. We will elaborate other attributes of *How* in in the following sections: protocols for communication between workers, hierarchical levels within the crowd, and the sequence of work. We distinguish two kinds of motivation in *Why*: the reasons for crowdsourcing a task, and the reasons for people to work on a task. For example, companies might use crowdsourcing to optimize profit [8], while researchers might do so to increase knowledge in a domain [13] or to understand an aspect of distributed cognition [22, 23].

Table 1. The dimensions of crowdsourcing (italics flag sections in the paper)

Broad dimension	Specific dimension	Example attribute
Who	Demographics	age, country
	Level of Expertise	novice, expert
What	Domain of the task	protein folding, image labeling
	Nature of the task	recognition, generation
	Output	sequence of protein folds
How	Incentives	contest prize for the best task
	Aggregation method	collection, combination
	Evaluation method	vote, expert opinion
	Visibility of outputs	opaque, transparent
	Communication	mediated through the tasks
	Levels of hierarchy	single, multiple
	Workflow	evaluation following collection
Why	Requester's motivation	profit, knowledge
	Worker's motivation	money, fun

Next, we focus on four areas in this design space that offer great promise as topics for future investigation. The authors have used crowdsourcing to perform generative, creative tasks [22, 23]. This topic often interests both industry and academe, so we address it first.

3 Nature of the Task: Creativity and the Crowd

Views of creativity reflect the times: the romantic era's mystique of the lone inventor inspired late in the night has been eclipsed by a view of the creative act as a social one, made by teams of closely knit people [24]. Art schools and design studios are social places: so are engineering institutes and research laboratories. The view, then, has moved from individual to group, and the groups are seen as cohesive and close knit. Nevertheless, people remain uneasy with the idea of collaborating across great distances. Indeed, some studies of geographically-distributed collaboration have identified a tradeoff between convenience and cost [25]. The Internet can reduce the cost of coordinating work among geographically-distributed, but otherwise traditional, teams. However, the Internet also makes possible a new model of collaboration: combining the work of isolated strangers.

Inspired by this new model of collaboration, we decided to assess the sequential application of crowds to both generate and aggregate design ideas. If this approach to idea generation works well, then many difficult design challenges might be attacked in parallel. If the approach has limits, then in finding these limits we will provide a more focused path for research on peer production (cf. [5]).

In one experiment, we combined the work of many designers working in parallel to create a composite design [26]. We found that taking elements from many different designs according a simple heuristic returned a strong design, as judged by experts. This suggests that Galton's insight into the crowd's ability to estimate [14] also applies to the crowd's ability to design, perhaps for the same reason: individual errors disappear, and strong connections reinforced. Conceptual combination [27] can be controlled at a finer scale, with pair-wise combination techniques such as those used in genetic algorithms [28-31]. We have performed experiments using these techniques [22, 23]. To do so, we needed to understand and implement mechanisms for connecting crowd output. We turn to these next.

4 Communication between Crowd Members

A member of an online crowd may have no connection to another member except through participation on the same project. On the other hand, in physical crowds, those in the immediate vicinity often influence one another. For example, members of a crowd at an outside concert will be aware of who is around, and will move in order to fill in gaps or create more breathing room. There may be no verbal interaction, but there may be awareness. Moreover, there is mediation: members of a crowd streaming toward an exit may affect members more distant through a wave of local adjustments.

Such awareness can exist electronically – for example, in virtual worlds. But in other cases, the crowd can't directly see others situated in space, whether cyber or physical. There have been many approaches to understanding how ideas flow and people interact in such environments. Coordination theory [32] focused on the resources through which people work toward goals. Electronic brainstorming looked at the way computers mediate communication [33, 34]. A general theme in such research is that team collaboration is not always useful. So collaboration between crowd members may or may not be a good idea.

In a strictly separated crowd, one might only let people interact through intermediate productive output. That is, one member of a crowd may be asked to edit the output of another crowd member. There is some coordination, but the interaction is very limited.

In more loosely separated environments, one might see more of what others have done, and one can modify – remix – that output. For example, in the Scratch environment, youths find examples of each other's work, modify it, and upload it [35]. In the Yahoo! Answers environment, users doesn't modify the contributions, but instead comment on them. Users make critical comments on answers if they think that the answers are wrong, and other users may provide new answers after reading the comments. In crowdsourcing environments, even very self-conscious individuals who have difficulty providing critical comments in the presence of others may feel comfortable doing so [36].

Both of these are forms of collaboration – commenting and remixing - can be combined. In Scratch, youths have engaged in a kind of recursive crowdsourcing – by using open-ended discussion forums, they initiate their own contests ask their own peers to enter, and provide their own prizes [37]. The above examples provide crowd members a way to react to or modify a single idea. What about providing participants with two ideas?

In a human *genetic algorithm* two ideas are combined [22, 30]. There is an integration of the crowds' ideas, but the integration works in one direction, with members never collaborating interactively with each other. The ideas can be textual, or involve images, and can be combined across many generations [23].

Games also provide a way for pairs of people to interact and integrate ideas. In *games with a purpose* the participants perform a collaborative task [12]. For example, in an image labeling game, they try to predict how the other person would label an image. Their responses provide a list of labels for images that are difficult for computers to come up with. These tasks involve a structured kind of collaboration, and are often restricted to pairs of people.

Larger groups can also communicate through games. Specifically, lab experiments have been conducted, in which participants are configured into social network structures and asked to perform tasks with known outcomes. For example, graph-coloring problems performed by a group connected one way will produce different results from a similar group connected another way [38]. Similar experiments have been run online, in which members of a crowd are connected in different ways, and given the opportunity to exchange information to aid in a search task [11].

Most of the previously described work doesn't allow people to interact with each other using natural language. Such interaction can itself be experimentally controlled: for example, one person can look at input and describe it to another [39]. There is opportunity to perform such work at large scales by providing crowd members with the ability to have structured dyadic conversations with each other around a goal, such as understanding a new concept or recognizing a threat. Moreover, experiments might also look at the ways crowd members evolve their own ways of communicating. For example, studies have shown that even when verbal and iconic communication are removed, pairs of people often invent their own task specific signaling systems [40].

What other coordination techniques can be used in open-ended tasks that lend themselves to creative outputs? In a project called *SwarmArt,* each member of the crowd can add a line to a drawing, and then adjust the strength of other participants' lines [41]. Each member of the crowd can see all previous members' inputs, and each member is both a generator and an evaluator of the input to date.

In sum, there are many ways crowds can be connected, as shown in Table 2.

Open-ended creative tasks might or might not benefit from the collaboration methods just discussed. The more one sees and hears of one's neighbors, the more one might fixate on the previous solutions. On the other hand, the more one sees, the more one might be inspired to produce a novel ideas triggered by someone else's idea. How much to reveal to members of the crowd, and when, is a question open for more research. Given the use of a strong collaboration method, it is possible to create simple, flat organizations, or complex, nested organizations, the next section's topic.

Table 2. Crowd connection schemes

Connection through:	Example action	Reference
virtual proximity	Approach someone in a virtual world	[42]
an overall task	Guess the weight of an animal	[14]
seeing two examples	Combine	[22, 23]
modifying another's work	Remix a computer program	[35], [6]
being a member of a team	Collaborate with team members	[43], [8], [13]
focus on one example	Critique	[35] [36]
playing a two-person game	Guess image labels	[12]
playing a many-person game	Solve NP-Complete problems	[38], [11]
dyadic conversation	Recognize threats	[39]
discussion boards	Code software	[35], [44]

5 Levels of Hierarchy

In organizations, tasks are broken up and assigned to different people. These people can sometimes in turn break up the tasks and assign them. Multiple levels of hierarchy allow an organization to scale work, and adapt flexibly to new demands. On the other hand, such hierarchies introduce bottlenecks.

Most paid crowdsourcing uses one level of hierarchy: the task originator asks people to contribute, and the originator performs the consolidation. Since workers are usually being paid *after* performing a task, they as workers may have little appetite for creating paid tasks for others to perform, which would introduce credit risk.

Still, it is possible to imagine recursive crowdsourcing, where members of the crowd ask others to accomplish a task. This would be useful in addressing open-ended creative problems: groups could create their own strategies for performing the task, and then assemble a crowd to follow their particular strategy.

Indeed, some sites are exploring similar structures. InnoCentive allows someone to form a team around an idea, and specify up-front design the compensation division between the leader and the team members [8]. By forming a new team, a leader participates in recursive crowdsourcing. The user-initiated contests of Scratch have a similar feel: there is an overall running contest, the reward of which is to see one's project displayed on the front page. Unexpectedly, participants have created their own contests [37]. Indeed, some groups of Scratch users describe themselves as "companies" [45]. Furthermore, sites related to citizen science and volunteerism allow for subgroups to be formed, creating at least two levels of hierarchy [46]. Finally, the winning team in the red balloon contest encouraged people to recruit others to find a set of balloons, and compensated them based on the structure of the recruiting tree [47].

In order to study the use of hierarchy in crowds, one needs ways of creating recursive crowdsourcing experiments. If money is involved, then perhaps some combination of pre and post payment might be used, to both reduce credit risk and to provide incentive for the completion of subtasks. For non-monetary crowdsourcing, ways of making announcements and coordinating work are needed, available to not just the initial requesters, but also to the workers who may want to also crowdsource their own tasks.

This area of research, while complex, would allow for the solution of complex open-ended problems that might require not just many workers, but also many different approaches.

6 Workflow

Crowdsourcing uses workflow: tasks are split up, modified, and then completed. Most companies use workflow: for example, expense vouchers are routed to managers and accounting departments for approval and payment. There has been much research on how to describe workflows formally [48, 49]. In an alternative approach, Malone et al. suggest a metaphor: the subtasks of crowdsourcing are genes, and these can be combined to form genomes [19].

This metaphor suggests that a crowdsourcing task design involves not just picking one attribute for each different dimension, but might involve picking more than one attribute that might be realized in different stages of crowdsourcing activity. For example, Threadless uses the crowd to vote on T-shirts, but then uses the site operators, presumably experts, to make the final determination of what to manufacture [7].

One area for research might involve taking the metaphorical idea of a genome, and combine it with formal definitions of workflow. Then, the design space includes different ways of assigning and sequencing recognition, generation, and evaluation tasks.

The crowd itself might be useful in designing and evaluating ways to structure work on a particular topic. For example, the crowd may first participate in two alternative processes, and then suggest ways of combining them to create a more effective process.

7 Concluding Thoughts

Crowdsourcing can become many things. Right now, the space of possible ways of coordinating large numbers of people is relatively unexplored. There is an opportunity to structure the crowd in new ways. There is also an opportunity to understand more about collective cognition by designing crowd-based experiments. Several areas seem particularly promising for research. First, the crowd can perform open-ended, creative tasks. Through this, we may solve large-scale social problems, and gain insight into social creativity. Second, crowd communication can run the spectrum from total isolation to open collaboration, and through experiments we may begin to understand how much and what kind of collaboration is fruitful for specific tasks. Third, crowd members can potentially crowdsource, creating flexible ad hoc hierarchies: unleashing this potential may be important for handling creative tasks. Fourth, crowds can be organized in new and interesting ways by selecting dimensional attributes from the design space and designing workflows that allocate and sequence tasks accordingly. Exploring this space may yield new organizations; eventually, the crowd will design itself.

Acknowledgements. This research was supported by the National Science Foundation under awards IIS-0855995 and IIS-0968561.

References

1. Howe, J.: Crowdsourcing: Why the power of the crowd is driving the future of business. Three Rivers Pr. (2009)
2. Quinn, A.J., Bederson, B.B.: Human Computation: A Survey and Taxonomy of a Growing Field. In: CHI. ACM Press, New York (2011)
3. Kittur, A.: Crowdsourcing, collaboration and creativity. Crossroads, XRDS (2010)
4. Little, G., Chilton, L.B., Goldman, M., Miller, R.C.: Exploring iterative and parallel human computation processes. In: Proceedings of the ACM SIGKDD Workshop on Human Computation, pp. 68–76. ACM, New York (2010)
5. Benkler, Y.: The wealth of networks: How social production transforms markets and freedom. Yale Univ Press (2006)
6. Gulley, N.: Patterns of innovation: a web-based MATLAB programming contest. In: CHI 2001 extended abstracts on Human factors in computing systems, p. 338. ACM, New York (2001)
7. Brabham, D.: Moving the crowd at Threadless: Motivations for participation in a crowdsourcing application. Information, Communication & Society 13, 1122–1145 (2010)
8. Jain, R.: Investigation of Governance Mechanisms for Crowdsourcing Initiatives. In: AMCIS 2010 Proceedings, p. 557 (2010)
9. Hollan, J., Hutchins, E., Kirsh, D.: Distributed cognition: toward a new foundation for human-computer interaction research. ACM Transactions on Computer-Human Interaction (TOCHI) 7(2), 174–196 (2000)
10. Salganik, M.J., Dodds, P.S., Watts, D.J.: Experimental study of inequality and unpredictability in an artificial cultural market. Science 311(5762), 854 (2006)
11. Mason, W.A., Jones, A., Goldstone, R.L.: Propagation of innovations in networked groups. Journal of Experimental Psychology-General 137(3), 422–433 (2008)
12. Von Ahn, L., Dabbish, L.: Designing games with a purpose. Communications of the ACM 51(8), 58–67 (2008)
13. Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., Leaver-Fay, A., Baker, D., Popović, Z.: Predicting protein structures with a multiplayer online game. Nature 466(7307), 756–760
14. Galton, F.: Vox Populi. Nature, 450–451 (1907)
15. Gurnee, H.: Maze Learning in the Collective Situation. The Journal of Psychology 3, 437–443 (1937)
16. Knight, H.C.: A Comparison of the Reliability of Group and Individual Judgments. Master's Thesis, Columbia University (1921)
17. Estes, W.K., Maddox, W.: Risks of drawing inferences about cognitive processes from model fits to individual versus average performance. Psychonomic Bulletin & Review 12(3), 403 (2005)
18. Brooks, F.: The Design of Design: Essays from a Computer Scientist. Addison Wesley, New York (2010)
19. Malone, T.W., Laubacher, R., Dellarocas, C.: Harnessing crowds: Mapping the genome of collective intelligence. MIT Sloan School Working Paper 4732-09 (2010)
20. Boudreau, K., Lacetera, N., Lakhani, K.: Parallel search, incentives and problem type: Revisiting the competition and innovation link. Harvard Business School, Working Paper 09-041, 2008 (2008)

21. Deci, E.L., Koestner, R., Ryan, R.M.: A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychological bulletin 125, 627–668 (1999)
22. Nickerson, J.V., Sakamoto, Y.: Crowdsourcing Creativity: Combining Ideas in Networks. In: Workshops on Information in Networks (2010)
23. Yu, L., Nickerson, J.V.: Cooks or Cobblers? Crowd Creativity through Combination. In: CHI. ACM Press, New York (2011)
24. Amabile, T.: Creativity in context. Westview Press (1996)
25. Hinds, P., Kiesler, S.: Distributed work. The MIT Press, Cambridge (2002)
26. Nickerson, J.V., Corter, J.E., Tversky, B., Zahner, D., Rho, Y.J.: The Spatial Nature of Thought: Understanding Systems Design Through Diagrams. In: ICIS 2008 Proceedings, p. 216 (2008)
27. Thagard, P.: Conceptual revolutions. Princeton University Press, Princeton (1992)
28. Goldberg, D.E.: Genetic algorithms in search, optimization, and machine learning. Addison-Wesley, Reading (1989)
29. Holland, J.H.: Building blocks, cohort genetic algorithms, and hyperplane-defined functions. Evolutionary computation 8(4), 373–391 (2000)
30. Kosorukoff, A.: Human based genetic algorithm. In: IEEE International Conference on Systems, Man, and Cybernetics, vol. 5, pp. 3464–3469. IEEE, Los Alamitos (2002)
31. Gero, J.S.: Computational models of innovative and creative design processes. Technological Forecasting and Social Change 64(2-3), 183–196 (2000)
32. Malone, T.W., Crowston, K.: The interdisciplinary study of coordination. ACM Computing Surveys (CSUR) 26(1), 87–119 (1994)
33. Dennis, A., Williams, M.: Electronic Brainstorming. In: Group creativity: Innovation through collaboration, pp. 160–178 (2003)
34. Nunamaker, J.F., Dennis, A.R., Valacich, J.S., Vogel, D., George, J.F.: Electronic meeting systems. Communications of the ACM 34(7), 40–61 (1991)
35. Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., Millner, A., Rosenbaum, E., Silver, J., Silverman, B.: Scratch: programming for all. Communications of the ACM 52(11), 60–67 (2009)
36. Tanaka, Y., Mochizuki, T., Manalo, E., Kusumi, T.: Cultural differences between Asian students regarding judgments about using critical thinking. In: 14th International Conference on Thinking (2009)
37. Nickerson, J.V., Monroy-Hernandez, A.: Appropriation and Creativity: User Initiated Contests in Scratch. In: Hawaii International Conference on System Sciences (2011)
38. Kearns, M., Suri, S., Montfort, N.: An experimental study of the coloring problem on human subject networks. Science 313(5788), 824 (2006)
39. Voiklis, J.: A Thing Is What We Say It Is: Referential Communication and Indirect Category Learning. Ph.D., Columbia University (2008)
40. Galantucci, B.: An experimental study of the emergence of human communication systems. Cognitive Science: A Multidisciplinary Journal 29(5), 737–767 (2005)
41. Boyd, J.E., Hushlak, G., Jacob, C.J.: SwarmArt: interactive art from swarm intelligence. In: Proceedings of the 12th annual ACM international conference on Multimedia, pp. 628–635. ACM, New York (2004)
42. Boellstorff, T.: Coming of age in Second Life: An anthropologist explores the virtually human. Princeton University Press, Princeton (2008)
43. Bell, R.M., Koren, Y.: Lessons from the Netflix prize challenge. ACM SIGKDD Explorations Newsletter 9(2), 75–79 (2007)

44. West, J.: How open is open enough? Melding proprietary and open source platform strategies. Research Policy 32(7), 1259–1285 (2003)
45. `http://blog.scratch.mit.edu/2010/01/scratch-companies-by-guest-blogger.html` (2010)
46. Raddick, J., Lintott, C.J., Schawinski, K., Thomas, D., Nichol, R.C., Andreescu, D., Bamford, S., Land, K.R., Murray, P., Slosar, A.: Galaxy Zoo: an experiment in public science participation. Bulletin of the American Astronomical Society 38, 892 (2007)
47. Pickard, G., Rahwan, I., Pan, W., Cebrian, M., Crane, R., Madan, A., Pentland, A.: Time Critical Social Mobilization: The DARPA Network Challenge Winning Strategy. J Arxiv preprint arXiv:1008.3172 (2010)
48. van Der Aalst, W.M.P., Ter Hofstede, A.H.M., Kiepuszewski, B., Barros, A.P.: Workflow patterns. Distributed and parallel databases 14(1), 5–51 (2003)
49. Milner, R.: Communicating and mobile systems: the pi-calculus. Cambridge University Press, Cambridge (1999)

Developing Systems for the Rapid Modeling of Team Neurodynamics

Ronald H. Stevens[1], Trysha Galloway[1], Chris Berka[2], and Peter Wang[1]

[1] UCLA IMMEX Project 5601 W. Slauson Ave. #272,
Culver City, CA 90230, USA
[2] Advanced Brain Monitoring, Inc. 2237 Faraday Ave., Suite 100,
Carlsbad, CA 92008, USA
immex_ron@hotmail.com

Abstract. Cognitive Neurophysiologic synchronies (NS) are a low level data stream derived from EEG measurements that can be collected and analyzed in near real time and in realistic settings. We are using NS to develop systems that can rapidly determine the functional status of a team with the goals of being able to assess the quality of a teams' performance / decisions, and to adaptively rearrange the team or task components to better optimize the team. EEG-derived measures of engagement from Submarine Piloting and Navigation team members were normalized and pattern classified by self-organizing artificial neural networks and hidden Markov models. The temporal expression of these patterns were mapped onto team events and related to the frequency of team members' speech. Standardized models were created using pooled data from multiple teams and were used to compare NS expression across teams, training sessions and levels of expertise. These models have also been incorporated into software systems that can provide for rapid (minutes) after training feedback to the team and provide a framework for future real-time monitoring.

Keywords: Collaboration, EEG, Neurophysiologic synchrony.

1 Introduction

The integrated thinking of a team is often described by the dynamic construct of team cognition which reflects the interrelated cognitions, behaviors and attitudes that contribute to team performance [1]. One of the challenges for studying team cognition in real-time is the development of unobtrusive and relevant measures of team performance that can be practically implemented in real-world environments [2].

We have explored using the simultaneous expression of EEG-derived cognitive measures by different members of a team as an alternative to verbal communication streams for constructing teamwork models. In this approach the values of a cognitive measure at a particular point in time are aggregated across the team members into a vector that is then clustered / classified by artificial neural network (ANN) technologies [3,4]. This results in a series of patterns termed Neurophysiologic Synchronies (NS) which are defined as the second-by-second quantitative co-expression of the same neurophysiologic / cognitive measures by different members

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 356–365, 2011.

of the team. The cognitive measures we have modeled include High Engagement and High Workload which have been derived from EEG data streams [5]. We have reasoned that if NS expression is a meaningful dynamic construct then their expression should:

1. Be sensitive to long and short-term task changes;
2. Relate to some established aspects of team cognition, yet reveal something new;
3. Be usable as well as useful;
4. Distinguish novice / expert performance; and,
5. Be sensitive to the effects of training.

Prior studies have documented the dynamics of NS expression in response to long and short term changes in the task [4]. In those studies the NS models were generated individually for each training session, i.e. they were autologous models. While those modeling approaches were useful research tools there were limitations for their practical application to training activities. First, as new models had to be created for each task and team it was difficult to compare across sessions / teams or levels of expertise as the pattern designations changed for each new model. Also, without standardized models it would be difficult to begin to extend this analysis to real-time team modeling. In this study we have developed models using pooled data from multiple teams to develop a generic set of models that remove these limitations.

2 Tasks and Methods

2.1 Submarine Piloting and Navigation (SPAN)

These studies were conducted with navigation training tasks that are integral components of the Submarine Officer Advanced Course at the US Navy Submarine School, Groton, CT. The task is a high fidelity Submarine Piloting and Navigation (SPAN) simulation that contains dynamically programmed situation events which are crafted to serve as the foundation of the adaptive team training. Such events in the SPAN include encounters with approaching ship traffic, the need to avoid shoals, changing weather conditions, and instrument failure. There are task-oriented cues to guide the mission, team-member cues that provide information on how other members of the team are performing / communicating, and adaptive behaviors that help the team adjust in cases where one or more members are under stress or are not familiar with aspects of the unfolding situation.

Each SPAN session contains three segments. It begins with a Briefing segment where the overall goals of the mission are presented along with information on position, contacts, weather, sea state, etc. Major participants during the Briefing are the Navigator, the Contact Coordinator along with the Captain, Instructor and / or Evaluator.

The Scenario is an evolving task and is more dynamic than the Brief containing easily identified processes of teamwork along with others which are less well defined. One of the more obvious processes is the regular updating of the ship's position termed 'Rounds'. Here, three navigation points are chosen, usually visually, and the bearing of each to the boat is rapidly measured and plotted on a chart.

Interleaved with these deterministic events are situations that arise due to new ship traffic, increased proximity to hazards, equipment malfunctions, reduced visibility or similar events. In contrast to the regular updating of the submarine's position, these events can be regarded as perturbations to the regular functioning of the team and indicate interesting points where the resilience of the team may be tested. Some events appear rapidly like a man overboard, while others evolve over 5-10 minutes and are based, in part, on previous decisions.

In the Debrief section there is open discussion of what worked, what other options were available, and long and short term lessons. The Debrief is the most structured part of the training with team members reporting in order, beginning with the Navigator. Within this reporting structure there is often overlapping or underlying nested structures where specific events within the Scenario are discussed.

2.2 EEG Metrics

The EEG data acquired from the wireless headset developed by Advanced Brain Monitoring, Inc. uses an integrated hardware and software solution for acquisition and real-time analysis of the EEG [5, 6]. It has demonstrated feasibility for acquiring high quality EEG in real-world environments including workplace, classroom and military operational settings. The system contains an easily-applied wireless EEG system that includes intelligent software designed to identify and eliminate multiple sources of biological and environmental contamination and allow real-time classification of cognitive state changes even in challenging environments. The 9-channel wireless headset includes sensor site locations: F3, F4, C3, C4, P3, P4, Fz, Cz, POz in a monopolar configuration referenced to linked mastoids. ABM B-Alert® software acquires the data and quantifies alertness, engagement and mental workload in real-time using linear and quadratic discriminant function analyses (DFA) with model-selected PSD variables in each of the 1-hz bins from 1-40hz, ratios of power bins, event-related power (PERP) and/or wavelet transform calculations.

To monitor "mental workload" (WL) and "engagement" (E) using the B-Alert model EEG metrics, values ranging from 0.1-1.0, are calculated for each 1-second epoch of EEG. Simple baseline tasks are used to fit the EEG classification algorithms to the individual so that the cognitive state models can then be applied to increasingly complex task environments, providing a highly sensitive and specific technique for identifying an individual's neural signatures of cognition in both real-time and offline analysis. These methods have proven valid in EEG quantification of drowsiness-alertness during driving simulation, simple and complex cognitive tasks and in military, industrial and educational simulation environments, quantifying mental workload in military simulation environments, distinguishing spatial and verbal processing in simple and complex tasks, characterizing alertness and memory deficits in patients with obstructive sleep apnea, and identifying individual differences in susceptibility to the effects of sleep deprivation.

2.3 Experimental Protocol

In prior studies analyzing the dynamics of problem solving with individuals we used the raw EEG-E and EEG-WL data streams [7]. Studying team processes using these

EEG measures requires a normalization step, which equates the absolute levels of EEG-E or EEG-WL of each team member with his/her own average levels. This allows the identification of whether an individual team member is experiencing above or below average levels of EEG-E or EEG-WL and whether the team as a whole is experiencing above or below average levels. As described previously (Stevens et al, 2010a) in this normalization process the EEG-E levels are partitioned into the upper 33%, the lower 33% and the middle 33%; these are assigned values of 3, -1, and 1 respectively, values chosen to enhance visualizations.

The next step combines these values at each epoch for each team member into a vector representing the state of EEG-E for the team as a whole; these vectors are used to train unsupervised artificial neural networks to classify the state of the team at any point in time. In this process the second-by-second normalized values of team EEG-E for the entire episode are repeatedly (50-2000 times) presented to a 1 x 25 node unsupervised artificial neural network.

During this training a topology develops such that the EEG-E vectors most similar to each other become located closer together and more disparate vectors are pushed away. The result of this training is a series of 25 patterns that we call NS Patterns that show the relative levels of EEG-E for each team member on a second-by-second basis. A profile of a generic NS Pattern is shown in Figure 1 for a six person team. Here, team members 3 and 5 have above average levels of this neurophysiologic measure and the other team members are below average.

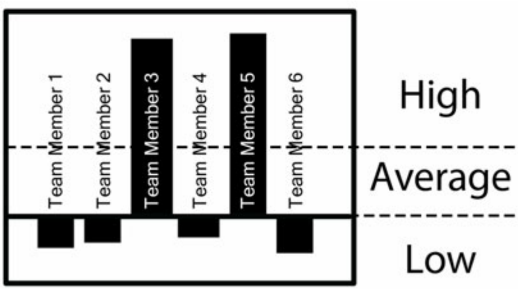


Fig. 1. Expression of a generic NS measure by members of a six-person team

NS Pattern expression can be thought of as output symbols from hidden states of a team, and if so the sequence may give additional information about those states. Hidden Markov modeling (HMM) would seem an appropriate approach for such modeling. The NS data stream for the combined team data was segmented into sequences of 10 to 240 seconds generating NS symbol arrays. HMMs were trained using these arrays assuming 5 hidden states as we have performed previously for modeling problem solving learning trajectories [8]. Training was for 500 epochs and resulted in a convergence of 0.0001. Next the most likely state sequence through the performance was generated by the Viterbi algorithm. The outputs of this subsequent modeling of NS Pattern streams by HMM are termed NS States.

3 Experimental Results

3.1 Detection of Long and Short-Term Task Changes by Autologous and Heterologous NS Models

For the generation of heterologous ANN and HMM models EEG-E data was pooled from 8 SPAN sessions from different experienced and novice navigation teams. This resulted in 31,450 team training vectors (~ 5.5 hours of teamwork) which were used as the training set. For all these sessions the position of each of the team members in the training vector was the same. This order was QMOW = Quartermaster on Watch in position 1, NAV = Navigator in position 2, OOD = Officer on Deck in position 3, ANAV = Assistant Navigator in position 4, CC = Contact Coordinator in position 5, and, RAD = Radar in position 6.

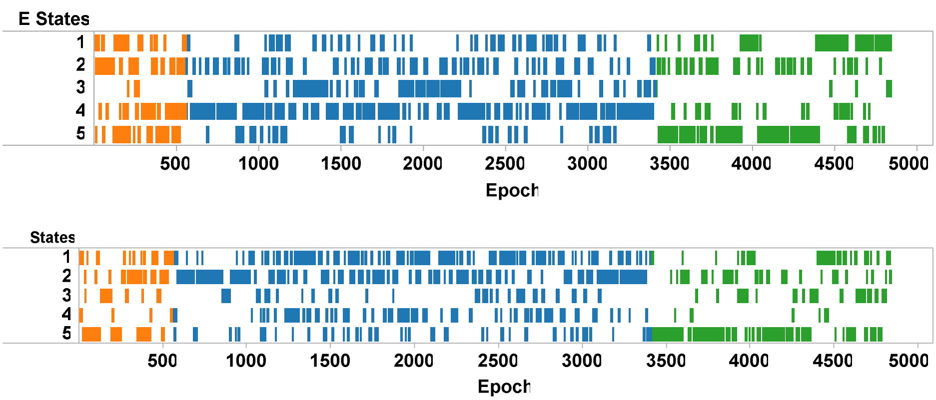


Fig. 2. Comparison of NS_E expression when modeled with autologous (top) or heterologous (bottom) ANN and HMM models. The data is shown for a Junior Officer navigation team at the early part of their SPAN training. The dark portion in the middle is the Scenario segment and the lighter portions to the left and right are the Brief and Debrief segments respectively.

Figure 2 compares the NS_E States following autologous and heterologous modeling of the same SPAN performance. Both models showed the NS_E state transitions at the Scenario / Debrief junction (epoch 3390) and at epoch 4400 of the Debrief. They also both showed a long period at the beginning of the scenario (epochs 590 – 1000) where a single state predominated and a period (3100 – 3385) at the end of the scenario where the same state predominated. These task-junction transitions have been observed in ten different SPAN sessions where autologous and heterologous modeling was conducted in parallel.

Another validation approach was to compare the Shannon entropy of the NS Pattern data streams obtained from each model. This metric is derived from information science and measures the level of uncertainty in a data stream [9]. The top of Figure 3 is a scatter plot of the levels of Shannon entropy for the NS_E values obtained from autologous and heterologous NS_E models. The histograms below the scatter plot highlight the peaks and valleys in entropy and show a strong concordance across both NS_E data streams. Combined, these data indicate that the heterologous NS_E models provide a close approximation of those obtained with autologous modeling.

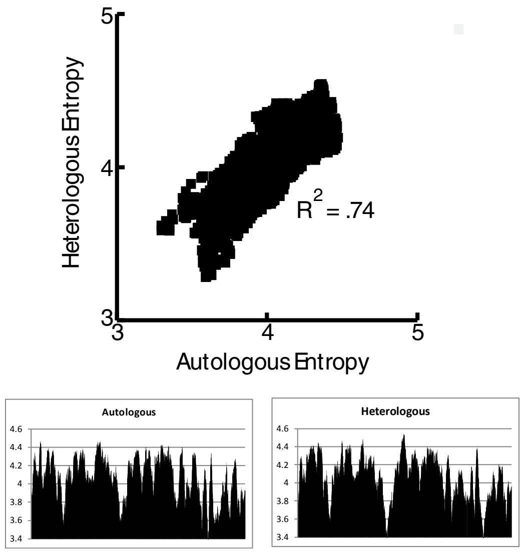


Fig. 3. Comparison of the Shannon entropy of NS_E Pattern expression from autologous (lower left) or heterologous (lower right) models. The top figure shows a scatterplot of the entropy from the two data streams.

3.2 Mapping NS_E Patterns to NS_E States for the Heterologous Models

In addition to documenting changes in NS expression in response to the changing task it is important to relate these changes to the cognitive measure itself. One result of the hidden Markov modeling is an emission matrix that maps different NS Patterns to NS States. This mapping provides a cognitive context across team members for the state changes associated with task events. These relationships are summarized in Figure 4 for NS_E Patterns and States. Each of the five (numbered 1-5) NS States is represented by a gray box containing the numbered NS Patterns most often associated with each State.

As expected from the complexity of the modeling, the associations of different NS Patterns with each NS State were not simple. In general however, NS E States 3 and 4 represent where many of the members of the team showed low EEG-E levels. The transition probabilities between these two states are very low (<0.03) suggesting that they are not subsets or close dynamic neighbors of each other. NS_E State 2 was the most frequently expressed NS E State and we refer to this as the normal operating mode (NOM) of the teams, as it is mainly expressed during the Scenario and less so during the Debrief. It is also a state where many of the team members expressed high

EEG-E. There are strong reciprocal transition probabilities between NS_E State 2 and NS_E State 1 (another state of high EEG-E expression), and lower probabilities with NS_E State 5.

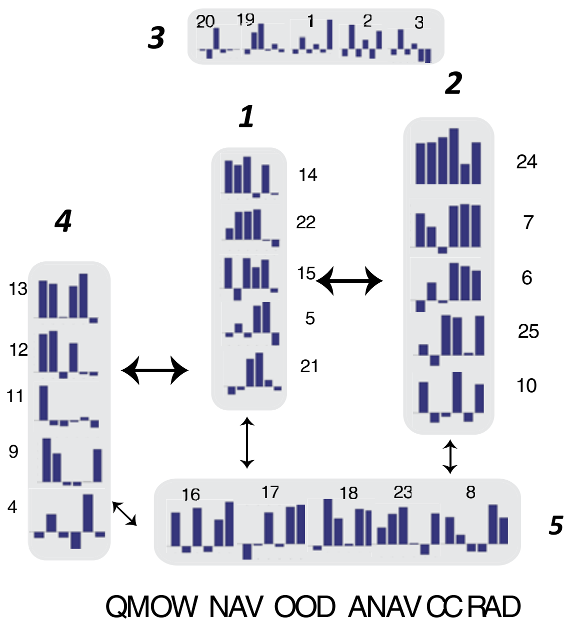


Fig. 4. Mapping of NS_E ANN patterns to HMM states. The team members associated with each bar in the histograms is shown below the figure. QMOW = Quartermaster on Watch, NAV = Navigator, OOD = Officer on Deck, ANAV = Assistant Navigator, CC = Contact Coordinator, RAD = Radar.

3.3 NS_E Expression across Teams and SPAN Sessions

One question that can be approached with the heterologous NS_E models is how consistently different NS_E States are used across teams and / or training sessions. Figure 5 shows the frequency distribution of NS_E for an expert (E2) and two Junior Officer teams (T4 and T5) that each performed two simulations and additional Junior Officer team that performed a single session (T1). The NS_E frequencies are separated into the Scenario, Debrief and Briefing segments of the simulation based on prior studies (such as Figure 2) that have shown there are often dynamic NS_E shifts at these segment junctions. For most teams the dominant NS_E States during the Scenario segment were 1 and 2. Referring to Figure 4, these states represent where most of the team is highly engaged. These appear to represent the normal operating mode for SPAN teams as their expression is diminished during the Debriefing segment and to a lesser extent in the Briefing segment. While there were slight differences in the NS_E State frequencies for E2, T4 and T5 the performance of team T1was different with NS_E State 4 dominating. Referring to Figure 4, this state is one where many of the team members' had low EEG-E.

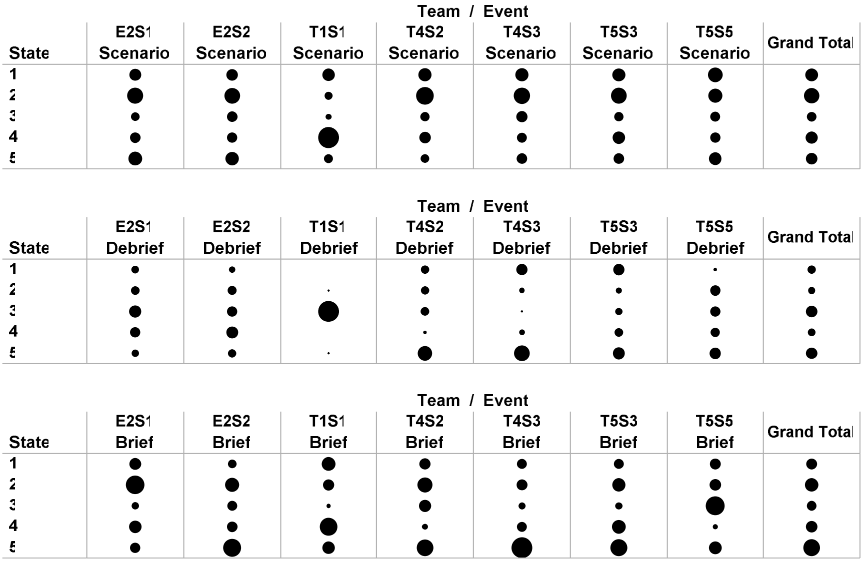


Fig. 5. Team NS_E state distributions across teams and sessions

The differences across teams were larger when comparing across the Debrief and Brief segments. Here there was proportionally higher expression of NS_E States 3 & 4 (teams with low EEG-E) for the expert team and NS_E State 5 for the Junior Officer teams.

3.4 Association of NS_E Expression with Speech

When team members interact the resulting communication stream contains information about knowledge, uncertainty, awareness of the situation, stress and other cognitive states. Their speech provides a detailed and dynamic representation of team cognition and is considered one of the gold standards for studying teamwork.

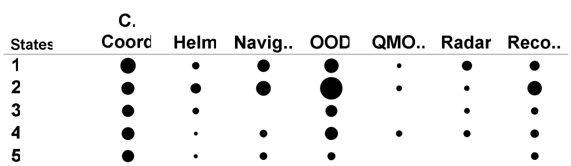


Fig. 6. NS_E expression while team members were speaking

Initial associations between team member's speech and NS_E expression is shown in Figure 6. For these studies we coded the speech of three teams on a second – by – second basis and pooled the data for cross tabulation analysis. The speech condition during the scenario resulted in higher than expected levels of NS_E State 3 while the non-speech condition had higher than expected levels of NS_E State 4. It is not clear if this is significant regarding teamwork as both are States where the overall team

EEG-E is low. The speech condition was eight times more common than the non-speech condition during the Debriefing segment and there were significantly fewer epochs where NS_E State 1 was expressed (data not shown).

Within the Scenario NS_E State expression was variable with NS_E States 1 & 2 preferentially being expressed when the Navigator (Navig), Officer of Deck (OOD) or Recorder (Rec) were speaking; while all five states were equally expressed when the Contact Coordinator (CC) was speaking.

4 Discussion

In our previous studies [3, 4, 10] NS expressions were derived from autologous datasets as we felt that such models may have the greatest sensitivity to small and large changes during the task. Such models were also necessary early in the project as there were few performances and datasets where there were team members in the same navigation positions. With more SPAN performances from Junior Officer and experienced teams we assembled a dataset of nearly 6 hrs. of teamwork and created standardized NS models.

Validation of the heterologous models was approached two ways; one using NS Patterns from ANN clustering of EEG-Engagement levels and one using NS States which provides a temporal component to the NS Patterns [10]. One of the most reproducible features of SPAN performances is the change in NS_E States at the junction between the Scenario and Debriefing. The heterologous and autologous models reproducibly detected these temporal features at this junction indicating an equivalent sensitivity of large task changes. A different form of validation drew on the concept of entropy from information theory which measures the degree of uncertainty in a data stream of symbols. In the present study we determined the entropy of both the autologous and heterologous NS_E data stream at 1 second intervals over a sliding window of the prior 90 seconds. These entropy profiles highlighted periods of high and low entropy modeled by both approaches. The strong concordance between the two models provides an additional validation of the sensitivity and specificity of the heterologous NS_E models.

It is currently difficult to say which model is the ‘right’ model. Heterologous datasets due to their larger size may not be sensitive to some combinations of EEG-E across team members due to their unique expression by a team. Similarly, autologous models may not have the repertoire of EEG-E combinations that would allow meaningful comparisons across teams. From a practical perspective both models seemed adequate for detecting shifts in NS_E expression in association with changes in the task and perturbations to the environment.

Prior to developing and validating the heterologous NS models only the first of the five usefulness criteria outlined in the introduction could be approached. As shown in this study, with the standardized models we can begin to compare NS expression across teams, training sessions and levels of expertise. Most recently these models have been incorporated into software systems that supply rapid (minutes) after training feedback to teams and provide a framework for future real-time adaptive monitoring and training.

Acknowledgements. This work was supported by The Defense Advanced Research Projects Agency under contract number(s) NBCHC070101, NBCHC090054. The views, opinions, and/or findings contained in this article/presentation are those of the authors and should not be interpreted as representing the official views or policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the Department of Defense. We wish to thank the sailors and staff at the Submarine Learning Center for their participation and support.

References

1. Warner, N., Letsky, M., Cowen, M.: Cognitive Model of Team Collaboration: Macro-Cognitive Focus. In: Proceedings of the 49th Human Factors and Ergonomics Society Annual Meeting, Orlando, FL, September 26-30 (2005)
2. Salas, E., Cook, N.J., Rosen, M.A.: On Teams, Teamwork, and Team Performance: Discoveries and Developments. Human Factors: The Journal of the Human Factors and Ergonomics Society 50(3), 540–547 (2008)
3. Stevens, R.H., Galloway, T., Berka, C., Sprang, M.: Can Neurophysiologic Synchronies Be Detected during Collaborative Teamwork? In: Proceedings: HCI International 2009, San Diego, CA, July 19-24, pp. 271–275 (2009)
4. Stevens, R.H., Galloway, T., Berka, C., Behneman, A.: A Neurophysiologic Approach for Studying Team Cognition. Interservice / Industry Training Simulation and Education Conference (I/ITSEC), Paper No. 10135 (2010)
5. Berka, C., Levendowski, D.J., Ramsey, C.K., Davis, G., Lumicao, M.N., Stanney, K., Reeves, L., Regli, S., Tremoulet, P.D., Stibler, K.: Evaluation of an EEG-Workload Model in an Aegis Simulation. In: Proceedings of the SPIE Defense and Security Symposium, Biomonitoring for Physiological and Cognitive Performance during Military Operations (2005)
6. Levendowski, D.J., Westbrook, P., Berka, C., et al.: Event-related potentials during a psychomotor vigilance task in sleep apnea patients and healthy subjects. Sleep 25, A462-A463 (2002)
7. Stevens, R.H., Galloway, T.L., Berka, C.: Allocation of time, EEG-Engagement and EEG-Workload as scientific problem solving skills are acquired. In: Proceeding Human Factors and Ergonomics Society, 51st Annual Meeting, Baltimore, MD, October 1-5 (2007)
8. Soller, A., Stevens, R.: Applications of Stochastic Analyses for Collaborative Learning and Cognitive Assessment. In: Hancock, G., Samuelson, K. (eds.) Advances in Latent Variable Mixture Models. Information Age Publishing (2007)
9. Shannon, C.E.: Prediction and entropy of printed English. The Bell System Technical Journal 30, 50–64 (1951)
10. Stevens, R.H., Galloway, T., Berka, C., Behneman, A.: Temporal sequences of neurophysiologic synchronies can identify changes in team cognition. In: Proceedings: Human Factors and Ergonomics Society 54th Annual Meeting, San Francisco, CA, September 27-October 1, pp. 190–194 (2010c)

Mapping Cognitive Attractors onto the Dynamic Landscapes of Teamwork

Ronald H. Stevens[1] and Jamie C. Gorman[2]

[1] UCLA IMMEX Project 5601 W. Slauson Ave. #272, Culver City, CA 90230, USA
[2] Psychology Department, Texas Tech University, Lubbock, TX 79409, USA
immex_ron@hotmail.com, jamie.gorman@ttu.edu

Abstract. The objective of this study was to apply ideas from complexity theory to derive new models of teamwork. The measures include EEG-derived measures of Engagement and Workload obtained from submarine piloting and navigation (SPAN) teams and communication streams from Uninhibited Air Vehicle Synthetic Task Environments (UAV-STE). We show that despite large differences in the data streams and modeling, similar changes are seen in the respective order parameters in response to task perturbations and the experience of the team. These changes may provide a pathway for future adaptive training systems as both order parameters could conceivably be modeled and reported in real time.

Keywords: Complexity, Teamwork, EEG, Neurophysiologic synchrony, Nonlinear dynamics.

1 Introduction

Teamwork is complicated, complex, and noisy. The ecological perspective of teamwork described by Cooke et al. [1] draws on this complexity to describe a dynamic view of the team, its' members, and the environment. Patterns of interaction and activity qualitatively emerge with the flow of the task, and perturbations to the teamwork and these patterns are characterized by fluctuations away from stable states. In this paper the concept of 'attractor landscapes' is proposed as a methodological approach to describe, explain, and visualize the dynamics of teamwork. In this approach individuals are not viewed as passive entities but rather as comprising a system capable of rich dynamics with the state of each member depending, in part, on the state of others. This synchronization of cognitive and communication components across the team provides a higher order system with its own dynamic properties as each individual attempts to achieve synchronization by adjusting his or her internal state or overt behavior in response to the evolving task and the state or behavior of the individuals with whom he or she is interacting.

How can we begin to model these adjustments and what can we learn that's new from these models? One approach is nonlinear dynamics which is a general theoretical approach for understanding complex systems and the linkages within and across subsystems independent of their specific behavioral or material substrate.

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 366–375, 2011.

When teamwork is viewed as a complex adaptive system there are multiple non-linear dynamic concepts that can be applied including self-organization, attractors, phase shifts, recurrence, entropy perturbations, and intrinsic dynamics.

This paper describes methodologies for modeling and visualizing teamwork in the context of cognitive attractors. The principles are organized around Haken'ssynergetics [2] and the developmental framework described and Smith [3]:

- *Define a level of analysis.* A recurring question for most teamwork research is determining the criteria for when, or if, a particular measure has been aggregated to an appropriate level and is being modeled at an appropriate temporal resolution. An appropriate team level of analysis may be influenced by variability in individual level properties from below and by organizational properties from above.
- *Identify patterns in behavior and define the order parameter.* There are many possible degrees of freedom of team member interaction. An order parameter—a collective variable—is a relatively low-dimensional variable that captures qualitative changes in teamwork. That is because the order parameter integrates team interactions that fluctuate critically at critical task points. Knowledge of the intrinsic constraints of team member interaction at critical task points can *a priori* define an order parameter, or, in application, we may distinguish patterns by introducing different manipulations (perturbations).
- *Describe the attractors of the system.* A dynamical system defined over teamwork is continuous, but it will have both repelling and attractive regions of its state space. In phase space the attractors define the qualitative changes in the stable teamwork patterns. The attractor landscape can be traversed by scaling a control parameter, which is an extrinsic parameter nonspecific to the order parameter, but which leads to qualitative change in the behavior of the order parameter.
- *Capitalize on dynamical similitude.* Dynamical models may or may not exist for a given order parameter and attractor landscape. If a model does not exist, then we may capitalize on dynamical similitude, which means simply that systems with different material substrates can share the same dynamics (e.g., teamwork and the dynamics of balance).
- *Identify phase transitions in teamwork:* Phase transitions are qualitative changes in dynamics due to changes in control parameters. Phase transitions occur when the underlying pattern of interaction shifts to another pattern under predictable conditions and may facilitate the identification of perturbations and antecedents to changes in the teams' dynamics.

2 Methods and Results

We highlight two approaches / tasks that draw on this framework and illustrate the advantages and challenges of applying a nonlinear dynamic approach to teamwork. The first approach models teamwork using neurophysiologic measures of the engagement of each person on submarine piloting and navigation (SPAN) teams. This is an empirical, exploratory study where the data stream is a set of non-numeric symbols called Neurophysiologic Synchronies (NS) that represent the relative levels of engagement of each person on the team [4-6].

For the second approach Gorman et al [7-8] developed a team coordination order parameter for three-person Uninhabited Air Vehicle (UAV) teams and in the resulting dynamical model the scaling of a control parameter, team-member familiarity, revealed qualitative changes in team coordination dynamics whereby team mixing resulted in a more stable attractor.

2.1 Submarine Piloting and Navigation (SPAN)

The goal of this project is to use neurophysiologic measures to rapidly determine the functional status of a team in order to assess the quality of a teams' performance / decisions and to adaptively rearrange the team or task components to better optimize the team. In this study the ideas of self-organization and attractor landscapes have been applied to derive new insights into the differences between novice and expert SPAN teams. The SPAN simulations contain dynamically programmed situations events that are crafted to serve as the foundation of the adaptive team training. Each SPAN session begins with a Brief presenting the goals of the mission. The more dynamic Scenario segment follows and contains easily identified processes of teamwork along with others which are less well defined. The Debrief follows and is highly structured with team members reporting on their overall performance. As shown previously [4, 6], there are often major shifts in NS expression at the junctions of these segments. The cognitive measure being studied is an EEG-derived measure of Engagement (EEG-E) defined by Advance Brain Monitoring's B-Alert® system [9, 10].The hypotheses of this study were:

- Multiple attractor basins (attractors) for engagement exist for SPAN teams; and,
- Some attractors are favored over others depending on the control patterns (i.e. task environment and team experience).

The B-Alert® system contains an easily-applied wireless EEG system that includes intelligent software that identifies and eliminates multiple sources of biological and environmental contamination and allows second – by -second classification of cognitive state changes such as Engagement. The EEG data streams for each person on the team are normalized and combined into vectors describing the EEG-E level of each person. They are used to train an unsupervised artificial neural network (ANN) that generates 25 NS clusters (symbols) representing the Engagement status of the team [4-6]. Each cluster displays a histogram showing the relative EEG-E level of each person. An example for a six person team is shown in Figure 1where persons 3 and 5 have high levels of a cognitive measure and the rest are low.

A topology also develops during this training where similar vectors cluster together and more disparate vectors are pushed away. For instance, NS_E Patterns 1-5 represent times where most of the team members had low levels of EEG-E while NS_E Pattern 24 represents times where most team members had high EEG-E. In a NS data stream the expression of these patterns represents second – by – second fluctuations of the engagement by different members of the team and provides an order parameter for these studies. The control parameter(s) for this study are team expertise and the task divisions (i.e. Brief, Scenario & Debrief).

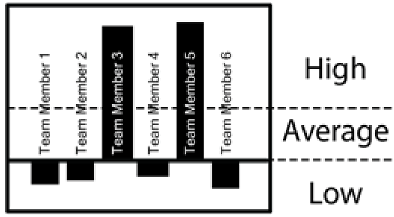


Fig. 1. Expression of a generic NS measure by members of a six-person team

The starting assumption was that many of the second-by-second changes in team NS_E would be small which would result in local transitions between NS_E Patterns. With the linear architecture of the self-organizing ANN this would appear in transition matrices as movement around / along a diagonal line. The thickness of the diagonal line confirms that there are transitions in local neighborhoods and these are more common than more distant transitions.

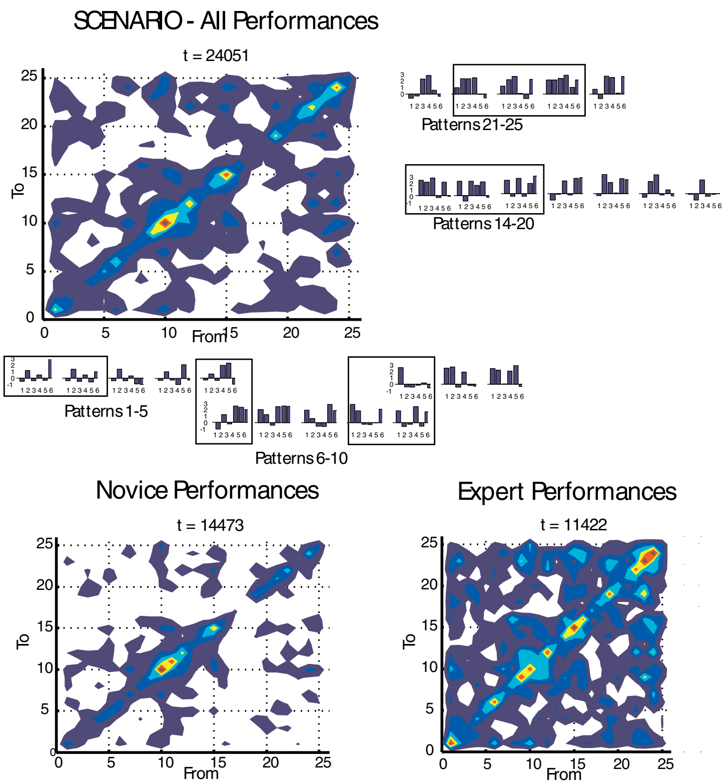


Fig. 2. NS_E transition matrices during the scenario segments of novices (left) and experts (right). The top figure shows the transition matrix of the combined dataset of five novice and five expert SPAN performances. Surrounding this figure are the 25 NS_E patterns resulting from ANN clustering. The boxes highlight patterns that are frequent in the matrix. The figures below show only the Scenario data for the novice (left) and expert (right) teams.

For novices, the areas with more frequent transitions were centered on NS_E Patterns10-11. Referring to Figure 2, these patterns were where many of the team members had low EEG-E. The most frequent patterns / transitions for experts clustered near NS_E 15 where most of the team showed above average EEG-E. A second cluster (attractor) centered near NS_E 22-25 where again the majority of the team showed high EEG-E. The expert teams also showed more minor transitions than novices as evidenced by the darker background contours throughout the matrix.

For both novice and expert groups the Debriefing segments showed transition matrices with restricted patterns of NS_E expression confined around the diagonal (Figure 3). The two groups however, showed reverse pattern of NS_E expression with those of the expert team members representing below levels of EEG-E, while those of the novices representing above levels of EEG-E for many team members.

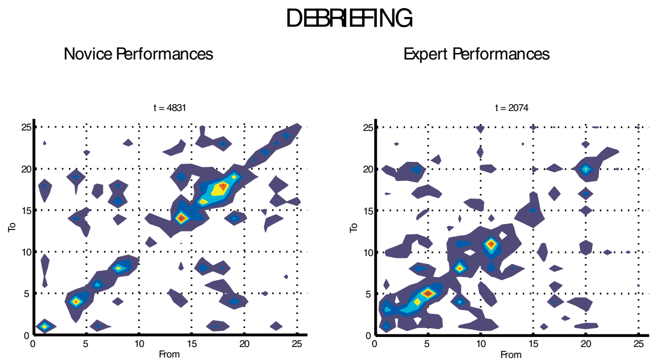


Fig. 3. NS_E transition matrices during the Debriefing of novices (left) and experts (right)

To capture the attractor dynamics of the NS_E Patterns transition matrix movies were created for each team where movie frames were updated every 8 seconds over a background history window of the prior 90 seconds. Two frames are shown in Figure 4 for the novice team T4S2. The left frame (epoch 1586) was where there was confusion in the team about contacting / avoiding another ship. Here the team oscillated between two attractors centered near NS_E 14-16 and NS_E 9-11. The right frame shows a period of diffuse attraction.

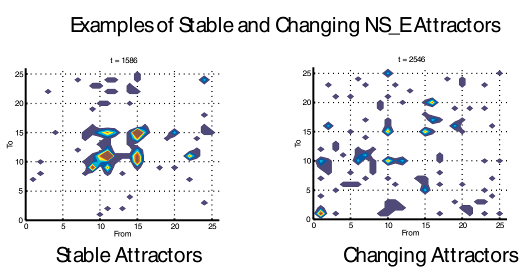


Fig. 4. Examples of stable (left) and changing (right) NS_E attractor states

2.2 UAV Team Coordination Dynamics

In the UAV task [11] three team members (a pilot, a navigator, and a photographer) interact over headsets to fly a simulated UAV over targets in order to take reconnaissance photos. This task is performed over a series of 40 min missions, each consisting of 11-12 targets. Each team members' computer work station displayed information specific to that team-member role as well as general flight information (e.g., current heading, altitude, and speed). Team members were seated in the same room with their backs to each other such that they only communicated verbally over the headsets. In this section we describe the methodological strategy for studying team coordination dynamics in this UAV task as well as results using that approach. We do so by summarizing the research originally reported in [7, 8], following the framework described in Section 1.

To identify an order parameter at the appropriate level of analysis (i.e., the team level), the functions that each team member performs just below that level of analysis were first identified. Recordings of previous UAV team communication revealed three primary functions: (1) the navigator sends target *information* (I) to the pilot; (2) the pilot and photographer *negotiate* (N) an appropriate airspeed and altitude; and (3) the photographer provides *feedback* (F) on the status of the target photograph. Team members dynamically combine these functions in a specific order via communication to photograph each target (I→N→F). To measure how this pattern of behavior changes over time, timestamps of these specific interactions at each target were collected. The coordination order parameter, called κ, is shown in Equation 1.

$$\kappa_t = \frac{\mathrm{time}(F_t) - \mathrm{time}(I_t)}{\mathrm{time}(F_t) - \mathrm{time}(N_t)} \qquad (t = 1, 2, \ldots \#\,\mathrm{targets}) \tag{1}$$

Because time cancels in the numerator and denominator of Equation 1, κ is a dimensionless (unit-free) measure of the relationship across the three primary functions at each target. Relative to the temporal arrangement of the components of coordination, $\kappa > 1$ is coordinated, $\kappa < 1$ is uncoordinated, and $\kappa = 1$ is indeterminate. Figure 5 shows κ trial series for (a) Intact and (b) Mixed (these terms are described below) team coordination dynamics. As shown in Figure 5, the κ order parameter fluctuates at the critical task points, which were defined as the targets themselves, as well as unexpected perturbations called roadblocks [7].

Although there has been over two and a half decades of research on coordination dynamics [12, 13, 14, 15], there was no previous research on *team* coordination dynamics. Therefore, dynamical similitude was employed to describe the attractors of UAV team coordination dynamics. If the globally stable attractor of team coordination was coordinated, then team members would never have to interact. To meet the demands of a continuously changing environment, however, team members have to interact. Like a balancing act, team coordination is continuous and effortful because it is the stabilization of an inherently unstable system [7]. Existing dynamical models of similar processes are found in postural dynamics (i.e., standing up straight; [16]) and manually balancing an inverted pendulum [17]. In those systems, the globally stable state is lying in a horizontal position on the ground; however, a metastable state emerges as the active components counter the forces that are pulling the upright human or pendulum toward the ground [7]. Based on dynamical similitude, then, the team coordination attractor may be summarized as follows: The globally stable state is uncoordinated. However, a metastable coordinated state emerges as team members interact to counter environmental forces (i.e., constantly varying task demands) that pull them toward the uncoordinated state.

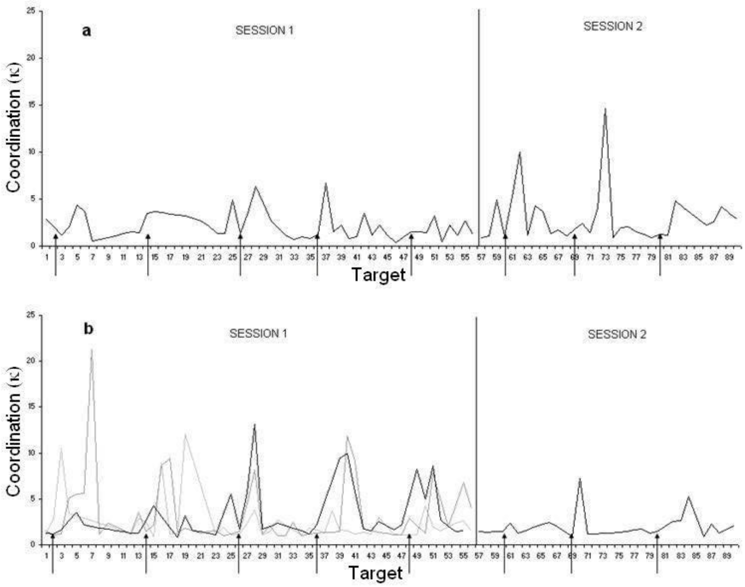


Fig. 5. Examples of κ fluctuations at critical points in the task: critical targets are indexed on the x-axis and roadblock perturbations are indexed by arrows on the x-axis (reprinted from [7])

The UAV team coordination dynamics were evaluated in an experiment in which team membership either stayed the same or changed following a retention interval [7]:After initially acquiring proficiency in the UAV task with one set of team members, participants would go away during a retention interval and, upon returning, either continue to work with the same members (Intact) or with completely different members (Mixed). (Participants maintained their same role on the team in both conditions.)The scaling of this team familiarity control parameter revealed qualitatively different team coordination dynamics.

To examine those differences in Intact vs. Mixed team coordination dynamics, attractor reconstruction [18] was performed on the κ trial series. Attractor reconstruction is a method for unfolding a scalar time series (e.g., Figure 5) into an appropriate dimension in which to view the dynamical system (i.e., the attractor) that produced the time series. As shown in Figure 6a, Intact teams' reconstructed attractor looks quite different than Mixed teams' reconstructed attractor (Figure 6b). In postural (or inverted pendulum) dynamics, short-term drift away from center is countered by long-range correction back to upright. Similar to those dynamics, Intact team coordination dynamics were centered on a small region of the reconstructed space (i.e., the phase space), near the origin (Figure 6). As with postural stabilization, explorations away from this small region of phase space were countered by long-range corrections back to this small, preferred region (i.e., the larger orbits moving away from, and then returning to, the origin in Figure 6a). On the other hand, the Mixed teams did not rigidly correct back to one small region of the reconstructed space: The Mixed team attractor consistently explored more of the phase space, and there was no correction back to a small, preferred region of phase space. Thus, the attractor landscape can be altered with the scaling of a team familiarity control parameter. Accordingly, the observed team coordination dynamics were not *encoded* by the level of familiarity; the scaling of the control parameter simply moved teams through the coordination attractor landscape.

Further analyses revealed that the stability of these attractors (the resistance to perturbation) was significantly correlated with successfully working through roadblock perturbations (denoted by the arrows on the abscissa in Figure 5; see [7] for details) such that higher stability was associated with adapting to unexpected changes in the task environment. Furthermore, the Mixed team attractor was significantly more stable than the Intact team attractor, suggesting that Mixed teams were more adaptive than Intact teams.

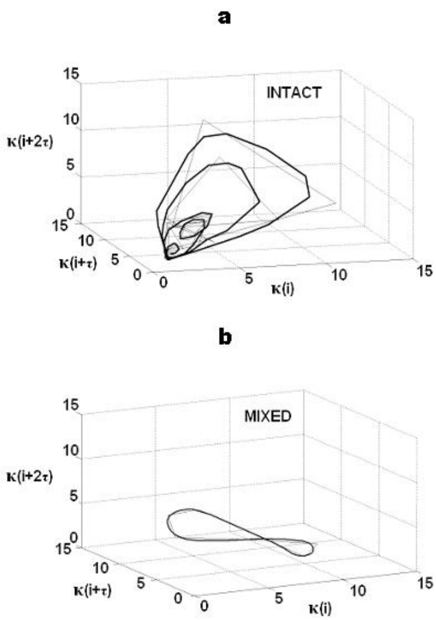


Fig. 6. Intact (a) and Mixed (b) reconstructed attractors (reprinted from [7])

In a following experiment, elements of the I→N→F coordination process of some teams were purposefully perturbed during task acquisition [8]. Though teams in that study were never mixed, those perturbations were intended to simulate the effects of increased interaction experience due to team mixing. As anticipated, this *perturbation training* led to more adaptive teams: Perturbation-trained teams exhibited significantly higher performance under novel task conditions than teams trained using standard methods; namely, cross-training and procedural training [8]. In this way, the

Intact vs. Mixed nonlinear dynamics results led to the development of a new team training method for team adaptation, and those predictions of enhanced team adaptation were borne out in the follow-up team training study.

3 Discussion

A goal of most training activities in complex environments is to be able to rapidly determine the functional status of a team in order to assess the quality of a teams' performance / decisions, and to adaptively rearrange the team or task components to better optimize the team. One of the challenges in accomplishing this goal is the development of rapid, relevant and reliable models for providing this information to the trainers and trainees.

While the two examples in this paper share task similarities, both being realistic and complex teamwork situations, the data streams and modeling approaches are very different. Despite these differences, the results show parallels in response to similar control parameters (Table 1).

Table 1. Model properties of the SPAN and UAV systems

	SPAN	UAV
Data Stream	EEG	Communication
Models	Symbolic	Numeric
Order Parameter	Neurophysiologic Synchronies (NS)	Coordination (κ)
Control Parameter(s)		
a) expertise	Novice / Expert Differences	Learning & Retention Differences
b) perturbations	NS State Changes	κ Fluctuations
Adaptive Models	Near Real-time	Near Real-time

The differences in the cognitive attractors seen in both of the systems with the changing experience of the team could provide a metric for following the efficacy of team training over time. The fluctuations of the order parameter within a task may also provide a pathway for future adaptive training systems as both could conceivably be modeled and reported in real time.

Acknowledgements. This work was supported by The Defense Advanced Research Projects Agency under contract number(s) NBCHC070101, NBCHC090054. The UAV research was funded by grants to Dr. Nancy J. Cooke from the Air Force Office of Scientific Research (FA9550-04-1-0234) and the Air Force Research Laboratory (FA8650-04-6442). The views, opinions, and/or findings contained in this article/presentation are those of the authors and should not be interpreted as representing the official views or policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the Department of Defense.

References

1. Cooke, N., Gorman, J., Rowe, L.J.: An ecological perspective on team cognition. In: Salas, E., Goodwin, J., Burke, C.S. (eds.) Team Effectiveness in Complex Organizations: Cross-disciplinary Perspectives and Approaches. SIOP Organizational Frontiers Series, pp. 157–182. Taylor & Francis, Abington (2009)
2. Haken, H.: Synergetics: An introduction. Springer, Berlin (1977)
3. Thelen, E., Smith, L.B.: A dynamic systems approach to the development of cognition and action. MIT Press, Cambridge (1994)
4. Stevens, R., Galloway, T., Berka, C., Behenman, A.: Identification and application of neurophysiologic synchronies for studying team behavior. In: Proceedings of the 19th Conference on Behavior Representation in Modeling and Simulation, pp. 21–28 (2010a)
5. Stevens, R.H., Galloway, T., Berka, C., Behneman, A.: A neurophysiologic approach for studying team cognition. Interservice / Industry Training Simulation and Education Conference (I/ITSEC), Paper No. 10135 (2010b)
6. Stevens, R.H., Galloway, T., Berka, C., Behneman, A.: Temporal sequences of neurophysiologic synchronies can identify changes in team cognition. In: Proceedings: Human Factors and Ergonomics Society 54th Annual Meeting, San Francisco, CA, September 27-October 1, pp. 190–194 (2010c)
7. Gorman, J.C., Amazeen, P.G., Cooke, N.J.: Team coordination dynamics. Nonlinear Dynamics, Psychology, and Life Sciences 14, 265–289 (2010)
8. Gorman, J.C., Cooke, N.J., Amazeen, P.G.: Training adaptive teams. Human Factors 52, 295–307 (2010)
9. Berka, C., Levendowski, D.J., Cvetinovic, M.M., Petrovic, M.M., Davis, G., et al.: Real-time analysis of EEG indexes of alertness, cognition, and memory acquired with a wireless EEG headset. International Journal of Human-Computer Interaction 17(2), 151–170 (2004)
10. Levendowski, D.J., Berka, C., Olmstead, R.E., Konstantinovic, Z.R., Davis, G., Lumicao, M.N., Westbrook, P.: Electroencephalographic indices predict future vulnerability to fatigue induced by sleep deprivation. Sleep 24(abstract supplement), A243–A244 (2001)
11. Cooke, N.J., Shope, S.M.: Synthetic Task Environments for Teams: CERTT's UAV-STE Handbook on Human Factors and Ergonomics Methods, pp. 46-1-46-6. CLC Press, LLC, Boca Raton, FL (2005)
12. Amazeen, P.G., Amazeen, E.L., Turvey, M.T.: Breaking the reflectional symmetry of interlimb coordination dynamics. Journal of Motor Behavior 30, 199–216 (1998)
13. Fuchs, A., Jirsa, V.K. (eds.): Coordination: Neural, behavioral and social dynamics. Springer, Heidelberg (2008)
14. Haken, H., Kelso, J.A.S., Bunz, H.: A theoretical model of phase transitions in human hand movements. Biological Cybernetics 51, 347–356 (1985)
15. Kelso, J.A.S.: Phase transitions and critical behavior in human bimanual coordination. American Journal of Physiology: Regulatory, Integrative and Comparative 15, R1000–R1004 (1984)
16. Collins, J.J., De Luca, C.J.: Open-loop and closed-loop control of posture: A random-walk analysis of center-of-pressure trajectories. Experimental Brain Research 95, 308–318 (1993)
17. Treffner, P.J., Kelso, J.A.S.: Dynamic encounters: Long memory during functional stabilization. Ecological Psychology 11, 103–137 (1999)
18. Abarbanel, H.D.I.: Analysis of observed chaotic data. Springer, New York (1996)

Behavioral and Brain Dynamics of Team Coordination Part II: Neurobehavioral Performance

E. Tognoli[1], A.J. Kovacs[1], B. Suutari[1], D. Afergan[2,3], J. Coyne[2], G. Gibson[2], R. Stripling[4], and J.A.S. Kelso[1,5]

[1] Center for Complex Systems and Brain Sciences, Florida Atlantic University, Boca Raton, FL
[2] Naval Research Laboratory, Washington, DC
[3] Strategic Analysis Inc., Arlington, VA
[4] Office of Naval Research, Arlington, VA
[5] Intelligent Systems Research Center, University of Ulster, Derry, N. Ireland
{tognoli,kovacs,suutari,kelso}@ccs.fau.edu, dafergan@sainc.com, {coyne,gibson}@itd.nrl.navy.mil, roy.stripling@navy.mil

Abstract. In this study, pairs of subjects performed a team-intensive task with the shared goal of clearing a virtual room from threats. The neurobehavioral dynamics of both subjects was analyzed to identify signatures of efficient team work. An ecologically valid task of room clearing was designed and a novel analysis framework was developed to address the challenge of understanding complex, continuous social processes at both behavioral and brain levels. A companion paper detailed the design of the neurobehavioral task and its associated dynamical analysis framework. In this paper, we present candidate neuromarkers for efficient room clearing and discuss key theoretical issues relating to successful team coordination.

Keywords: Neuromarkers, EEG, neurobehavioral dynamics, social behavior, complexity.

1 Introduction

Team coordination is critical for general human performance, and all the more when members' survival and safety depend upon efficient cooperation, such as when Marines neutralize dangers in a confined urban environment. During such tasks, a host of behavioral, cognitive and social processes have to be coordinated in space and in time in a context-dependent fashion. The goal of this study was to quantify the dynamics of neurobehavioral processes unfolding during a room clearing task that imposes high demand on both individual and team coordination. In a companion paper [1], we have presented a behavioral room clearing task. As a reminder, during this task, pairs of subjects navigate a series of rooms in a virtual environment while executing formal behavioral and mental operations that take place during real room clearing performance: stacking, tap signaling, entry, orientation to corner of domination, pieing, friend/enemy detection, shoot-no shoot decision making and restacking. The task requires the recruitment of sociocognitive, perceptual and

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 376–382, 2011.

attentional processes that must be coordinated within and across the brains of both team members. The goal of this study was to decipher the neurobehavioral organization of these processes from continuous dual-EEG recordings. For details on the task and analysis, please refer to [1]. In the following, we present neurobehavioral results obtained from a group of 18 subjects (9 pairs). Further, from the theory of Coordination Dynamics [2], we discuss key theoretical issues that shed light on successful team coordination.

2 Neurobehavioral Performance

Examples of behavioral trajectories for complete building clearing are illustrated in Figure 1. A spatial map is shown in (A) and corresponding temporal courses (avatars' rate of movement) in (B). Subjects complied sufficiently well with training instructions to yield efficient behavioral performance for most pairs. They generally performed well-phased series of motion and gaze behavior (Figure 1B).

For each room, motion was typically divided into two cleanly separated phases (entry-to-corner and restacking). Occasionally, team members lost coordination (not shown). In this case, it was generally observed that coordination was not reinstated before the onset of the subsequent building, suggesting that in our novice teams, behavioral flows are more strongly determined by the time scale of the building, than that of the rooms. To remediate such detrimental performance, resetting would be required at shorter time scales: during restacking (time scale of room) and immediately after the onset of disorganized behavior (time scale of instantaneous behavior).

Analysis of movement coordination also suggested that novice subjects were not necessarily weakly coupled: a large number of low-performance trials (variable "time to clear room") showed that subjects' movements were strongly determined by their partners': for instance, one avatar initiating a movement only when the partner initiates (in-phase coordination) or terminates his movements (anti-phase coordination). In contrast, in time-effective trials, subjects were able to phase movements more flexibility, exhibiting a coordination pattern with tendencies for the avatars to come together at the same time as to maintain a degree of individual autonomy. Coordination dynamics dubs this complementary tendency metastability [2-4]. Metastable forms of coordination have been shown to enhance complexity [5,6]. Recent work in our laboratory suggests that several neural mechanisms exist to dissolve coordinated behavior. One is increased activity in phi1, the first component of the so-called 'phi complex'--a neuromarker of social coordination and lack thereof

[7]. Another is an instantaneous neuromarker of behavioral segregation which is observed in steady-state social coordination at the transition from coordinated to uncoordinated behavior [8]. Interestingly, at the transition to independent behavior, the brains of interacting subjects are strongly coupled by a mechanism called "synchronized brain transitions" [8-9]. In fact, between-brain coupling is as strong at the transition to independent behavior as it is at the onset of coordinated behavior. This suggests that efficient independent behavior is achieved by virtue of exchange of information and coupling of the subjects' brains, in agreement with the theory of metastability [2-4].

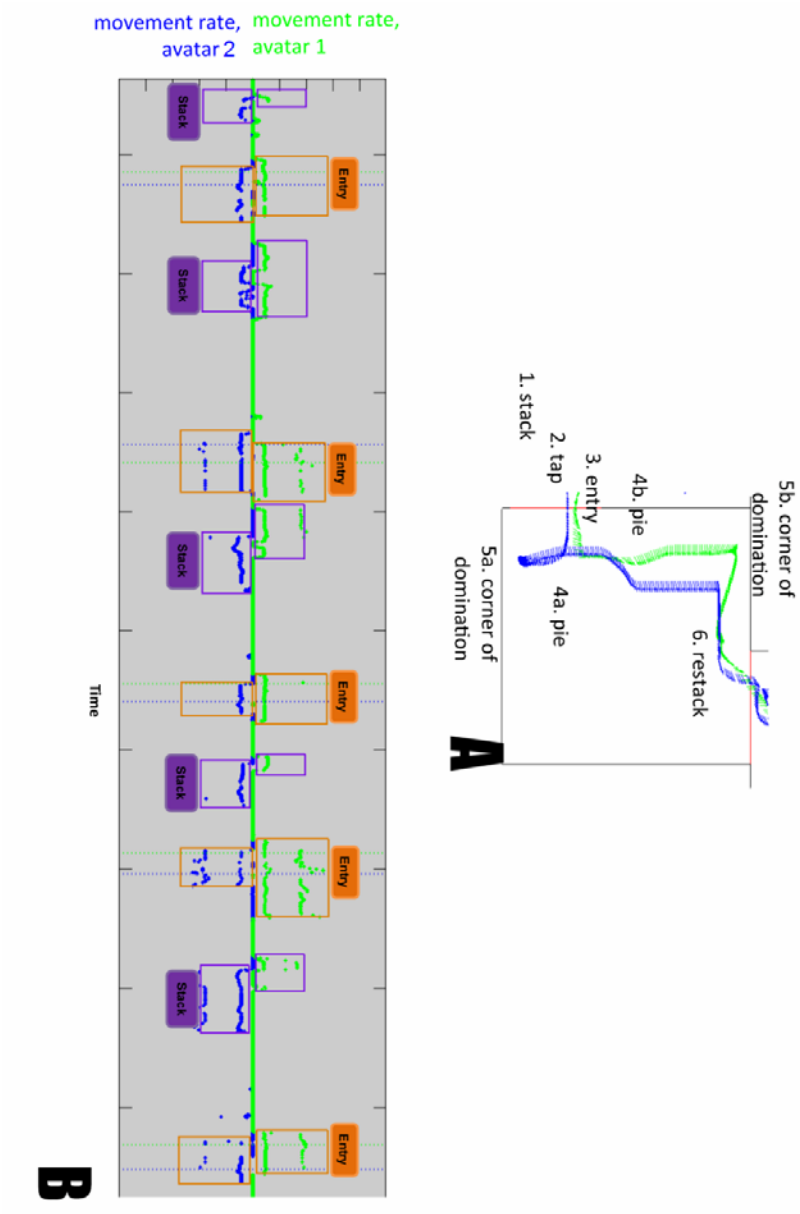


Fig. 1. Spatial view of the green and blue avatar paths in a room with the behavioral markers of interest annotated (A). Corresponding temporal view of the rate of change of movement for the entire building (B) for well-executed room clearing.

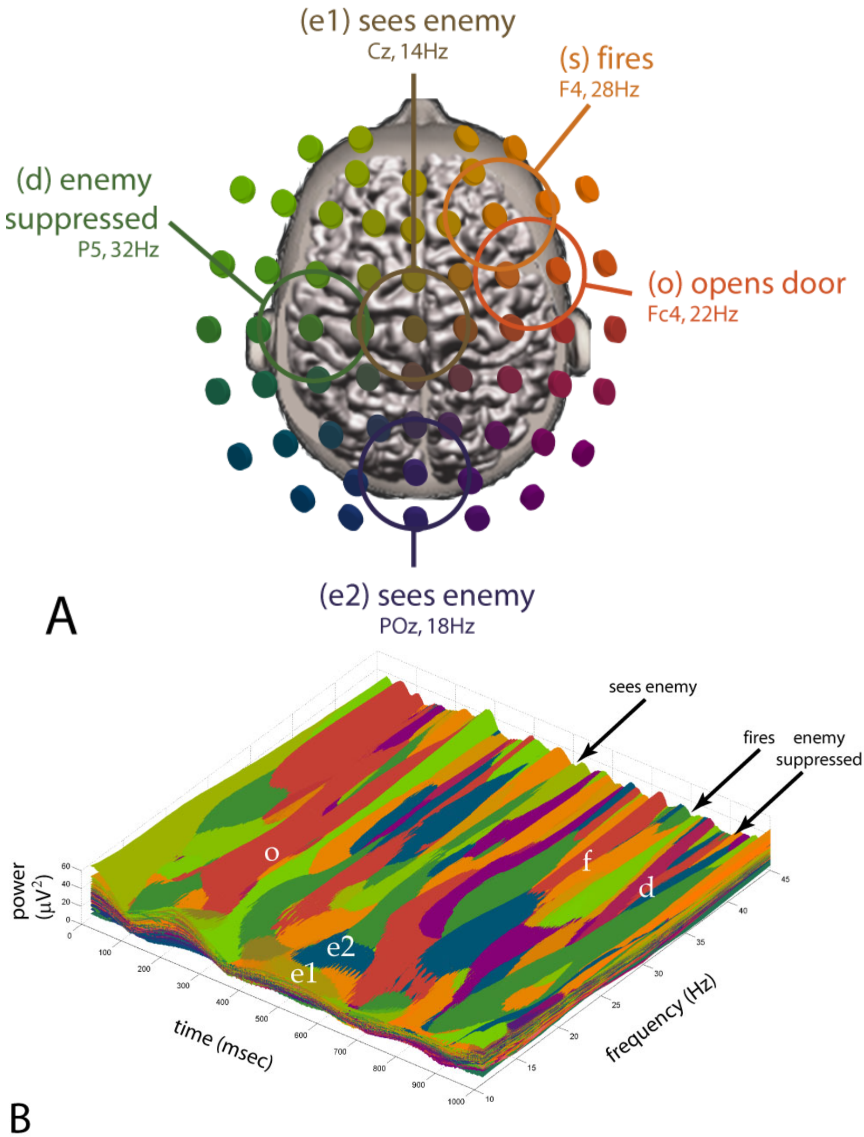


Fig. 2. Candidate neuromarkers for a variety of behaviors observed after entry in a hostile room, shown with their typical topography and frequency (A) along with their time distribution as observed from time-frequency-power plots (B). The colors in (B) reflect the spatial distribution of power as illustrated in (A). In this single trial (B), 1000 ms of EEG data are displayed, with the agent uncovering the enemy avatar and firing in less than 300 msec.

In previous studies, we identified several neuromarkers of social [7,10] and attentional behavior [7,11] in the 10Hz EEG frequency band. In the virtual environment created for the present study, performance of room clearing ("time to clear room") was bounded by experimental parameters (e.g. maximal avatar velocity, room dimension).

Execution of all behavioral, cognitive and social processes that led to task completion (see section 2.2 from companion paper [1]) was typically performed in 2 or 3 seconds, a time scale too short for EEG's slow frequencies alone to be useful information carriers. As a consequence, we extended our analysis to the beta and gamma frequency ranges (14-30Hz and 30-60Hz respectively). Analysis was conducted at three time scales: at the level of instantaneous behavior using either spatio-spectral or spatio-temporal techniques (neuromarkers type I); for durations of time that match the execution of goal-driven behaviors (e.g. reaching a corner of domination: neuromarkers type II); for periods extending over completion of the task (neuromarkers type III). Analysis suggests candidate neuromarkers for a number of individual and social behaviors (see e.g., type I neuromarkers, Figure 2A). These neuromarkers can readily be seen in a single trial from the spatio-temporo-spectral graphs (Figure 2B). Because of the covariation of many behavioral, cognitive and social processes during task performance, further work is ongoing to assess the specificity and reliability of candidate neuromarkers. This further work is guided by the relevant time scales of behavioral processes, and their correspondence with the duration of neural processes (Figure 2B).

Neuromarkers of type II-III emphasize the importance of time scales for team coordination. For instance, we identified a beta-gamma complex over the left rolandic region appearing while leaders moved toward the corners of domination. These neural complexes appeared only in well-coordinated trials (best half of 'coordinated entry' variable, see section 2.6 from companion paper [1]), and we did not detect any particular behavior that was immediately served by such neural patterns.

We hypothesize that the beta-gamma complex that occurs during the 5^{th} behavioral phase of room clearing (see figure 1A) is a remnant of earlier behavior during the 3^{rd} (entry) phase. Such diachronic phenomena that emerge across multiple frequency bands in the brains of team members illustrate the multiple time scales that have to be considered in explaining realistic social behavior (Figure 2B). Even though preliminary, the present data argue against two time-worn features of classical neuroscience paradigms: impoverished contexts for task performance ("all other things being equal") and short, iterative, experimenter-controlled stimulus-response manipulations. In contrast, the ecologically valid task that we have developed retains the essence of the real situation and is already capturing meaningful neurobehavioral signatures of brain self-organization during team work.

3 Concluding Remarks

The notion that individuals often cooperate in a reciprocal and synergistic way to achieve a common goal -joint action as it is sometimes called- is currently a topic of much interest in social neuroscience. The idea can be traced back to efforts aimed at identifying functional synergies in complex human movements that require the coordination of many degrees of freedom [12]. Here we have presented preliminary findings in an ecologically valid task aimed at understanding the coordinated action of two team members committed to accomplishing a common goal. To accommodate for emergent complexity, both in terms of the many processes to be coordinated and the multiple frequencies on which they occur, both neurally and behaviorally, we developed a novel framework that rests on the combined analysis of, and intimate relation between, behavioral dynamics and real-time, continuous EEG.

Crucially, such a framework is necessary to track the emergence of neurobehavioral processes on multiple space and time scales. This, in turn, allows team work to persist both during behavioral phases of strong informational exchange (such as entry), and during independent phases (such as reaching the corner of domination). In the absence of direct informational exchange, social coordination can be accomplished by virtue of memory processes. In his classic work "The Strategy of Conflict" [13], the Nobel Laureate Thomas Schelling describes experiments in which people coordinate their future behavior by virtue of shared knowledge, even though there is no interaction whatsoever between the individuals. We hypothesize that such phenomena are important components in the training of coordinated teams. We have previously shown that individual behavior can be modified following mutual information exchange even though the source of coupling is no longer available: a persistence of 'social memory' in individual agents [14]. We have also shown that sensorimotor (mu) rhythms of the human brain are modulated by memory requirements imposed on leaders and followers in social imitation tasks [10]. The present results on team coordination reveal diachronic neurobehavioral processes during coordinated room clearing (see section 2). The outlook is exciting for determining how such neuromarkers can be used to predict team performance in real time and thereby improve training.

Acknowledgments. The technical support of William McLean is acknowledged. This work is supported by the US Office of Naval Research Contract N000140510117. JASK and ET are also supported by NIMH Grant MH080838, NSF Grant BCS0826897and the Davimos Family Endowment for Excellence in Science.

References

1. Tognoli, E., Kovacs, A.J., Suutari, B., Afergan, D., Coyne, J., Gibson, G., Stripling, R., Kelso, J.A.S.: Behavioral and brain dynamics of team coordination part I: task design (this issue)
2. Kelso, J.A.S.: Dynamic patterns: the self-organization of brain and behavior. The MIT Press, Cambridge (1995)
3. Tognoli, E., Kelso, J.A.S.: Brain coordination dynamics: True and false faces of phase synchrony and metastability. Progress in Neurobiology 87, 31–40 (2009)
4. Kelso, J.A.S., Tognoli, E.: Toward a complementary neuroscience: metastable coordination dynamics of the brain. In: Kozma, R., Perlovsky, L. (eds.) Neurodynamics of Cognition and Consciousness, pp. 39–59. Springer, Berlin (2007)
5. Friston, K.J.: Transients, metastability, and neuronal dynamics. Neuroimage 5, 164–171 (1997)
6. Sporns, O.: Complex neural dynamics. In: Jirsa, V.K., Kelso, J.A.S. (eds.) Coordination Dynamics: Issues and Trends, pp. 197–215. Springer, Berlin (2004)
7. Tognoli, E., Lagarde, J., DeGuzman, G.C., Kelso, J.A.S.: The phi complex as a neuromarker of human social coordination. Proc. Natl. Acad. Sci. USA 104, 8190–8195 (2007)

8. Benites, D., Tognoli, E., DeGuzman, G.C., Kelso, J.A.S.: Brain coordination dynamics: Continuous EEG tracking of the neural functional organization in a social task. Psychophysiology 47, S75–S75 (2010)
9. Tognoli, E., DeGuzman, G.C., Kelso, J.A.S.: Interacting humans and the dynamics of their social brains. In: Wang, R., Gu, F. (eds.) Advances in Cognitive Neurodynamics (II), pp. 139–143. Springer, Heidelberg (2010)
10. Suutari, B., Weisberg, S., Tognoli, E., Kelso, J.A.S.: Neuromarkers of individual and social Behavior (submitted)
11. Calderon, R.: Brain computer interface and neuroprosthetics (Master's Thesis), Florida Atlantic University (2008)
12. Kelso, J.A.S.: The nature of human interlimb coordination. Science 203, 1029–1031 (1979)
13. Schelling, T.: Strategy of conflict, 2nd edn. Harvard University, London (1980)
14. Oullier, O., De Guzman, G.C., Jantzen, K.J., Lagarde, J., Kelso, J.A.S.: Social coordination dynamics: Measuring human bonding. Social Neuroscience 3, 178–192 (2008)

Feature Selection in Crowd Creativity

Lixiu Yu and Yasuaki Sakamoto

Howe School of Technology Management, Stevens Institute of Technology, USA
{lyu3,ysakamot}@stevens.edu

Abstract. Crowdsourcing is emerging as a wellspring of creative designs. This paper examines the mechanisms that support collective design. A sequential combination system is described: one crowd generates designs, and another crowd combines these designs. Previous experiments showed that the combined designs were judged more creative than the initial designs. The current work extends this previous research by examining the combination process of the designs more closely, looking at how features of the designs were selected and integrated into later designs. Participants preferred atypical features to typical ones for integration, and given a choice, selected practical but less atypical features over impractical but more atypical features. We conclude that crowds attend to both novelty and practicality of the features, and that the presence of atypical yet practical features contributes to the increased creativity of the combined designs.

Keywords: Crowdsourcing, collective creativity, combination, feature selection.

1 Introduction

Creativity has been studied in a variety of domains from two perspectives. First, who can perform creative activities? Second, what is the underlying process of creative thought? Traditionally, creativity has been regarded as unexplainable, an ability possessed only by a small number of talented people (e.g., [1]). In recent years, however, web-enabled technology has allowed non-experts to contribute to collective creative activities [2-4]. Specifically, crowdsourcing makes it possible to generate creative ideas by assembling large numbers of workers in a short amount of time [5, 6]. Our work has been examining how such crowds can be assembled and organized to generate creative ideas [7, 8]. In particular, we have developed and tested a sequential combination system: A crowd generates designs, and another crowd combines these designs. This past work has shown that combined ideas become more creative than the initial ideas, enabling crowds to improve the designs. The current work extends our past work by systematically studying the combination process: how feature selection in the combination process contributes to creative designs. By better understanding the combination process, we seek to improve the design of our system.

Our focus on the combination process is based on the popular conjecture that combination is a key process underlying creativity. For example, a Darwinian perspective considers combination as part of creativity by through an analogy to natural selection; ideas are combined at random, and some of them lead to discoveries

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 383–392, 2011.

(e.g., [9]). Following the spirit of the Darwinian view, management researchers explain innovation in terms of recombination of existing knowledge [10, 11]. Similarly, cognitive scientists claim that the process of conceptual combinations, in which separate ideas or concepts are merged, underlies much of creativity [e.g., 12-14]. Past experiments have examined how the nature of idea pairs, such as the similarity of ideas to be combined, influences creativity [15-18], and how the superimposition of two images affects the creation of art [19]. Nevertheless, little is known about which features people will choose to integrate, and how this selection contributes to the creativity of the design.

To address this issue, we conduct two experiments that examine which features people integrate. We make predictions based on schema research in cognitive psychology (e.g., [20]). By performing these experiments, and by better understanding the combination process, we extend the combinatorial conjecture of creativity. Furthermore, our results will help us fine-tune the design of our combination system. Continued work along this line will augment collective creativity, and may one day enable crowds to solve many challenging design problems.

In the remainder of the paper, we first introduce the sequential combination system and describe previous work on the combination process of crowd creativity. Then we present two experiments.

2 Background

The current work is built on Yu and Nickerson's past work, which showed that atypical features from the initial ideas are preserved in the combined ideas [7]. Before detailing this result, we describe the sequential combination system.

2.1 The Sequential Combination System

In our sequential combination system, crowds generate, evaluate, and combine visual designs in sequence, as summarized in Figure 1. The system uses the idea of a genetic algorithm [21, 22], implemented with humans [23]: The computers manage the process of finding pairs of designs to combine, based on crowds' evaluations of the designs, and the crowds perform the actual combining of the designs. In the design community, a variety of computer-based and human-assisted genetic algorithms have been tried, with some success [23]. Recently, other researchers have successfully conducted large-scale experiments using crowds, in which humans perform tasks as if they were computers [e.g. 4, 24, 25]. In our system, one crowd creates a first generation of designs, and then other crowds create successive generations by combining the designs made by previous crowds. In between, different crowds evaluate the designs from each generation. A pair of designs for combination is selected using tournament selection [26]: two ideas are randomly selected, and the one receiving a higher evaluation score by the crowd wins; this procedure is repeated to find another winner, and these two winners are chosen to form a pair of parents.

This organizational process is integrated with Amazon's Mechanical Turk [27, 28] and the Google Docs drawing application [29]. The solicitation and management of participants are handled through Mechanical Turk. All participants receive nominal compensation. When participants engage in design generation, they are directed to a Google drawing page. This same page, in later generations, presents the designs to be combined. The drawing tool provides many menu choices including a freehand sketch option, a vector line, text, and pull-down shape palette.

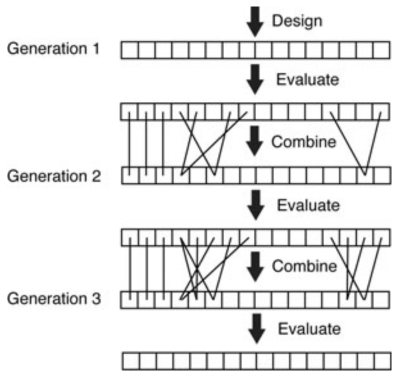


Fig. 1. The generations of the experiment

2.2 Test of the Sequential Combination System

The system was tested on two design problems. The first was an open-ended problem, designing a chair for children. Figure 2 shows the solutions to this problem and the combination process. The second problem, the design of alarm clocks, was more tightly specified: constraints of cost and safety were added. Figure 3 shows the solutions to this problem and the combination process.

In total, 1207 participants were involved in designing the chair, and 540 participants were involved in designing the clock [7]. Participants were asked to present their design ideas through sketches. One crowd solved the design problem by producing sketches independently, another crowd evaluated the sketches, and yet another crowd combined the design sketches generated by the previous crowd. Both of the design problems were run through three generations.

The creativity of the combined designs from Generation 3 was compared to the creativity of the initial designs from Generation 1, using the method explained in [7]. Generation 3 had about twice as many creative designs as Generation 1 for both problems. To find out why the combination process led to increased creativity, Yu and Nickerson took a closer look at the combination process in the chair experiment. What they found was that the number of atypical features increased almost three times in the later combined designs. From this observation, Yu and Nickerson concluded that people selected atypical features during combination, and that the selection of atypical features contributed to the increased creativity of the combined designs. Here we systematically test the idea that people select atypical features for combination. We draw on schema research from cognitive psychology to build our hypotheses.

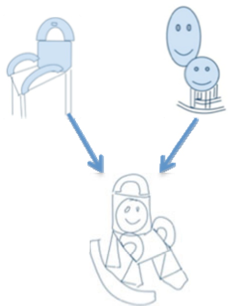
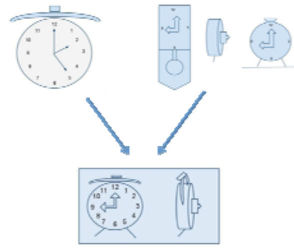
	

Fig. 2. Examples of chair designs. The chair on the bottom is combined from the chairs above.	**Fig. 3.** Examples of clock designs. The clock on the bottom is combined from the clocks above.

2.2 Schema Effects

A central issue in the schema research is how people process schema-consistent and schema-inconsistent information. A schema is a general knowledge structure that provides a set of expectations based on prior experience [30-32]. For example, a schema for a chair might include straight legs and a square back; that is, it might look like a dining room chair in your house. Schema-consistent information is information that fits one's existing schema; schema-inconsistent information is information that violates one's schema. For the chair schema, for example, a square back is schema-consistent; a flower back is schema-inconsistent.

A strong finding in the schema research is that people are attracted to schema-inconsistent information [33-35]. For example, if you see a chair with a flower back in your office, you will sit on it, play with it, or at least remember it. This effect is a special case of novelty effects, in which people attend to unfamiliar and surprising events and objects [36]. Based on this work, we hypothesize that people select atypical features to use in their designs.

However, the schema research predicts that not all atypical features will be entertained; people tend to ignore information irrelevant to the schema. Thus, people will stay away from atypical features, if they are not relevant to the schema. For designing chairs, features of the chair are irrelevant to the chair if they lack practicality; for example, a back of the chair that cannot support the person sitting on the chair is irrelevant to the chair in the sense that it cannot be used. Thus, we hypothesize that people will not select impractical features, even if they are atypical.

3 Method and Results

To explore the pattern of feature selection in the combination process, two experiments were conducted. In both experiments, participants were recruited from Mechanical Turk. In the preference experiment, participants were asked to judge which of the three chairs is most atypical (preference experiment A), and to select which of the three types of chair backs they would like to incorporate in their design (preference experiment B). In the production experiment, participants were asked to produce designs by combining two given designs.

3.1 Preference Experiment

In the preference experiment, three chairs were designed that differ only with respect to their backs: an often-seen chair with a typical square back, an atypical and practical flower back, and an atypical but impractical back composed of sharp triangles. Figure 4 shows the chairs used in the preference experiment. The position of the chairs was randomized in the experiments.

Altogether 80 participants took part in preference experiment A. Their ages ranged from 20 to 61, with a mean of 32. Sixty-three percent of participants were males. In preference experiment A, the crowd was asked:

> *Which chair is most atypical?*

There were also 80 participants in preference experiment B. Their ages ranged from 21 to 57, with a mean of 33. Sixty percent of participants were males. In preference experiment B, a different crowd was asked:

> *The chairs below are designs from your fellow workers. If you are designing a new chair, which back of the chairs would you like to incorporate in your design?*

Our two hypotheses were (1) the participants would select atypical features (2) but only when those features were practical. Thus, we predict that the atypical, practical back will be preferred to the typical back and preferred to the atypical, impractical back.

We first checked to make sure that chair backs differ in typicality. When asked to choose the most atypical chair in experiment A, 42 participants picked the atypical, impractical chair; 26 chose atypical, practical chair; and 12 selected typical chair, $\chi^2(2, N = 80) = 16.90$, $P < 0.01$. More people thought the atypical, impractical chair was more atypical compared to the atypical, practical chair. Then, if people select features based only on novelty, the atypical, impractical back should be chosen more often.

Our main interest was the crowd's responses in experiment B. Out of 80 participants, 53 participants picked the atypical, practical back, 10 picked the atypical, impractical back, and 17 picked the typical back, $\chi^2(2, N = 80) = 39.93$, $P < 0.01$. Figure 5 shows the proportions of participants for the three options. As predicted, the crowd selected atypical features that are also practical, even though more people thought the atypical, impractical chair was more atypical than the atypical, practical chair. Thus, the crowd values both novelty and practicality.

Fig. 4. Chairs presented in the preference experiment, left to right: the chair with the typical back, the chair with the atypical, practical back and the chair with the atypical, impractical back

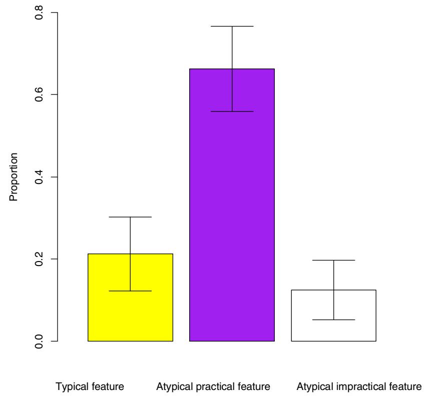


Fig. 5. The proportions of the selected features

3.2 Production Experiment

In the production experiment, we examine whether the atypical, practical features will be integrated into the combined designs when participants are asked to recreate their own designs. We are also interested in whether the combined designs with atypical, practical features are judged more creative than designs with only typical features.

Two chairs were designed. Each chair consisted of both atypical, practical features and typical features. The two atypical practical features were a flower back and rocking chair legs. The two typical features were a square back and straight legs with crossing bars. Figure 6 shows the chairs presented in the production experiment.

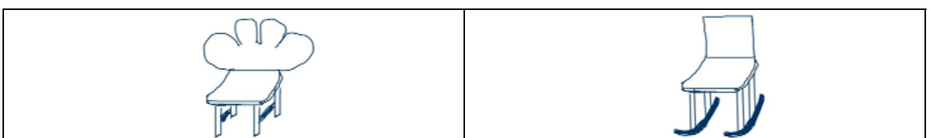


Fig. 6. The two chairs used in the production experiment as parents: on the left is parent chair 1, with a flower back and straight legs with crossing bars; on the right is parent chair 2 with a square back and rocking chair legs

Overall 94 participants took part in the experiment, and 94 sketches were collected. Among 94 participants, the ages ranged from 18 to 59 with a mean of 30. Fifty-two percent of the participants were males.

In the production experiment, two chairs were presented on the right side of the drawing interface. The participants were asked to design a chair for children by combining the presented chairs. Our prediction was that the atypical, practical features—the flower back and the rocking chair legs—would be integrated more often into the combined chairs than the typical features—the square back and the straight legs with crossing bars. We also predicted that the designs with atypical, practical features would be judged more creative than the designs with only typical features.

Figure 7 shows four examples of the combined chairs. Most combined chairs conformed to the features of the initial two chairs, but some incorporated new features that were not present in the initial chairs. The coding of typical features and atypical features was performed by two raters: they counted the number of square backs, straight legs with crossing bars, flower backs, and rocking chair legs in the combined chairs. Inter-rater reliability, measured by a Kappa score, was 0.71.

As predicted, atypical features were integrated more often in the combined designs than typical features, $\chi^2(1, N = 40) = 22.50$, $P < 0.01$. Figure 7a shows a combination of the atypical feature from each parent: the flower back from parent chair 1 and rocking chair legs from parent chair 2. Figure 7b shows an example of a combination of the typical features of the initial two chairs: the square back and the straight legs with crossing bars. There were two other types of combination: a new chair consisting of a mix of atypical and typical features as shown in Figure 7c, and a new chair with new features added as shown in Figure 7d.

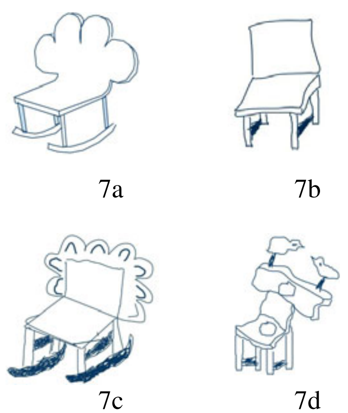


Fig. 7. Examples of combined chairs

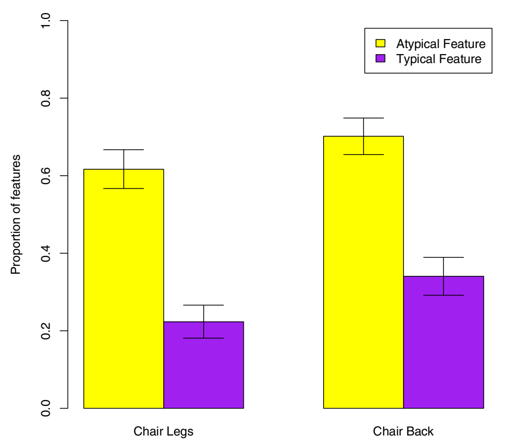


Fig. 8. Proportion of typical features and atypical features

In order to test the hypothesis that atypical, practical features are selected more often than typical features, the occurrences of typical features were compared to those of atypical, practical features. As shown in Figure 8, the atypical, practical features were selected much more often than the typical ones, $\chi^2(1, N = 94) = 23.21$, $P < 0.01$ for the chair back, and, $\chi^2(1, N = 94) = 28.30$, $P < 0.01$ for the chair legs.

We then examined whether combined chairs with only atypical features (atypical chairs) were judged more creative than the combined chairs with only typical features (typical chairs). Out of 94 combined chairs, 31 chairs were classified as atypical chairs and five chairs were composed of typical chairs. Two raters coded the classification, and inter-rater reliability, measured by a Kappa score, was 0.96.

A standard measure of creativity is a count of the number of original ideas that pass a threshold on the practicality dimension: Only ideas that are considered practical should be given credit for originality [7, 37]. A crowd of 400 participants was asked to evaluate the chairs on originality and practicality. We used the approximate mean value of the ratings across all designs, 4.0, as the threshold. The number of creative designs was four among atypical chairs and there were no creative designs in typical chairs. Although this result was consistent with our hypothesis, the count was too small to draw conclusions from.

In summary, the results from preference and production experiments show that the atypical features are more likely to be selected and integrated into the new combined designs, but only when they are also practical. The presence of atypical features might contribute to the creativity of the designs. In contrast, the typical features, which were judged low in originality, were dropped during the combination process. Therefore, the combination process plays a role in enhancing originality while maintaining practicality. The crowd's affinity for atypical, practical features is a source of creativity in the sequential combination system.

4 Discussion and Conclusions

Everyone can be creative [38], and crowdsourcing systems can facilitate an individuals' ability to generate creative solutions. The previous and present work investigated how such augmentation might happen. Previous work demonstrated that the crowd was able to generate creative designs through a combination process. The current work performed an analysis on the initial and combined designs, and investigated how features of designs propagated through the combination process. The main finding was that atypical, practical features were favored in the combination process. These features are both novel and useful, which should improve the creativity of the combined designs.

The contributions of the findings are two fold. Theoretically, they provide experimental evidence to the combinatorial conjecture of creativity: Creative designs emerge through the combination of two discrete designs. The special role of the atypical features in the combination process provides insight into the crowd's creative thought. Practically, the findings suggest a way of designing a crowdsourcing system that can perform creative tasks: the system can be designed to encourage the generation of more diverse ideas which in turn might increase the number of atypical ideas. It can also incorporate a process to favor the atypical and practical ideas for selection and integration.

The present work describes a large space that deserves future study. First, the reported studies were performed within a population consisting mostly of novices. Experts might have different opinions on how to integrate features in designs and may assess the creativity of the designs differently. Thus, experts might be invited to test the external validity of the crowd's creativity. Second, the present analyses only looked at the combined features clearly present in the parent designs. However, we noticed a large number of emergent features in the combined designs that could not be traced back to any features of the parent designs. It is possible that these emergent features were somehow primed by the features in the parent designs. Since we observed that the most creative designs tended to have such emergent features, the emergence of these features in crowd designs deserves more attention.

In conclusion, the experiments open a door to crowd creativity. The combination system, and its output, provides a way to study collective intelligence, and in particular how and why it occurs. The crowd is able to perform creative work by infusing diverse ideas, filtering out ordinary ideas, and then integrating unique ideas.

Acknowledgement. Funding for this research was provided by the National Science Foundation, awards IIS-0855995 and IIS-0968561. We thank our colleague Jeffrey V. Nickerson for his ideas. This paper is taken from a dissertation to be submitted in partial fulfillment of the requirements for the Ph.D. degree at Stevens Institute of Technology.

References

1. Koestler, A.: The Act of Creation. Penguin, Arkana (1964)
2. Benkler, Y.: The Wealth of Networks: How Social Production Transforms Markets and Freedom. Yale University Press, New Haven (2006)
3. Kittur, A.: Crowdsourcing, collaboration and creativity. XRDS: Crossroads (December 2010)
4. Little, G., Chilton, L.B., Goldman, M., Miller, R.C.: Exploring iterative and parallel human computation processes. In: ACM SIGKDD Workshop on Human Computation, pp. 68–76 (2010)
5. Maher, M.L.: Design Creativity Research: From the Individual to the Crowd. Design Creativity (2010)
6. Quinn, A.J., Bederson, B.B.: Human Computation: A Survey and Taxonomy of a Growing Field. In: CHI. ACM Press, New York (2011)
7. Yu, L., Nickerson, J.V.: Crowd creativity through combination. In: Proceedings of CHI 2011. ACM, New York (2011)
8. Nickerson, J.V., Sakamoto, Y.: Crowdsourcing Creativity: Combining Ideas in Networks. In: Workshop on Information in Networks. NYU (2010)
9. Campbell, D.T.: Blind variation and selective retention in creative thought as in other knowledge processes. Psychological Review 67, 380–400 (1960)
10. Fleming, L.: Recombinant uncertainty in technological search. Management Science 47(1), 117–132 (2001)
11. Nelson, R.R., Winter, S.G.: An evolutionary theory of economic change. Belknap Press/Harvard University Press, Cambridge (1982)
12. Simonton, D.K.: Creativity in science: Chance, Logic, Genius and Zeitgeist. Cambridge University Press, New York (2004)
13. Thagard, P.: Conceptual revolutions. Princeton University Press, New Jersey (1992)
14. Thagard, P., Stewart, T.C.: The AHA! experience: Creativity through emergent binding in neural networks. Cognitive Science 35, 1–33 (2011)
15. Ward, T.B.: Cognition, creativity and entrepreneurship. Journal of Business Venturing 19, 173–188 (2004)
16. Estes, Z.C., Ward, T.B.: The emergence of novel attributes in concept modification. Creativity Research Journal 14, 149–156 (2002)

17. Kunda, Z., Miller, D.T., Claire, T.: Combining social concepts: the role of causal reasoning. Cognitive Science 14, 551–577 (1990)
18. Wilkenfeld, M.J., Ward, T.B.: Similarity and emergence in conceptual combination. Journal of Memory and Language 45, 21–38 (2001)
19. Sobel, R.S., Rothenberg, A.: Artistic creation as stimulated by superimposed versus separated visual images. Journal of Personality and Social Psychology 39(5), 953–961 (1980)
20. Bower, G.H., Black, J.B., Turner, T.J.: Scripts in memory for text. Cognitive Psychology 11, 177–220 (1979)
21. Goldberg, D.E.: Genetic Algorithms in Search, Optimization and Machine Learning. Kluwer Academic Publishers, Boston (1989)
22. Holland, J.H.: Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor (1975)
23. Kosorukoff, A.: Human based genetic algorithm. In: Proc. IEEE Conference on Systems, Man, and Cybernetics, pp. 3464–3469 (2001)
24. Kittur, A.: Crowdsourcing, collaboration and creativity. XRDS: Crossroads (December 2010)
25. Raykar, V.C., Yu, S., Zhao, L.H., Valadez, G.H., Florin, C., Bogoni, L., Moy, L.: Learning from crowds. Journal of Machine Learning Research 11(7), 1297–1322 (2010)
26. Fogel, D.: Evolutionary Computation: Towards a New Philosophy of Machine Intelligence. IEEE Press, Piscataway (1996)
27. Kittur, A., Chi, E.H., Suh, B.: Crowdsourcing user studies with Mechanical Turk. In: Proc. CHI 2008, pp. 453–456. ACM Press, New York (2008)
28. Amazon Mechanical Turk, `https://www.mturk.com/mturk/welcome`
29. Google Docs Drawing Application,
`http://docs0.google.com/demo/`
`edit?id=scACRQaIm3t83kVISWPhWfrqx#drawing`
30. Graesser, A.C., Nakamura, G.V.: The impact of a schema on comprehension and memory. In: Bower, G.H. (ed.) The psychology of learning and motivation, vol. 16, pp. 59–109. Academic Press, New York (1982)
31. Brewer, W.F., Treyens, J.C.: Role of schemata in memory for places. Cognitive Psychology 13(2), 207–230 (1981)
32. Taylor, S.E., Crocker, J.: Schematic bases of social information processing. In: Higgins, E.T., Herman, C.P., Zanna, M.P. (eds.) Social cognition: The Ontario symposium, vol. 1, pp. 89–134. Erlbaum, Hillsdale (1981)
33. Bower, G.H., Black, J.B., Turner, T.J.: Scripts in memory for text. Cognitive Psychology 11, 177–220 (1979)
34. Goodman, G.S.: Picture memory: How the action schema affects retention. Cognitive Psychology 12, 473–495 (1980)
35. Hastie, R., Kumar, P.A.: Person memory: Personality traits as organizing principles in memory for behaviors. Journal of Personality and Social Psychology 37, 25–38 (1979)
36. von Restorff, H.: Analyse von Vorgängen in Spurenfeld: I. Über die Wirkung von Bereichsbildungen im Spurenfeld [Analysis of processes in the memory trace: I. On the effect of group formations on the memory trace]. Psychologische Forschung 18, 299–342 (1933)
37. Finke, R., Smith, S., Ward, T.: Creative Cognition: Theory, Research, and Applications. MIT Press, Cambridge (1996)
38. Bohm, D.: On Creativity. Routledge, London (1998)

Part IV

Augmented Cognition for Learning

Augmented Cognition Methods for Evaluating Serious Game Based Insider Cyber Threat Detection Training

Terence S. Andre[1], Cali M. Fidopiastis[2], Tiffany R. Ripley[1], Anna L. Oskorus[1], Ryan E. Meyer[1], and Robert A. Snyder[1]

[1] TiER1 Performance Solutions
6 E 5th St., Suite 400 Covington, KY 41011, USA
[2] University of Alabama-Birmingham
1530 3rd Avenue South, Birmingham, Alabama 35294, USA
{t.andre,t.ripley,a.oskorus,r.meyer,
r.snyder}@tier1performance.com, cfidopia@uab.edu

Abstract. DoD investments into cyber threat defense are ongoing; however, little attention is paid to training personnel to detect and prevent threats to cyber networks that come from internal sources. Supervisors need to know what behavioral signs to watch for that might indicate an employee intends to commit an insider crime. Monitoring employee workstations is proving an ineffective means of determining insider threats. Training is needed to provide examples of the numerous ways cyber threats are achieved. An interactive role-play game environment may provide an appropriate instructional delivery system to train supervisors. Such a training system should employ instructional support features, aids, and feedback to the trainer and the trainee. The training system should also provide adaptive learning pathways to facilitate accelerated learning where individual assessments show mastery of specific content. Creating such a system not only requires appropriate training materials, but also a means to assess the systems efficacy. Augmented cognition methods and techniques for evaluating the cognitive state of a learner provide a real-time, objective means of evaluating training delivery and content. In this paper we discuss our efforts to assess learner engagement using psychophysiological measures.

Keywords: accelerated learning; adaptive training; learner engagement, psychophysical measures.

1 Introduction

Accelerated learning remains a major, yet unattained, goal of education systems [1], government [2] [3], business [4] and the military [5]. In the military, in fact, the need "to dramatically accelerate the transition from novice to expert in key military tasks" is seen as absolutely crucial to the continued effectiveness of U.S. fighting forces around the globe [6]. On the educational front, in a system that requires 25-75% fewer school days than most other countries do, rapid increases in learning rate might prove to be the only factor that can offset the U.S.'s continued decline in science and math

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 395–403, 2011.

test scores in comparison with those of other nations. Thus, for many reasons, technological applications that substantially and reliably accelerate learning would have near unlimited market demand and potentially great social impact. The greatest social impact for accelerated learning may be upon us in the area of cyber security. The growing reliance on technological infrastructures has made organizations increasingly vulnerable to cyber attacks, especially those from inside the organization [7]. Effective training of key personnel is needed so that organizations have proactive measures in place to detect and prevent intrusions before breaches can do significant damage. Accelerated learning, implemented through customized learning pathways and interactive role-playing games may be the right approach to effectively train cyber network supervisors so they can detect behavioral warning signs of an insider threat.

Scenario-based, role-playing games are believed to increase student engagement [8]. However, measures of engagement are typically subjective. Psychophysiological measurements such as EEG and eye-tracking sensors have led to significant advances in scientists' understanding of the kind of learning retained [9] [10] [11]. Through accelerated learning approaches and innovative ways to measure learner engagement, we expect to establish new guidelines for learning applications, just as one might use a user-in-the-loop iterative strategy when designing adaptive training systems [12].

2 Accelerated Learning

Most managers have a goal of getting employees' knowledge and skills up to speed as quickly as possible and then maintaining their knowledge and skills over time. Effective organizational performance is often correlated with employee knowledge and skill. The issue of speed is especially important in cyber security training as the magnitude of the overall threats and the frequency of attacks continues to grow. Thus, there is increasing pressure on organizations to train their employees both faster and better in order to reduce “time to value” [13]. The most common name for this highly desired and valuable goal is accelerated learning.

Educational technology appears to have the greatest potential effect on learning speed through courseware that provides adaptive (thereby accelerated) learning [14]. In addition, scenario-based, role-playing games appear to be effective for those with higher levels of knowledge or skills at the outset and increase student engagement for all learners [15].

2.1 TiER1 Approach to Accelerated Learning

Definitions of accelerated learning typically differ depending upon the specific context in which the definition is used. A broad spectrum of accelerated learning definitions is represented in Table 1.

Working from the base of these varying perspectives, TiER1 has operationally characterized accelerated learning as the ***reduction of learner time required to meet learning objectives in a training event***.

Table 1. Definitions of Accelerated Learning

Definitions of Accelerated Learning	Authors
Paraconscious mental activity that can create conditions to automate and use memory, brain, and intellectual reserves of people effectively	Lozanov (1978) [16]
Providing effective training in a short period of time	Gill and Meier (1989) [17]
Making a superlink between the right and left brain	Linksman (1996) [18]
Adapting and learning new skills quickly	Lawlor and Handley (1996) [19]
The ability to absorb and understand new information quickly and retain information	Rose and Nicholl (1997) [20]
Changing behavior with increasing speed	Russell (1999) [21]
Fast learning	Lynn, Akgun, and Keskin (2003) [22]
Learning faster and smarter to keep up with change, call on new knowledge, and apply new skills	Landale (2004) [23]
Rapidized training; getting individuals to achieve high levels of proficiency at a rate faster than ordinary; the idea of making learning more immune to decay	Hoffman, et al. (2010) [13]

To accomplish this goal, we approached acceleration within our system in two ways:

1. Adapting the amount of content learners need to cover based on their current level of proficiency
2. Increasing the efficiency of the learning process

Our team is developing XL-CITR (Accelerated Learning for Cyber Insider Threat Reduction), a comprehensive training solution that integrates best practices in instructional design, competency-driven learning objectives, content filtering for accelerated learning, performance feedback, and learning management. The design of the study will be unique in that it will 1) be able to assess both the individual and the combined effects of learning acceleration and game-based interaction, and 2) incorporate objective psychophysiological measurements using principles of augmented cognition (e.g., real-time, unobtrusive, and quantitative measures). The focus of this paper is on our approach for measuring learner engagement.

3 Augmented Cognition

A goal of augmented cognition research is to use psychophysiological measures such as electroencephalography (EEG), pupil diameter, and gaze tracking to identify—in real-time—perceptual, attentional, and cognitive workload states that may be helpful or detrimental to Warfighter performance [24]. Recent investigations within military-relevant training environments suggest that EEG signatures of attention, memory, and workload can be validly assessed during learning [25]. Furthermore, EEG measures offer a reliable means to accurately quantify key aspects of information processing [26] [27] [28]. These studies suggest that changes in EEG power spectra, as well as event-related EEG changes, are identifiable and correlate with levels of skill acquisition in simple and complex tasks. Thus, these measures are useful in matching mental capabilities to task and learning requirements so that mastery over the subject matter is maximized while mental effort is minimized [29].

The goal of the study run by augmented cognition researchers from the Interactive Simulation Laboratory (*i*Sim Lab) at the University of Alabama at Birmingham is to compare the cognitive state of learners who receive Accelerated Learning for Cyber Insider Threat Training (XL-CITR) training (e.g., fundamentals, quizzes, game-based interaction, and feedback) versus those who train using equivalent modules of a conventional training (CONVEN) implementation. An augmented cognition framework, using EEG and eye tracking sensors, will be used to assess the workload and the engagement levels of learners as they progress through abbreviated, generalized versions of XL-CITR and CONVEN, allowing design elements to be mapped onto intra- and inter-participant psychophysiological differences. That is, the experiment will establish individual baselines for the psychophysiological response measures and identify changes from baseline data during randomly assigned on-set and off-set exposure to the various program design elements. We expect the pattern of these changes in workload and engagement provide enough sensitivity to guide subsequent program and role-play game design.

3.1 Equipment

EEG. The wireless EEG sensor set to be used in this study was developed by Advanced Brain Monitoring (ABM). The system combines a 1.5 V battery-powered headset with a sensor placement system, following international standards. The full system is a lightweight, easy-to-apply cap that can acquire and analyze six to nine channels of high-quality EEG data. The sensors require no scalp preparation and provide a comfortable and secure sensor-scalp interface for 8 to 12 hours of continuous use. Sensor site locations on the system include the following: F3, F4, C3, C4, P3, P4, Fz, Cz, and POz. These sites can be combined in bi-polar or monopolar configurations (referenced to mastoids). The head pack contains miniaturized electronics that amplify, digitize, and transmit the EEG data in most environments, including environments with high electromagnetic interference.

Eye Tracking. Blink rate, pupil dilation, and scan path are convergent measures for mental workload [30]. The Arrington Research ViewPoint monocular eye tracker system will be used to track eye movements during the study. The eye tracker system uses an infrared (IR) light source to shine IR light on the eye, and the IR camera captures frames of images of eye as the person wearing the system performs different tasks, as shown in Figure 1. The eye camera and eye illuminator are located proximal to the eye; therefore, a head tracker is not required to accurately identify the pupil center. Using a bright pupil technique, the illuminator projects near-infrared light through the pupil such that the light reflects off the retina. The reflection created by the returning light traversing the cornea produces a bright pupil effect. This optical set-up allows the eye camera, which captures pictures of the eye every 17 ms (60 Hz), to estimate the eye pupil center and to discriminate the corneal reflection (CR). These two parameters are then used to estimate the point of gaze of the operator. The accuracy of the point of gaze estimation by the Arrington eye tracker is less than 1 degree visual angle (roughly a 7 mm error when the participant is seated at 76 cm from the display, as in this study). Further, using the captured eye images, the image processing algorithms can distinguish between iris and pupil areas and fir a circle or ellipse to calculate pupil diameter in real-time.

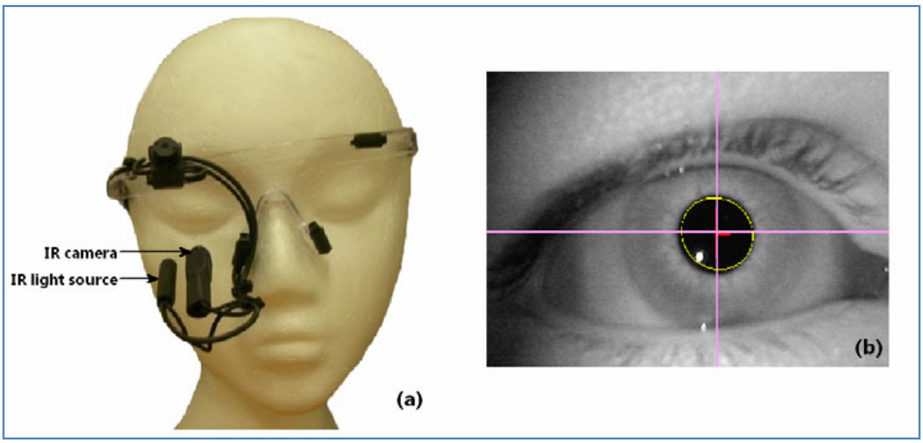


Fig. 1. (a) Arrington Research ViewPoint eye tracker system with infrared camera and light source. (b) Pupil detection using the light source and camera.

3.2 Subjective Self-assessments

An assessment of program quality and cognitive load surveys will be used in this study. Participants will respond twice, once for the XL-CITR module and once for the CONVEN module. These subjective measures should support the psychophysiological measures of workload.

3.3 Participants

A total of 32 male and female volunteer participants from the local military population (active duty, retired, or government-civilian) in Colorado Springs will participate in this study. Participants will be recruited through the local base newspapers and business groups. Participants in this study will have half of the training content delivered by the XL-CITR system and half by CONVEN. During the training, participants will wear a lightweight cap that will record psychophysiological measures as indicated in the equipment description Blink rate, pupil dilation, and scan path (of gaze) will be tracked with a remote camera.

3.4 Method

Participants will be fitted with the 9-channel, wireless EEG made by ABM. An impedance check of all sensors will then be performed. Once the sensors have settled to under 20 mV, a set of baseline tasks (a 3-choice vigilance task, an eyes-open rest task, and an eyes-closed task) will be performed to ensure that the EEG system is working properly. Once the EEG baseline is complete, the participant will undergo a calibration procedure to align the eye tracker coordinate system with that of the monitor displaying the XL-CITR content. In the calibration procedure, participants direct their gaze to 17 predefined points, one at a time, on the display, while the eye tracker maps the participants' eye position relative to the monitor.

Participants will be randomly assigned to one of four experimental conditions (GRP 1-4), balancing order of presentation and content set:

- GRP1: XL-CITR Module A, then conventional multi-media Module B
- GRP2: XL-CITR Module B, then conventional multi-media Module A
- GRP3: Conventional multi-media Module A, then XL-CITR Module B
- GRP4: Conventional multi-media Module B, then XL-CITR Module A

[The content of Module A will be the same across all conditions. The same holds true for the content of Module B.]

3.5 Dependent Measures

The dependent measures for this study include the EEG measures (EEG percent workload, EEG percent engagement, frequency changes at each EEG location, percent fixation for an area of interest), the eye tracking measures (blink rate, pupil diameter), subjective assessment of program quality, and the perceived workload based on the cognitive load questionnaire.

Testable Assumptions (TA) and Hypotheses (H)

- SII-TA1: The EEG measures of engagement will be sensitive enough to discern which aspects of the XL-CITR (e.g., fundamentals, quizzes, role-play game, and feedback) positively influence participants' focused attention.
- TA2: The EEG measures of workload will be sensitive enough to pinpoint such workload issues as how participants experience each component of the role-play game. As participants master the learning material, their workload will decrease while their performance will improve.

- TA3: The blink rate and pupil dilation will provide a real-time convergent means to reinforce findings of the EEG measures. Those persons who master the task will make fewer saccades and more fixations toward relevant information.
- TA4: The cognitive load survey will provide an overall measure of cognitive workload that will assist in assessing the psychophysiological measures.
- H1: Average scores on psychophysiological measures of engagement: XL-CITR > CONVEN.

4 Conclusion

We believe new opportunities exist for researchers who want to quantify learner engagement and design environments that have a significant impact on learning performance. Our work in accelerated learning and learner engagement is just now examining a few of the many variables that might show a positive outcome. Future studies are planned to generate a set of best practices that can be adapted by large organizations who are interested in improving the overall learning experience for their workforce.

Acknowledgements. This material is based upon work supported by the Air Force Research Laboratory (AFRL) under Contract No. FA8650-10-C-6061. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of AFRL.

References

1. Moe, T., Chubb, J.: Liberating learning: Technology, politics and the future of American education. Wiley, Hoboken (2009)
2. Jacobson, S.: Can we control training waster?. Frontline Magazine, 30–31 (March-April 2005)
3. Office of Management and the Budget. Report to Congress on the Implementation of the E-Government Act of 2002 (2005), Retrieved from `http://www.whitehouse.gov/OMB/inforeg/2004_egov_report.pdf`
4. Persaud, A.: A model of new product success and accelerated learning in collaborative new product team. Innovation: Management, Policy and Practice 6(2), 199–219 (2004)
5. Quinkert, K.A., Morrison, J.E., Fletcher, J.D., Moses, F.L., Roberts, E.J.: The Army Science of Learning Workshop. ARI Research Note 2007-02. Arlington, VA: US Army Research Institute for the Behavioral and Social Sciences (DTIC No. ADM001922) (2007)
6. Department of Defense. Department of Defense fiscal year (FY), budget estimate (2010), Retrieved from `http://www.darpa.mil/Docs/2010PBDARPAMay2009.pdf`
7. Randazzo, M.R., Keeney, M., Kowalski, E., Cappelli, D., Moore, A.: Insider Threat Study: Illicit Cyber Activity in the Banking and Finance Sector. National Threat Assessment Center, U.S. Secret Service, and CERT® Coordination Center/Software Engineering Institute, Carnegie Mellon (August 2004), Retrieved from `http://www.secretservice.gov/ntac/its_report_040820.pdf`
8. Bergeron, B.: Developing serious games. Charles River Media, Hingham (2005)
9. Fox, A.: The brain at work. HR Magazine 53(3), 36–43 (2008)

10. Rock, D., Schwartz, J.: The neuroscience of leadership (2006), Retrieved from http://www.strategy-business.com/press/freearticle/06207
11. Rock, D., Schwartz, J.: Why neuroscience matters to executives (2007), Retrieved from http://www.strategy-business.com/li/leadingideas/li00021
12. Fidopiastis, C., Nicholson, D.: User-in-the-loop adaptive system design: Information fusion examples for visualizing and measuring cognitive states. In: Applied Ergonomics International Conference, Las Vegas, NV, July 14-17 (2008)
13. Hoffman, R.R.: Accelerated proficiency and facilitated retention: Recommendations based on an integration of research and findings from a working meeting. Report on Grant FA8650-09-2-6033, 1-352 (2010)
14. National Institute on Standards and Technology. Adaptive learning systems (2005), Retrieved from http://www.atp.nist.gov/atp/97wp-lt.htm
15. Orvis, K.A., Horn, D.B., Belanich, J.: Task difficulty and prior videogame experience: Their role in performance and motivation in instructional videogames. ARI Technical Report 1202. U.S. Army Research Institute for the Behavioral & Social Sciences, Arlington, VA (June 2007)
16. Lozanov, G.: Suggestology and outlines of suggestodedy. Gordon and Breach Science Publishers, Inc., London (1978)
17. Gill, M.J., Meier, D.: Accelerated learning takes off. Training and Development Journal, 63-65 (1989)
18. Linksman, R.: How to learn anything quickly: An accelerated program for rapid learning. Citadel Press, Secaucus (1996)
19. Lawlor, M., Handley, P.: The creative trainer: Holistic facilitation skills for accelerated learning. McGraw-Hill, East Windsor (1996)
20. Rose, C., Nicholl, M.J.: Accelerated learning for the 21st century. Dell Publishing Group, New York (1997)
21. Russell, L.: Fortifying strategic decisions with shadow teams: A glance at product development. Competitive Intelligence Magazine 2, 9–11 (1999)
22. Lynn, G., Akgun, A.E., Keskin, H.: Accelerated learning in new product teams. European Journal of Innovation Management 6(4), 201–212 (2003)
23. Landale, A.: Mental agility training helps CSA staff to navigate rapid change. Human Resource Management International Digest 12(2), 20–23 (2004)
24. Schmorrow, D., Kruse, A.: Augmented cognition. In: Bainbridge, W.S. (ed.) Berkshire Encyclopedia of Human-Computer Interaction, vol. 1, pp. 54–59. Berkshire Publishing Group, Great Barrington (2004)
25. Berka, C., Levendowski, D., Lumicao, M., Yau, A., Davis, G., Zivkovic, V.: EEG correlates of task engagement and mental workload in vigilance, learning and memory tasks. Aviation Space and Environmental Medicine 78(5), pp. B231-B244 (2007)
26. Berka, C., Levendowski, D.J., Cvetinovic, M., Petrovic, M.M., Davis, G.F., Lumicao, M.N., Popovic, M.V., Zivkovic, V.T., Olmstead, R.E., Westbrook, P.: Real-time analysis of EEG indices of alertness, cognition and memory acquired with a wireless EEG headset. Special issue of the International Journal of Human-Computer Interaction on Augmented Cognition 17(2), 151–170 (2004)
27. Berka, C., Levendowski, D., Davis, G., Lumicao, M.N., Ramsey, C.K., Stanney, K., Reeves, L., Tremoulet, P.D., Regli, S.H.: EEG indices distinguish spatial and verbal working memory processing: Implications for real-time monitoring in a closed-loop tactical Tomahawk weapons simulation. In: Schmorrow, D. (ed.) Foundations of augmented cognition. Proceedings of the 1st International Conference on Augmented Cognition, pp. 405–413. Lawrence Erlbaum Associates, Mahwah (2005)

28. Poythress, M., Russell, C., Siegel, S., Tremoulet, P.D., Craven, P.L., Berka, C., Levendowski, D.J., Chang, D., Baskin, A., Champney, R., Hale, K., Milham, L.: Correlation between expected workload and EEG indices of cognitive workload and task engagement. In: Proceedings of 2nd Annual Augmented Cognition International Conference, San Francisco, CA (2006)
29. Feldon, D.F.: The implications of research on expertise for curriculum and pedagogy. Educational Psychology Review 19(2), 91–110 (2007)
30. Sciarini, L.W., Fidopiastis, C., Nicholson, D.: Toward a modular cognitive state gauge: Assessing spatial ability utilization with multiple physiological measures. In: Proceedings of the 53rd Annual Conference of the Human Factors and Ergonomics Society, vol. 53(3), pp. 146–150. Human Factors and Ergonomics Society, Santa Monica (2009)

Ongoing Efforts towards Developing a Physiologically Driven Training System

Joseph Coyne[1], Ciara Sibley[1], and Carryl Baldwin[2]

[1] Naval Research Laboratory,
4555 Overlook Ave SW, Washington, DC, USA
[2] George Mason University,
Fairfax, VA, USA
{Joseph.Coyne,Ciara.Sibley}@nrl.navy.mil, cbalwi4@gmu.edu

Abstract. There have been a number of successes of real-time application of physiological measures in operational environments such as with the control of remotely piloted vehicles (RPV). More recently, similar techniques have been investigated within the context of improving learning. A major challenge of the learning environment is that an individual's ability to perform the task, and thus their workload experienced during the task, are constantly changing. Cognitive Load Theory provides insight into how workload interacts with learning. One aspect of this theory is that as information is learned it reduces working memory demands. This paper discusses results from an RPV training study investigating the effects of workload and learning on pupil diameter. Specifically, pupil diameter decreased overtime as the task difficulty was held constant, and increased as new information was presented. The results of these studies are discussed in terms of how they can be used in a physiologically driven adaptive training system.

Keywords: Augmented Cognition, Pupil Diameter, Training, Workload.

1 Introduction

The introduction of the terms Neuroergonomics [1] and Augmented Cognition in recent years has signaled a renewed interest in applying measures of the brain to improving performance at work. Neuroergonomics refers to an interdisciplinary area of research where findings and techniques from neuroscience are used to better understand how the brain functions at work. The goal of this field is to create a better work environment which is informed by neuroscience. Augmented Cognition also has an emphasis on applying neuroscience to work but has been more specifically focused on using neurophysiological signals as inputs for closed loop systems. The goal of these closed loop systems is to detect when individuals are overloaded or underloaded and adjust their environment accordingly. The ability to create a close looped system, however requires that we have metrics that are sensitive to changes in cognitive load in near real-time. There have been a number of different types of sensors used with EEG being among the most popular due to its high temporal resolution.

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 404–412, 2011.

1.1 EEG and Workload

EEG has been one of the most extensively used real-time physiological measures of mental workload and several techniques have been emerging which show promise. Pope et al [2] developed a function utilizing a simple algorithmic formula based upon Beta power divided by Alpha plus Theta Power. This formula called the Engagement index has been applied to an adaptive closed loop simulated piloting task [3] called the MAT-B and was demonstrated to improve performance in a vigilance task. The formula does not require any individual calibration which likely reduces its sensitivity but increases its ease of use.

Advanced Brain Monitoring's (ABM) B-Alert system [4] uses a proprietary method of classifying both workload and engagement based upon discriminant function analysis. The workload index is reported to track processes that generally considered to be executive functioning whereas engagement is associated with more attentional and sensory processing resources. The system applies its classification to 1 second segments of EEG data. The ABM system requires EEG data be collected from participants on a simple vigilance task over 15 minutes to establish individualized coefficients for the DFA. The metrics have been demonstrated to successfully track workload in digit span and mental arithmetic tasks [4], however others have found it to be insensitive to more complex spatial processing tasks [5].

Ongoing research at Wright Patterson Air Force Base has applied Artificial Neural Networks (ANN) to classify high and low workload in a simulated UAV task [6, 7]. The ANN incorporates EEG and a number of other physiological signals including ECG and Pupillometry. Once the ANN has been trained using physiological data collected on segments of high and low workload on the task it can accurately assess high and low workload on untrained segments with 85-90% accuracy. However, the ANN's accuracy in distinguishing high and low workload drops when used on different days and different tasks which restricts its potential applicability to the use in operational settings.

Additionally there has been some recent success in using single trial Event Related Potentials (ERP's) [8]. ERPs are typically averaged across a number of trials and components such as P300 (a peak in amplitude occurring around 300ms after a stimulus has been presented) have been found to distinguish low and high workload. Moving to a single trial analysis increases the amount of noise but also increases the potential applicability of ERP as a real-time operational metrics of workload. However single trial ERP is still developing and its application requires precise millisecond timing between the task environment and EEG system. Although the approaches to classifying EEG data vary they have definitely shown promise as a means of assessing mental workload in real-time. However there are additional time costs to using EEG in an operational environment such as set up time and time to train the classification algorithms which may limit its operational utility. Other physiological metrics such as pupillometry also show promise as a potential real-time metric of cognitive load and require less set up time.

1.2 Pupillometry and Workload

Although not a direct measure of brain activity, pupil diameter has a long history of being tied to different cognitive processes. An association has been found between increased pupil dilation and activation of the middle frontal gyrus, which has been associated with central executive and high demand functions [9]. Pupil diameter has been shown to steadily increase as workload or working memory demands increase in a large variety of different simple and complex tasks [5, 10-13]. Averaged pupil diameter has been found to be sensitive to multiple levels of workload through increasing levels of difficulty of the task. There is still some conflicting evidence however about what happens to pupil diameter during overloaded conditions, as some evidence suggests pupil diameter levels off [12] while others have actually found pupil diameter to drop when task demands exceed available resources [10].

Additionally, pupil diameter has been found to be linked with fatigue by pupil size decreasing over the course of experimental sessions [11] as well as with motivation, with individuals demonstrating larger pupil diameters when an incentive to perform is provided [13]. While pupil diameter is typically averaged and used as a post hoc assessment of workload, there have been several investigators looking at applying pupil diameter to real time or over smaller time windows [14]. Marshall's index of cognitive activity [14] is a real time gauge of workload based upon applying a proprietary wavelet analysis to pupil diameter. Other researchers [5] have used average pupil diameter and maximum pupil diameter over short time windows.

Pupillometry has the potential to be a valuable index of mental workload in an operational setting. Although pupil diameter also varies with other more tonic psychological phenomenon such as fatigue and incentive, there is still evidence to suggest that it can still detect more short term changes in workload while these other phenomena are occurring [5, 13]. Technology for eye tracking systems has improved such that they are now completely off the head and require less than a minute to calibrate to an individual. These advancements in eye tracking systems make it more viable to investigate questions about pupillometry under various conditions.

1.3 Closed Loop Physiological Systems

Real-time adaptive automation involves the allocation of responsibilities or aiding to a human-machine system during a task and based on an input metric (EEG, pupillometry, performance) that is being analyzed in real-time. In recent years, researchers at Wright Patterson Air Force Base have had success using artificial neural networks (ANN) to classify operator state with accuracies up to 85% for an individual [6, 7].The researchers fed EEG and eye tracking data in real-time into an ANN and showed a 50% improvement in performance on a RPV target identification task when using adaptive automation in the task to slow down vehicle speed when workload was classified as high, and speeding it up when the ANN classified workload as low. The research serves as one of the best examples of how real-time physiological assessment and monitoring can be used to improve performance in an operational environment. The success has fueled interest in applying physiological sensors as real-time inputs to other closed loop systems and moving it from the operational domain to a training domain.

1.4 Applying Physiology to Learning

The goal within the operational environment is to identify periods of time where the operator is overloaded and then provide automation or aiding to reduce workload. The training environment however is dynamic and as trainees acquire skills, they utilize different brain regions and demonstrate reduced workload while performing the same task [15]. Cognitive Load Theory (CLT) [16] is a theory of learning which states that learning is essentially, processing and organizing information in working memory and storing it in long term memory. The organization of information in working memory is a cognitively demanding process and the largest bottleneck in the learning process. Overloading an individual's working memory capacity results in an inability to transfer information to long term memory and thus increased time to learn the task. Once information is learned and stored in long term memory the demands on working memory (cognitive load) are greatly reduced. The goal of adaptive training would therefore be to identify when an individual has learned a specific level of a task or information and then present them with additional information to learn.

Presently there are no systems that monitor task knowledge in real-time via physiology and then adapt training material. One approach to measuring skill mastery is being performed by researchers at EGI, where they are investigating the use of single trial event related potentials in specific brain regions during a language learning task. The present research addresses an alternative approach, measuring working memory load or workload in real-time with pupillometry, as an individual learns a task.

As an individual acquires knowledge (i.e., stores it in long term memory), they become less reliant on working memory and eventually the task becomes automatic. Therefore an individual who is learning a task should demonstrate both higher performance and lower workload once they have mastered the task. The present experiment trained individuals on a simulated RPV intelligence, surveillance and reconnaissance (ISR) task. Trainees had to calculate the vehicle's direction of movement based upon the UAV heading and the apparent target direction of travel on the simulated video feed. Skill level in the task was manipulated by increasing the amount of mental rotation necessary to calculate direction of travel.

2 Method

2.1 Participants

Thirteen participants (18-30 years, M = 22, SD = 4.10) from the George Mason University volunteered to engage in a UAV training simulation experiment in exchange for course credit. Four participants' data were excluded due to problems with missing data.

2.2 Materials

Virtual Battlespace 2 (VBS2) was used to construct simulated UAV video files that were played back in a separate application created for this experiment. VBS2 is a high-fidelity, three-dimensional virtual training system used for experimental and military training exercises. In addition, the Tobii X120 off the head unit was used to collect eye tracking data. The unit sat in front of the participant and just below the surface of the monitor running the simulation. The system recorded both eyes at 60 samples per second. Neuroscan was used to collect EEG data at 500 samples per second. EEG data however is not considered in this paper.

2.3 UAV Desktop Simulation

After receiving a brief PowerPoint training about the task, participants engaged in a UAV desktop simulation in which he or she was trained to report information on moving vehicles as seen from a UAV (see Image 1). Participants were asked to identify and report heading information about the target vehicles crossing the screen. At the beginning of each experimental block, examples of each target to identify were presented and participants were expected to learn to recognize each by name (ID task). An example of one of the vehicles is shown in Image 2. For each experimental trial, participants were given the heading of the UAV and were asked to estimate the heading of the vehicle on the ground (heading task). A graphical depiction of a compass facing due north with 30 degree increments was provided to the participant for reference. After entering the target heading estimation and the identity of the target, participants were then asked to rate their perceived mental effort in calculating the target heading and identifying the target.

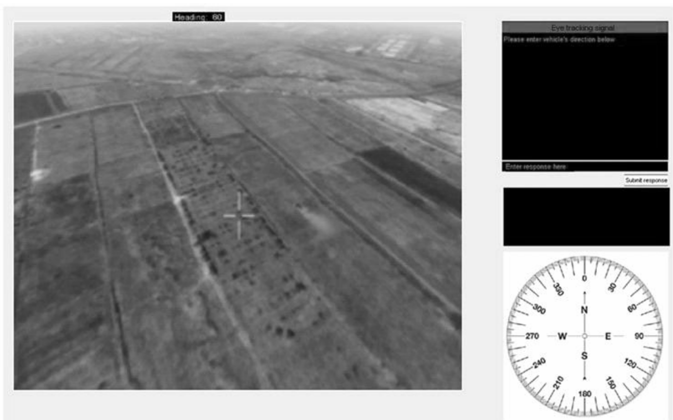


Fig. 1. Interface for the experiment

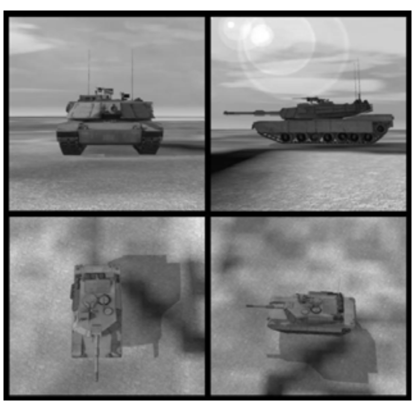


Fig. 2. Picture of an M1A1 that participants were expected to recognize by name

The difficulty of the task progressed over three blocks of trials. Only one vehicle was shown on the screen at a time and a total of twenty vehicles were shown within each block. Difficulty was manipulated by varying the UAV heading as well as the possible target heading. Additionally, the ID task difficulty was manipulated by increasing the number of vehicles the participant learned for each block from two to four to six. For example, the easiest level (block one) showed the UAV heading at only 0 degrees and the target's heading could be either 0, 90, 180 or 270 degrees. During this level, participants were only expected to learn two vehicles.

The most difficult level (block three) showed the UAV heading at any 30 degree increment and changed after each target, and the target heading could also be any 30 degree increment. During this final block, participants were presented with six vehicles and were expected to recognize all six of them by name.

2.4 The Experiment

Written informed consent was obtained from all participants and the participants were then introduced to the experimental tasks. This experiment took place over one day with a duration of approximately two to three hours, including EEG preparation, eye tracking calibration, training and experimental trials. After being prepped for EEG, participants reviewed a PowerPoint training on the task and then began the experiment. All participants completed blocks one, two and three in the same order from easiest to hardest.

3 Results

The researchers considered one of the most cognitively demanding parts of the simulation to be during the target heading calculation. Therefore, this analysis primarily focused on the pupillometry data during this task. This was done in order to compare pupil dilation during the high demand parts of the task to other less demanding parts. Additionally, in order to investigate the affect of learning on pupil dilation, the first and last three trials of each difficulty block were analyzed and compared to each other. We hypothesized that as the participant began to learn the material, it would become less challenging throughout that difficulty block, and that his/her pupil size would get closer to baseline levels towards the end of the block.

3.1 Maximum Pupil Size and Difference Scores

To best capture the period of mental effort during the heading calculation task, we took the maximum pupil size (an average of the top five pupil sizes) during each part of the task. Figure one depicts the maximum pupil sizes during the first and last three trials of each block during the ID and heading task. According to performance data, participants focused their efforts on the heading task, which is confirmed in the pupillometry since pupil size is greater during the heading task than it is during the ID task; both of which are within seconds of each other.

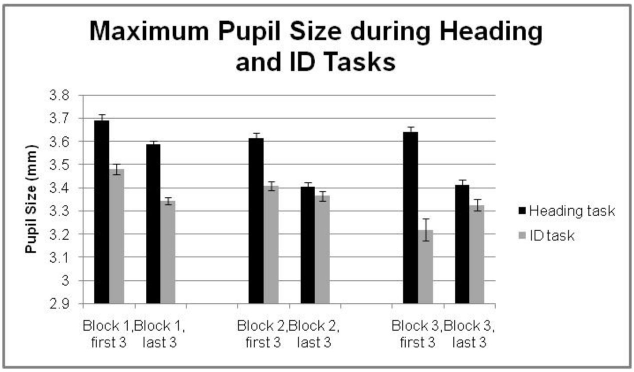


Fig. 3. Pupil dilation is sensitive to mental effort and is highest during the most demanding part of the task

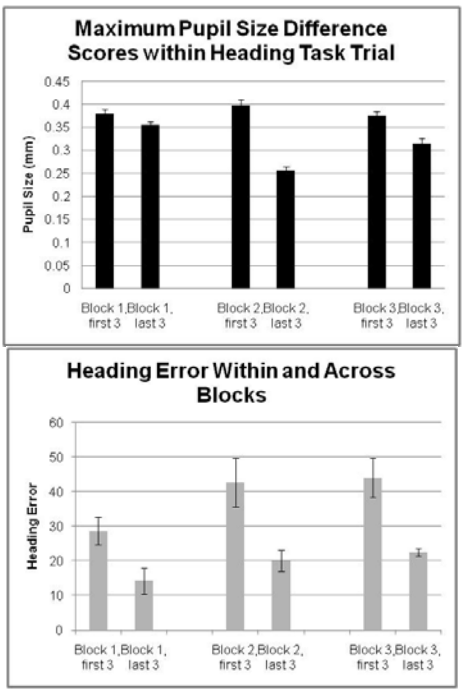


Fig. 4. Maximum pupil difference size is highest at the beginning of the new block of difficulty, and then attenuates with learning (left figure). This trend follows with the performance data (right figure) showing participants' error reducing by the end of each block.

Difference scores were then calculated for pupil dilation by calculating each individual's average pupil size during each individual trial and subtracting that from the maximum pupil size data. This was done in order to show change from the average and reduce factors caused by individual variability. Positive scores mean that the pupil size was greater than average, while negative scores indicate pupil size was smaller than average. Figure 2 (left) shows not only that pupil size is far above average (average being zero), but also that that pupil size attenuates towards the end of each block, but then jumps back up at the beginning of the next difficulty block. This pattern also corresponds with performance data (figure 2, right) where the participants' performance becomes significantly better towards the end of each block.

A paired-samples t-test was conducted to compare the pupil size for the first three compared to the last three trials across each block. There was a significant difference in the pupil sizes in the first part of the block (M=0.38, SD= 0.16) compared to the second part of the block (M=0.31, SD= 0.14), $t(45)= 2.40$, $p= 0.01$.

4 Discussion and Implications

The results of the present study demonstrate that pupil diameter is sensitive to both phasic and tonic changes of workload. Phasic, short term sensitivity is evidenced by large increases in pupil diameter as individuals moved from the search phase (easiest portion of the task) to the mental calculation phase (hardest phase). Pupil diameter was also sensitive to tonic changes in workload as evidenced by the gradual decrease in pupil diameter from the beginning of the block to the end of the block. The fluctuations of pupil diameter during the heading task between and across blocks matches what was expected based upon Cognitive Load Theory. That is, within each block of difficulty pupil diameter significantly decreased when comparing the beginning of the block to the end of the block. These changes in workload also corresponded with increases in performance from the beginning of the block to the end of the block. The combined increase in performance and decrease in workload suggests that information was transferred to long term memory and the burden on working memory was reduced i.e., information was being learned. Additionally, as new information was presented e.g., from block 1 to block 2 pupil diameter once again significantly increased.

Although the present study was not a closed loop adaptive training, it did provide evidence to suggest that pupil diameter could be used to drive such a system. Unlike the operational environment, developing a closed loop training system has many unique challenges beyond the simple identification of sensitive metrics; although even within metric selection, learning presents some unique challenges. For example, how do you train an ANN when workload fluctuates as a function of learning? Identifying high and low workload may need to be done with a different task which may impact the sensitivity of the ANN. This is in contrast to the operational environment, where we simply identify periods of overload and turn some additional automation on, or identify periods of underload and turn some automation off. Changes in learning appear to be more gradual and present challenges that need to be investigated. For example, determining what threshold level suggests an individual is ready to learn new material is something that still needs to be identified. This threshold may also change as the goal of learning shifts from acceptable performance to retention. Future studies will investigate these and other questions with the intention of eventually closing the loop in a training environment.

Acknowledgements. This work was funded by the Office of Naval Research's Human Performance Training and Education Program Grant number: N0001410WX20539.

References

1. Parasuraman, R.: Neuroergonomics: Research and Practice. Theoretical Issues in Ergonomics Science 4, 5–20 (2003)
2. Pope, A.T., Bogart, E.H., Bartolome, D.S.: Biocybernetic system evaluates indices of operator engagement in automated task. Biological Psychology 40, 187–195 (1995)
3. Freeman, F.G., Mikulka, P.J., Prinzel, L.J., Scerbo, M.W.: Evaluation of an adaptive automation system using three EEG indices with a visual tracking task. Biological Psyhology 50, 61–76 (1999)
4. Berka, C., Levendowski, D., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T., Olmstead, R.E., Tremoulet, P., Craven, P.L.: EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning, and Memory Tasks. Aviation, Space, and Environmental Medicine 78, B231–B244 (2007)
5. Sibley, C., Coyne, J.T., Cole, A., Gibson, G., Baldwin, C.L., Roberts, D., Barrow, J.: Adaptive training in an Unmanned Aerial Vehicle: Examination of several candidate real-time metrics. In: Karwowski, W., Salvendy, G. (eds.) Applied Human Factors and Ergonomics, Taylor & Francis, Boca Raton (2010)
6. Wilson, G.F., Russell, C.A.: Real-Time Assessment of Mental Workload Using Physiological Measures and Artifiicial Neural Networks. Human Factors 45, 635–643 (2003)
7. Wilson, G.F., Russell, C.A.: Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined adaptive aiding. Human Factors 49, 1005–1018 (2007)
8. De Lucia, M., Michel, C.M., Murray, M.M.: Comparing ICA-based and single-trial topographic ERP analyses. Brain Topography 23, 119–127 (2010)
9. Spinks, J.A., Zhang, J.X., Fox, P.T., Gao, J.H., Tan, L.H.: More workload on the central executive of working memory, less attention capture by novel visual distracters: Evidence from an fMRI Study. NeuroImage 23, 517–524 (2004)
10. Poock, G.K.: Information processing vs. pupil diameter. Perceptual & Motor Skills 37, 1000–1002 (1973)
11. Khaneman, D., Peavler, W.S.: Incentive effects and pupillary changes in association with learning. Journal of Experimental Psychology 79, 187–196 (1969)
12. Peavler, W.S., McLaughlin, J.P.: Pupil size, information overload, and performance differences. Psychophysiology 11 (1974)
13. Heitz, R.P., Schrock, J.C., Payne, T.W., Engle, R.W.: Effects of incentive on working memory capacity: Behavioral and pupillometric data. Psychophysiology 45, 119–129 (2008)
14. Marshall, S.P.: Identifying cognitive state from eye metrics. Aviation, Space, and Environmental Medicine 78, B165–B175 (2007)
15. Luu, P., Shane, M., Pratt, N.L., Tucker, D.M.: Corticolimbic mechanisms in the control of trial and error learning. Brain Research 1247, 100–113 (2009)
16. Sweller, J.: Cognitive load during problem solving: Effects on learning. Cognitive Science: A Multidisciplinary Journal 12, 257–285 (1988)

A Hierarchical Adaptation Framework for Adaptive Training Systems

Sven Fuchs, Angela Carpenter, Meredith Carroll, and Kelly Hale

Design Interactive, Inc. 1221 E Broadway,
Oviedo, FL 32765, USA
{sven,angela,meredith,kelly}@designinteractive.net

Abstract. Real-time adaptation is challenging in both operational and training environments, as the system must be able to identify what, why, and when mitigation is needed, and how best to mitigate to optimize the human-system interaction. Training systems have additional complexities, as the sole goal is not to optimize performance as in operational environments, but to optimize training, which may involve more error allowance for learning opportunities. This paper outlines a proposed hierarchical adaptation framework for adaptive training systems, involving diagnoses of learning state, performance, and expertise. It will also discuss candidate approaches to obtaining the necessary measurements using physiological and neurophysiological processes, provide some guidance for designing strategies for optimal adaptation, and highlight current challenges and future research areas.

Keywords: Adaptive Training, Augmented Cognition, Training Systems.

1 Introduction

Adaptive training has great potential across many domains, including airport baggage screening, which involves human review of electronic images/video of bags in search of explosives and other dangerous items. Screening is a repetitive visual search task that often has a very low probability of encountering a threat, but extremely high consequences if a serious threat is missed. Due to the importance of screening accuracy, screeners are required to complete extensive training both before going on the job and while employed. The unrepresentative ratio of threat targets during training (to ensure exposure within constrained timeframe) and repetitiveness of current training practices over time degrades its effectiveness, and current evaluation methods are unable to quantitatively measure details of officer state such as hazardous cognitive states (e.g., fatigue, boredom, inattention) that could negatively impact training outcomes and identify root cause(s) of performance errors (i.e., scan vs. recognition error) and influencing factors (e.g., threat type, location, orientation, clutter in bag, etc.). Instead, instructors using traditional training methods are often limited to observable behavioral metrics (e.g., detections, false alarms).

To more completely identify all obstacles to training effectiveness in this domain and others, new training systems are required that 'peer into the mind of the users' to capture perceptual and cognitive processes not otherwise accessible, and provide

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 413–421, 2011.

quantitative metrics to evaluate trainee state throughout a training session. This would result in a more complete measurement suite that allows a comprehensive diagnosis process to identify root cause(s) of training deficiencies/inefficiencies that can in turn provide directed feedback for addressing issues, thus individualizing training to optimize training effectiveness and efficiency. Such a training system could provide in-action feedback, where training scenarios can adapt in real-time to provide individualized training targeted at specific error patterns. Furthermore, such a training system could provide a comprehensive After Action Review (AAR) display that summarizes training strengths, deficits, and provides playback options (e.g., freeze in action, step through action) to highlight training opportunities. A thorough understanding of when to adapt, what to adapt, and how to adapt to optimize training is required to ensure that deficiencies/inefficiencies are targeted at appropriate times to focus training and avoid negative adaptation consequences such as distraction or confusion. There is a need to develop tools that support instructors and screeners to address when errors occurred, what errors occurred and why these errors occurred to increase the diagnostic value of training assessment.

Currently, no operational training systems incorporate this level of advanced performance measurement. While many testbeds have been developed to study these tools in a laboratory setting, few have addressed the many theoretical and technical challenges associated with developing a fieldable system. First, the sensor hardware must be usable and noninvasive while maintaining high levels of accuracy and reliability. Second, the system must facilitate near real time correlation of sensor data with scenario and behavioral events. Third, the system must facilitate the processing of very granular process level data into meaningful actionable diagnoses. Fourth, the system must respond to these diagnoses in a manner that ideally addresses the training gaps and learning state of the trainee.

This paper will outline a proposed adaptation framework for diagnoses of learning state, performance, and expertise, discuss candidate approaches to obtaining the necessary measurements using physiological and neurophysiological processes, and provide some guidance for designing strategies for optimal adaptation. Current challenges and future research areas are also discussed.

2 A Training Adaptation Framework

Adaptive training systems are different from their operational counterparts in that performance optimization is not the only goal of the adaptation. In a training setting, the objective is to train efficiently and effectively, which often requires the user to make mistakes as an acceptable and valuable part of learning. Such "mistake" opportunities may be lost (mitigated or prevented) in a performance-optimizing operational system. While adaptations are used operationally to mitigate problem states (a "treat the symptom" approach), adaptations should be used in training to address root causes of problem states, as training allows time and flexibility to drill down to that level. Conversely, while good performance is the goal of operational systems, it may signal the need for adaptation in a training environment if the trainee has mastered a skill and is ready to move on.

In order to effectively adapt a scenario without disrupting learning, use of a hierarchical adaptation framework is proposed. At the highest priority level (Level 1), it should be ensured that the situation affords learning and that learning can occur. Given this, it is first necessary to mitigate negative learning states such as drowsiness and distraction. Given the example vigilance task, these states are anticipated to occur and could conceal true performance issues and prevent training opportunities. The second priority (Level 2) is to address skill deficiencies, guiding trainees to acquire all necessary knowledge and skills, preventing them from practicing bad habits, and perpetuating incorrect performance or error patterns. Once performance is adequate, the third priority (Level 3) is to target development of expertise, providing trainees with practice opportunities and instruction designed to move them up the expertise continuum to automated performance. Based on these three diagnostic categories (i.e., Learning State, Skill, and Expertise), it is necessary to identify how data would be obtained and integrated to diagnose adaptation trigger points. Candidate approaches are discussed in the following section.

3 Measurement and Diagnosis

Given the 'unobservable' perceptual and cognitive processes involved in many complex tasks, it may be necessary not only to assess outcomes such as accuracy of response, but also to assess process measures at a cognitive level. To realize such an adaptive training solution, advances in both measurement/diagnosis and in training science could be utilized that address the framework's information requirements and enable effective and efficient adaptation to optimize training. A comprehensive suite of metrics and diagnostic capabilities is needed to drive such individualized feedback to optimize training effectiveness and efficiency. Advances made in psycho-physiological sensing technology have demonstrated success in evaluating cognitive state in near real-time. Numerous sensors exist to capture process measures of cognitive activity, including behavioral, physiological and brain-based technology. These technologies provide insights that go beyond outcome performance and beyond the capabilities of an observing instructor.

According to the proposed training adaptation framework, adaptations to training should be made with the goals of first enhancing readiness for learning to ensure the information provided during training can be effectively processed by the trainee, second, improving knowledge and skills to allow development of skilled performance, and third, increasing expertise levels to boost efficiency and effectiveness of both training and trainee performance. In the following subsections, the three diagnostic approaches – learning state diagnosis, skill diagnosis, and expertise diagnosis – will now be discussed in detail.

3.1 Learning State (Readiness) Diagnosis

With the emergence of neurophysiological and physiological measurement technology that allows for real-time assessment of perceptual and cognitive processing, many unobservable processes become accessible. Specifically, some cognitive states that are measurable via electroencephalography (EEG), including workload and engagement, can provide neurophysiological measures of the unobservable aspects of cognition [1, 2]. The following list outlines specific cognitive states that generally negatively affect the readiness for training by reducing attentional resources that facilitate learning and retention. Thus, it may be possible to utilize certain neurophysiological cognitive state metrics to detect issues with readiness to learn:

- Workload: High cognitive workload is expected when performing in a knowledge-based control mode because no automaticity guides the process [3, 4]. In rule-based control mode, rules are consciously retrieved from memory and applied to gathered information, also causing increased cognitive processing demands. Experts using skill-based techniques, however, employ automated routines that require fewer cognitive resources. Thus, it is expected that the assessment of cognitive workload can contribute to the identification of the trainee's control mode.
- Engagement: Because of high task demands, novice and journeyman trainees are expected to exhibit higher levels of engagement than expert trainees because studies have shown a trend for decreasing EEG engagement with increasing task proficiency [3, 5].
- Distraction: Distraction is a state characterized by a lack of clear and orderly thought and behavior, where a trainee becomes involved somewhere other than the cognitive tasks of interest [6]. Expert performers have an exhaustive mental model of the task or situation so that very few situations cause distraction. Confusion is one element of distraction. In rule-based decision makers, confusion may stem from the conscious selection of rules and difficulties in applying them to the situation at hand. Naïve trainees are expected to show relatively high levels of confusion because their mental models are more likely to be incorrect or insufficient so that new situations may cause a mismatch.
- Drowsiness: Sleep disorders are common and can have deleterious effects on performance [7, 8, 9]. In fact, loss of sleep can accumulate over time and result in a ''sleep debt,'' which can lead to impairments in alertness, memory, and decision making. Individuals with chronic accumulation of fatigue are often unaware of the impact on their performance.

3.2 Skill Diagnosis

Skill development is typically measured using performance metrics that are built into the training environment. Typical examples are "time to engage" or "number of hits". However, these metrics do not provide information about why performance breakdowns occur, so that traditional training often addresses symptoms by offering repeat practice instead of addressing the root cause(s) of errors and/or error patterns. Using the baggage screener example, it would not be possible to provide proper feedback without knowing whether a threat was missed due to improper visual scanning patterns (i.e., the trainee did not look at the threat or did not dwell on the threat long enough to allow cognitive processing of threat) or recognition issues (i.e., screener visually interrogated threat, but did not correctly identify it as such).

For training involving perceptual tasks (which is the case for most training environments), the addition of eye tracking can add a level of richness in evaluation that cannot be accomplished by only monitoring overt behavioral responses. Processes that can be detected by eye tracking but are otherwise hard to capture are those associated with perceptual-cognitive evaluation (i.e., perceptual-cognitive processes that do not necessarily have an immediate observable response). Instead, the first response to a changed situation is often information gathering, evaluation and decision making, all of which are non-behavioral activities. The addition of eye tracking enables the collection of real-time quantitative and qualitative data indicative of the perceptual-cognitive process between a stimulus and the (behavioral) response.

Eye tracking also aids in the assessment of perception through measurement of visual attention via gaze, scan path, and fixation data. The metrics can identify whether and what Areas of Interest (AOIs) were gazed upon, how often, and the duration of the visual attention. By examining attention shifts during an observation task (e.g., to relevant AOIs after receiving a command), eye tracking data can provide contextual information (i.e., what is or is not being attended to) to enhance performance diagnosis. Additionally, eye tracking performance criteria can be developed to determine if a trainee is applying effective scanning strategies, looking at the correct cues, and if they are completing the perceptual tasks efficiently.

Physiological metrics could also be combined with other measures to provide a more holistic picture of performance. Continuous monitoring of cognitive state and performance metrics can provide indications of task vigilance without altering the task flow, especially during low-frequency target detection tasks like baggage screening. Visual scan and focal patterns could be identified to indicate whether trainees (1) scanned appropriate areas, (2) dwelled on potential threats, and (3) combined with behavioral outcomes, determine whether appropriate threats were correctly identified. Results of this analysis can be used to drive discrete training interventions that address scan vs. detect vs. recognition errors. One could also develop an individualized neural profile of brain states coincident with low and high performance periods to obtain insights into perceptual and cognitive processing. Assessing underlying skills associated with identifying relevant information versus distractors could drive error pattern understanding to identify key parameters that should be the focus of future training.

3.3 Expertise Diagnosis

To develop expertise, it is necessary to train beyond simple skill acquisition. Practicing skills from different angles and acquiring strategies for more efficient processing is necessary to reach a level of proficiency at which the processes related to each skill have become easily accessible and can be executed in an automated manner. Expertise diagnostics could also replace the instructor's ability to aggregate all available information to avoid burdening the user with irrelevant information or inappropriate training pace or complexity.

It is proposed that physiological indicators can be used in conjunction with behavioral metrics and performance indicators to identify levels of expertise. Prior research [10] has identified a domain general control network that supports cognitive tasks involved in the learning of new tasks and is comprised of lateral frontal (goal processing and task switching), medial frontal (process monitoring, decision making, conflict management), and poster parietal (attentional control) brain regions. In general, these areas "drop-out" after extended training. This has been described [11] as a "processing efficiency change" that predicts increased cortical idling (alpha activity) with the development of expertise. As a trainee transitions from novice to advanced learner in a given task domain, cortical support transitions from the control network to task-specific processing regions that are more efficient and characterized as automatic processing. Accordingly, EEG activity reflects this change with increased alpha activity in the control network and more focal beta and gamma activity (slight decrease compared to novices due to processing efficiency) in task-specific regions. This is consistent with the neural efficiency hypothesis of expertise since experts deliver greater performance with seemingly less high frequency cortical activation. The widespread increases in alpha activity essentially "quiets" the brain which increases the salience of the focal higher frequency activation, even though it is relatively low levels, which supports expert task performance. Based on these findings, the hypothesis for driving expertise evaluation is that low frequency control network activity will dramatically increase with expertise whereas high frequency task-specific activity will slightly decrease.

Identification of expertise levels has important implications for implementing an adaptive training solution that considers trainee progress (i.e., transformation), and actively and effectively transitions trainees from one expertise level to the next one until proficiency is reached. Experts are said to learn more from their mistakes than from correct responses [12], and thus "tough cases" that are meant to instigate errors will be used in order for trainees to learn how to handle complexity, overcome knowledge shields (i.e., rationalization of misunderstandings), and guide them in acquiring more adaptive knowledge that can be used to address the unexpected (i.e., tough cases) [13]. Once trainees have reached a high level of competency for a particular set of procedural skills, training can be automatically refocused on other skill areas to ensure optimal use of training resources. At this level refresher training may also occur to maintain proficiency.

4 Adaptation Strategies

Once mechanisms are in place to support the collection of data and calculation of diagnostics required for the hierarchy levels, adaptation strategies are necessary to enable the closed-loop adaptive training system to react to these insights. Specifically, adaptation triggers and training strategies should be developed to (1) address non-optimal learning states (e.g., drowsiness, distraction) that require mitigation, (2) address skill-related deficiencies/inefficiencies which need remediation, and (3) adapt to changes in expertise. The goal should be to develop considerate, context-sensitive adaptations that exhibit minimal cognitive cost.

Generally, two types of adaptation strategies can be distinguished: real-time interventions that are triggered during the course of training, and between-scenario adaptations that are based on more global or composite measures and are typically invoked after a training section or lesson is completed. Real-time mitigations need to seamlessly integrate into the training scenario so as to guide/aid as needed to optimize learning and training transfer. Examples are feedback cues, instructional messages, and adaptation of training difficulty. Adaptations occurring between scenarios can be used to alter the training course, training method, or the skill set being trained.

5 Future Work

The above framework and approaches constitute an early conceptual approach to accomplishing truly adaptive training that has the potential to revolutionize training effectiveness and efficiency. However, many of the outlined approaches and strategies are still at infancy and advancements in several areas are needed to fully realize the potential of dynamic training adaptation.

One important consideration is the level of individualization. Multiple factors will impact the effectiveness of mitigation strategies implemented, including the training context and scenario content, and individual differences, as some strategies may be more effective for a given individual compared to another trainee. For example, some trainees may be visual learners, and thus respond optimally to a visual representation of corrective action, while others may best learn via spoken instruction/correction. Thus, the ideal mitigation framework would be adaptive itself, and learn over time which strategies work best for a given individual, thereby continuously optimizing the individual training experience.

Other challenges are the technical limitations of the measurement equipment. Physiological sensors must become cheaper and less intrusive to gain broad user acceptance while maintaining or even improving accuracy. Time-synching of multiple data streams, as well as practical issues in diagnosis design such as classifier development, state/skill/expertise threshold establishment, and pattern identification are additional areas with tremendous improvement potential.

Extending to team training settings, adaptation design challenges further increase, as some adaptation strategies (e.g., adapting timing parameters of tasks) may not be effective in a team training environment. For example, slowing down the pace of a scenario designed to assist one trainee in understanding the intricacies of the training may negatively impact team members' performance by slowing down the flow of information and/or changing the order of information flow which may result in negative training transfer. Therefore, the impact of adaptations on team settings must be considered and new team-wide adaptation strategies must be developed.

6 Conclusions

Advances in adaptive training are addressing some of the limitations of current performance assessment and training. It has become apparent that a one size fits all philosophy does not make for successful training. Valuable simulation time is lost and training quality is compromised when trying to force all students through the exact same curriculum on the same schedule. Likewise, relying on expert assessment of student performance within a dynamic environment such as a simulation is not always effective or reliable. It is easy for even the expert observer to miss important details of a student's performance in a rapidly unfolding scenario.

This insight has led to a wide-spread increase of research into adaptive training approaches, and the framework described in this paper is the result of such efforts. Although considerable conceptual and technological challenges exist, the identification of limitations is an important step towards addressing them. Once acknowledged, future efforts can be focused appropriately to address them in a targeted manner, so that the full potential of adaptive training can soon be realized.

Acknowledgments. This material is based upon work supported in part by the Office of Naval Research (ONR) under SBIR contracts N00014-09-M-0385 and FA8650-08-M-6826, and the Department of Homeland Security (DHS) under SBIR contract N10PC20028. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views or the endorsement of ONR and DHS.

References

1. Dorneich, M.C., Whitlow, S.D., Mathan, S., Ververs, P.M., Erdogmus, D., Adami, A., Pavel, M., Lan, T.: Supporting Real-time Cognitive State Classification on a Mobile Individual. J. Cog. Eng. & Decision Making 1(3), 240–270 (2007); Special Issue on Augmented Cognition: Past, Present, and Future
2. Levonian, E.: Retention Over Time in Relation to Arousal During Learning: An Explanation of Discrepant Results. Acta Psychological 36(4), 290–321 (1972)
3. Berka, C., Levendowski, D.J., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T., Olmstead, R.E., Tremoulet, P.D., Craven, P.L.: EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning, and Memory Tasks. Aviat. Space Environ. Med. 78(5, suppl.), pp. B231–B244 (2007)
4. Klein, G.A.: Seeing the Invisible: Perceptual-Cognitive Aspects of Expertise. In: Rabinowitz, M. (ed.) Cognitive Science Foundations of Instruction, pp. 203–226. Erlbaum, Mahwah (1992)
5. Stevens, R., Galloway, T., Berka, C.: Allocation of time, EEG-engagement and EEG workload resources as scientific problem solving skills are acquired in the classroom. In: Proceedings of 3rd Augmented Cognition International, held in conjunction with HCI International 2007, Beijing, China, July 22–27, Springer, Heidelberg (2007)
6. Poythress, M., Russell, C., Siegel, S., Tremoulet, P.D., Craven, P., Berka, C., Levendowski, D.J.: Correlation Between Expected Workload and EEG Indices of Cognitive Workload and Task Engagement. In: Schmorrow, D., Stanney, K., Reeves, L. (eds.) Augmented Cognition: Past, Present and Future, pp. 32–44. Strategic Analysis, Inc., Arlington (2006)
7. Berka, C., Levendowski, D.J., Cvetinovic, M., Petrovic, M.M., Davis, G.F., Lumicao, M.N.: Real-time Analysis of EEG Indices of Alertness, Cognition and Memory Acquired with a Wireless EEG Headset. Int. J. of Human-Computer Interaction 17(2), 151–170 (2004)
8. Berka, C., Levendowski, D.J., Ramsey, C.K., Davis, G., Lumicao, M.N., Stanney, K., et al.: Evaluation of an EEG-Workload Model in an Aegis Simulation: Biomonitoring for Physiological and Cognitive Performance during Military Operations. Int. Soc. for Opt. Eng. 5797, 90–99 (2005)

9. Neri, D. F., Dinges, D. F., Rosekind, M.R.: Sustained Carrier Operations: Sleep loss, Performance, and Fatigue Countermeasures. Moffett Field, California: NASA Ames Research Center (1997), `http://human-factors.arc.nasa.gov/zteam/PDF_pubs/Nimitz1997.pdf` (accessed July 24, 2009)
10. Schneider, W., Chein, J.M.: Controlled & Automatic Processing: Behavior, Theory, and Biological Mechanisms. Cog. Sci. 2(7), 525–559 (2003)
11. Kelly, A.M.C., Garavan, H.: Human Functional Neuroimaging of Brain Changes Associated with Practice. Cerebral Cortex 15(8), 1089–1102 (2005)
12. Sonnentag, S.: Excellent Performance: The Role of Communication and Cooperation Processes. App. Psych.: An International Review 49(3), 483–497 (2000)
13. Hoffman, R.R., Feltovich, P.J., Fiore, S.M., Klein, G., Ziebell, D.: Accelerated Learning? IEEE Intelligent Systems 24(2), 18–22 (2009)

Developing and Automating a Prototype for Assessing Levels of Student Involvement

Curtis Ikehara and Martha Crosby

University of Hawaii at Manoa, Department of Information and Computer Sciences,
1680 Easst-West Road, Honolulu, Hawaii 96822, USA
{cikehara,crosby}@hawaii.edu

Abstract. The proposed project objective is to develop and automate a methodological technology for objectively measuring a student's affective states, cognitive states and levels of involvement during computer-mediated instruction. Passive devices will record gaze activity, facial expressions and body motions while students are doing computer mediate instruction. From these measurements, a sensor fusion classification algorithm will be developed to provide an automated assessment of affective states, cognitive states and levels of involvement of the student. This automated assessment system will be validated using student interviews and rater observations. The system will provide detailed categorized information never before available to researchers. For the instructor, a large class could be equipped and assessed in real-time so that an instructor can appropriately focus attention to improve the learning environment or for student evaluation during instruction and for self-evaluation of instructional strategies after instruction.

Keywords: Cognition, real-time passive sensors, computer mediated instruction, gaze, body motion, facial expression, affective, student involvement.

1 Introduction

The project objective is to develop and automate a methodological technology for objectively measuring a student's affective states, cognitive states and levels of involvement during computer-mediated instruction. The system will be developed as a relatively inexpensive technology that provides teachers with just-in-time (JIT) data for evaluation of a student's behaviors in relation to learning tasks. Research will be conducted in classroom and computer laboratory settings where students participate in computer-mediated instruction (CMI).

This system will be at the technology forefront of automatic in-class student assessment which currently does not exist and once developed can be used in an augmented cognition system that enhances student learning. The system will be validated using student interviews and rater observations. Among the benefits of the assessment system is that it will provide detailed categorized information never before available to researchers that will lead to a greater understanding of how students learn in-class. The system will automatically provide a complete record of the cognitive states, affective states and levels of involvement while a student is learning

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 422–431, 2011.

curriculum through computer delivered instruction. Also, with the advance of low cost technology, a large class could be equipped and assessed in real-time so that an instructor can appropriately focus attention to improve the learning environment for individual students and small groups. This technology will advance the tools available to the teacher for student evaluation during instruction and for self-evaluation of instructional strategy.

The project's research goal is to develop and evaluate the robustness of the assessment system as a measure of individual student behavior leading to learning. Researchers will create a modified computer equipped with video camera and passive sensing devices that will record the level of involvement of each student using the computer. The video camera and passive sensing devices used will record gaze activity, facial expressions and body motions during curriculum instruction. The system will process the data through a sensor fusion classification algorithm to provide an assessment of the student's affective states, cognitive states and the level of involvement.

The automatic assessment system will be applicable to a variety of educational settings including formal and informal learning environments as well as classroom or laboratory environments. It will be effective for a variety of instructional applications including adult education, individual instruction, computer-mediated tutoring, individualized lessons, and distance education. It is anticipated the system will be applicable across age groups.

The project will develop research practices and evaluation methods for determining student involvement with educational content through computer mediation. It is anticipated that potential findings will contribute to the research knowledge base in four core areas: (1) cognitive and computer science (human-computer interactions); (2) education research and practice, (3) psychology (student affective traits) and (4) provide a validated system that can use augmented cognition techniques to improve learning. Essentially, the project will apply clinical research practices in computer science to assist educators with tools for evaluating student involvement during content instruction, with the goal of adapting instructional practice for the improvement of student achievement.

The project will address the following research questions:

1. What is the level of student involvement with curriculum?
2. To what degree is the level of student involvement correlated to performance?
3. Can computer analysis be used to provide a measure of the level of student involvement?
4. How reliable is the automated measurement?

1.1 Integration into the Learning Environment

Should this automatic method of assessment using physiological data be successful, a system can be developed to present this assessment information to the instructor in real-time or after the class. Methods of presentation could include the use of a color map of different levels of attention or involvement as shown in Figure 1. Also, a composite view of students (see Figure. 2) with an overlay (see Figure 3) will provide the instructor specific details on individual students as well as the class as a whole.

Instructors who use the automatic rating system information can monitor class performance. Also, the automatic rating system may help less experienced instructors by identifying a critical point where student intervention would improve student involvement (see Figure 4).

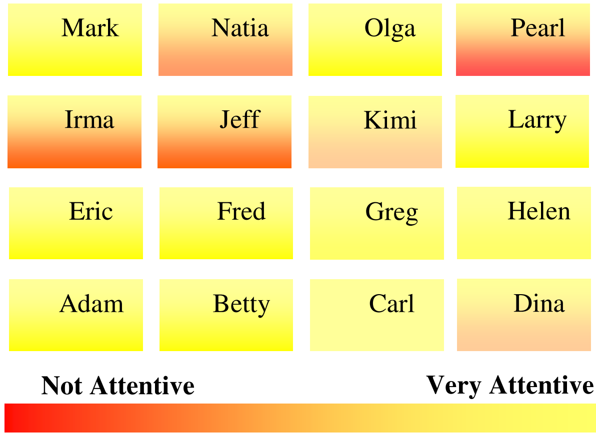


Fig. 1. A color map view of all of the students. Irma and Jeff are talking distracting Natia, while Pearl is looking out the window.

2 Background

Fred Newmann, author of Student Engagement and Achievement in American Secondary Schools [1], states that engaged students make a "psychological investment in learning. They try hard to learn what school offers. They take pride not simply in earning the formal indicators of success (grades), but in understanding the material and incorporating or internalizing it in their lives" (pp. 2–3). Every instructor knows the non-verbal expressions a student makes when involved and or disengaged. In a large class, it is nearly impossible to assess the level of involvement or disengagement for every student.

Fig. 2. A composite view of all students in the class

Chen [2] used multiple monitors in a mock classroom situation with an overlay of gesture information. A survey of instructors indicated they felt the tool could be useful if implemented in a real classroom. In this research proposal, the passive physiological sensing devices used will record gaze activity, facial expressions and body motions during computer mediated curriculum presentation. Andrassi [3] in his summary, links eye movement to problem solving, reading efficiency and the allocation of attentional resources. Eye blinks were related to affective displays of a startle response.

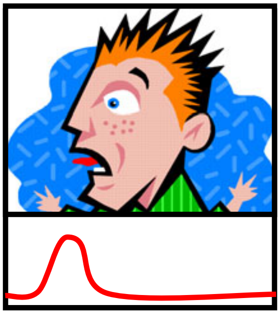


Fig. 3. Overlay graph of affective state

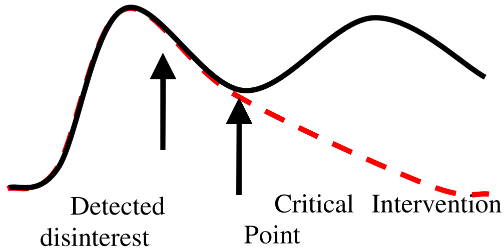


Fig. 4. Level of student interest and intervention point

Our research team has considerable experience in collecting these physiological measures [4, 5 & 6] and Ikehara and Crosby [7] used the pressures applied to a computer mouse to identify users. Crosby, Iding and Chin [8] used fixation duration and saccade to predict background complexity and number of targets.

Previous research has used facial expression and body motion measures to assess behavioral, cognitive and affective aspects of human behavior. Colmenarez, Xiong & Huang [9] developed an algorithm that can extract facial features in real-time and also gestures such as shaking and nodding. Kapoor & Picard [10] describes a vision based system that can detect head shakes and nods with a high recognition rate. Busso, Deng, Yildirim, Bulut, Lee, Kazemzadeh, Lee, Neumann, & Narayanan [11] successfully did emotion or affect recognition using video although they required facial markers. Initially data recorded from each student will be rated by experienced instructors and cross validated with the associated student's rating for items such as attention/inattention, perplexed/comprehend, useful/useless and involved/disinterested. Reliably rated items will be correlated to the student's performance history (e.g., math & science grades and standardized test scores) to obtain rated items of significance to performance.

Crosby and Ikehara successfully used classification algorithms like discriminant analysis and neural networks to identify patterns in the pressures applied to a computer mouse to identify people [8]. An automated rating system for significant items will be developed using a data fusion classification algorithm. To determine the robustness of the algorithm, to improve the automatic acquisition of data and because of the limited number of subjects, a second and third year of experiments will be needed to test and improve the algorithm that automatically classifies physiological measures into affective states, cognitive states and levels of involvement.

3 Methods

The Gantt chart in Figure 5 shows the phases of the project. Data collection is done three times with the first data collection focusing on algorithm development, the second data collection focusing on an automatic data preparation algorithm and the third data collection focusing on the integration and testing of a fully automatic system.

Development Timeframe

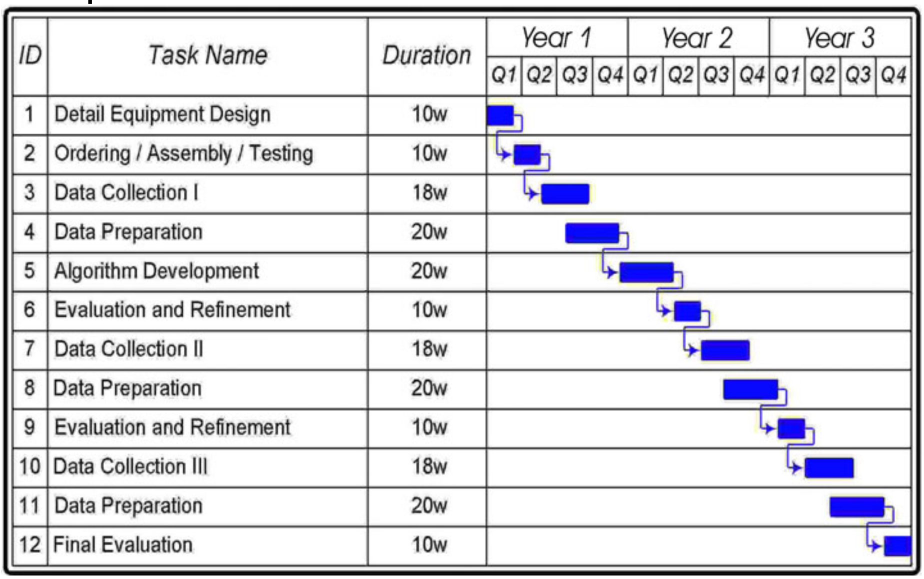


Fig. 5. The Gantt chart shows the phases of the project

3.1 Curriculum Description

The Curriculum Research & Development Group (CRDG) of the University of Hawaii has developed several curricula for teaching science, technology, engineering and mathematics (STEM) subjects that could be useful for this project, including a computer-based, self-contained algebra curriculum (X-Power Interactive) adapted from CRDG's research-based Algebra: A Process Approach textbook. The individualized instructional feature of X-Power Interactive makes the program particularly appropriate for piloting the automatic assessment system. Other curricula utilizes virtual manipulatives and simulations and will be used as appropriate for developing, piloting, and testing the automatic assessment system.

3.2 Automatic Assessment System

Detailed Equipment Design. The automatic assessment system consists of 30 student workstations. Each student station consists of one video camera, video-converter and capture, vibration isolated camera mounting and customized passive sensor positioning system. The equipment design phase will take approximately 10 weeks and include determining how data will be collected and stored and the procedures for assembly and testing.

Ordering / Assembly / Testing. Each of the 30 individual student stations will need to be ordered, assembled and tested. This should take approximately 10 weeks. Other logistics issues handled during this period include equipment handling and storage. Also, to calibrate the assessment system, a short 15 to 30 minute computer-mediated presentation will be developed. Up to 20 pilot subjects will be used to test and evaluate software, hardware and procedures.

Thirty computers in a classroom will be equipped with passive physiological sensing devices that will automatically measure the affective states, cognitive states and levels of involvement of each student. Students will be doing computer mediated instruction.

Data Collection I. Data will be collected from students taking a class during the CRDG Summer Programs designed specifically for this project. Data from passive devices will record from a representative sample of curriculum presentation, gaze activity, facial expressions and body motions during curriculum presentation. Passive devices include the video cameras and custom sensors to support the collection of body motion. After selected curriculum presentations, the researcher will select students to interview. A student will view a split screen containing the student's facial expressions and the presentation on the screen (see Figure 6). The researcher will ask the student to describe what the student was thinking and feeling at various times during the presentation in reference to expressions the student made. There will be a structured interview with a standard set of closed and open-ended questions. Also, the researcher will allow the student to interject comments and insights.

At the end of the summer session all relevant performance evaluations, homework and examinations will be collected for all students. An exit survey will be given to each student to assess how he or she felt about the subject (e.g., algebra), the educational experience (e.g., computer tutoring) and the experimental experience (e.g., monitoring and researcher questions).

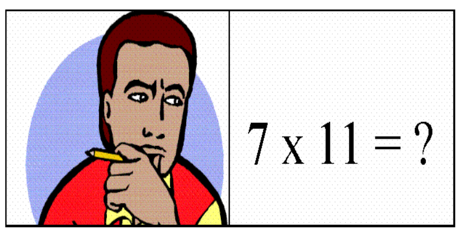


Fig. 6. A split screen view of the student musing over a problem

At least four-to-six raters with instructional experience will be recruited to view the student videos and rate the changing cognitive states, affective states and level of student involvement during the learning episodes. Raters will be trained to a common standard of rating. The large volume of recorded video will require a significant amount of time for the ratersto review. Therefore, priority will be assigned to those videos that have been previously viewed by the student and researcher. The data will be compiled for the algorithm development phase as shown in Figure 7. Gaze activity, facial expressions and body motions viewed by students and raters will generate the data of affective states, cognitive states and levels of involvement. Student performance will be used as a reference and should correlate with the level of student involvement (see Figure 8).

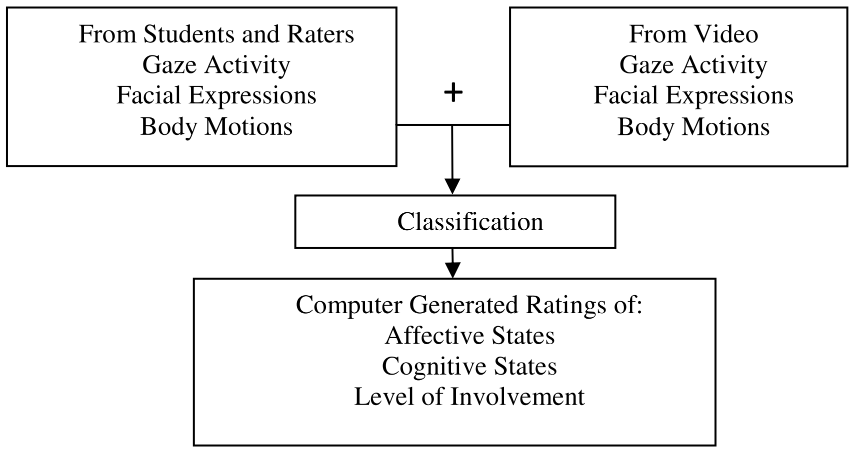


Fig. 7. Data from students and raters are used in combination with the video collected to train the rating algorithm to produce ratings highly correlated to students and raters

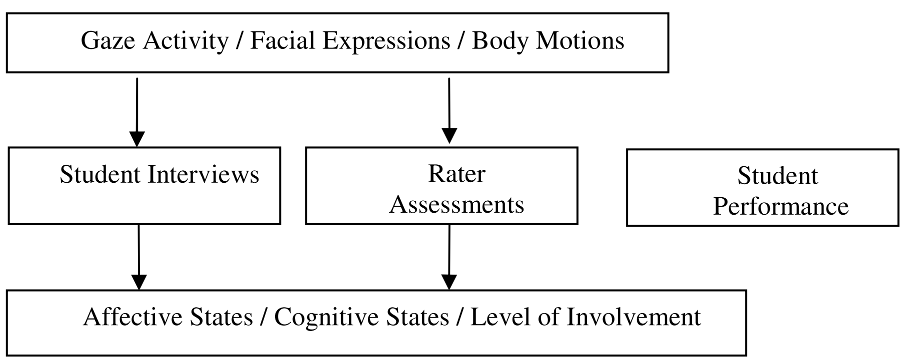


Fig. 8. Gaze activity, facial expressions and body motions viewed by students and raters will generate the data of affective states, cognitive states and levels of engagement

The student and rater data will facilitate the determination of the level of involvement with the curriculum. Also, it will be possible to use the data to determine how much student involvement is correlated to performance.

Data Preparation. Data preparation can proceed as soon as data is collected, so this phase can overlap with data collection. Data collection personnel and data preparation and algorithm development personnel will be separate team members. This will prevent the introduction of bias into the algorithm development phases of the project. This phase is expected to take 20 weeks to select a variety of exemplars of gaze activity, facial expressions and body motions that correspond to affective states, cognitive states and the level of student involvement from the entire student data set collected over the weeks.

Algorithm Development. An algorithm that uses recorded gaze activity, facial expressions and body motions to identify affective states, cognitive states and the level of student involvement will be developed using student interviews, rater assessments and student performance data. Student interviews, rater assessments and student performance data will be compiled to provide a training data set for a sensor fusion classification algorithm based on gaze activity, facial expressions and body motion. Subsequently, the algorithm (see Figure 7) will be able to produce ratings of affect states, cognitive states and student involvement comparable to students and raters.

Evaluation and Refinement. Subject data will be randomly assigned to two groups, training and test groups. The training group's data will be used to develop the data fusion classification algorithm. The test group's data will be used to evaluate the accuracy of classification. This evaluation will guide the data collection in the next phase where modifications to how the physiological data is collected (e.g., moving the position of the cameras) may be necessary to improve classification outcomes. Refinement of the algorithm will involve using both group's data to produce a refined data fusion classification algorithm to be tested with the next set of data collected.

Data Collection II. Data collection will proceed similarly to the previous data collection (see Figure 7) guided by issues discovered during the first evaluation and refinement of the algorithm. The focus of the second data collection is an automatic data preparation algorithm.

Data Preparation. Data preparation will proceed similarly to the previous data collection guided by issues discovered during the first evaluation and refinement of the algorithm. This data preparation phase will concentrate on the automatic extraction of data exemplars which is necessary for the development of a fully automatic system.

Evaluation and Refinement. Using the data fusion classification algorithm developed during the first refinement of the algorithm, the entire first year's subject data as the training group and the entire second year's subject data as the test group, an evaluation of the accuracy of the classification algorithm will be performed. Also, an evaluation of the efficacy of the automatic data preparation algorithm will be conducted. As in the previous refinement, this evaluation will determine modifications needed in the data collection of the next phase. Refinement will involve using the two years of data to produce a second refinement of the data fusion classification algorithm.

Data Collection III. Data collection will proceed similarly to the previous data collections guided by issues discovered during the previous evaluation and refinement of the algorithm and focusing on the integration and testing of a fully automatic system.

Data Preparation. Data preparation will use the automatic data preparation algorithm. It is expected that this algorithm will require improvements to accurately segment exemplars from the data reliably. Improvements to the automatic data preparation will occur during this phase.

Final Evaluation. The final evaluation will use the refined data fusion classification algorithm developed during the second refinement of the algorithm, the automatic data preparation algorithm, the past two years of subject data as the training group and the third year of subject data as the test group. Evaluation of the accuracy of the classification algorithm using the automatic data preparation algorithm will be conducted. A successful system will be able to reliably produce automatic ratings of affective states, cognitive states and levels of involvement, based solely on video and custom body motion sensors. These automatically produced rating should be highly correlated to the student and rater evaluation of affective states, cognitive states and the levels of involvement.

A method to cross check the assessment outcome is to verify if higher levels of student involvement will correlate positively with higher levels of student performance. Also, it would be interesting to determine how affective and cognitive states vary with student performance.

4 Conclusions

This project will build on current knowledge of human physiological states as indicators of the cognitive states of the individual. Several studies of the project authors have shown correlation between sensory data collected and the human-computer interactions of participants. This project will extend that knowledge to determine if student cognitive and affective states correlate to performance during computer-based instruction. Secondly, the project will determine if the system for measuring student cognitive and affective traits can be reliably automated.

The project will develop a model for collaborative partnerships consisting of a diversity of university clinical researchers and educational practitioners/researchers. As university researchers and educators focus on developing effective measurement indicators, the project will validate and then codify the results for application and extension in future studies. The study promises to add to the knowledge base in physiological psychology, curriculum development, and teaching and learning. Project results are expected to provide evidence that clinical research and evaluation methods can be used to inform instructional practice when educators and researchers collaborate on methods development and their application. The study holds the potential for providing curriculum developers of computer-mediated instruction a practical method for evaluating its effectiveness with students in a variety of circumstances. Likewise, it holds the potential for providing teachers with an effective methodology for just-in-time evaluation of student involvement in instruction with the ability to adapt instruction to individual students. Also, assuming a validated system is developed, augmented cognition techniques to improve learning can be guided by this system.

References

1. Newmann, F.M.: Student engagement and achievement in American secondary schools. Teachers College Press, New York (1992)
2. Chen, M.: Visualizing the pulse of a classroom. In: Proceedings of the eleventh ACM international conference on Multimedia, MULTIMEDIA 2003. ACM, New York (2003)

3. Andrassi, J.L.: Psychophysiology Human Behavior and Physiological Response, Lea, 4th edn (2000)
4. Crosby, M.E., Auernheimer, B., Ikehara, C.A.C.: Physiological data feedback for application in distance education. PUI (2001)
5. Vick, R.M., Ikehara, C.: A Methodological Issues of Real Time Data Acquisition from Multiple Sources of Physiological Data. In: Proceedings of the Hawaii International Conference on System Sciences, Kona, Hawaii (2003), `http://dlib2.computer.org/conferen/hicss/1874/pdf/187450129a.pdf?`
6. Ikehara, C., Crosby, M.E.: A Real-Time Cognitive Load in Educational Multimedia. In: Proceedings of the 2003 World Conference on Educational Multimedia, Hypermedia & Telecommunications, Honolulu, HI, June 2003 (2003a)
7. Ikehara, C., Crosby, M.E.: A User Identification Based on the Analysis of the Forces Applied by a User to a Computer Mouse. In: Proceedings of the Hawaii International Conference on System Sciences, Kona, Hawaii (2003b), `http://dlib2.computer.org/conferen/hicss/1874/pdf/187450130a.pdf?`
8. Crosby, M.E., Iding, M.K., Chin, D.N.: Visual search and background complexity: Does the forest hide the trees? In: Bauer, M., Gmytrasiewicz, P.J., Vassileva, J. (eds.) UM 2001. LNCS (LNAI), vol. 2109, p. 225. Springer, Heidelberg (2001)
9. Colmenarez, A.J., Xiong, H.Z.T.-S.: Facial Analysis from Continuous Video with Applications to Human-Computer Interface, 1st edn. International Series on Biometrics. Springer, Heidelberg (2004)
10. Kapoor, A., Picard, R.W.: A Real-Time Head Nod and Shake Detector. In: Proceedings of the 2001 PUI, Orlando FL (2001)
11. Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C.M., Kazemzadeh, A., Lee, S., Neumann, U., Narayanan, S.: Analysis of Emotion Recognition using Facial Expressions, Speech and Multimodal Information. In: ICMI 2004 (October 2004)

Considering Cognitive Traits of University Students with Dyslexia in the Context of a Learning Management System

Carolina Mejía[1], Alicia Díaz[2], Juan E. Jiménez[2], and Ramón Fabregat[1]

[1] Institute of Informatics and Applications, University of Girona, 17070 Girona, Spain
{carolina.mejia,ramon.fabregat}@udg.edu
[2] Department of Psychology and Education, University of La Laguna, 38207 Canary Islands, Spain
{adiazm,ejimenez}@ull.es

Abstract. This paper studies the cognitive processes involved in reading among Spanish-speaking university students with dyslexia, and proposes to evaluate these processes to identify specific cognitive traits. On this basis, an automated battery for the assessment of cognitive processes was designed to be included in a learning management system (LMS). To integrate this battery into the LMS, a web service architecture that works independently of the LMS was designed. The assessment battery has been built based on a multimodal communication mechanism that delivers evaluation tasks using the visual, auditory, and speech communication channels of human-computer interaction.

Keywords: Dyslexia, cognitive traits, user model, university students, multimodal communication.

1 Introduction

The use of learning management systems (LMS) to support the teaching-learning process is becoming increasingly important. Therefore, it has become necessary to consider particular student characteristics, such as learning disabilities (LD), within the context of these systems. Dyslexia is a very common LD in education. It requires that teachers pay special attention and provide suitable resources to intervene with and assist affected students during their learning process.

Several studies provide examples of the work that has been done with dyslexia in children: identifying populations of children with dyslexia, evaluating their cognitive processes to determine specific deficits, and creating intervention programs to address the learning deficits presented [1] [2] [3]. Many of those programs have been supported by information and communication technologies (for example, software) that tend to increase student motivation and personalize the learning process [4] [5] [6].

Our research is focused on university students with dyslexia: a population that has been studied very little according to [7] [8] [9]. In the past two decades research in this area has shown that LDs persist into adulthood [10] [11] [12]. For this reason, it

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 432–441, 2011.

has become necessary to study the cognitive processes that can be altered in university students and to identify the particular cognitive traits of each student and how these deficits can be treated in this population.

According to [9], appropriate tools in Spanish cannot be found to assess these cognitive processes in the population of adult dyslexics. As a result, the research conducted in [13] consisted of the Spanish adaptation of an instrument, the UGA Phonological/Orthographic Battery developed at the University of Georgia [14]. Using as references that work and the analysis made in [15] of instruments to identify LDs, we have designed and built a new battery to assess all the cognitive processes involved in reading, making use of web-based technology so that it can be integrated into a LMS using web services.

Once the battery has been integrated into an LMS, the variables corresponding to the student cognitive traits are stored in a user data model. That data model allows us to represent the information of a previously designed user model [15]. This user model is formed by four submodels: personal profile, learning style, cognitive traits, and cognitive performance. In this paper we focus on the cognitive trait submodel, designed to store the results of the evaluation of each cognitive process (phonological awareness, orthographic processing, lexical access, processing speed, verbal working memory, and semantic processing) in the LMS. Based on this user model information, we designed an architecture with some adaptation mechanisms for intervention and assistance tasks in an LMS for students that present some type of cognitive deficit, to improve their cognitive and academic performance, and to personalize the learning process according to the specific cognitive traits of each one of them.

This paper is structured as follows. In the second section we explain what is generally considered to be dyslexia (or reading disabilities) and the cognitive processes involved in reading. The third section proposes an assessment battery for those cognitive processes. In the fourth section a user model based on student cognitive trait information is presented. Finally, in the fifth section we draw some conclusions and enumerate proposals for future work.

2 Dyslexia: Learning Disabilities in Reading

Reading is considered the basis of the educational process since most of the knowledge transmitted during academic development relies on the written language. That is why, from the very first years of schooling, learning to read correctly is considered a basic tool for academic development. Furthermore, when we refer to reading as the basis of the educational process, we mean it not only in terms of academia but also the importance it has in a general sense. The way we access most of the information in our environment is also connected with written language because we are immersed in the so-called information society, where activities (including productive, economic, educational, and cultural ones) are regulated through communication and information. And learning to read correctly is essential for the development of the individual in this society. When students have difficulty acquiring this skill, their academic performance and general personal development are affected. These consequences make it necessary to study the reading disabilities (RD) also known as dyslexia.

The most accepted definition of the term dyslexia was proposed by [16]: "Dyslexia is a specific learning disability that is neurobiological in origin. It is characterized by difficulties with security and/or fluent word recognition and by poor spelling and decoding abilities. These difficulties typically result from a deficit in the phonological component of language that often is unexpected in relation to other cognitive abilities and the provision of effective classroom instruction. Secondary consequences may include problems in reading comprehension and reduced reading experience that can impede growth of vocabulary and background knowledge." According to this definition, dyslexia is an LD that may pose a number of difficulties in the various processes involved in reading.

The acquisition and development of reading depends on two types of factors: external and internal. External factors refer to the presence of a tutor/teacher who conducts sequential reading instruction; unlike oral language, reading requires formal teaching by an instructor. Internal factors refer to the development of certain cognitive processes that facilitate reading. These processes include phonological awareness, orthographic processing, lexical access, processing speed, semantic processing, and working memory.

- *Phonological awareness* is the ability to separate the units into which speech can be divided: the phonemes or sounds that make up the words. This is a major deficit in dyslexia and is characterized by difficulty in acquiring, consolidating, and automating phonological processes [17].
- *Orthographic processing* involves recognizing the word as an orthographic pattern and retrieving its pronunciation from memory (via the visual route). Although research in this process has received less attention than phonological processing [13], it is important to note that people with dyslexia present a deficit in orthographic processing [18], probably due to a deficit in phonological processing [12] [19].
- *Lexical access* is the process involved in obtaining the meaning of written words. This can occur over two routes [20]: one that directly connects graphic signs with meaning (visual route) and another that transforms the graphic signs into their corresponding sounds and uses those sounds to access the meaning (phonological route). This process is essential for proper reading performance and its impairment is considered a major deficit in dyslexia [21].
- *Processing speed* refers to the speed in which stimuli are processed. Slowness in naming familiar visual stimuli may be related to dyslexia [22] [23]. When a person reads a series of processes similar to those carried out in tasks measuring processing speed (attention to the stimulus, visual processes, access and retrieval of phonological labels, activation and integration of semantic information, etc.) are required.
- *Working memory* is the ability to temporarily retain information in memory, work with it or operate on it, and produce a result. Working memory is important in reading because readers have to decode and recognize words as they remember the meaning of what they have read. It has been suggested that the underlying deficit in dyslexia is in verbal working memory and that that can be attributed to difficulties accessing or using phonological structures [24].

- *Semantic processing* refers to understanding and interpreting written information. This processing involves the extraction of meaning from text and the integration of information in memory. This process involves readers' background knowledge about what they are reading (a text), which will facilitate a mental representation of the entities evoked by the text [25].

All these processes are essential for reading comprehension to be successful.

3 Assessment Battery for Cognitive Processes

For students with dyslexia, conducting a proper diagnosis, and understanding what their real deficits are requires a thorough analysis of their problem. Tests to detect deficits must be administered and the results studied to establish the foundations upon which different learning adaptations can be based to achieve personalized learning. To enhance the learning process, it is important to identify students' cognitive traits. We focus on detecting cognitive traits associated with dyslexia, and take into account the failure of specific cognitive processes involved in reading to design and build an assessment battery that detects deficits in the cognitive processes mentioned in the previous section.

The battery involves tasks to assess phonological awareness, orthographic processing, lexical access, processing speed, verbal working memory, and semantic processing, all of which are necessary to identify dyslexia in university students. Our purpose is to create a complete battery that assesses all cognitive processes involved in reading using web-based technology so that it can be integrated into an LMS using web services.

The battery has a modular design (see Fig. 1) to facilitate interaction between the different modules. For each type of battery user a different interface is presented depending on the permissions and tasks that can be developed. Figure 1 presents the architecture of the battery illustrating the components and their relationships. The components are: 1) *assessment modules*, 2) *management modules*, 3) a *web server* that stores the modules and allows communication between users and the battery by means of a browser, and 4) a *database* where the data from the users, results, history, etc. can be stored.

The battery has eight modules, each one designed with functions for each user type. Since the battery is a software tool designed to be used in the university context, we identified three types of users in this context: *Experts*, or users responsible for performing activities related to the creation of tasks, the evaluation of each cognitive process, the definition of the guidelines to present the results of students and teachers, the provision of recommendations that teachers could follow for each student with cognitive deficits, and the checking of student results; *Teachers*, or users responsible for scheduling and activating the battery in their classes, checking the results report of the students, and viewing the recommendations given by experts for each student with cognitive deficits; and *Students*, or users that complete the battery evaluation tasks (activated by the teachers) and check their results report.

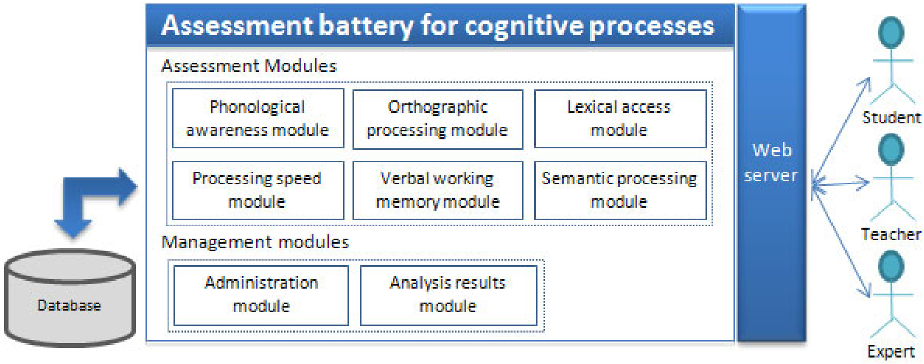


Fig. 1. Architecture of the assessment battery for cognitive processes

The assessment modules (see Fig. 1) are independent modules designed to bring together the different assessment tasks for each cognitive process, as shown in Table 1. The selection of the tasks used to assess each cognitive process is based on research work carried out by some of the authors of this paper [26] [13] [6].

Table 1. Assessment tasks for each cognitive process

Modules	Tasks
Phonological awareness	• Segmentation into syllables • Number of syllables • Segmentation into phonemes • General rhyme • Specific rhyme • Phonemic location • Omission of phonemes
Orthographic processing	• Homophone/pseudohomophone choice • Orthographic choice
Lexical access	• Reading words and pseudowords
Processing speed	• Visual speed test
Verbal working memory	• Verbal working memory test
Semantic processing	• Reading expository and narrative texts

To implement the tasks shown in Table 1 we rely on a multimodal architecture [27] that allows the student to communicate with the battery through different modes according to the specific objective of each assessment task. The tool uses modes of interaction for inputs and outputs that allow the combined use of spoken and written language and other devices like the keyboard and the mouse. Figure 2 depicts the channel alternatives for communication between the student and the battery.

For student information input, the battery includes an automatic component of speech recognition that converts human speech into syllables or individual words, insertion of written words and characters for specific commands and use of the mouse device. The battery gives students instructional output information or guidance (data output) using output mechanisms such as text on screen, graphical representation, recorded audio, and synthesized voice.

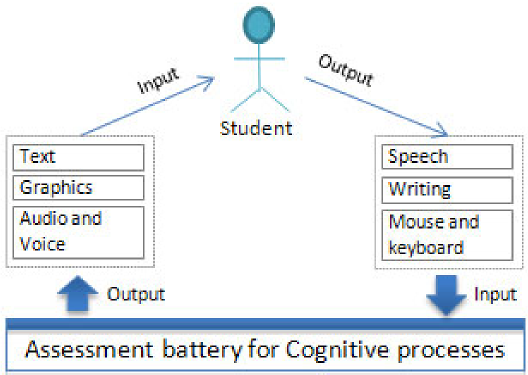


Fig. 2. Multimodal communication input and ouput

Management modules (see Fig. 1) are designed to facilitate the tool's administration and use and the analysis of results by specialists (user type: *experts*). The administration module is designed and implemented for the exclusive use of a subject matter expert (psychologist, pedagogue, or counselor); this module allows the creation and/or edition of different assessment tasks needed to identify cognitive deficits in students. The analysis results module is created to design and deliver the results report of the students after they have completed the assessment tasks. It automatically generates an individual report for each student with: 1) the overall cognitive performance, 2) the identification of specific cognitive deficits present, 3) the diagnosis of the presence or absence of dyslexia, and 4) some recommendations to follow in each particular case. This module can be accessed by experts, teachers, and students: experts who are responsible for the content of the reports to be delivered, teachers who wish to know student results and recommendations for each case, and students that want to see their personal result report.

The battery has been designed considering the use of standard technology and characteristics of reusability, interoperability, accessibility, and extensibility, to facilitate its integration into the structure of an LMS.

4 Student's Cognitive Traits Model

The user model we propose corresponds to information related to students' cognitive traits that can be stored and used in the LMS. This model is one of the four submodels of a main user model presented in [15]. The cognitive trait model identifies variables related with each aforementioned cognitive processes, allowing us to represent information about the student's LD. Descriptions of some of the elements of this model can be matched and related with the guidelines of the IMS Accessibility for Learner Information Profile Specification (IMS-AccLIP) [28].

The identified variables allow student information corresponding to each cognitive process assessed with the battery to be stored in the LMS (see Fig. 3). These variables have assigned the results of each of the assessment exercises from the tasks described in Table 1 and let to know whether or not students have a cognitive deficit. These results include the time students take to solve each exercise, right and wrong answers and other information particular to each task.

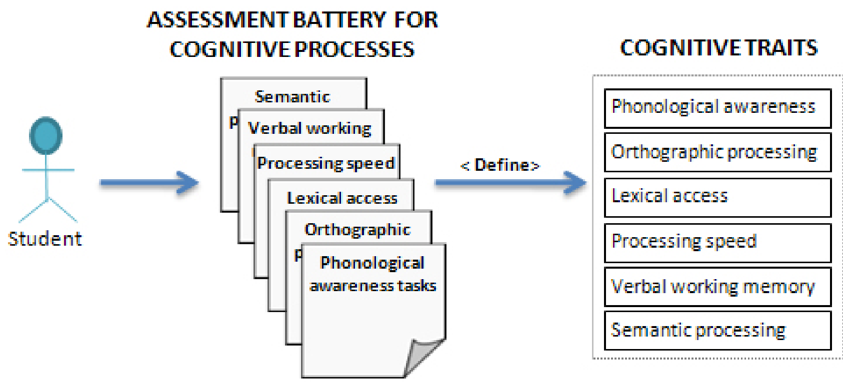


Fig. 3. Cognitive traits model

Whether or not a student has a specific cognitive deficit can be determined by calculating the scales of each assessment task performed by that student. To carry this out, a representative population of university students from different academic programs and levels will be convened to complete all the assessment tasks in the battery. Then, the results of the university student sample will be used to calculate the means, standard deviations, and percentiles, and a table of scales will be defined to identify how successfully each task confirms the presence or absence of a cognitive deficit.

Based on this information from the user model, we designed an architecture with adaptation mechanisms to intervene with and assist, through an LMS, students that present some type of cognitive deficit, in order to improve their cognitive and their academic performance. The proposed adaptation mechanisms for intervention tasks are based on adaptive methods for visual, auditory, and speech multimodal communication channels. Many studies regarding the training of cognitive processes [29] [30] [31] [32] indicate that student performance can be improved significantly if intervention tasks use both acoustic and visual modalities. Moreover, the use of assistive technology (for example, speech recognition systems, screen readers, talking spell checkers, and proofreading programs) [33] [34] is considered for personalized assistance tasks.

In summary, the design of the architecture considers: 1) a user model formed by the four previously mentioned submodels (personal profile, learning style, cognitive traits, and cognitive performance), 2) some accessibility guidelines for students with dyslexia [35], 3) e-learning standards, especially those related to the storage of the user model information [36], and 4) an adaptation engine that handles the delivery of contents and tasks adapted to the particular needs and preferences of students with dyslexia. The adaptation engine has adaptation rules that can be applied to personalize the intervention and assistance tasks that students complete to improve their cognitive performance. These rules will be evaluated considering the type and difficulty level (easy, medium and hard) of the task as well as the appropriate support tools, depending on the specific deficit of each student.

5 Conclusion and Future Work

This work aims to provide tools in the context of an LMS that support the learning process of university students with dyslexia. A battery for the assessment of cognitive processes involved in reading has been designed and implemented. This battery provides teachers and specialists with guidance on intervention and assistance for students with deficits that limit their learning performance.

The battery has a modular design to facilitate communication between the modules and interaction between users and the tool. Moreover, the battery's tasks are supported by a multimodal architecture that allows students to communicate with the battery through different modes (visual, auditory, and speech) according to the specific objective of each assessment task.

In future work the battery will allow us to retrieve the scales of each assessment task completed by university students. With those scales the battery will automatically identify whether or not the student presents a cognitive deficit. Furthermore, we will design and develop adaptation mechanisms in an LMS to provide suitable resources to intervene with and assist affected students during their learning process.

Acknowledgments. The authors would like to thank the Spanish Ministry of Science and Innovation for financial support through the A2UN@ project (TIN2008-06862-C04-02/TSI). Thanks to the Scholarship Programme of the University of Girona, reference BR08/09.

References

1. Nicolson, R.I., Fawcett, A.: Automaticity: A new framework for dyslexia research? Cognition 35, 159–182 (1990)
2. Metsala, J.L.: The development of phonemic awareness in reading disabled children. Applied Psycholinguistics 20, 149–158 (1999)
3. Guzmán, R., Jiménez, J.E., Ortiz, M.R., Hernández-Valle, I., et al.: Evaluación de la velocidad de nombrar en las dificultades de aprendizaje de lectura. Psicothema 16, 442–447 (2004)
4. Wise, B.W., Olson, R.K.: Computer-based phonological awareness and reading instruction. Annals of Dyslexia 45, 99–122 (1995)
5. Barker, T.A., Torgesen, J.K.: An evaluation of computer-assisted instruction in phonological awareness with below average readers. J. Edu. Comp. Res. 13, 89–103 (1995)
6. Rojas, R.E.: Diseño y validación de un videojuego para el tratamiento de la dislexia. PhD thesis, University of La Laguna, p. 606 (2008)

7. Gregg, N.: Underserved and Unprepared: Postsecondary Learning Disabilities. Learning Disabilities Research and Practice 22, 219–228 (2007)
8. Sparks, R.L., Lovett, B.J.: College students with learning disability diagnoses: Who are they and how do they perform? J. Learn. Disab. 42(6), 494–510 (2009)
9. Jiménez, J.E., Gregg, N., Díaz, A.: Evaluación de habilidades fonológicas y ortográficas en adolescentes con dislexia y adolescentes buenos lectores. Infancia y Aprendizaje 27(1), 63–84 (2004)
10. Finucci, J.M., Gottfredson, L.S., Childs, B.: A follow-up study of dyslexic boys. Annals of Dyslexia 35, 117–136 (1986)
11. Johnson, D.J., Blalock, J.: Young Adults with Learning Disabilities. Grune & Stratton, Orlando (1987)
12. Bruck, M.: Word recognition and component phonological processing skills of adults with childhood diagnosis of dyslexia. Developmental Review 13, 258–268 (1993)
13. Díaz, A.: Perfiles cognitivos y académicos en adolescentes con dificultades de aprendizaje con y sin trastorno por déficit de atención asociado a hiperactividad, PhD thesis, University of La Laguna (2007)
14. Gregg, N., Coleman, C., Stennett, R., Davis, M., et al.: Sublexical and lexical processing of young adults with learning disabilities and attention deficit/hyperactivity disorder. In: Witruk, E., Friederici, A.D., Lachmann, T. (eds.) Basic Functions of Language, Reading and Reading Disability, pp. 329–358. Kluwer Academic Publishers, London (2002)
15. Mejía, C., Fabregat, R., Marzo, J.L.: Including Student's Learning Difficulties in the User Model of a Learning Management System. In: XXXVI Conferencia Latinoamericana de Informática (CLEI 2010), Asunción, Paraguay (2010)
16. Lyon, G.R., Shaywitz, S.E., Shaywitz, B.A.: A definition of dyslexia. Annals of Dyslexia 53, 1–14 (2003)
17. Jiménez, J.E.: A reading-level match study of phonemic processes underlying reading disabilities in a transparent orthography. Reading and Writing: An Interdisciplinary Journal 9, 23–40 (1997)
18. Farmer, M., Klein, R.: The evidence for a temporal processing deficit linked to dyslexia: a review. Psychonomic Bulletin and Review 2, 460–493 (1995)
19. Share, D.C., Stanovich, K.E.: Cognitive processes in early reading development: A model of acquisition and individual differences. Issues in Education: Contributions from Educational Psychology 1, 1–7 (1995)
20. Coltheart, M., Rastle, K.: Serial processing in reading aloud: Evidence for dual-route models of reading. J. Exper Psych.: Hum. Perc. Perf. 20(6), 1197–1211 (1994)
21. Jiménez, J.E., Hernández-Valle, I.: Word identification and reading disorders in the Spanish language. J. Learn. Disab. 32, 267–275 (2000)
22. Fawcett, A.J., Nicolson, R.I.: Naming speed in children with dyslexia. J. Learn. Disab. 27, 641–646 (1994)
23. Wimmer, H., Mayringer, H., Landerl, K.: The double-deficit hypothesis and difficulties in learning to read a regular orthography. J. Educ. Psych. 92(4), 668–680 (2000)
24. Bar-Shalom, E.G., Crain, S., Shankweiler, D.: A comparison of comprehension and production in good and poor readers. Applied Psycholinguistics 14, 197–227 (1993)
25. Fayol, M.: A propos de la compréhension. In: ONL (ed.): Regards sur la lecture et ses apprentissages, pp. 85–102. Montluçon (1995).
26. Jiménez, J.E., Antón, L., Díaz, A., Estévez, A., et al.: Sicole-R: Un sistema de evaluación de los procesos cognitivos en la dislexia mediante ayuda asistida a través del ordenador [Software]. University of La Laguna (2007)

27. W3C Multimodal Interaction Framework, W3C NOTE (May 06, 2003), http://www.w3.org/TR/2003/NOTE-mmi-framework-20030506/
28. IMS Learner Information Package Accessibility for LIP, Version 1.0 Final Specification, IMS Global Learning Consortium, Inc. (2003), http://web4all.atrc.utoronto.ca/IMS_docs/IMS%20ACCLIP%20documents/imsacclip_usecasesv1p0.pdf
29. Mayer, R.E., Moreno, R.: A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. J. Educ. Psych. 90, 312–320 (1998)
30. Brünken, R., Steinbacher, S., Plass, J.L., Leutner, D.: Assessment of cognitive load in multimedia learning using dual-task methodology. Experimental Psychology 49(2), 109–119 (2002)
31. Mayer, R.E., Fennell, S., Farmer, L., Campbell, J.: A personalization effect in multimedia learning: Students learn better when words are in conversational style rather than formal style. J. Educ. Psych. 96, 389–395 (2004)
32. Holzinger, A., Kickmeier-Rust, M., Wassertheurer, S., Hessinger, M.: Learning Performance with Interactive Simulations in Medical education: Lessons learned from results of learning complex physiological models with the HAEMOdynamics SIMulator. Computers & Education 52(1), 292–301 (2009)
33. Lancaster, P.E., Schumaker, J.B., Deshler, D.D.: The development and validation of an interactive hypermedia program for the teaching a self-advocacy strategy to students with disabilities. Learning Disability Quarterly 25, 277–302 (2002)
34. Mejia, C., Fabregat, R.: Towards a Learning Management System that Supports Learning Difficulties of the Students. In: XI Simposio Nacional de Tecnologías de la Información y las Comunicaciones en la Educación (CEDI 2010), Valencia, Spain (2010)
35. WAI Web Accessibility Initiative, W3C, http://www.w3.org/WAI/intro/people-use-web.php
36. IMS Global Learning Consortium, http://www.imsglobal.org/specifications.html

Improving Students' Meta-cognitive Skills within Intelligent Educational Systems: A Review

Alejandro Peña[1,2,3], Michiko Kayashima[4],
Riichiro Mizoguchi[3], and Rafael Dominguez[2]

[1] WOLNM
[2] ESIME-Z & CIC -National Polytechnic Institute
31 Julio 1859 # 1099-B, Leyes Reforma, DF, 09310, Mexico
[3] Institute of Scientific and Industrial Research, Osaka University
8-1 Mihogaoka, Ibaraki, Osaka, 567-0047 Japan
[4] College of Humanities, Tamagawa University, Japan,
6-1-1 Tamagawagakuen, Machida, Tokyo, 194-8610 Japan
apenaa@ipn.mx, kayasima@lit.tamagawa.ac.jp,
miz@ei.sanken.osaka-u.ac.jp

Abstract. Metacognition aims at monitoring and regulating one's thinking devoted to problem-solving processes and learning habits among others cognitive tasks. Hence, individuals engaged in better acquisition of domain knowledge achieve higher scores when they are bewaring of how to exploit their metacognitive faculties. Thus, we present a review of some models and methods with the purpose to understand what metacognition is and know how stimulate metacognitive skills. In addition, we propose a Metacognition-Driven Learning paradigm as a reference to guide the design of Intelligent Educational Systems oriented to improve students' metacognitive skills.

Keywords: Metacognition, metacognitive skills, metacognitive models, Metacognition-Driven Learning, Intelligent Educational Systems.

1 Introduction

John Flavell of Stanford University was a pioneer researcher of the metacognition field. He coined the metacognition term and set the earliest formal model for metacognition. Flavell's publications show the influence of the work achieved by Jean Piaget [1]. Such a work accounts the notion of *intentionality* to presuppose: "A kind of deliberative and goal-driven thinking that plans a sequence of actions". Later on, Flavell states the *metamemory* term to label: "The individual's skill to manage and monitor the input, storage, search and retrieval of mental contents of her own memory" [2]. Soon afterwards, Flavells defines the metacognition concept as: "In any kind of cognitive transaction with the human and the non-human environment, a variety of information processing activities may go on…" [3].

Flavell also claims: "Metacognition concerns to the active monitoring, regulation and orchestration of information processes in relation to cognitive objects on which

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 442–451, 2011.

they bear". In consequence, metacognition is intentional, foresighted, conscious, purposeful and devoted to fulfill a goal. However, other viewpoints argue that metacognition does not necessary evoke awareness [4, 5].

Metacognition is a cognitive faculty of human beings, whose activity is driven by a sort of metacognitive skills, such as: reflection [6], self-awareness [7], self-monitoring [8], self-regulation [9], self-assessment [10], self-management [11] and so on. They organize and supervise cognitive activities that an individual performs. Metacognitive skills hold knowledge, strategies, experience, goals and tasks [12]. They enable individual to set goals, make plans, initiate tasks, monitor the development of mental activities, estimate the likely achievement of goals, detect deviation on tasks, correct cognitive processes, and keeping track the effect of one's behavior on others.

Metacognition skills are independent of subject-domain. Although an individual holds little background, once she masters a skill, she is able to apply it across domain. Some people naturally develop metacognitive skills, but others need external advice [13]. Because metacognition plays a critical role in successful learning, it is necessary to stimulate students' metacognitive skills. So Intelligent Educational Systems (IES) should pursue a twofold objective: engage learners to be aware of their own metacognitive activity and demonstrate how students can be taught to better apply their cognitive resources by the use of metacognitive skills.

Therefore, in order to provide a useful reference to tailor effective ELS the remainder of the paper embraces the following subjects: A survey of formal models for metacognition and a review of approaches for stimulating metacognitive skills are depicted respectively in sections 2 and 3. Afterwards, we present our Metacognition-Driven Learning paradigm that holds a model and a method. Hence, a hierarchical-loop metacognitive model is outlined in section 4; whereas, a three-stage method is set in section 5. Conclusions section is devoted to describe some contributions of the proposal and identify the future work to be accomplished.

2 A Sample of Metacognitive Models

In this section we try to respond the question: How can we shape metacognition? Wherefore, by a profile of representative models we try to describe metacognition. We focus on the items, relationships, hierarchies and flows of cognitive information to identify involved components and understand their role as follows.

2.1 Essentials Metacognitive Phenomena

The "Formal Model for Metacognitive Monitoring" proposed by Favell in 1979 is a starting point to depict metacognition [12]. He claimed that: "Knowledge, experience, goals-tasks and strategies are four types of phenomena that hold some kind of relationship to support metacognitive activity" [14].

Metacognitive knowledge is a sort of phenomenon that contains individual's knowledge or beliefs about the factors that bias cognitive activities. It embraces three sorts of variables: person, task and strategy. The first variable concerns with the individual's knowledge and beliefs about himself as a thinker or learner, and what she believes about other people's thinking processes. It is split into three versions: intra-individual (i.e. assertions about the interests, propensities, aptitudes, abilities of oneself), inter-individual (i.e. comparisons between people in a relativistic manner), universal (i.e. generalizations an individual sets about learning and learners in general). The task variable heads the person in the management of a task, and provides expectations about the success that it is likely to meet. The strategy variable represents goals and criteria for selecting cognitive processes to apply in their fulfillment.

The second class of phenomena, metacognitive experiences, is a cognitive awareness that is relevant to one's thinking processes. Such a class is a stream of consciousness process in which other information, memories, or earlier experiences may be recalled as resources in the process of solving a current cognitive problem.

The metacognitive goals and tasks compound the third class of phenomena to depict the desired outcomes or objectives of a cognitive activity. It cares about comprehension, committing facts to memory, producing something, improving one's knowledge about something. The emergence in the child of awareness of the flow of time, and awareness of future time could support the ability to form metacognitive goals.

The fourth class of phenomenon corresponds to metacognitive strategies. They are ordered processes used to control one's own cognitive activities and to ensure the accomplishment of a cognitive goal. They enable a person to oversee her own learning process, plan and monitor ongoing cognitive activities, and to compare cognitive outcomes with internal or external standards.

2.2 Metacognitive Model Based on Knowledge and Regulation

Essentially, Brown acknowledges two cognitive components of metacognition: knowledge and regulation [15]. He shapes a model, where knowledge of cognition depicts activities that involve conscious reflection on ones cognitive skills and activities, and regulation of cognition concerns with self-regulatory mechanisms during an ongoing attempt to learn or solve problems. Both components are closely related, each feeding on the other recursively, although they can be easily distinguishable as follows.

Knowledge of cognition represents stable and frequently fallible information that an individual holds about her own cognitive processes. It aims individual retreats and contemplates her cognitive activities as target of thought and reflection, usually referred to as “knowing that”.

Regulation of cognition is devoted to regulate and monitor cognitive performance. It embraces a sort of processes oriented to: plan activities, select strategies, predict outcomes, undertake the solution of a problem, oversee cognitive activities, test results, qualify outcomes against criteria of efficiency and effectiveness, and reconsider alternative courses of action. These activities are relatively unstable, not necessary statable and relatively age independent (i.e. task and situation dependent) [16].

2.3 Hierarchical Metacognitive Models

There are other models that organize cognitive processes into several interrelated levels with the purpose to distinguish metacognition from cognition. One of them corresponds to the “Meta-level/object-level” model outlined by Nelson and Narens [17]. At meta-level, regulation skill is carried out to modulate cognitive activity that occurs at object-level. Cognitive activity undertakes the achievement of a mental goal or takes over the individual-external world interaction. It is accomplished at object-level. In addition, meta-level holds a cognitive model of the object-level. Such a model is structured according to specific metacognitive principles. Moreover, the meta-level monitors the object-level by bottom-up information (e.g., source monitoring in memory retrieval, error detection) that comes from the object-level. In consequence, meta-level controls the object-level by top-down information flows (e.g., resource allocation, error correction, planning, inhibitory control, conflict resolution) in order to initiate, modify or terminate cognitive actions.

Another double-level model is the "Executive Function" model set by Norman and Shallice [18]. The executive function involves the ability to monitor and control the information processing necessary to produce voluntary action. The executive function model holds an executive system at top level and a set of schemas at bottom level. The executive system contains a model of the perceptual and cognitive functions existent at the bottom level. Schemas are basic units of thought and action. They can be exogenously activated by environmental cues (e.g. automatic processes) and endogenously triggered or inhibited by input from the executive system (e.g. voluntary process). Thus, schema selection depends on perceptual information (i.e. a sensorial bottom-up flow) and attentional modulation (i.e. a control top-down flow).

A four-tier model for metacognition is the one proposed by Tobias and Everson [19]. They claim that: "Metacognition is the ability to monitor, evaluate, and make plans for one's learning". Thereby, they design a hierarchical metacognitive model to take over learning. The model is organized into four ascending dependency levels. At the bottom appears the knowledge monitoring component. It is the ability of an individual for knowing what she knows and knowing what she does not know. At second layer the evaluation learning item is found. It holds criteria for determining the degree of satisfaction achieved according to former expectancies. Selection of strategies is an element stated at third level. It represents the attempt to set or adjust the course of action according to some guidelines. At the top of the hierarchy, the planning component is allocated. It defines the path of actions to be accomplished under the strategy's criteria to carry out the pending learning goals.

The "Hierarchy Model of Skills" tailored by Kayshima and Inaba highlights a semantic net of skills to define common and particular attributes [20]. They set at the top the skill class to associate a sort of attributes to generalize the properties of any skill. At second level, two kinds of skills are found; one corresponds to motor skills (i.e. perceptual-motor skill such as type a keyboard) and the other represents cognitive skills (i.e. a skill achieved in the mind). This is split into basic cognitive skills (i.e. they fulfill cognitive activities to meet a goal. Their target is found at the outside-world of the individual) and metacognitive skills (i.e. they take over the performance of basic cognitive skills. The inner-world of the person is their target).

3 Two Sorts of Metacognitive Approaches

Once metacognition has been characterized by the collection of models earlier stated, we wonder: How can we stimulate metacognition? The answer to such a question is given by the exposition of two kinds of metacognitive approaches. One is devoted to depict metacognitive methods oriented to develop individuals' metacognitive skills. Another introduces some IES applications that stimulate metacognitive skills.

3.1 Metacognitive Methods

"ASK to THINK – TEL WHY" is a method to stimulate self-regulation skill [21]. It is a peer tutoring method that structures interaction between tutor-tutee roles, where peers exchange roles. The interaction follows a question-asking process. When an individual plays the tutor role, she only asks five types of questions to acquire and organize domain knowledge. Individuals who perform the learner role only provide answers. Thus, tutor gradually acquires the ability to make suitable questions to learners. These experiences are cognized by tutor to improve her self-regulation skill.

The "Whole-Class Discussions" is a method oriented to promote reflection skill [22]. The method encourages the discussion of a given problem between learners by the moderation of the tutor. Tutor does not show the solution to learners neither leads them to find it. She monitors the learners' discussion as input to regulate and provides suggestions to learners to regulate their own thinking. Thus, learners act to reflect their intention to communicate about their reasoning.

Other metacognitive methods are: "Kitchen Sink" set by Schonfeld [23], "Supporting Learners to Develop their Self-regulation Skill" designed by Kayashima and Inaba [24], and "Reciprocal Teaching" proposed by Palinscar and Brown [25].

3.2 Intelligent Educational Systems Oriented to Metacognition

In the IES arena there is tendency to build applications devoted to exercise students' metacognitive skills, such as scaffolding, tools to recreate learning environments and functionalities in intelligent tutoring systems (ITS). A sample of them is given next.

Ecolab is software scaffolding that aims students to improve their seeking and task selection skills [26]. Story-station is an agent-based scaffolding tool to stimulate metacognitive skills involved at writing. It uses animated agents to advice student about issues of writing [27]. Se-Coach is a ITS that provides several levels of prompting and scaffolding to progressively improve students' self-explanation skill [28].

Betty's Brain is a multi-agent environment that heads student to learn by teaching. It aims student to reflect about the domain knowledge provided to a learner agent, Betty, by evaluating its responses and explanations to quizzes and queries [29]. Collect-UML is a ITS that stimulates students to acquire collaborative skills. It accounts a collaboration model represented as a set of meta-constraints [30]. Plan Externalization is a workgroup tool that facilities the design of a problem-solving process. Where a solver sketches her plan and others monitor and question the solution [24].

4 The Model of the Metacognition-Driven Learning

We propose a Metacognition-Driven Learning paradigm to guide the design of IES. It embraces a model and a method. The former is a "hierarchical-loop metacognitive" model; whereas the later is a workflow composed by three stages. In this section, we describe the three levels of the model and a sequence of class activities.

4.1 A Hierarchical-Loop Metacognitive Model

Based on the formal metacognitive models prior introduced, we tailor a hierarchical-loop metacognitive model to identify three levels of activity, as it is sketched in

Figure 1. At the bottom of the hierarchy, the external level is allocated and the cognitive level is found in the middle tier. At the top, the metacognitive level is stated. Moreover, a loop of activities is achieved at each level and a cognitive information flow goes bottom-up and backs top-down between couples of levels [31].

At external level, the individual interacts with her outside-world, sets goals to achieve and deals with problems to be solved. She uses her motor skills to behave and socialize with others every day. The surrounding objects and the performance of her physical actions are perceived by her senses. The cognized information is an input to her basic cognitive skills. In consequence, they trigger physical actions as output.

At cognitive level, the individual uses her basic cognitive skills to achieve mental or physical goals. They are activated as loops of cognitive activities, such as: *observation* of sensory and perceptive stimuli from the outside-world, *rehearsal* to hold information needed by other cognitive activities on short-term cognitive memory, *evaluation* of external conditions, *virtual execution* of candidate cognitive activities, *selection* of the cognitive activity to be achieved through physical activities. The execution and results of cognitive activities is an input to the metacognitive skills. As a result, they control the behavior of cognitive activities as output [32].

At metacognitive level, the individual uses her metacognitive skills to assess and regulate the performance of cognitive skills and assure the accomplishment of goals. They essentially apply similar cognitive activities, as the basic cognitive skills do, but the difference is the concerned source and target. Hence, *observation* monitors the course of the activity and the results achieved at cognitive level, *rehearsal* maintains information requested by other metacognitive activities on short-term metacognitive memory, *evaluation* qualifies the activity and outcomes fulfilled at cognitive level, *virtual execution* simulates the activation of metacognitive activities, *selection* chooses the metacognitive activity to be carried out for controlling cognitive skills.

4.2 Class Cognitive Activities

Metacognition is a matter of generalization of knowledge, strategies, experiences, tasks and goals that is applied to a wide sort of purposes and issues. The acquisition, tuning and application of such metacognitive resources are fulfilled by abstraction, modification and instantiation metacognitive activities.

The abstraction focuses on key attributes of a given object or activity to shape a class of metacognitive resources. Such a class embraces conceptual items that depict pieces of knowledge, main guides of strategies, key issues gained from experiences and procedural descriptions of how to carry out a task and meet a goal.

The modification refines conceptual items of the class in order to improve its generality and usefulness. It applies three tasks to manipulate classes, such as: *addition* devoted to create a new class, *modification* oriented to update conceptual items of a given class, *deletion* aimed to unlearn conceptual items and eliminate the whole class.

The instantiation applies a given class to a current matter. So it provides specific values to the conceptual items of a class for tailoring an instance. The instance reveals the way to deal with a present object or activity based on metacognitive resources.

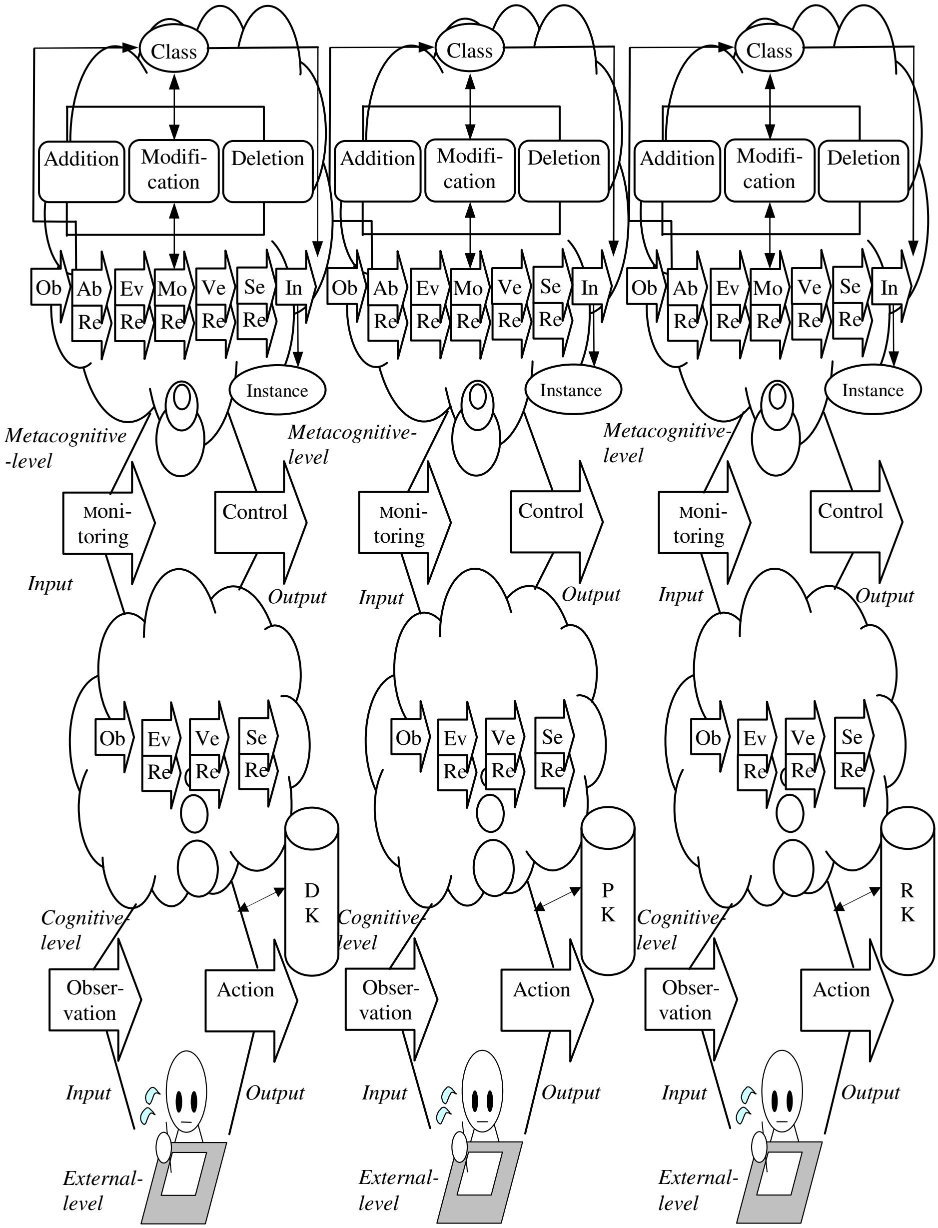


Fig. 1. Metacognitive-Driven Learning paradigm. It holds a model and a method. The model contains three levels: metacognitive, cognitive and external. The method have three stages: cognitive, associative and autonomous, Activities' id at metacognitive and cognitive levels means: *Ob*: observation; *Ab*: abstraction; *Re*: rehearsal; *Ev*: evaluation; *Mo*: modification; *Ve*: virtual execution; *In*: instantiation; *DK*: declarative knowledge; *PK*: procedural knowledge; *RK*: refined knowledge.

5 The Method of the Metacognition-Driven Learning

The Metacognition-Driven Learning paradigm also embraces a method for stimulating metacognitive skills. It follows the workflow set by Anderson [33]. As it is shown in Figure 1, the sequence begins with the cognitive stage, continues with the associative stage and ends with the autonomous stage. As soon as the individual develops the stages, she progressively acquires, evolves and extends her skills´ knowledge.

The cognitive stage enables individual to acquire knowledge by objectivism practice. The outcome is a kind of *declarative knowledge* of the skill. As a result, the individual had stated concepts that depict the nature, purpose and process of the skill.

The associative stage applies constructivism practice by problem-solving exercises. During the stage, the individual faces a diversity of problems and seeks a solution. As consequence, she gains *procedural knowledge* from the experiences.

The autonomous stage encourages the individual to deal with more domain problems of increasing complexity. Little by little, the individual is challenged to evaluate, contrast and pursue optimization of problem-solving process. During the stage, the individual's *knowledge* of the skill is gradually refined.

The Metacognitive-Driven Learning makes up a model and method into just one paradigm. Its sequence of three stages is pictured in Figure 1 from left to right. At each stage, the individual takes the corresponding lectures to produce a version of skill's knowledge. Three types of activities are concurrently performed in their corresponding level. The ascending information flow represents an input between a pair of levels; whereas, a descending one corresponds to an output [34].

6 Conclusions

In this paper we have reviewed a set of representative works to understand what metacognition is and how it can be stimulated. In addition, we have outlined a Metacognition-Driven Learning paradigm with the aim to inspire the design of IES. Our model offers a holist viewpoint to structure levels of activities and flows of information to monitor and control activities.

We acknowledge the importance of enhancing students´ learning faculties. So we assert: Individuals who are able to use their metacognitive skills increase their opportunities to accomplish educational and personal goals. As a future work, we plan to develop a computer-based prototype to implement our paradigm and develop a trial to collect empirical evidence. According to the results a new version will be outcome.

Acknowledgments. The first author gives testimony of the strength given by his Father, Brother Jesus and Helper, as part of the research projects of World Outreach Light to the Nations Ministries (WOLNM). The research holds a partial support from grants: CONACYT-SNI-36453, CONACYT 118962, CONACYT 118862, SIP-20110398, SIP-EDI: DOPI/3189/08, COFAA-SIBE.

References

1. Flavell, J.H.: The Developmental Psychology of Jean Piaget. D Van Nostrand, N.Y (1963)
2. Flavell, J.H.: First Discussant's Comments: What is Memory Development the Development of? Human Development 14, 272–278 (1971)
3. Flavell, J.H.: Metacognitive Aspects of Problem Solving. In: Resnick, L.B. (ed.) The Nature of Intelligence, pp. 231–236. Erlbaum, Hillsdale (1976)
4. Reder, L., Schunn, C.D.: Metacognition does not Imply Awareness: Strategy Choice is Goverened by Implicit Learning and Memory. In: Reder, L.M. (ed.) Implicit Memory and Metacognition, pp. 45–78. Erlbaum, Mahwah (1996)
5. Kentridge, R.W., Heywood, C.A.: Metacognition and Awareness. J. Consciousness and Cognition 9, 308–312 (2000)
6. Brown, A.: Metacognition, Executive Control, Self-Regulation and other more Mysterious Mechanisms. In: Weinert, F.E., Kluwe, R.H. (eds.) Metacognition, Motivation, and Understanding, pp. 65–116. Lawrence Erlbaum Associates Publishers, Hillsdale (1987)
7. Gregory, S.: Promoting General Metacognitive Awareness. J. Instructional Science 26(1-2), 113–125 (1998)
8. Barry, J.Z., Andrew, S.P.: Self-Monitoring During Collegiate Studying: An Invaluable Tool for Academic Self-Regulation. J. New Directions for Teaching and Learning 63, 13–27 (1995)
9. Dale, H.: Schunk: Metacognition, Self-Regulation and Self-Regulated Learning: Research Recommendations. J. Educational Psychology Review 20(4), 463–467 (2008)
10. Roll, I., Ryu, E., Sewall, J., Leber, B., McLaren, B.M., Aleven, V., Koedinger, K.R.: Towards Teaching Metacognition: Supporting Spontaneous Self-Assessment. In: Ikeda, M., Ashley, K.D., Chan, T.-W. (eds.) ITS 2006. LNCS, vol. 4053, pp. 738–740. Springer, Heidelberg (2006)
11. William, P.: Rivers: Autonomy at All Costs: An Ethnography of Metacognitive Self-Assessment and Self-Management among Experienced Language Learners. The Modern Language Journal 85(2), 279–290 (2001)
12. Flavell, J.H.: Metacognition and Cognitive Monitoring: A New Area of Cognitive-Developmental Inquiry. J. American Psychologist 34, 906–911 (1979)
13. Kayashima, M., Inaba, A.: Towards Helping Learners Master Self-Regulation Skills. In: Int. C. Artificial Intelligence in Education, Sydney, pp. 1–10 (2003)
14. Flavell, J.H.: Speculation About the Nature and Development of Metacognition. In: Weinert, F., Kluwe, R. (eds.) Metacognition, Motivation, and Understanding, pp. 21–29. Lawrence Erlbaum, Hillsdale (1987)
15. Brown, A.: Knowing when, where, and how to Remember: A Problem of Metacognition. In: Glaser, R. (ed.) Advances in Instructional Psychology, pp. 77–165. Lawrence Erlbaum Associates Publishers, Hillsdale (1978)
16. Gama, C.A.: Integrating Metacognition Instruction in Interactive Learning Environments. PhD Thesis, University of Sussex (2004)
17. Nelson, T.O., Narens, L.: Why Investigate Metacognition. In: Metcalfe, J., Shimamura, A.P. (eds.) Metacognition: Knowing about Knowing, MIT Press, Cambridge (1994)
18. Norman, D.A., Shallice, T.: Attention to Action. Willed and Automatic Control of Behavior. In: Davidson, R.J., Schwartz, G.E., Shapiro, D. (eds.) Consciousness and Self Regulation, pp. 1–17. Plenum, New York (1986)
19. Tobias, S., Everson, H.T.: Knowing what you Know and what you don't: Further Research on Metacognitive Knowledge Monitoring. In: College Board Research Report, pp. 2–3. College Entrance Examination Board, New York (2002)

20. Kayashima, M., Inaba, A.: Difficulties in Mastering Self-regulation Skill and Supporting Method for It. In: Int. C. Artificial Intelligence in Education, Sydney, pp. 1–8 (2003)
21. King, A.: Discourse Patterns for Mediating Peer Learning. In: O'Donnel, A.M., King, A. (eds.) Cognitive Perspectives on Peer Learning, pp. 87–115. Lawrence, New Jersey (1999)
22. Wood, T., Cobb, P., Yackel, E.: The Nature of Whole-Class Discussion. J. Research in Mathematics Education 6, 55–122 (1993)
23. Schoenfeld, A.H.: What's all the Fuss about Metacognition? In: Shoenfeld, A.H. (ed.) Cognitive Science & Mathematics Education, pp. 189–215. Lawrence Erlbaum Assoc. (1987)
24. Kayashima, M., Inaba, A.: The Model of Metacognitive Skill and How to Facilitate Development of the Skill. In: Int. C. Computers in Education, Hong Kong, pp. 277–285 (2003)
25. Palincsar, A.S., Brown, A.: Reciprocal Teaching of Comprehension - Fostering and Comprehension – Monitoring Activities. J. Cognitive and Instruction 1(2), 117–175 (1984)
26. Luckin, R., du Boulay, B.: Ecolab: The Development and Evaluation of a Vygotskian Design Framework. Int. J. Artificial Intelligence and Education 10(2), 198–220 (1999)
27. Wiemer-Hastings, P., Glasswell, K.: StoryStation: Agent-based Scaffolding of Metacognitive Processes for Writing. In: Int. C. Artificial Intelligence on Education, Brighton, pp. 1–7 (2009)
28. Conati, C., VanLehn, K.: Toward Computer-Based Support of Meta-Cognitive Skills: A C. Framework to C. Self-Explanation. J. Artificial Intelligence in Education 11, 398–415 (2000)
29. Biswas, G., Roscoe, R., Jeong, H., Brian Sulcer, B.: Promoting Self-Regulated Learning Skills in Agent-based Learning Environments. In: Int. C. Computers in Education, Hong-Kong, pp. 67–74 (2009)
30. Baghaei, N., Mitrovic, A.: From Modeling Domain Knowledge to Metacognitive Skills: Extending a Constraint-based Tutoring System to Support Collaboration. In: Conati, C., McCoy, K., Paliouras, G. (eds.) UM 2007. LNCS (LNAI), vol. 4511, pp. 217–227. Springer, Heidelberg (2007)
31. Kayashima, K., Inaba, A., Mizoguchi, R.: What is Metacognitive Skill? In: Proc. World C. Educational Mutimedia, Hypermedia, Telecommunications, Switzerland, pp. 2660–2665 (2004)
32. Kayashima, M., Inaba, A., Mizoguchi, R.: What Do You Mean by to Help Learning of Metacognition? In: Proc. Artificial Intelligence in Education, pp. 346–353 (2005)
33. Anderson, J.: Acquisition of Cognitive Skill. Psychological Review 89(4), 369–406 (1982)
34. Kayashima, M., Mizoguchi, R.: A new Perspective for Metacognition-driven Learning. In: Proc. Int. C. Computers in Education, pp. 1–10. Kuala Lumpur, Malaysia (2010)

Interactive Neuro-Educational Technologies (I-NET): Development of a Novel Platform for Neurogaming

Giby Raphael, Adrienne Behneman, Veasna Tan, Nicholas Pojman, and Chris Berka

Advanced Brain Monitoring, Inc. (ABM), 2237, Faraday Ave,
Suite 100, Carlsbad, CA, 92008, USA
graphael@b-alert.com

Abstract. The advances in sophisticated, immersive and highly engaging video gaming technology have resulted in the introduction of "serious gaming" as platforms for training. A virtual environment that mimics reality as closely as possible is an effective instructional medium and also serves as a performance improvement/evaluation platform. However, the current methodologies suffer from several limitations: 1) conventional qualitative evaluation techniques that are removed from the trainee's actual experience in both time and context 2) open loop platforms fail to support adaptive training and scenarios or leverage repeatability to accelerate training 3) failure to adapt to individual's current psychophysiological state, limiting skill acquisition rates 4) multi-person tasks that lack tools for objective assessment and prediction of team cohesion or performance. As part of our initiative to invent a suite of Interactive Neuro-Educational Technologies (I-NET), we have developed a Neurogaming platform that will help resolve many of these limitations.

Keywords: EEG, Neuroergonomics, Neurosensing, Augmented Cognition.

1 Introduction

Advances in the fields of computer graphics and artificial intelligence, combined with the availability of sophisticated multi-core gaming hardware have resulted in the application of "serious gaming" as platforms for training simulations in industry, academia and military. The virtual environment provides a safe, controlled and cost effective setting to educate and evaluate users without the dangers associated with real life scenarios, especially for some industrial and military applications. The military has been a strong proponent of the gaming technology and uses sophisticated simulators for enhancements in training, safety, as well as to analyze military maneuvers and battlefield positions [1,2]. The "Educate to Innovate" campaign launched by the President of the United States in 2009 aimed at harnessing the power of interactive games to improve technological, mathematical, scientific and engineering abilities of American students. A wide variety of games were thus introduced in order to provide engaging exercises, improve retention of complex concepts, and create a rewarding learning experience for students [3]. Although this new generation of training technology is engaging and popular, it represents only an initial step towards a true revolution in creating successful instructional delivery systems.

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 452–461, 2011.

Conventional methods for evaluating instructional design such as subjective reports, performance metrics and expert observations are mostly qualitative in nature and are removed from the trainee's experience in both time and context. Optimized environments that leverage brain-behavior relationships are known to improve the efficiency of learning (Neuroergonomics theory, [4,5]). Increasing evidence suggests that physiological correlates of attention, alertness, cognitive workload, arousal, and other fundamental constructs essential to training can be identified to further improve learning efficacy. The convergence of recent advances in ultra-low power consumer electronics, ubiquitous computing and wearable sensor technologies enables real-time monitoring of these cognitive and emotional states providing objective, timely, and ecologically valid assessments of psychophysiological states associated with learning. Our previous work has revealed specific EEG correlates of the stages of skill acquisition in simple learning and memory tasks, as well as in more cognitively complex and challenging test environments. Unique event-related EEG signatures detected during various stages of skill acquisition were evaluated to assess participants' ability to reflect aspects of learning across tasks and environments [6,7,8]. Such quantitative assessments will enable guidance of the user through distinct stages of skill development as well as provide timely evaluation and mitigation to improve the efficiency of learning.

Practice is accepted as a ubiquitous strategy to accelerate skill development. Repetition alone however, does not ensure success and repeated poor technique can lead to performance deficiencies and/or stress injuries. Instructional strategies and feedback are believed to be critical in accelerating skill learning. Recent investigations have suggested that skill learning may be dependent upon the availability of cognitive resources including attention and working memory and that the speed and efficacy of learning may be affected by either state or trait differences in these cognitive capacities [9,10]. Closed-loop and adaptive training platforms that incorporate real time sensing of cognitive and emotional state of the trainee (Neurosensing) and tailor the information delivery to an individual's or team's evolving skill level can considerably enhance the learning experience [11,12]. We have successfully employed similar closed-loop systems previously in implementing EEG-based drowsiness alarms in a driving simulator [13] and EEG-workload based Aegis radar and Tactical Tomahawk Weapons simulations [14], however the physiological thresholds and mitigations employed were specific to the application and not transferable to a general training platform.

The process presented here is an initial step towards a dynamic training system that will adaptively incorporate the psychophysiologic state of the user in order to optimize training. The interactive training platform we intend to develop incorporates three modules: 1) Event-locked extraction of physiological metrics, 2) Closed-loop mitigation of tactical scenarios based on real time physiological metrics, and 3) Real time evaluation and mitigation of team neurodynamics. This paper presents a pilot study addressing the first module: event-locked extraction of physiological metrics. We will then discuss work done to address Modules 2 and 3, and how the three modules may be integrated into an adaptive training platform to increase training efficacy in individuals and teams. The platform was developed as part of our initiative to invent a suite of Interactive Neuro-Educational Technologies (I-NET) to accelerate skill learning and novice-to-expert transition.

2 Methods

2.1 Participants

Twenty-three participants (10 females and 13 males, mean age 25.87 years, range 19-40 years) were recruited from local colleges and newspaper/online advertisements. All participants had normal or corrected to normal vision and reported no history of neurological problems. No participants that had undergone formal marksmanship training were admitted to the study. Informed consent was obtained from all participants in accordance with the guidelines and approval of the Biomedical Research of America Institutional Review Board.

2.2 Data Acquisition

Electroencephalographic (EEG) and Electrocardiographic (EKG) data were collected using the wireless B-Alert® X10 EEG sensor headset developed by Advanced Brain Monitoring (ABM). Nine Ag/AgCl EEG electrodes were located at F3, Fz, F4, C3, Cz, C4, P3, POz, P4, according to the international 10-20 system. All EEG channels were referenced to linked reference electrodes located behind each ear on the mastoid bone. EKG was recorded with electrodes placed on the clavicle and opposite lower rib. Data was sampled at 256 Hz.

2.3 Paradigm

The popular military gaming platform - Virtual Battle Space 2 (VBS2), Tactical Warfare Simulation running on Real Virtuality 2 simulation engine, developed by Bohemia Interactive was used to create the gaming scenarios. In collaboration with Laser Shot Inc., we developed five custom combat scenarios using VBS2. The scenarios had realistic settings and contexts in which participants (acting as soldiers) were required to make deadly force decisions. In order to mimic reality as closely as possible, the game room was equipped with life-size projection of threats, stereo sound delivered via earbuds and other paraphernalia found in the battlefield environment. The participant used a demilitarized "airsoft" replica of an M4 rifle that interacted with the game using a wireless laser-based training system from Laser Shot Inc. The M4 was mounted with an EOtech holographic weapon sight, or red dot scope, commonly used in combat environment for quick target acquisition. Sandbags were provided to support the weight of the weapon, to both simulate combat firing procedures and reduce the effect of muscle fatigue on performance.

Participants were initially given marksmanship instructions (power point presentation) and requested to undergo a set of training tasks. Training addressed the fundamentals of marksmanship (aiming, breath control, trigger control, etc.), and the Rules of Engagement applicable in the testing scenarios. Five testing scenarios were administered in a randomized order for each subject. Each scenario was set in a unique environment that replicated typical fire fighting situations for soldiers (e.g., checkpoint, market, etc. in Afghanistan). In order to avoid excessive fatigue in participants, the scenarios were designed from a fixed point of view and lasted only 3-4 minutes. Throughout the scenario a mixture of enemy and friendly units, both stationary and moving, appeared at varying distances. Participants were instructed to evaluate threats and eliminate all enemy units as quickly as possible.

2.4 Data Analysis

ABM B-Alert® software was used to acquire, filter and analyze the physiological signals transmitted by the headsets in real time. The software identifies and eliminates multiple sources of environmental and physiological contamination such as power frequency hum, eye-blinks, EMG, etc. as well as other artifacts such as spikes, excursions, saturations, etc. using patented wavelet based signal processing algorithms. The software also incorporates patented algorithms for real-time classification of EEG based Engagement and Workload. These metrics are calculated using quadratic and linear discriminant functions that analyze Power Spectral Densities (PSD) of EEG frequency bins ranging from 1-40 Hz on a second-by-second basis [15]. Simple baseline tasks (completed on a prior study visit) were used to fit the EEG Engagement and Workload algorithms to the individual, so that the cognitive state models provide a highly sensitive and specific technique for identifying an individual's neural signatures of cognition.

An External Sync Unit (ESU) was used to synchronize the physiological signals with events in the game as well as user responses (rifle shots). Synchronization in the windows environment is dependant on the windows task scheduler and cannot guarantee an upper bound for user level tasks. The ESU is a general purpose data integration platform that can synchronize multi-source digital data (serial and/or parallel port protocols) with physiological signals from B-Alert® headsets acquired via Bluetooth protocol to millisecond level precision. ABM's automated analysis tools were used to extract various performance metrics in order to identify, 1) psychophysiological states associated with learning and skill acquisition, 2) cognitive factors that influence decision making, and 3) relevant physiological measures that distinguish top performers.

The analysis of Event Related Potentials (ERPs) offer excellent temporal resolution for tracking the flow of information from sensory processing, detection and identification of relevant objects and decision-making. ERPs and Event Related Engagement / Workload / Heart Rate were derived by time-locking to the presentation of the test bed stimuli (one-second post-stimulus), and to the one second epochs prior to and following user shots. ERPs were then plotted for the two seconds surrounding each shot (one second pre-shot, one second post-shot), at each sensor site and for each single trial shot event. Single trial ERP waveforms were averaged within and then across participants to compute grand means. Before averaging, data that included artifact such as eye blinks, excursions or excessive muscle activity were rejected on a trial-by-trial basis using automated in-house software [15].

As a normalization step, single trial event-related Engagement / Workload / Heart Rate data were z-scored to each individual's average level of each metric. This allows the identification of whether an individual was experiencing above or below average Engagement / Workload / Heart Rate in relation to a given event.

3 Results

3.1 Event Related Potentials

Fig.1 below illustrates the averaged time series at the vertex (Cz) for missed shots (shots that did not hit a target) versus kill shots (shots that hit and killed an enemy target). Greater positivity from 875 ms to 250 ms before the shot distinguished kill shots from missed shots. Peak amplitude preceding kill shots was significantly greater than the peak amplitude preceding missed shots ($t(22) = 2.92$, $p<.01$). Eighteen of the 23 subjects (78%) showed this distinction, with peak amplitude preceding kill shots on average 5.37 µV greater than the peak amplitude preceding missed shots. This appears similar to findings reported by Konttinen and Lyytinen, 1998 [16], in which pre-shot slow potential positivity was associated with increased rifle stability. Alternatively, the pre-shot positivity for kill shots could be associated with target identification and recognition with a greater attenuation to enemy targets preceding kill shots (and presumably absent preceding missed shots). All subjects exhibited a characteristic *post-shot* positive potential beginning at the time of the shot and peaking between 100 and 250 ms post-shot. This initial positive component was most clearly seen in the central channels, and did not differentiate shot types (e.g., missed shots versus kill shots). However a late positive component differentiated missed shots from kill shots as early as 250 ms post-shot and was sustained for windows in excess of 900 ms post-shot. Maximal miss/kill differences were seen in central and parietal channels, and varied from 490-850 ms post-shot. Peak amplitude of the late positive component (between 490-850 ms post-shot) was significantly higher following kill shots than following misses ($t(20) = 4.84$, $p<.0001$). The peak amplitude of the late component following kill shots was on average 7.39 µV greater than the peak amplitude following missed shots. This later positivity is likely a P300 or Late Positive component, possibly in response to the visual cue of the enemy death in the simulation environment (absent in the case of missed shots).

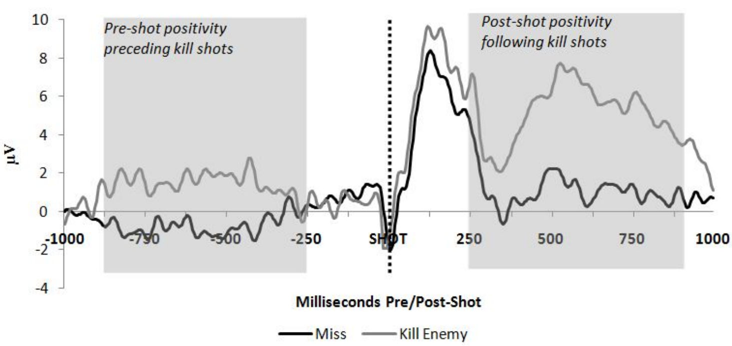


Fig. 1. Grand mean (n=23) ERP waveforms for missed shots versus shots that killed an enemy, at vertex site (Cz), for one-second pre- and post-shot

3.2 Other Event Related Metrics

EEG-Engagement one second pre-shot distinguished missed shots from those that killed an enemy (see Fig.2(a)). Below average engagement was associated with missed shots. This difference did not reach statistical significance ($P=0.07$), however suggests differing neurophysiologic states for the preparation of successful versus unsuccessful shots. A subsample of subjects (n=6) had instances of "friendly fire" (shooting a civilian or comrade). Fig.2(b) shows normalized EEG-Engagement and EEG-Workload in the one second pre-shot, averaged for those six subjects.

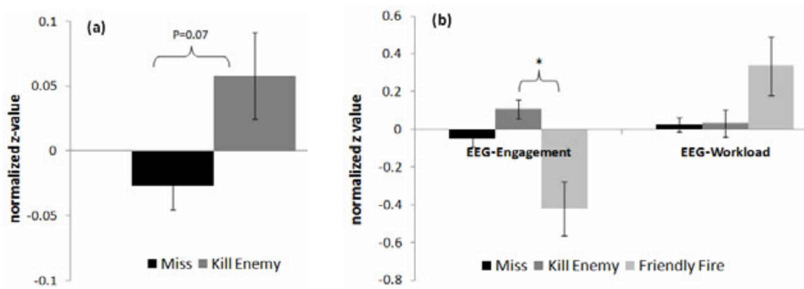


Fig. 2. (a) Grand means for normalized (within individuals) EEG-Engagement in the one second preceding missed shots versus kill shots (n=18; performance outliers removed). (b) Group means for the six subjects that had instances of friendly fire errors. Normalized EEG-Engagement and EEG-Workload in the one second pre-shot for misses, kills, and friendly fire.

Friendly fire errors were marked by below-average EEG-Engagement (about 0.4 standard deviation *below* average), and above average EEG-Workload (about 0.3 standard deviation *above* average). EEG-Engagement preceding friendly fire errors was significantly lower than EEG-Engagement preceding shots that killed an enemy ($t(4) = 2.92$, $p<.05$). No other differences reached significance. Due to the low number of friendly fire errors in the dataset, these results are only relevant to the six subjects that made this type of error and may not generalize across a larger population. A test bed with a greater number of civilians or comrades (or with enemy and friendly forces that are harder to distinguish from each other) would provide better opportunity for studying the neurophysiology associated with that type of error.

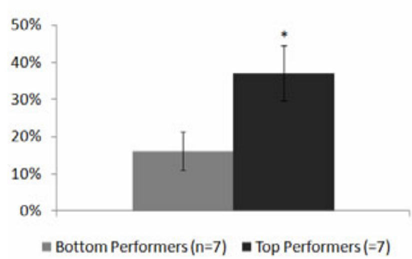


Fig. 3. Average percent pre-shot Engagement across all shots, for bottom vs. top performers

3.3 Top vs. Bottom Performers

Each of the thirteen VBS2 performance metrics were z-scored to the population (23 subjects), and then aggregated into a summary performance measure to determine the overall top (n=7) and bottom (n=7) performers. Pre-shot EEG-Engagement, EEG-Workload, and HR (raw values; not normalized within individuals) were compared between the two groups. Fig.3 illustrates that top performers had nearly 2 times higher pre-shot EEG-Engagement than bottom performers (t (12) = -2.32, p<0.05). This finding highlights the importance of Engagement in the combat pre-shot state.

4 Discussion

In this study we developed a platform which allows for event-locked extraction of physiological metrics in a tactical training environment. Our preliminary results suggest that physiological signatures may distinguish elements of good and poor performance that could be used to accelerate the efficiency of learning in individuals and to improve performance of teams. The interactive training platform under development incorporates three modules: 1) Event-locked extraction of physiological metrics, 2) Closed-loop mitigation of tactical scenarios, based on real time physiological metrics and 3) Real time evaluation and mitigation of team neurodynamics. The study presented above addresses the first module. Below, we discuss efforts taken to address the second two modules.

4.1 Closed-Loop Mitigation of Tactical Scenarios

The concept of a comprehensive closed-loop module was tested independently in order to incorporate the mitigation parameters and strategy derived through research studies [17]. Automated adaptive training was incorporated based on both physiological (EEG, EKG, GSR) and non-physiological (performance, subjective training, expert observations) metrics. The Synchronous Operational Psychophysiological Sensor Suite (SyKron) developed by the University of Central Florida's ACTIVE laboratory was used to integrate, synchronize, as well as analyze physiological signals from ABM EEG headsets as well as other non-physiological inputs such as performance, subjective rating etc. The data logging and playback features of SyKron were used to facilitate iterative assessment of adaptive mitigation and threshold development. General Purpose Real-Time Mitigation Engine (GPRIME) developed by the Warfighter Human-Systems Integration Laboratory at the U.S. Naval Research Laboratory (NRL) was used to close the loop by providing real time modification of the game. GRPIME is a software platform that can support streaming data from multiple IP addresses, allowing for mitigations to be triggered by data variables streaming from multiple computers on a local network. GPRIME receives processed real-time (or near real-time) physiological inputs along with subjective and/or performance data from SyKron as variables to create Boolean logic (If, And, Or, >, =, etc.) rules that are saved and evaluated in real-time to assess when it is appropriate to perform a mitigation. When the streaming data inputs meet the threshold rules, pre-recorded keyboard and mouse click macros are triggered to modify the training scenario in VBS2.

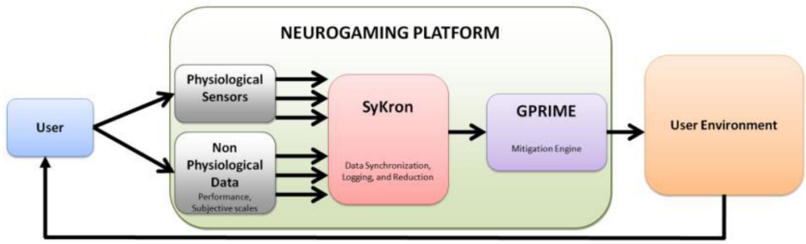


Fig. 4. Closed-loop Adaptive module for real-time mitigation (adapted from Berka et.al, 2010)

4.2 Real Time Evaluation of Team Neurodynamics

In order to develop a tightly controlled platform for investigating team neurodynamics in a simulation environment, we developed a team training module using desktop based simulators. Working with Discovery Machine Inc., ABM designed a three person teaming scenario, called 'Safe Passage,' within VBS2. The scenario was designed to evaluate and train team members to work together to successfully complete a team mission, emphasizing team communication and friendly fire avoidance. The primary mission objective of the team was to safely escort a convoy through enemy territory. Team members were assigned individual tasks, each with unique roles essential to meeting the mission's objectives.

Pilot Study: ABM recruited two teams of three people, with each team performing six iterations of the teaming mission. Each team member was fit with a wireless X10 B-Alert headset which allowed EEG-Engagement, EEG-Workload and Heart Rate to be calculated on a second-by-second basis. All computers were programmed to follow the Network Time Protocol (NTP) such that the physiological parameters, game events, user responses, etc. from all team members were time synchronized for analysis by real-time as well as offline algorithms.

The automated online program developed by Stevens et.al [18, 19] was used to generate EEG-based neurosynchrony (NS) profiles for team performance in real-time. A summary of the layered analytic approach used by the program is as follows: The data flow is organized into collection, processing, modeling and analysis modules. The data collection and processing modules are included in ABM B-Alert software where EEG is decontaminated and Workload (WL) and Engagement (E) values from each of the team members are calculated on a second-by-second basis. The values of WL and E are then normalized and statistically partitioned. The values at each second for each team member are then combined into a vector representing the state of the team (EEG-E) as a whole. A trained 1×25 node unsupervised artificial neural network develops a topology and outputs a linear series of 25 team EEG-E patterns that are termed as neurophysiologic synchronies (NS). A 'hit' frequency map showing the number of times each node was expressed during a performance was then created, which when aligned with training context can provide significant insight into team dynamics and potentially be used for real time mitigation of tasks. Predictive models developed using Hidden Markov Modeling by analyzing the correlations and persistence of the NS could also be potentially used to indicate effective/ineffective team compositions. Data analysis for this pilot study is in progress.

The final integrated Neurogaming platform will incorporate automated event logging and synchronization with physiological data and closed-loop mitigations for individuals and teams. Even though the scenarios discussed above were specific to military applications, the scope of the Neurogaming platform can be extended to many other medical, educational and industrial applications.

Acknowledgment. This work was supported by The Defense Advanced Research Projects Agency (government contract number NBCHC090054). The views, opinions, and/or findings contained in this article are those of the author and should not be interpreted as representing the official views or policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the Department of Defense. Approved for Public Release, Distribution Unlimited. The development of the closed-loop module was funded by the Office of the Secretary of Defense SBIR Award # N00014-09-M-014.

References

1. Ackerman, R.K.: Navy Researchers Target, Virtually. Signal 60(11)
2. Alpert, M.: My Virtual War. Scientific American.com. January 22 (2006)
3. Anderson, et al.: Critters in the classroom: a 3D computer-game-like tool for teaching programming to computer animation students. In: Proceedings in SIGGRAPH (2007)
4. Kramer, A., Parasuraman, R.: Neuroergonomics: Application of Neuroscience to Human Factors. In: Caccioppo, J.T., Tassinary, L.G., Berntson, G.G. (eds.) Handbook of Psychophysiology, 3rd edn., pp. 704–722. Cambridge University Press, New York (2007)
5. Parasuraman, R.: Neuroergonomics: The Brain at Work. Oxford University Press, Oxford (2005)
6. Berka, C., et al.: Objective Measures of Situational Awareness using Neurophysiology Technology. In: Schmorrow, D., Stanney, K., Reeves, L. (eds.) Augmented Cognition: Past, Present and Future, pp. 145–154. Strategic Analysis, Inc, Arlington (2006)
7. Stevens, R., Galloway, T., Berka, C.: Integrating EEG Models of Cognitive Load with Machine Learning Models of Scientific Problem Solving. In: Schmorrow, D., Stanney, K., Reeves, L. (eds.) Augmented Cognition: Past, Present and Future, pp. 55–65 (2006)
8. Berka, C., et al.: Real-time Analysis of EEG Indices of Alertness, Cognition and Memory Acquired with a Wireless EEG Headset. International Journal of Human-Computer Interaction 17(2), 151–170 (2004)
9. Best, J.B.: Cognitive Psychology, 3rd edn. West Publishing Company, New York (1992)
10. Solomon, R.M., Horn, J.M.: Post-shooting traumatic reactions: A pilot study. In: Reese, J., Goldstein, H. (eds.) Psychological Services in Law Enforcement, pp. 383–393. United States Government Printing Office (1986)
11. Schmorrow, D., Stanney, K.M., Wilson, G.F., Young, P.: Augmented cognition in human-system interaction. In: Salvendy, G. (ed.) Handbook of human factors and ergonomics, 3rd edn. John Wiley, New York (2005)
12. Stripling, R., Coyne, J.T., Cole, A., Afergan, D., Barnes, R.L., Rossi, K.A., et al.: Automated SAF Adaptation Tool (ASAT). In: Augmented Cognition Conf., pp. 346–353 (2007)
13. Berka, C., Levendowski, D., Westbrook, P., Davis, G., Lumicao, M., Ramsey, C., et al.: Implementation of a Closed-Loop Real-Time EEG-Based Drowsiness Detection System: Effects of Feedback Alarms on Performance in a Driving Simulator. In: Paper presented at the 1st International Conference on Augmented Cognition, Las Vegas, NV (2005)

14. Berka, C., Levendowski, D., Ramsey, C.K., Davis, G., Lumicao, M., Stanney, K., et al.: Evaluation of an EEG-Workload Model in an Aegis Simulation Environment. In: Paper presented at the Proceedings of SPIE Defense and Security Symposium, Biomonitoring for Physiological and Cognitive Performance during Military Operations, Orlando, FL (2005)
15. Berka, C., et al.: Real-time Analysis of EEG Indices of Alertness, Cognition, and Memory Acquired with a Wireless EEG Headset. International Journal of Human-Computer Interaction 17(2), 151–170 (2004)
16. Konttinen, N., Lyytinen, H., Viitasalo, J.: Rifle-balancing in precision shooting: behavioral aspects and psychophysiological implication. Scandinavian J. Med. Sci. Sports (1998)
17. Berka, C., Pojman, N., Trejo, J., Coyne, J., Cole, A., Fidopiastis, C., Nicholson, D.: Neurogaming: Merging Cognitive Neuroscience & Virtual Simulation in an Interactive Training Platform. In: 1st International Conference on Neuroergonomics Florida (2010)
18. Stevens, R.H., et al.: Can Neurophysiologic Synchronies be Detected during Collaborative Teamwork? In: Proceedings in HCI International, San Deigo (July 2009)
19. Stevens, R.H., Galloway, T., Berka, C., Behneman, A.: A Neurophyiologic Approach for Studying Team Cognition. In: Proceedings in ITSEC (2010)

Learning in Virtual Worlds: A New Path for Supporting Cognitive Impaired Children

Laura A. Ripamonti and Dario Maggiorini

Dipartimento di Informatica e Comunicazione, Università degli Studi di Milano
via Comelico, 39 I-20135 Milano, Italy
{ripamonti,dario}@dico.unimi.it

Abstract. We have adopted the serious game perspective to design, develop, and test a prototypal application, in a virtual world, aimed at teaching children affected by Down Syndrome how to read a clock. The main idea has been to offer them a new and intriguing learning environment to reduce the sense of frustration they often are burdened with during educational activities. In particular, an approach based on serious gaming has been coupled with the Feuerstein's method, which is currently spreading as an effective support to teaching activities aimed at impaired kids. The prototype has been developed adopting a play-centric process and has been tested with a group of children who were unable to read the time.

Keywords: serious games; videogames; healthcare; virtual world; usability; augmented cognition, Down Syndrome.

1 Introduction

Nowadays, women tend to have their first child at an older age than in the past. This is positively correlated with the increase in the percentage of newborns with Down syndrome (DS), as shown in Tab. 1. Although all the efforts made by scholars and researchers in the medical field, it is not yet possible to prevent nor to cure DS. The only viable way to help these children to increase the quality of their lives (i.e., from the independence and autonomy point of view) is through appropriate "rehabilitation" therapies, aimed at supporting them to achieve more advanced cognitive, linguistic, and motional capabilities [5].

In particular, the most serious difficulties afflicting children with DS are related to the cognitive side [17, 34]: hence, it is particularly relevant to improve the methodologies aimed at developing cognitive capabilities and skills. Among several possible approaches to this problem, one whose diffusion has increased during the last decade is rooted into the theories developed by Reuven Feuerstein [2, 10, 11, 12]. Feuerstein's methodology is now adopted not only in the pedagogical field, but also in clinical environments, aiming at stimulating cognition in DS children. Its major advantage is to avoid the iterative processes typical of more "traditional" rehabilitative approaches, and to focus on stimulating a problem solving attitude, exploiting the innate capability of human brain to self-modify, evolve, and adapt – even under disadvantaged conditions. The application of the Feuerstein method requires the presence of a special

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 462–471, 2011.

"mediator", that is to say a person in charge of interpreting the child experiences when faced with a new task, and able to help her in addressing her efforts toward a specific goal. The importance of the mediator shines even clearer when observed through the lenses of the *mirror neurons* theory [3, 14, 16, 29]. According to this theory, in our brain there is a certain amount of neurons whose job is to learn how to accomplish a certain task, simply by observing others performing the appropriate actions. In the same vein, they allow us to "feel what's on a person mind"; thus, also supplying the substrate for empathy. In other words, the observer's neurons "mirror" what is going on in the brain of the person she is observing, by activating exactly as if the action was undertaken by her. Hence, effective mediators should be able to stimulate DS children mirror neurons, transforming an action they see into something they are able to do.

Table 1. Relation between mother's age and probability of having a baby with DS [15]

Mother age	DS risk	Mother age	DS risk
20	1/1527	36	1/280
25	1/1352	37	1/218
28	1/1113	38	1/167
30	1/895	39	1/128
32	1/659	40	1/97
34	1/446	44	1/30
35	1/356		

2 Learning through Serious Games

Learning patterns have changed radically in the last decades [36], and further developments are easily foreseeable: younger are "native speakers" in the language of digital media, and (video)games – serious or not – are one of the "Digital Natives" ways of interaction that spreads and evolves at an impressive pace. New generations are experiencing new forms of computer and videogame entertainment – and impaired kids are no exception! – and, as Prensky underlines, this new forms of entertainment has shaped their preferences and abilities while offering an enormous potential for their learning, both as children and as adults [25]. The idea of exploiting videogames as a teaching media has been "formalized" in 2002: in that year, the Woodrow Wilson Center for International Scholar in Washington, D.C. founded the Serious Games Initiative. Although we instinctively tend to associate "*serious games*" to videogames, this term was already in use long before computers and electronic devices became a way to convey entertainment [1]. In spite of their old story, no generally accepted and shared definition of what a Serious Game (SG) should be exists yet, anyway, in our opinion, the one provided by Zyda seems good enough for our goals: *Serious Game: a mental contest, played with a computer in accordance with specific rules, that uses entertainment to further government or corporate training, education, health, public policy, and strategic communication objectives* ([36], p.26). It is important to enlighten that SGs are not only a way to instruct people about something, but also way to convey knowledge within a context that is motivationally rich.

As we pointed out in the previous section, kids affected by DS feel often constrained and frustrated while interacting with other people. As a consequence, it is particularly relevant to explore and design new interaction patterns, able to overcome – at least partially – their difficulties. We hypnotized that SGs could be of help in overcoming the above limits; actually, kids affected by DS are exposed to new media and videogames rampaging diffusion exactly as their non-affected counterparts. Moreover, the usefulness of games as learning tools is a well known phenomenon, that becomes of overwhelming relevance in the first years of our life [6, 7, 27], and it is demonstrated that SGs are able to guarantee high learning effectiveness in quite short time [30, 31]; e.g., they are adopted as training tools in medical, military, and business fields. Moreover, as [30] point out, SGs are effective learning tools, able to train cognitive functions, such as memory, and to teach how to exert control over emotions [21, 22, 35]. This latter characteristic seems particularly relevant when dealing with kids affected by DS, which often experience more difficulties in managing their emotions, due also to the frustration and disappointment they feel when faced with tasks they can only hardly accomplish, but that they perceive as easily accessible to kids of the same age.

2.1 Coupling SGs with Feuerstein Methodology

Due to the peculiarities presented by kids affected by DS, we have decided to couple the serious game approach with Feuerstein methodology, which is based onto Reuven Feuerstein's theories of Structural Cognitive Modifiability (SCM) and of Mediated Learning Experience (MLE). SCM is *"the unique propensity of human beings to change or modify the structure of their cognitive functioning to adapt to the changing demands of a life situation."* [9]. The theory is rooted into the idea that human development is not only biological but also socio-cultural. That is to say, the development of cognitive capabilities and higher mental processes can be positively influenced by the human-environment interactions surrounding the kid. In particular, direct exposure and mediated learning experience are the two ways we can exploit to affect cognitive capabilities. At the very heart of SCM lies the theory of MLE [8], which is a quality of human-environment interactions at the basis of human modifiability. MLE is more than a mere pedagogical approach and it entails the shaping of cognitive process as a by-product of cultural transmissions. Due to its effectiveness, Feuerstein's methodology is more and more adopted with the aim of stimulating cognition in DS children. The application of the Feuerstein method requires the presence of a special "mediator", that is to say a person in charge of interpreting the child experiences when faced with a new task, and able to help her in addressing her efforts toward a specific goal. The importance of the mediator is further emphasized by the *mirror neurons* theory [3, 14, 16, 29]: mirror neurons provide effectively their functions thanks to the fact that they are "able" to bridge the actual and the virtual aspects of the world that surrounds us [4]. This latter characteristic seems particularly promising from the point of view of exploiting virtual playgrounds as environments for teaching how to accomplish simple "practical" tasks.

3 A Playcentric Approach

To prove our intuition, we have designed, developed, and tested the prototype of a SG aimed at teaching to kids affected by DS how to read the time. To reach this goal, we have adopted a user-centered approach. In particular, we have a bit simplified the Playcentric Approach to design often adopted in the game industry [13]. Fig.1 summarizes the steps of the approach, each of which should also go through an iterative process (generate ideas, formalize ideas, test ideas, evaluate results), that puts the focus on the players' feedbacks and game experience.

In particular, we went through steps 1, 4, 5, and 6 (shaded in gray in Fig. 1). Step 2 and 4 collapsed because in the virtual environment we choose to adopt as a development platform (see §4) it is very easy, quick, and cheap to create and modify content. Step 3 has been unnecessary: we had no subject that should be convinced to produce the game (for the same reason Step 7 has not been completed). The main critical point has been the first step of the process: brainstorming. We involved in the project the NGO "Associazione Vivi Down", whose activities aim at helping families with DS kids. Together with a psychologist of the association, we produced a list – with priorities – of the most critical teaching activities. Among these activities, the one found to be more effectively supported through a SG was "reading the time". Once the player experience goals were set ("*learn how to read the time in an intriguing way*") we explored different possible game concepts, keeping in mind which constraints derived from the peculiarities of our intended audience (see §5). All the subsequent phases went through the iteration process, and all the design choices have been validated by the psychologist. Actually, in spite of what the playcentric approach requires, no "player" has been involved during the development of the prototype, mainly due to the practical and legal difficulties of involving impaired minors. Moreover, it would have been quite difficult to define a player sample really representative, due to the fact that people affected by DS may present particularities that make each case different from all the others.

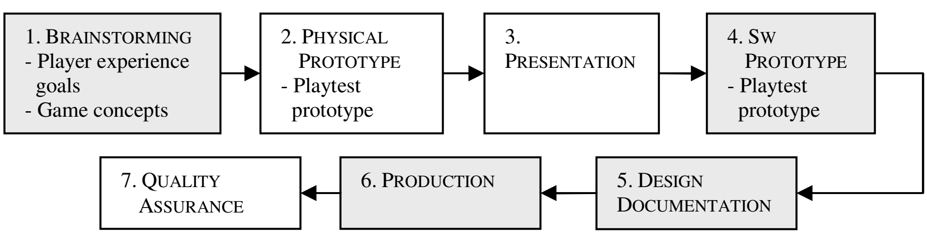


Fig. 1. The Playcentric approach to game design [13]

4 Choosing a Technical Solution for Developing the SG

Once the goal of the SG has been clearly defined, we had to select the most appropriate technical environment for developing the prototype. Several different alternatives have been explored, and the final choice has been a virtual world, namely Second Life (SL). We have chosen to develop our project in SL for several reasons, among which the most relevant are technical considerations and the quite "long" and rich tradition of medical research taking place into this virtual world (see, e.g., Fig.2, Fig.3, and [33]).

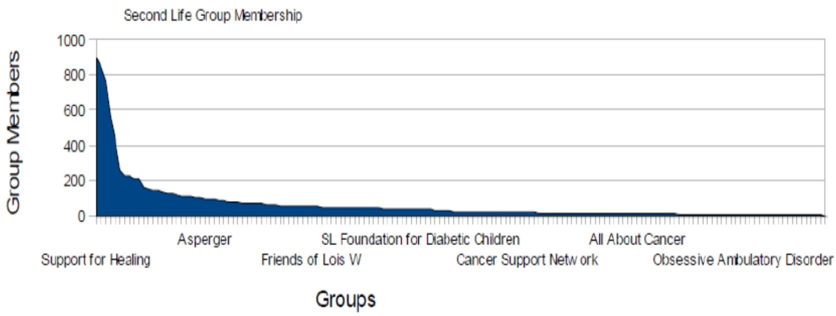


Fig. 2. Users (thousands) of the major support groups in SL [19]

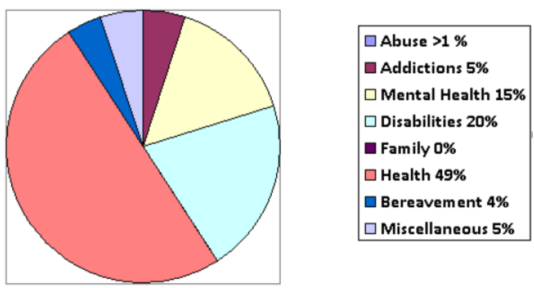


Fig. 3. Distribution of the groups active in the health/medicine areas in SL

4.1 Several Positive and Negative Features of SL

The main positive points of SL, from our perspective, are:

– a built-in intuitive and powerful modelling tool, that allows for a quick creation and editing of 3D complex objects;
– the availability of a scripting language (LSL – Linden Scripting Language) to easily create software applications and add interaction to objects;
– a user-friendly interface, with a smooth and short learning curve: this is very relevant if the target users are cognitive impaired people;
– the possibility to heavily customize avatars: this is of great help in increasing the immersion and identification sensations [28].

Unfortunately, SL also has several negative technical features, among which the most important is the fact that it is hard (if not impossible) to guarantee an high degree of privacy unless subscribing expensive contracts with Linden Lab (the company behind the virtual world). This is obviously a minus due the peculiar type of users we are dealing with. From a less technical point of view, SL offers opportunities that could be of great importance when the aim is developing a SG for impaired kids:

- the access to the virtual world, as user, is for free;
- SL has not pre-defined content, but every object it contains (from the bigger building to the tiniest pebble) has been created by its users;
- SL has a solid tradition in supporting medical and scientific activities: a great number of groups is are active in this area (see Fig. 2 and 3);
- SL is not a “game” in the common sense, but a real virtual world, characterised by a strong emphasis on social interactions and community building;
- a certain sensitivity for the DS is present among SL members: in 2009 the first virtual edition of the renown Buddy Walk took place in the virtual world. The Buddy walk (www.buddywalk.org) has been established in 1995 by the National Down Syndrome Society in the US.

Anyway, it is also true that maintaining a proprietary space in SL is quite costly, hence unaffordable for non-profits and NGOs (the typical subjects that offer support to families with impaired members, at least in Italy).

5 Developing a Prototype

According to Koster’s suggestions [18], we have tried to develop a sequence of actions aimed at giving hints about the cognitive patterns hidden in the game at a pace able to maintain high the players’ level of interest. In particular, we have considered the fact that people with DS tend to metabolize better new information and concepts when they are relevant, repeated and well structured/organized [24]. To simplify the task we avoided trying to teach also how to read seconds: it would have added a lot of difficulties compared to the utility of acquiring such capability. Learning objectives have been studied to be simple and hierarchically ordered: the first step is to learn how to read hours, then minutes, then the two together and, as a last more difficult step, to put in the correct position the hands, according to a specific hour communicated by the mediator. This implied to design clocks at different level of difficulty (e.g., representing only hours, only minutes and both) and facing problems like: recognizing correctly numbers (not always so easy for the younger kids), the distinction between A.M. and P.M. hours, among minutes and hours hands, among night and day, etc. In Tab.2 we have summarized several among the main difficulties we needed to face when designing the game, and how they have been addressed during the design phase.

Last but not least, since concepts seems to be learned more effectively through SGs when the are embedded into a meaningful context [37], the clocks have been put into a virtual version of the Vivi Down Centre (the NGO that hosted our project), in a building near the seaside. Unexpectedly we observed that kids involved in the testing phase were very attracted by this “natural” environment, and were delighted to bath their avatars into the digital sea. This further intriguing element – although unforeseen – could possibly have increased the efficacy of our SG.

The final prototype has been tested with a small pilot group of kids affected by DS, which were totally unable to read the hours before the experiment. The test has been conducted at the presence of a mediator that was in charge of helping the children accordingly to the Feuerstein approach. That is to say she triggered a feedback circle through which the kid could understand the strategy her action were supposed to follow, and hence could be able to continue the test by herself, self-mediating her actions in the virtual world through her avatar [2, 20].

Table 2. Difficulties experienced by kids and subsequent design choices (see also Fig. 4)

DIFFICULTY	DESIGN CHOICE
Distinguishing hours and minutes hands	Hands are represented with a fork and a spoon of different colours: both objects are easily recognizable and kids are familiar with them.
Distinguishing between hours and minutes	Three different clocks have been built: one for the hours, another for the minutes and the last one for hours and minutes together.
Distinguishing between AM and PM	A small window beside the clock shows the sun and the moon in different moments of the day/night.
Reading correctly hours that are not "precise"	A sign hanging over the clock shows the correct time in written form.
Experimenting	Two buttons (+ and -) beside the clock allow the user to change the time showed by the hands (and refresh its written form).
Following a learning path	The clocks have been appropriately positioned in a building, whose exploration implied encountering the clocks in the correct order: hours, minutes, and both.

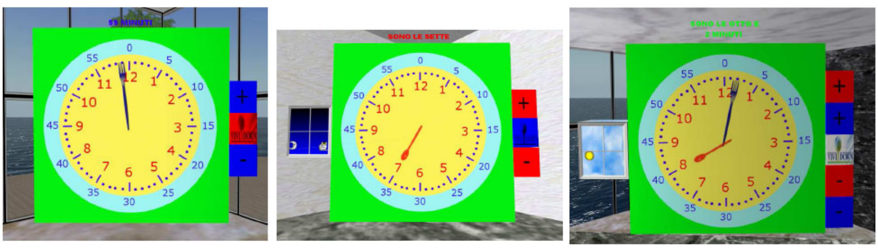


Fig. 4. From left to right: the minute clock, the hours clock and the complete clock. Notice the time in written form over the dial, the window showing night/day hours and the plus/minus buttons.

6 Testing the Prototype with DS Kids: What We Found Out

The test of our prototype involved six kids, among which three males and three females, aged 8-18. All of them were unable to read the hours prior the test, although educators had already tried to tech them the use of the clock, adopting traditional approaches. Only one kid at the time was allowed to interact with the digital environment, in order to help her to keep focused on what was going on. Each kid interacted with the SG for a minimum of 30 minutes to a maximum of one hour. It is important to underline that no parent and/or caretaker was allowed in the test room, except the mediator. This has been a precise choice, since too often parents tend to influence the kid choices – even unintentionally –, through their behaviour, both explicit and implicit. The kids have been left free to explore the digital environment by themselves. The mediator, according to the Feuerstain method, had the only task to stimulate the kid to (self)ask questions and to compare objects in order to discover relations among them. She also kept trace of the reactions and difficulties showed by the kid while interacting with the SG. As summarized in Tabs. 3, 4, and 5, the results seems encouraging, since the major part of the kids managed with little or no help to interact with the clocks showing minutes and hours. More difficulties aroused when the two concepts (hours and minutes) were brought together (Tab. 5). Nonetheless it has been interesting to notice that, in a limited time (30 to 60 minutes), almost half of the kids grasped the concept of reading the time, in spite of the fact they were nearly unable to understand it when explained through more traditional approaches.

Table 3. Results obtained by kid interacting with the hours clock

	Autonomously	Little help	Major help	Major difficulties
Using buttons	3	1	2	0
Distinguish among numbers (hours)	1	2	2	1
Reading hours	2	2	1	1
Distinguish among the 24 h	2	2	1	1
Read/tell correctly the time (12h)	2	2	1	1
Read/tell correctly the time (24 h, using window)	2	2	1	1
Tot.	*12*	*11*	*8*	*5*

Table 4. Results obtained by kid interacting with the minutes clock

	Autonomously	Little help	Major help	Major difficulties
Using buttons	3	2	1	0
Recognize minutes on the clock	3	1	1	1
Reading correctly minutes	3	1	2	0
Tell correctly the time (minutes)	3	0	1	2
Tot.	*12*	*4*	*5*	*3*

Table 5. Results obtained by kid interacting with the complete (hours and minutes) clock

	Autonomously	Little help	Major help	Major difficulties
Using buttons (hours and minutes)	0	3	2	1
Distinguish hours numbers	1	2	1	2
Recognize (minutes) numbers and signs	1	2	1	2
Distinguish between minutes/hours hands	5	0	0	0
Reading the time (o'clock)	3	0	2	1
Reading the time (whichever)	1	1	3	1
Tell the time (hours and minutes - 12h)	2	0	2	2
Tell the time (hours and minutes - 24h)	2	0	2	2
Tot.	*15*	*8*	*13*	*11*

An interesting side-effect we registered is the interest awaked in the kids by the virtual environment. Not only they bathed their avatars in the digital ocean, but they also wanted to know how to modify their appearance. It is also important to notice that they had little or no difficulties in understanding how to interact with the virtual environment, e.g., how to move the avatar.

7 Conclusions and Future Work

Although the test group was quite small, the results of this first testing phase have been positive: all the kids have been able to learn how to read the time in the virtual environment with fewer difficulties than those they experienced in the actual world. They also appreciated a lot the virtual environment, that tackled their curiosity and which they liked to explore actively with very few difficulties. This seems to enforce the hypothesis that a teaching approach rooted into the serious game paradigm could be of help for cognitively impaired people. Further development of the prototype and of the test could involve more children at the same time, also remotely connected and interacting with the mediator directly into the virtual environment.

Aknowledgements. We wish to thanks greatly Dr. Stefania Fois, Dr. Paola Borroni, the Associazione Vivi Down Onlus in Milano (Italy), and the six kids that experimented our prototype: without their passion, curiosity and dedication this work would have not be possible.

References

1. Abt, C.: Serious Games. The Viking Press, New York (1970)
2. Ambrosetti, U., Gualandri, V.: Inquadramento clinico, chirurgico e riabilitativo della persona con Sindrome di Down, Omega edizioni (2008)
3. Buccino, G., Vogt, S., Ritzl, A., Fink, G.R., Zilles, K., Freund, H.J., Rizzolatti, G.: Neural Circuits Underlying Imitation Learning of Hand Actions: an Event-Related fMRI study. Neuron (2004)
4. Canarini, F., William, J.B.: I mediatori in educazione speciale: mezzi, strumenti e metodiche, Franco Angeli (2008)
5. Cembrani, F.: Disabilità e libertà dal bisogno, Erikson Trento (2005)
6. Din, F.S., Calao, J.: The Effects of Playing Educational Videogames on Kindergarten Achievement. Child Study Journal 2, 95–102 (2001)
7. Durik, A.M., Harackiewicz, J.M.: Achievement Goals and Intrinsic Motivation: Coherence, Concordance, and Achievement Orientation. Journal of Experimental Social Psychology (2003)
8. Feuerstein, R., Feuerstein, S.: Mediated Learning Experience: A Theoretical Review. In: Feuerstein, R., Klien, P.S., Tannenbaum, A.J. (eds.) Mediated learning experience (MLE); Theoretical, psychosocial, and learning implications, pp. 3–51. Freund, London (1991)
9. Feuerstein, R., Feuerstein, S., Falik, L., Rand, Y.: Dynamic Assessments of Cognitive Modifiability. ICELP Press, Jerusalem (1979/2002)
10. Feuerstein, R., Klein, P., Tannenbaum, A.: Mediated Learning Experience, Theoretical, Psychosocial and Learning Implications. Freund, Tel Aviv and London (1991)
11. Feurestein, R., Rand, Y., Hoffman, M.B., et al.: Instrumental Enrichment: an Invention Program for Cognitive Modificability. University Park Press, Baltimore (1980)
12. Feuesrtein, R., Rand, Y., Hoffman, M.: The Dynamic Assessment of Retarded Performers: the Learning Potential Assessment Device (LPAD). University Park Press, Baltimore (1979)
13. Fullerton, T.: Game Design Workshop: a Playcentric Approach to creating Innovative Games. Morgan Kaufmann Publishers, Elsevier (2008)
14. Gallese, V.: The "Shared Manifold" Hypothesis: From Mirror Neurons to Empathy. Journal of Consciousness Studies 8(5/7), 33–50 (2001)

15. Geleherter, T.D., Collins, S.C., Ginsburg, D.: Genetica medica, Masson Ed (1999)
16. Iacoboni, M., Woods, R.P., Brass, M., Bekkering, H., Mazziotta, J.C., Rizzolatti, G.: Cortical Mechanisms of Human Imitation. Science 286, 2526–2528 (1999)
17. Kasari, C., Sigman, M.D.: Expression and Understanding of Emotion in Atypical Development: Autism and Down syndrome. In: Lewis, M., Sullivan, C. (eds.) Emotional Development in Atypical Children, pp. 351–360. Plenum, New York (1978)
18. Koster, R.: A Theory of Fun for Game Design. Paraglyph Press (2005)
19. Madara: Statistics for Second Life support groups (2008), `http://www.hcii2011.org/files/typeinst.doc`
20. Meadows, M.S.: I, avatar, The Culture and Consequences of Having a Second Life. New Riders Press, Berkeley (2008)
21. Michael, D., Chen, S.: Serious games: Games that Educate, Train and Inform. Thomson Course Technology, Boston (2006)
22. Mitchell, A., Savill-Smith, C.: The Use of Computer and Video Games for Learning: A review of the literature, Learning and Skills Development Agency (2004)
23. Nadel, L.: Down Syndrome in Neurobiological Perspective. In: Epstein, C.J. (ed.) The neurobiology of Down Syndrome, pp. 239–251. Raven Press, New York (1986)
24. Prensky, M.: Don't bother me mom, I'm Learning! Paragon House Publisher (2005)
25. Ritterfeld, U., Weber, R.: Video Games for Entertainment and Education. In: Vorderer, P., Bryant, J. (eds.) Playing Video Games – Motives, Responses, and Conseguences, Lawrence Erlbaum, Mahwah (2006)
26. Ripamonti, L.A., Di Loreto, I., Maggiorirni, D.: Handbook of Research on Socio-Technical Design and Social Networking Systems. IGI Global (2009); ISBN: 978-1-60566-264-0
27. Rizzolatti, G., Craighero, L.: Mirror neuron: a neurological approach to empathy. Springer, Heidelberg (2005)
28. Squire, K., Jenkins, H.: Harnessing the Power of Games in Education. Insight 3(1), 5–33 (2003)
29. Susi, T., Johannesson, M., Backlund, P.: Serious Games – An Overview. Technical Report HS-IKI-TR-07-001, University of Skovde, Sweden (2007)
30. Toro-Troconis, M., Mellström, U., Partridge, M., Meeran, K., Barrett, M., Higham, J.: Designing Game-Based Learning Activities for Virtual Patients in Second Life. Journal of CyberTherapy & Rehabilitation 1(3), 227–239 (2008)
31. Wishar, J.G.: The Development of Learning Difficulties in Children with Down's Syndrome. J. Intellect Disabilities Res. 37, 389–403 (1993)
32. Wong, W.L., Shen, C., Nocera, L., Carriazo, E., Tang, F., Bugga, S., Narayanan, H., Wang, H., Ritterfeld, U.: Serious Video Game Effectiveness. In: ACE 2007, Salzburg, Austria (2007)
33. Zyda, M.: From Visual Simulation to Virtual Reality to Games. Computer, vol. 38(9). IEEE Computer Society Press, Los Alamitos (2005)
34. Van Eck, R.: Digital Game-Based Learning: It's Not Just theDigital Natives Who Are Restless. EDUCAUSE Review 41(2) (2006)

Part V

Augmented Cognition and Interaction

A Longitudinal Study of P300 Brain-Computer Interface and Progression of Amyotrophic Lateral Sclerosis

Nathan A. Gates, Christopher K. Hauser, and Eric W. Sellers

East Tennessee State University, Department of Psychology
East Tennessee State University
807 University Pkwy
Johnson City, TN 37614-1700
{gatesn,sellers}@etsu.edu, hauser@goldmail.etsu.edu

Abstract. BCI can provide communication for people locked in by amyotrophic lateral sclerosis (ALS). Empirical examination of how disease progression affects brain-computer interface (BCI) performance has not been investigated. This pilot study uses a longitudinal design to investigate changes in P300-BCI use as ALS disability increases. We aimed to (a) examine the relationship between BCI accuracy and the ALS/Functional Rating Scale and (b) examine changes in the event-related potential (ERP) components across time. Eight subjects have been enrolled in the study. BCI accuracy was measured and ERP components were assessed by a principal component analysis (PCA). Two subjects have been followed for an average of nine-months, and BCI accuracy is 99.6%. While many research obstacles remain, these preliminary data help elucidate the relationship between BCI performance and disease progression.

Keywords: Amyotrophic lateral sclerosis, electroencephalogram, brain-computer interface, P300 event-related potential, assistive communication.

1 Introduction

Brain-computer interface (BCI) can restore communication for severely disabled individuals when all other communication options fail to produce reliable communication alternatives. Currently, people with locked-in syndrome (LIS) caused by amyotrophic lateral sclerosis are the population most likely to benefit from BCI, given the relatively slow communication rates BCIs offer. Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease affecting the lateral columns and anterior horns of the spinal cord [1]. The resulting atrophy to the muscle tissue will eventually lead to complete paralysis and anarthria within 3-5 years after diagnosis [2]. Before people enter LIS they typically choose to accept artificial ventilation. Once people with ALS enter LIS or complete (CLIS) social isolation can become severe. LIS is defined as near-complete paralysis with the retention of voluntary eye movements, eye blinks and limited facial twitches [3]. CLIS is a condition whereby all motor control is lost [4]. A number of electroencephalographic (EEG) methods have been used in attempts to restore communicative ability for people with ALS. For example: sensorimotor rhythms (SMR; [5, 6]); slow cortical potentials (SCPs; [7-9]);

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 475–483, 2011.

steady state visual evoked potentials (SSVEP; [10-12]); and event-related potentials (ERPs; [13-15]). Currently, ERP methods (i.e., P300-based BCIs) have shown to be the most promising BCI for people with LIS for daily communication, environmental control, and regained social interaction [16].

Nonetheless, the relationship between BCI performance and disease progression has received little attention. When CLIS has been reached before BCI use begins, people with ALS are unable to use BCI technology [3]. For this subset of the population, who are unable to communicate by any means, a definitive causal explanation cannot be determined [17]. In addition, no LIS person successfully using a BCI has progressed to CLIS. A hypothesis put forth by [18] is that people who can effectively use a BCI are less likely to enter CLIS.

A few studies have begun to look at how BCI performance is affected by factors such as mood and disease severity. Silvoni et al. [19] investigated the relationship between clinical status, age, and P300-BCI performance in a sample of 21 individuals with ALS. Data were collected in an auditory oddball and a 4-choice visual oddball task. One year later the same methods were repeated in a follow-up study which showed no significant differences in P300 morphology. However, only five of the original 21 participants completed the second part of the study. Thus, the attrition rate of this study limits the generalizability of the findings, which is a common obstacle for ALS research. Another study investigated the effects of motivation and psychological state on BCI performance [20]. Six participants with advanced ALS were chosen for the study and assigned to either a SMR-based BCI group with a block of 20 sessions or a P300-BCI group with a block of 10 sessions. Questionnaires assessing quality of life and severity of depression were completed before each session to determine mood and motivation. The authors concluded that motivational factors relate to BCI performance in individual subjects; however, they did not report if ALS progression had effects on waveform morphology or ERP components.

Farwell and Donchin introduced the first P300-BCI in 1988 [15]. In their paradigm a 6x6 matrix of characters were presented on-screen and participants are instructed to attend to a target item. Items are flashed in groups of rows and columns, the row-column paradigm (RCP). A recent modification to the original P300-BCI presents stimuli in a quasi-random method whereby characters appear to flash independently of one another [21]. This alternative to the RCP has been termed the checkerboard paradigm (CBP). Research has indicated that the CBP optimizes user performance by reducing the number of errors inherent to the RCP, thereby increasing the speed and accuracy of the BCI speller device [21].

Principal component analysis (PCA) has been widely used for several decades as a method to analyze ERP data (e.g. [22-25]). PCA is a data driven method for analyzing ERP waveforms, which removes ambiguity associated with traditional and/or subjective component identification methods (i.e., peak and area measures; [26]). The largest source of covariance in a PCA of EEG data are assumed to be ERP components, which are characteristic features of the ERP waveform spread across multiple time points and electrodes [27]. The P300 ERP component is a positive deflection usually observed over posterior parietal sites (common average, nose, linked mastoid references) and is time-locked to a personally relevant or meaningful event. The classic paradigm used to elicit a P300 is the oddball paradigm [28], where rare deviant stimuli (i.e., targets) occur at random among a series of higher frequency standard stimuli (i.e., non-targets; [29]). In the current study we present preliminary findings from a longitudinal study examining P300-BCI use throughout disease progression. We aimed to: 1) examine the relationship between BCI performance and level of disability in ALS and 2) examine ERP component structure across time.

2 Methods

2.1 Participants

Eight subjects diagnosed with ALS (El Escorial probable) have been enrolled in the study (5 male, 3 female). Subjects were recruited with the assistance of the Kingsport and Knoxville, Tennessee chapters of the ALS Association, and the Duke ALS Clinic. After enrolling in the study two subjects were deceased after completing one and two sessions, respectively. Three additional subjects have completed three or fewer sessions. One subject completed more than three sessions; however, recording problems prevented data analyses. The remaining two subjects completed eight and four sessions, respectively. The data from these two subjects will be the primary focus of this study. The study was reviewed and approved by Institutional Review Board of East Tennessee State University, and each subject gave informed consent to participate in the study.

2.2 Paradigm

The level of physical disability was assessed in each participant using the revised ALS functional rating sale (ALSFRS-R; [30]). The ALSFRS-R scale includes categories for respiratory, autonomic, and motor functions.

Subjects completed a 21 item copy-spelling calibration phase without feedback, then a 14 item online copy-spelling task, with feedback, for an additional 14 items. The procedure was conducted at intervals of approximately 1.6 months. The P300-BCI used the checkerboard paradigm (see [21] for a complete description) with a 6x6 matrix [15]. Subjects sat approximately 1 m from a computer monitor that displayed the matrix of characters. The subjects were instructed to attend to the desired character and count the number of times the item flashed. Following collection of calibration data a stepwise linear discriminant (SWLDA; see below for description) classifier was derived and then used during the online classification portion of the session [31]. The interstimulus interval (ISI) was 62.5 ms with a stimulus flash duration of 187.5 ms (a stimulus onset asynchrony [SOA] of 250 ms), for copy-spelling and online feedback conditions. For each item selection, each item in the matrix was flashed 20 times in quasi-random groups of four or five items before the classifier chose the single item with the highest SWLDA score.

2.3 Data Acquisition

16-channel EEG was recorded (right mastoid reference left mastoid ground) at 256 Hz and bandpass filtered (range = 0.05 to 30 Hz) using a g.tec biosignal amplifier (g.USBamp version 2, Guger Technologies). Impedances were reduced to below 10.0

kΩ before recording. Electrodes Fz, Cz, Pz, Oz, P3, P4, PO7, and PO8 were used for online BCI operation [16]. BCI2000 software was used for stimulus presentation, and data collection.

2.4 Online Feature Classification and Performance

A SWLDA algorithm was used to determine spatiotemporal features of the EEG that accurately discriminate target and non-target flashes (MATLAB version 7.6 R2008a, stepwisefit function was used to derive the classifier). Classifiers were derived using the eight electrode montage described in [16], sixteen feature from each of the eight electrodes (see above), and each of the 21 item selections (i.e., 2688 features 1/6th target items), were submitted to the SWLDA algorithm (from 0 ms from stimulus onset to 800 ms post onset). The resulting SWLDA coefficients were then used for online classification. The SWLDA coefficients for each item in the matrix were summed and the item with the highest score was selected and presented to the participant as feedback.

2.5 Offline ERP Analysis

Blink artifacts were removed using a single value decomposition procedure on the continuous data (NeuroScan, Inc. [2003]; El Paso, TX). Average target and non-target waveforms were computed from artifact-free EEG epochs (855 ms, 55 ms prestimulus baseline) for all 35 characters of each recording session. Averaged scalp potential amplitude data were submitted to a temporal principal component analysis (PCA) derived from the covariance matrix followed by unrestricted Varimax rotation of the covariance loadings [26, 32, 33]. The data matrix was comprised of 219 variables (timepoints, -55 to 800 ms) and 8,960 cases including 8 sessions (or fewer cases for the subjects with fewer than eight sessions), 35 characters, 2 conditions (target/non-target), and 16 electrode sites. PCA factor loadings and factor score topographies were used as a data driven method of selecting appropriate time windows and electrode locations for statistical analyses of ERP components. PCAs were run separately for each participant. A repeated-measures Analysis of Variance (ANOVA) was used to investigate potential changes in ERP component amplitude across time (i.e., session). Greenhouse-Geisser epsilon (ε) correction was used to compensate for violations of sphericity when appropriate.

3 Results

3.1 Online Performance and ALSFRS-R

Mean online accuracy was computed for each of the eight subjects, regardless of the number of completed sessions. Overall accuracy was high for all eight subjects (mean, 95.83%, SE, 2.84). Mean ALSFRS-R was 24.8 (SE, 8.30). Subject 1 decreased from 38 to 19 and Subject 2 remained at 23 throughout the study.

3.2 ERP Component Findings

Averaged ERP waveforms for Subjects 1 and 2 show comparable and stable component structure, predominantly characterized by a large bilateral central-to-frontal distributed positivity (peak latency approximately 200 ms) and later negativity (peak latency approximately 500 ms). PCA factors were highly comparable between the participants and revealed two large variance factors that corresponded to the temporal and spatial characteristics of the positive and negative peaks observed in the scalp potentials (see Fig.1, left panels). The factor score topographies corresponding to the early positivity was highly similar between the two subjects (P227 and P188 for Subject 1 and 2, respectively), showing a broad central-to-frontal distribution with a Cz maxima. However, the factor score topographies corresponding to the later negativity slightly differed in scalp distribution between participants; specifically, Subject 1 showed a Cz maxima (N493) whereas Subject 2 showed a Pz maxima (N552). Figure 1 (right panels) show the electrode sites for the subjects that most concisely represent the activity for the early positive (P227 and P188) and late negative (N493 and N552) components. Windows selected for analyzing each component's mean amplitude were guided by the factor loadings (see Fig. 1, left panels).

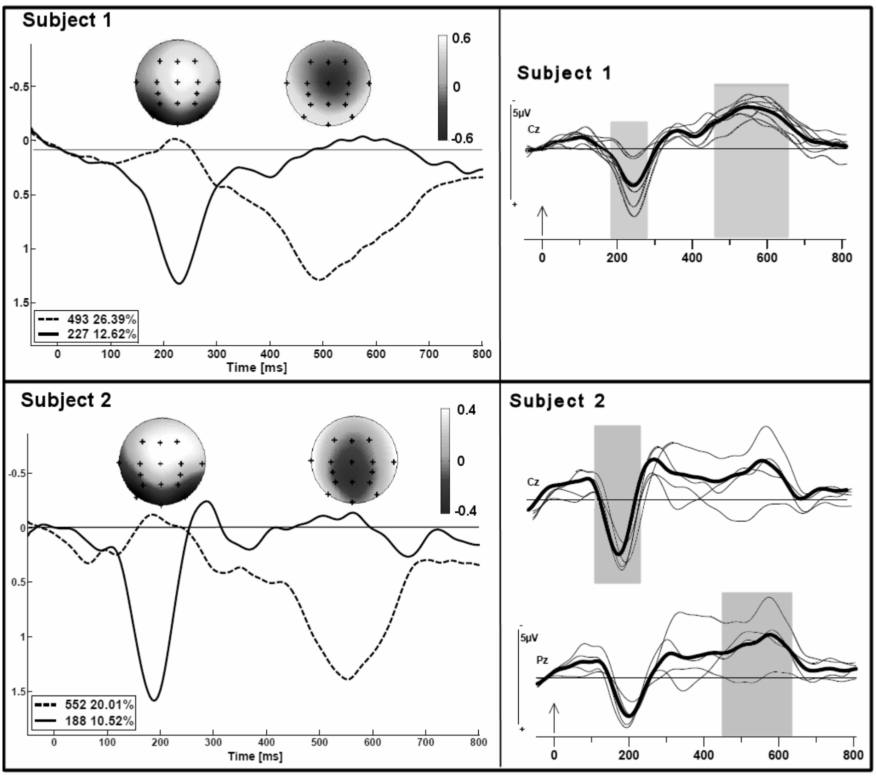


Fig. 1. PCA factor loadings and factor score topographies for subject 1 (*top left*) and subject 2 (*bottom left*) reveal two high variance and highly comparable temporospatial. factors that unambiguously correspond to an early positive and late negative deflection in the surface potential waveforms. Temporal windows used for analyzing peak amplitudes were obtained from the PCA loadings for each subject. The average windows and electrodes used for statistical analyses are shown for subject 1 (*top right*) and subject 2 (*bottom right*), bold line represents mean morphology across all sessions.

Mean amplitude of the early positive component for Subject 1 (P227, 180 – 280 ms window) varied significantly across session ($F_{(1,7)}$ = 9.42, p = 0.001, ε = 0.22), whereas mean amplitude for the later negative component did not (N493, 450-660 ms window; $F_{(1,7)}$ = 1.69, p = 0.15, ε = 0.05). Conversely, mean amplitude of the early positive component for Subject 2 (P188, 125 – 250 ms window) did not vary across recording session ($F_{(1,3)}$ = 2.39, p = 0.09, ε = 0.07), whereas the later negative component showed significant variation across time (N550, 450 – 650 ms window; $F_{(1,3)}$ = 38.29, p < 0.001). Significant deviations in ERP component amplitude were observed across sessions for both subjects. Given the lack of variability in accuracy and ALSFRS-R scores, it was not possible to statistically ascertain the relationship between component amplitude and disease progression.

4 Discussion

This research investigates how BCI performance (i.e., accuracy) changes as disability related to ALS becomes more severe. We proposed that BCI performance would remain stable as disability (as measured by ALSFRS-R) increases. Eight subjects have participated in the study to date. Only one subject has shown a substantial decrease in ALSFRS-R score from 38 to 19 over the course of 13 months. As functional ability decreased BCI performance remained 99.13% throughout eight sessions. These are novel and interesting results that may have implications for how BCI protocols should be developed and when to begin using BCI technology.

Offline ERP component analyses revealed a highly comparable and simple component structure between and within subjects. Although the relative amplitudes for components explaining the greatest variance differ across recording sessions, accuracy remained stable. This suggests that overall functional decline does not reflect a systematic decline or disturbance the brain mechanisms responsible for successful BCI communication. Furthermore, the findings that (a) mean amplitude did not consistently vary between subjects for ERP components and that (b) subjects performed at an optimal level across sessions suggests that neither component is solely responsible for successfully detecting targeted characters within the BCI speller matrix. Moreover, this later finding further demonstrates the robustness of the SWLDA classification algorithm for use with P300-based BCIs.

The current findings should be validated with a much larger sample of subjects, and it is essential for subject recruitment to begin soon after a diagnosis of ALS. The problematic nature of research with an ALS population is demonstrated by the limited number of subjects typically reported in BCI studies. In the current study the problem of subject identification and retention is exacerbated because we are following the subjects for many months. A strong relationship with augmentative and alternative communication specialists is an essential factor to help increase enrollment. Also, people with ALS may be reluctant to enter a study soon after diagnosis, when BCI use offers them no direct benefit. In fact, the typical BCI user or subject in a study is a person for whom all other assistive technology has already failed (i.e., they already have LIS).

Nonetheless, this study is an important first step in understanding how disease progression affects BCI use and can potentially lead to determining the optimal time to begin use of BCI technology for severely disabled people.

Acknowledgments. We thank the participants in this study for their time and effort. We thank Gerry E. Frye for helpful comments and manuscript preparation. We also thank Charles L. Brown, III (New York State Psychiatric Institute) for sharing his 'Disaver' waveform plotting software. This work has been supported by NIBIB and NINDS, NIH (EB00856); NIDCD, NIH (R21 DC010470-01); NIDCD, NIH (1 R15 DC011002-01).

References

1. Rowland, L.P., Shneider, N.A.: Amyotrophic lateral sclerosis. The New England journal of medicine 344, 1688–1700 (2001)
2. Turner, M.R., Scaber, J., Goodfellow, J.A., Lord, M.E., Marsden, R., Talbot, K.: The diagnostic pathway and prognosis in bulbar-onset amyotrophic lateral sclerosis. Journal of the neurological sciences 294, 81–85 (2010)
3. Kubler, A., Birbaumer, N.: Brain-computer interfaces and communication in paralysis: extinction of goal directed thinking in completely paralysed patients? Clin Neurophysiol. 119, 2658–2666 (2008)
4. Bauer, G., Gerstenbrand, F., Rumpl, E.: Varieties of the locked-in syndrome. Journal of neurology 221, 77–91 (1979)
5. Wolpaw, J.R., McFarland, D.J.: Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. Proc. Natl. Acad. Sci. U S A 101, 17849–17854 (2004)
6. McFarland, D.J., Miner, L.A., Vaughan, T.M., Wolpaw, J.R.: Mu and beta rhythm topographies during motor imagery and actual movements. Brain Topogr. 12, 177–186 (2000)
7. Anders, S., Eippert, F., Wiens, S., Birbaumer, N., Lotze, M., Wildgruber, D.: When seeing outweighs feeling: a role for prefrontal cortex in passive control of negative affect in blindsight. Brain 132, 3021–3031 (2009)
8. Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kubler, A., Perelmouter, J., Taub, E., Flor, H.: A spelling device for the paralysed. Nature 398, 297–298 (1999)
9. Birbaumer, N., Kubler, A., Ghanayim, N., Hinterberger, T., Perelmouter, J., Kaiser, J., Iversen, I., Kotchoubey, B., Neumann, N., Flor, H.: The thought translation device (TTD) for completely paralyzed patients. IEEE Trans. Rehabil. Eng. 8, 190–193 (2000)
10. Bin, G., Gao, X., Yan, Z., Hong, B., Gao, S.: An online multi-channel SSVEP-based brain-computer interface using a canonical correlation analysis method. Journal of neural engineering 6, 46002 (2009)
11. Bin, G., Lin, Z., Gao, X., Hong, B., Gao, S.: The SSVEP topographic scalp maps by canonical correlation analysis. In: Conf. Proc. IEEE Eng. Med. Biol. Soc., pp. 3759–3762 (2008)
12. Cecotti, H.: A self-paced and calibration-less SSVEP-based brain-computer interface speller. IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society 18, 127–133 (2010)

13. Sellers, E.W., Donchin, E.: A P300-based brain-computer interface: initial tests by ALS patients. Clin Neurophysiol. 117, 538–548 (2006)
14. Donchin, E., Spencer, K.M., Wijesinghe, R.: The mental prosthesis: assessing the speed of a P300-based brain-computer interface. IEEE Trans. Rehabil. Eng. 8, 174–179 (2000)
15. Farwell, L.A., Donchin, E.: Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalogr Clin Neurophysiol. 70, 510–523 (1988)
16. Sellers, E.W., Vaughan, T.M., Wolpaw, J.R.: A brain-computer interface for long-term independent home use. Amyotroph Lateral Scler 11, 449–455 (2010)
17. Murguialday, A.R., Hill, J., Bensch, M., Martens, S., Halder, S., Nijboer, F., Schoelkopf, B., Birbaumer, N., Gharabaghi, A.: Transition from the locked in to the completely locked-in state: A physiological analysis. In: Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology, December 8 (2010)
18. Birbaumer, N., Cohen, L.G.: Brain-computer interfaces: communication and restoration of movement in paralysis. J. Physiol. 579, 621–636 (2007)
19. Silvoni, S., Volpato, C., Cavinato, M., Marchetti, M., Priftis, K., Merico, A., Tonin, P., Koutsikos, K., Beverina, F., Piccione, F.: P300-Based Brain-Computer Interface Communication: Evaluation and Follow-up in Amyotrophic Lateral Sclerosis. Front Neurosci. 3, 60 (2009)
20. Nijboer, F., Birbaumer, N., Kubler, A.: The influence of psychological state and motivation on brain-computer interface performance in patients with amyotrophic lateral sclerosis - a longitudinal study. Frontiers in neuroscience 4 (2010)
21. Townsend, G., LaPallo, B.K., Boulay, C.B., Krusienski, D.J., Frye, G.E., Hauser, C.K., Schwartz, N.E., Vaughan, T.M., Wolpaw, J.R., Sellers, E.W.: A novel P300-based brain-computer interface stimulus presentation paradigm: moving beyond rows and columns. Clin Neurophysiol. 121, 1109–1120 (2010)
22. Donchin, E.: A multivariate approach to the analysis of average evoked potentials. IEEE Trans. Biomed. Eng. 13, 131–139 (1966)
23. Donchin, E.: Discriminant analysis in average evoked response studies: the study of single trial data. Electroencephalogr Clin Neurophysiol 27, 311–314 (1969)
24. Spencer, K.M., Dien, J., Donchin, E.: A componential analysis of the ERP elicited by novel events using a dense electrode array. Psychophysiology 36, 409–414 (1999)
25. Dien, J., Spencer, K.M., Donchin, E.: Localization of the event-related potential novelty response as defined by principal components analysis. Brain Res. Cogn. Brain Res. 17, 637–650 (2003)
26. Kayser, J., Tenke, C.E.: Optimizing PCA methodology for ERP component identification and measurement: theoretical rationale and empirical evaluation. Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology 114, 2307–2325 (2003)
27. Donchin, E., Coles, G.H.: While an undergraduate waits. Neuropsychologia 29, 557–569 (1991)
28. Fabiani, M., Gratton, G., Karis, D., Donchin, E.: Definition, Identification, and Reliability of Measurement of the P300 Component of the Event Related Potential. Advances in Psychophysiology 2, 1–78 (1987)
29. Squires, K.C., Donchin, E., Herning, R.I., McCarthy, G.: On the influence of task relevance and stimulus probability on event-related-potential components. Electroencephalogr. Clin. Neurophysiol. 42, 1–14 (1977)

30. Cedarbaum, J.M., Stambler, N., Malta, E., Fuller, C., Hilt, D., Thurmond, B., Nakanishi, A.: The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS Study Group (Phase III). J. Neurol. Sci. 169, 13–21 (1999)
31. Draper, N.R., Smith, H.: Applied regression analysis, 2nd edn. Wiley, New York (1981)
32. Kayser, J., Tenke, C.E.: Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns: II. Adequacy of low-density estimates. Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology 117, 369–380 (2006)
33. Kayser, J., Tenke, C.E.: Trusting in or breaking with convention: towards a renaissance of principal components analysis in electrophysiology. Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology 116, 1747–1753 (2005)

Discovering Context: Classifying Tweets through a Semantic Transform Based on Wikipedia

Yegin Genc, Yasuaki Sakamoto, and Jeffrey V. Nickerson

Center for Decision Technologies, Stevens Institute of Technology
Castle Point on Hudson, Hoboken, NJ 07030 USA
{ygenc,ysakamot,jnickerson}@stevens.edu

Abstract. By mapping messages into a large context, we can compute the distances between them, and then classify them. We test this conjecture on Twitter messages: Messages are mapped onto their most similar Wikipedia pages, and the distances between pages are used as a proxy for the distances between messages. This technique yields more accurate classification of a set of Twitter messages than alternative techniques using string edit distance and latent semantic analysis.

Keywords: Text classification, Wikipedia, semantics, context, cognition, latent semantic analysis.

1 Introduction

Humans are experts in recognizing new and useful messages while ignoring others. They do this by extracting meaning from messages, categorizing messages with related meaning into the same topics, and noticing information that does not fit any existing categories. Attempts to automate this fundamental ability of cognition using semantic models still leave room for improvement (e.g. [1]).

We study how we can categorize messages streaming through Twitter. These messages, called tweets, come in at a rate of more than 600 a second [2], and are often cryptic. If we can find ways of categorizing messages and recognizing new and useful topics in this noisy environment, we may provide automated tools with pragmatic uses: Twitter functions as a large sensor system, and can increase our awareness of our surroundings (e.g. [2-5]).

In an organizational context, the act of decrypting confusing or novel information is called sense-making, and is accomplished in part by asking other people what they think [6]. In artificial intelligence, this act is related to the understanding of context. For example, tweets might be looked up in Wikipedia, and the closest entry to a tweet found [7]. This technique is a promising instance of a larger idea, in which machine algorithms inform themselves by seeking out contextual information [8-12].

In the present paper, we apply several semantic models to tweet categorization. We introduce a Wikipedia-based classification technique. Extending the insight of Michelson and Macskassy [7], we develop a technique for calculating semantic distances between messages based on the distances between their closest Wikipedia

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 484–492, 2011.

pages: in effect, we regard Wikipedia as a transform space in which measurements can be made. We next describe this classification technique, and two other techniques we will use for comparison.

2 Classification Techniques

Classifying tweets is not an easy task because statistical methods of text classification have difficulty on short texts [13]. Moreover, if emerging topics of conversation are regarded as signal, the vast majority of tweets would be characterized as noise.

Some past studies of tweet classification have examined the use of specific features, such as emoticons [14] and author profiles [15], in improving the classification performance. Other studies have regarded tweets as a window into customer perception [16]; then, the challenge becomes recognizing sentiment. In contrast to these past work, we are interested in categorizing tweets in order to detect topics, which requires the ability to cluster tweets without a priori knowing which features will be important.

Recent work on extracting topics from short texts relies on knowledge bases to find context that is not in the texts. For example, Stone et al. used Wikipedia as training corpus to improve the ability of statistical methods to discover meanings of short texts [17]. Similarly, Gabrilovich and Markovitch used concepts derived from Wikipedia to identify the semantic relatedness of texts [9]. Wikipedia was also used by Michelson and Macskassy [7]: since we build on their model we will discuss it in the next section, after giving an overview of our own process.

Tweets are classified in three steps. First, between-tweet distances are calculated using one of the techniques described next. We map the tweets onto two-dimensional planes using multidimensional scaling (MDS) of the between-tweet distances. MDS helps us interpret the underlying relationships in the data, by allowing us to visually examine the clustering of tweets, similarity between clusters, and the size and internal consistency of the clusters. Then we use discriminant function analysis to measure how well each technique can predict the category memberships of the tweets [18]. Discriminant function analysis predicts a categorical dependent variable, in this case the tweet category, by one or more independent variables, in this case the two-dimensional MDS solution (i.e., the x and y coordinates).

There are many ways of building a between-tweet distance matrix. We picked String Edit Distance [19] and Latent Semantic Analysis [20] to compare with the Wikipedia-based algorithm presented next.

2.1 A Semantic Transform Using Wikipedia

Michelson and Macskassy have developed a model that discovers topics of interests of Twitter users based on their tweets: capitalized non-stop words in tweets are linked to Wikipedia pages; then, topics are derived from socially tagged categories listed in the linked Wikipedia pages [7]. Specifically, topics are discovered by traversing the tree structure of the taxonomy of Wikipedia. We apply a similar technique to a different end. We are interested not in determining the topic of a particular user's set of posts, but instead, understanding topic emergence across many users (e. g. [3, 21]).

In order to do so, we will create a distance matrix between tweets. As shown in Fig. 1, there are two stages: (1) we map tweets to Wikipedia pages, and then (2) compute the distance between the Wikipedia pages as a measure of semantic distance between the tweets. More formally, we regard Wikipedia as a transform space, in which we measure the between-tweet distances:

$$d(message_1, message_2) \propto d(T(message_1), T(message_2)),$$

where d is a measure of semantic distance, and T is a transformation function mapping a message to a page in Wikipedia. The transformation is worth performing for two reasons. First, humans categorize Wikipedia pages based on their meanings, and thus the Wikipedia networks likely reflect semantic networks in human brains. This can be helpful because we are dealing with tweets, short texts humans write online. Second, Wikipedia pages are mapped to categories that are named by the crowd. These categories can serve as topics, eliminating the problem of inferring the meaning of latent topics in Latent Semantic Analysis and other statistical methods.

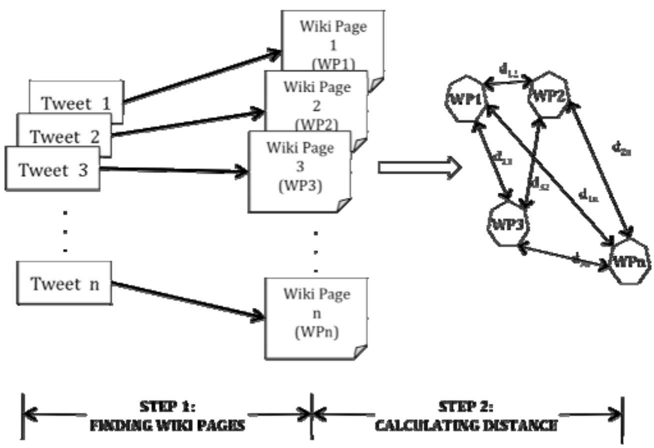


Fig. 1. The two steps involved in calculating distances between tweets using Wikipedia. We use the distance between the two associated Wikipedia pages as an indicator of the distance between the two tweets.

Finding Associated Wikipedia Pages. To associate a tweet to a Wikipedia page, we first identify a set of words for this tweet. The word set includes all the words in the tweet after eliminating certain words in the English stop-words list provided in the LSA package for R [23]. For each word, we check to see if there is a direct page dedicated to the word, and if there is a disambiguation page. The disambiguation page provides a precise mapping to the right page, leading to more accurate distance measures. Then a list of candidate pages for the tweet is found by aggregating each page associated with each word of the word set. We compute a score for each candidate page by counting the number of occurrences of the words in the word set. The page with the highest score is selected as the associated Wikipedia page for the tweet. This process is visualized in Fig. 2.

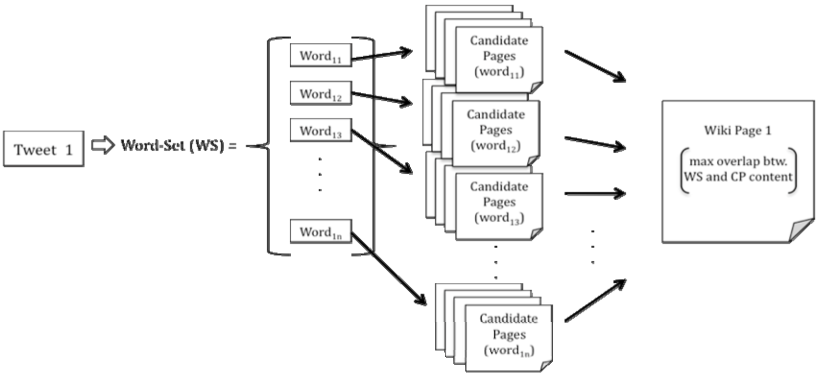


Fig. 2. Finding a Wikipedia page associated with a tweet

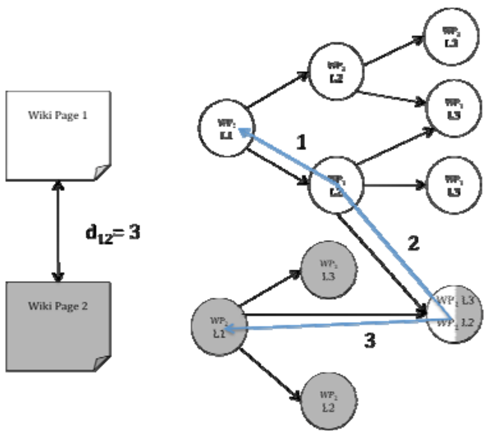


Fig. 3. Calculating the distance between two Wikipedia pages

Calculating Distances. The distance of two Wikipedia pages is calculated based on the link between the categories associates with these two pages. Categories of the Wikipedia pages are linked to one another in a graph structure. A category can be linked to multiple parent categories. We capture the network structure of categories for each Wikipedia page for five levels. We compute the semantic distance between the two Wikipedia pages by finding the length of the shortest path from a category of one page to a category of the other page. Fig. 3 shows an example.

2.2 String Edit Distance

The String Edit Distance (SED) method may work given that tweets are short: related tweets might contain the same set of keywords. In the SED method, the distance between two tweets is found by calculating the number of edits it takes to transfer one tweet to another, also called Levenshtein distance. As an example, the Levenshtein distance between “kitten” and “sitting” is 3 since (1) ‘k’ is replaced by ‘s’ (2) ‘e’ is replaced by ‘i’ and (3) ‘g’ is added to the end. In calculating the distance between two tweets, we normalize their distance value by dividing it by the string length of the longer tweet. If tweets in the same category tend to use the same keywords, within-category SED should become smaller than between-category SED. We used Bibiko’s R package [22] for calculating the between-tweet distances.

2.3 Latent Semantic Analysis

The Latent Semantic Analysis (LSA) method is a broadly applied text processing technique [20]. LSA represents a set of tweets in a term by tweet matrix. A row in the matrix is a unique term. A column in the matrix is a tweet. Each matrix cell contains the frequency of each term within each tweet. This term-by-tweet matrix goes through singular value decomposition. Like principal component analysis, the factors are ordered by the amount of variance they capture in the original matrix. By using only the most influential factors, one can create an approximation of the original matrix, which removes the noise associated with the particular text sample, and uncovers somewhat abstract commonalities in word usage patterns. The approximated matrix yields a vector representation of terms with dimensionality equal to the number of factors included. The pair-wise similarities of all tweets are calculated by taking the vector cosine of the two tweets’ vectors.

Essentially, LSA exposes the similarity relations among related words by measuring how often these words appear together. In doing so, it reduces the dimensionality from thousands (i.e., the number of unique words in all documents) to hundreds. Tweets, however, are much shorter than typical documents used in LSA.

We used the LSA package for R by Wild [23]. We did not use the stemming option that reduced the words to the word-stems; we did use the English stop-words list provided in the package. An approximated term-by-tweet matrix was obtained. Because our analyses included 100 tweets or less, we used the few dimensions whose sum of singular values equaled or exceeded half of sum of singular values of all dimensions. Cosine similarities, which ranged between -1 and 1, were transformed to distances by subtracting these similarities from one and adding an epsilon value.

3 Method and Results

As an initial test, we applied the SED, LSA, and Wikipedia models to two sets of tweets that can be easily classified into three categories by humans. The first set had 45 tweets, consisting of 15 tweets from each of the three events that occurred at the time of our data collection: (1) death of J. D. Salinger, an American author, (2) an earthquake in Haiti, and (3) the release of iPad, Apple’s tablet computer. In the first set, the categories of all tweets were known.

The second set included all the tweets from the first set, and an additional 55 tweets that were randomly selected from the same time period. The addition of randomly sampled tweets tested the robustness of the classification techniques in a noisier environment.

The tweets were pre-processed by replacing any non-alphanumeric characters with the space character. The tweet-by-tweet distance matrix was obtained for each technique as described previously. A two-dimensional solution from multidimensional scaling of the distance matrix was used to predict the category membership of tweets in discriminant function analysis. We compared the techniques by looking at their accuracies of classifying tweets based on true positives and true negatives, using leave-one-out cross validation.

Table 1. Accuracy (hit plus correct rejection) of classifying 45 tweets with known categories

Technique	J. D. Salinger	iPad	Haiti
String Edit Distance	.67	.13	.60
Latent Semantic Analysis	.67	.73	.80
Wikipedia	.93	.87	.80

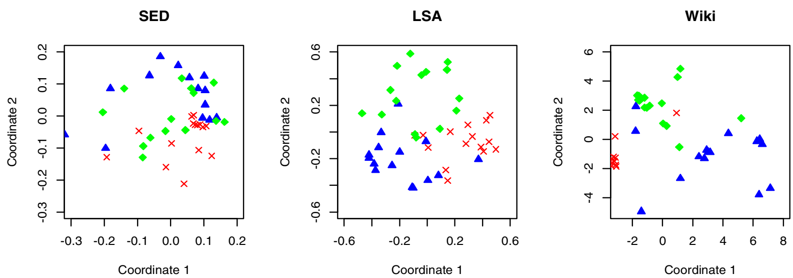


Fig. 4. Forty-five tweets with known categories mapped onto two-dimensional planes using multidimensional scaling of the between-tweet distances based on String Edit Distance, LSA and Wikipedia. An *x* is a tweet about J. D. Salinger and a triangle is a tweet about the iPad.

3.1 Tweets with Known Categories

Table 1 summarizes the classification accuracy of each technique. LSA and the Wikipedia model clearly performed better than SED. As shown in the Fig. 4, the Wikipedia distance measure yielded better-delineated clusters of related tweets than LSA. Interestingly, the Wikipedia model yielded a tight cluster for the tweets about J. D. Salinger, which generally discussed the unique topic of the author' s death; on the other hand, tweets about the iPad addressed a looser assortment of topics and, thus, clustered more loosely. In the Wikipedia distance space, there was one J. D. Salinger tweet that was far apart from other J. D. Salinger tweets. This tweet contained neither the author's name nor the title of his best-known book, which other tweets mentioned.

3.2 Adding Randomly Sampled Tweets

Table 2 summarizes the performances of LSA and Wikipedia for the data set containing randomly selected tweets. SED was dropped because of its weak performance in the first data set. When randomly sampled tweets were added, the Wikipedia model clearly outperformed LSA.

LSA's performance significantly deteriorated; that is, LSA was very sensitive to the addition of the other tweets. LSA processes terms in relation to what other terms appear in the corpus; thus LSA is highly affected by the context. On the other hand, the Wikipedia distance measure is robust to the addition of the randomly selected tweets, because the distance calculation is based on the Wikipedia page that best matches the topic of a given tweet, and the matching pages for existing tweets will not be affected by the introduction of new tweets. As can be seen in Fig. 5, the categories derived from the Wikipedia distance show clear separation, in contrast to the categories derived from LSA.

There was one randomly sampled tweet that was close to the tweets about J. D. Salinger. We thought this would be another tweet about the author, but it was not. This randomly sampled tweet was about the football player, “Warner.” This tweet resulted in a short Wikipedia distance to tweets about J. D. Salinger, because the author is associated with Warner Books and Warner Brothers.

Table 2. Accuracy (hit plus correct rejection) of classifying 45 tweets with known categories when 55 randomly sampled tweets are added

Technique	J. D. Salinger	iPad	Haiti
Latent Semantic Analysis	.60	.60	.20
Wikipedia	.93	.87	.73

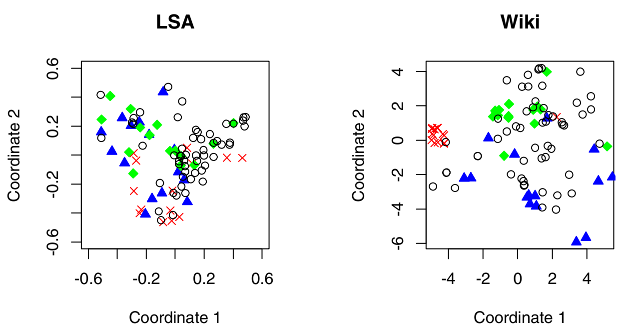


Fig. 5. Forty-five Tweets with known categories and 55 randomly selected tweets mapped onto two-dimensional planes using multidimensional scaling of the between-tweet distances based on LSA and Wikipedia. An open circle is a randomly sampled tweet, an *x* is a tweet about J. D. Salinger, and a triangle is a tweet about the iPad.

4 Discussion and Future Directions

A unique aspect of our technique is that it uses Wikipedia as its knowledge base to calculate between-tweet distances. Tweets and Wikipedia pages are both socially constructed artifacts. This, we think, allowed our technique to simulate the way humans categorize.

For our system to be used in production, the classification technique will need to be able to classify texts into events in near-real time. This, we think, is achievable: Once the novel tweet is mapped into the semantic space, the classifier can predict its category membership based on its similarity to other transformed tweets in the space.

Also, the classification technique will need to be adaptive. Although the randomly sampled tweets were treated as noise in the current work, seemingly useless tweets may actually contain useful information depending on the context. Future work might weight the value of information based on what topics are being discussed. Such classification models with selective attention mechanisms have been successful in simulating human classification behavior [24, 25].

In addition, the method for calculating distances using Wikipedia can be improved. The distances were purely based on the number of steps from one Wikipedia category to the other. The number of Wikipedia categories in each level, which was ignored in the current work, could be used to normalize the distances. In addition, LSA could be used to provide another distance measure between the Wikipedia categories, thus combining LSA and Wikipedia. The integration of probabilistic topic modeling techniques (e.g., [26]) might also be considered.

To recapitulate, when monitoring world events, the volume of tweets presents us with a problem: there is too much information to pay attention to. We are interested in looking at only the novel and useful information. In order to do so, we need ways of flagging emerging topics of interest, without a priori knowing what the topics will be. We suggest here an approach: short tweets are used to find longer passages in Wikipedia. These longer passages have already been linked to other Wikipedia passages. Thus, the distance between tweets can be approximated by the link distance measure between their corresponding Wikipedia pages. In an exploratory study, we showed this technique produced better classification accuracy than two other techniques, String Edit Distance and Latent Semantic Analysis. This work is an instance of a broader approach: by tapping Wikipedia and other living artifacts of social computing, computational methods might provide results that better serve humans.

Acknowledgments. We thank Valentin Leon-Bonnet and colleagues in the Center for Decision Technologies for their ideas.

References

1. McNamara, D.S.: Computational Methods to Extract Meaning From Text and Advance Theories of Human Cognition. Topics in Cognitive Science 3(1), 3–27 (2011)
2. Twitter blog. (2010), `http://blog.twitter.com/2010/02/measuring-tweets.html`
3. Sankaranarayanan, J., Samet, H., Teitler, B.E., Lieberman, M.D., Sperling, J.: Twitterstand: News in tweets. In: Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 42–51. ACM, New York (2009)
4. Demirbas, M., Bayir, M.A., Akcora, C.G., Yilmaz, Y.: Crowd-sourced Sensing and Collaboration Using Twitter. In: 11th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), IEEE Computer Society Press, Los Alamitos (2010)
5. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes Twitter users: real-time event detection by social sensors. In: Proceedings of the 19th International Conference on the World Wide Web, pp. 851–860. ACM, New York (2010)
6. Weick, K.E.: Sensemaking in organizations. Sage Publications, Inc., Thousand Oaks (1995)
7. Michelson, M., Macskassy, S.A.: Discovering users' topics of interest on twitter: A first look. In: Proceedings of the Workshop on Analytics for Noisy, Unstructured Text Data (2010)

8. Macskassy, S.A.: Leveraging contextual information to explore posting and linking behaviors of bloggers. In: International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 64–71. IEEE, Los Alamitos (2010)
9. Gabrilovich, E., Markovitch, S.: Wikipedia-based semantic interpretation for natural language processing. Journal of Artificial Intelligence Research 34, 443–498 (2009)
10. Bratus, S., Rumshisky, A., Magar, R., Thompson, P.: Using domain knowledge for ontology-guided entity extraction from noisy, unstructured text data. In: Proceedings of The Third Workshop on Analytics for Noisy Unstructured Text Data, pp. 101–106. ACM, New York (2009)
11. Strube, M., Ponzetto, S.P.: WikiRelate! Computing semantic relatedness using Wikipedia. In: Proceedings of the National Conference on Artificial Intelligence, p. 1419. AAAI Press, MIT Press (2006)
12. Carlson, A., Betteridge, J., Kisiel, B., Settles, B., Hruschka Jr, E.R., Mitchell, T.M.: Toward an architecture for never-ending language learning. In: Proceedings of the Twenty-Fourth Conference on Artificial Intelligence (2010)
13. Phan, X.H., Nguyen, L.M., Horiguchi, S.: Learning to classify short and sparse text & web with hidden topics from large-scale data collections. In: Proceeding of the 17th International Conference on World Wide Web, pp. 91–100. ACM, New York (2008)
14. Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford (2009)
15. Sriram, B., Fuhry, D., Demir, E., Ferhatosmanoglu, H., Demirbas, M.: Short text classification in twitter to improve information filtering. In: Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval, pp. 841–842. ACM, New York (2010)
16. Jansen, B.J., Zhang, M., Sobel, K., Chowdury, A.: Twitter power: Tweets as electronic word of mouth. Journal of the American society for information science and technology 60(11), 2169–2188 (2009)
17. Stone, B., Dennis, S., Kwantes, P.J.: Comparing Methods for Single Paragraph Similarity Analysis. Wiley Online Library, Chichester (2010)
18. Venables, W.N., Ripley, B.D.: Modern applied statistics with S. Springer, Heidelberg (2002)
19. Ristad, E.S., Yianilos, P.N.: Learning string-edit distance. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(5), 522–532 (2002)
20. Dumais, S.T., Landauer, T.K.: A solution to Platos problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. Psychological review 104, 211–240 (1997)
21. Cataldi, M., Di Caro, L., Schifanella, C.: Emerging topic detection on Twitter based on temporal and social terms evaluation. In: Proceedings of the Tenth International Workshop on Multimedia Data Mining, pp. 1–10. ACM, New York (2010)
22. Bibiko, H.-J.: R code for Levensthein distance (2006)
23. Wild, F.: Latent Semantic Analysis Package in R (2010)
24. Kruschke, J.K.: ALCOVE: An exemplar-based connectionist model ot category learning. Connectionist psychology: a text with readings 99(1), 107 (1999)
25. Love, B.C., Medin, D.L., Gureckis, T.M.: SUSTAIN: A network model of category learning. Psychological Review 111(2), 309–332 (2004)
26. Griffiths, T.L., Steyvers, M.: Finding scientific topics. Proceedings of the National Academy of Sciences of the United States of America 101(1), 5228 (2004)

Toward a Wearable, Neurally-Enhanced Augmented Reality System

David H. Goldberg[1], R. Jacob Vogelstein[2],
Diego A. Socolinsky[3], and Lawrence B. Wolff[1]

[1] Equinox Corporation, New York, NY, USA
[2] Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA
[3] Equinox Corporation, Baltimore, MD, USA
{david,diego,wolff}@equinoxsensors.com,
jacob.vogelstein@jhuapl.edu

Abstract. Augmented reality systems hold great promise, but as they become more complex they can become more challenging to use. Incorporating neural interfaces into augmented reality systems can dramatically increase usability and utility. We explore these issues in the context of Equinox Corporation's Night REAPER™ system—an augmented reality system for dismounted warfighters. We describe the current Night REAPER system and then survey some of the potential enhancements and unique design challenges associated with the addition of a neural interface. Signals, sensors, and decoding techniques for the system's brain-machine interface are discussed.

Keywords: augmented reality, brain-machine interface, wearable systems.

1 Introduction

Augmented reality (AR) systems are becoming increasingly widespread, and have countless consumer, industrial, medical, and military applications. The proliferation of inexpensive sensors has greatly increased the quantity of information that can be incorporated into an AR system. While this development holds great promise, the increased complexity of AR systems poses a challenge for usability. Each sensor and operating mode is accompanied by a combinatorial explosion of configurations from which the user must select. Compounding this problem is the notion that in many environments, such as the operating room or the battlefield, AR systems are most useful if they can be managed in a hands-free manner. Furthermore, the cognitive load associated with operating a complex AR system can distract the user from the very task that it is meant to facilitate.

Incorporating a neural interface into AR systems can greatly increase their usability and utility. A user can potentially control the system with their brain activity, permitting quick, hands-free execution of tasks such as choosing from one of several discrete options. Beyond this, there is also the possibility of a deeper connection between brain and machine. We can imagine designing a system where the user's brain and an onboard computer communicate bi-directionally and work synergistically to solve problems that neither could solve on its own. For example, in a nighttime

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 493–499, 2011.

surveillance task, the AR system could use a night vision sensor to make subjects visible, while the user's brain performs the more challenging task of detecting a subject of interest. The system observes a correlate of the detection in the user's brain activity and then zooms the camera in on the subject. Interestingly, the use of brain activity may not even require the user's conscious awareness.

We are particularly interested in wearable AR systems that are untethered and permit the user to move freely through a dynamic environment. Obviously, such systems are subject to strict size, weight, and power constraints. They require miniaturizing and making portable the brain-machine interface, as well as integrating it with the head-mounted display of an AR system. In this paper we will introduce a specific application—the Night REAPER™ system for dismounted warfighters—and describe how it can potentially be enhanced by a neural interface. We will also describe some of the unique design challenges that arise in this application. That said, the neurally-enhanced sensory augmentation concepts we will discuss are general and are applicable to a wide array of applications, including systems that facilitate the operation of vehicles or the performance of computer assisted-surgery.

2 Night REAPER Augmented Reality System

Over the past six years, Equinox Corporation has been developing the Night REAPER™ (Rapid Engagement Aim Point viewER) system, a wearable AR system for dismounted warfighters (Figure 1). Through the use of advanced signal and image processing, it harnesses the strengths of intensified night vision imaging and thermal imaging to deliver a detailed situational picture. Unlike other image fusion platforms, Night REAPER combines inputs from sensors that are located separately and moving independently of one another. Using a head-mounted display (HMD), the user can see the imagery from his thermal weapon sight (TWS) co-registered in real-time over a helmet-mounted intensified night vision field of view. This allows the user to seamlessly transition from navigation to target detection, identification, and, ultimately, engagement. In navigation mode, the user relies primarily on his wide-field head-mounted night vision device, while using the Night REAPER's ability to overlay the TWS output as a "thermal flashlight" that allows quick detection of targets (Figure 2). If the target is to be engaged, Night REAPER can provide assistance in the form of range information and other ballistic calculations. All the while, the user has access to additional sources of information such as maps and network-centric assets that can be displayed at will.

2.1 Vision for a Neurally-Enhanced Night REAPER System

We propose a neural interface for the Night REAPER system that allows for a seamless integration of advanced sensors, network-centric assets, and computational capabilities with the user's senses. The user manages the system with their brain activity, bypassing the need to provide explicit input. Furthermore, brain activity is coupled to an on-board computer vision system, enabling tasks that the computer is unable to do alone. A block diagram of the proposed system is depicted in Figure 3.

Fig. 1. Current Night REAPER system with labeled components

Fig. 2. Night REAPER augmented reality view. The grayscale background represents the intensified imagery, and the color overlay represents the imagery from the thermal weapon sight.

Consider the target detection-identification-engagement sequence under a neurally-enhanced Night REAPER system. While the user is on patrol, he subconsciously spots a target of interest. Through the neural interface, the system becomes aware of this detection event and cues the user's conscious perception by highlighting the potential target on the HMD. Sensing that the user intends to identify the potential target, the system enhances the target's appearance. It also provides additional information, such as target GPS coordinates in conjunction with network-acquired information about friendly troop disposition in the area. When the decision to engage the target is made, the system seamlessly transitions to targeting mode, where perhaps the full TWS field of view is expanded to fill the HMD and ballistic computation assistance appears as part of the reticle. Throughout this entire process, the user is not required to shift his attention to system management. All system decisions are managed directly through a neural connection, reducing the cognitive load and increasing efficiency in a complex task.

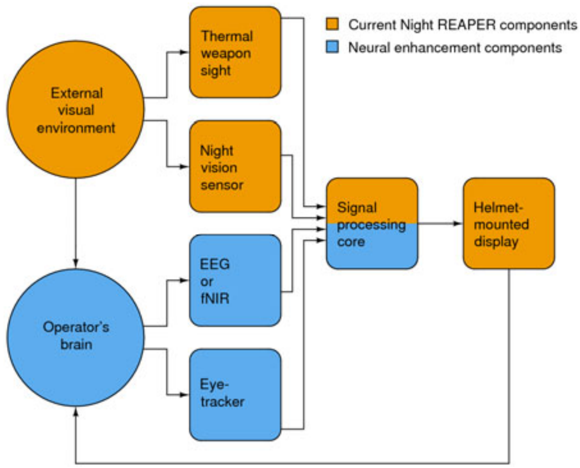


Fig. 3. Block diagram of proposed system, depicting current Night REAPER components (orange) and components required for a neural interface (blue)

2.2 Potential Neural Enhancements

Neural enhancements of the Night REAPER system can generally be divided into two categories—system management and brain-assisted sensory processing. System management generally encompasses the management of the augmented reality system with user-generated brain activity. This could include the switching of visualization modes (e.g., from helmet-mounted, intensified night vision to TWS). Brain-assisted sensory processing entails employing the user's brain to perform sensory processing tasks that are generally challenging for computer systems. This may include detection of novel objects in the environment, classifying a vehicle target, or tracking a target as it moves through the scene.

2.3 Constraints on the Neural Interface

Dismounted warfighter platforms are subject to strict size, weight, and power restrictions. For example, the total weight of helmet-mounted, body-worn and weapon-mounted components must be minimized because dismounted warfighters already carry more than 50 pounds of gear, which can limit their mobility [3]. The helmet-mounted weight is limited to 1.5–2 pounds, a requirement that arises from the fact that a combat helmet typically weighs 3–3.5 pounds [8], and guidelines intended to limit neck injury dictate that total head-supported weight not exceed 5 pounds [7]. The incorporation of a brain-machine interface (BMI) into such a system poses a unique set of challenges; in many cases, BMI research has assumed unlimited computing resources and an inexhaustible power supply. Furthermore, the sensors that transduce the neural signals must be integrated with a combat helmet while preserving the helmet's ballistic protection.

The power consumption requirements of the neural interface are indirectly dictated by the body-worn weight requirement—the user must carry enough batteries to supply the system for a typical three-day mission. This translates to 8 hours of operation on a single set of batteries, which imposes a tight constraint on power consumption.

The latency requirements of the neural interface are dictated by the nature of the task the user is performing. For system management tasks like mode selection, the time between when the user thinks the command and the change in mode should be comparable to a conventional button-press method. The latency requirement for brain-assisted sensory processing is highly task-dependent, ranging from hundreds of milliseconds for novelty detection, to seconds for a more challenging task.

3 Neural Interface for the Night REAPER System

The neural interface is shown in the context of the entire system in Figure 3. It can be further broken down into lower-level blocks, as shown in Figure 4. Brain activity is recorded by one or more sets of sensors, and then decoded into a form that the user interface control can understand. The user interface control then updates the HMD.

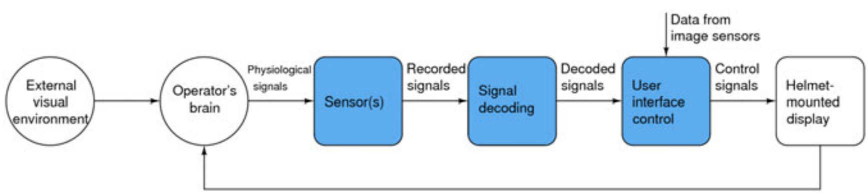


Fig. 4. Block diagram of neural interface to sensory augmentation system. Neural interface components are highlighted in blue.

3.1 Signals

A critical aspect of the proposed system is its reliance on intuitive neural control signals that do not require the user to shift their attention from the task at hand. The vast majority of studies on BMI to date have relied upon "artificial" strategies for neuromodulation that require significant mental effort, such as performing motor imagery tasks or mental calculations, but there are a few reports in the literature of BMI systems that operate on naturally occurring neural signals (reviewed in [1]). The ability to decode such signals is critical for system acceptance; a BMI that requires attention to be directed away from the mission would likely add to the user's cognitive burden, as opposed to relieving it.

3.2 Sensors

Many sensors have been considered for BMI, including electroencephalography (EEG), functional near infrared (fNIR) imaging, functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), and implanted microarrays [4]. The size, weight, and power constraints outlined above limit the types of neural acquisition systems that can be used by the system. For example, fMRI and MEG require room-sized sensors and controls that are by no means portable, and implanted microelectrodes are not reliable enough to be incorporated into non-medical fieldable systems. Consequently, in the near term, EEG and fNIR are the most likely candidates for sensors that can be straightforwardly integrated with a combat helmet.

The choice of EEG versus fNIR affects some system requirements but not others. For example, both EEG and fast optical signal (FOS)-based fNIR have similar bandwidth and sample rate requirements, as the FOS appears to directly reflect aggregated neural spike activity in real-time and can be used as a high-bandwidth signal akin to EEG [6]. However, EEG and fNIR will have differing effects on other system parameters such as the physical interface to the human user and the size, weight and power budget. The physical interface in particular merits scrutiny, as it is non-trivial to maintain a good connection between an electrode/optode and the scalp in freely-moving users. Active EEG electrodes help to eliminate movement artifacts, but the best EEG recordings are typically derived from preparations in which the scalp is abraded and contact is made through conductive gel. Neither of these considerations is easy to employ in a fieldable neural interface technology, but this problem has recently received much attention from the military and may be solved with novel electrode technologies in the near future [2]. In contrast to EEG, the use of fNIR eliminates motion artifacts and the need for both scalp abrasion and conductive gel, but introduces a problem of a low signal-to-noise ratio when considering the FOS.

3.3 Decoding

Due to the diversity of neural interfaces and BMI applications, there is no “standard” approach to decoding neural control signals [1]. Moreover, the specific decoding algorithm employed will depend heavily on the operational paradigm. For example, a decoder that runs in continuous time may look for unique spatiotemporal patterns in the multi-channel neural data, or simply watch for modulation of signal power in one or more frequency bands. Alternatively, if the decoder is queued to operate during intervals in which the user must choose between a finite number of options, event-related decoding strategies can be employed to identify event-related synchronization or desynchronization, or event-related potentials. Steady-state visual evoked potentials could also potentially be used if the selection alternatives are presented at different temporal intervals.

For any choice of decoding algorithm, there is a large spectrum of possible inputs to the algorithm. For example, some algorithms operate on the raw waveforms from one or more channels of recording, while others require data preprocessing to extract salient features such as signal power within a particular frequency band, or a projection of the raw data into a lower- or higher-dimensional subspace [1]. In some cases, features can be selected based on empirical or theoretical models of the expected neural response—a classic example is found in systems relying on event-related potentials such as the visual P300 signal, which typically employ a matched filter with a peak at 300 milliseconds post-stimulus [5]. However, it is more often the case that there is no underlying model for the expected neural activity, so a broad range of features is empirically evaluated through either manual or automated searches.

Statistical machine learning techniques can be employed to identify patterns in the inputs. Previous BMI research and development efforts have demonstrated the utility of a diversity of techniques including neural networks, linear discriminant analysis, Bayesian approaches, support vector machines (SVM), Kalman filters, and random forests [1]. As with the input features, there is often no principled way to ascertain which technique will be most effective on any particular dataset, so multiple alternatives should be evaluated and compared for both accuracy and efficiency.

References

1. Bashashati, A., Fatourechi, M., Ward, R.K., Birch, G.E.: A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. Journal of Neural Engineering 4(2), R32–R57 (2007)
2. Boland, R.: Army uses advanced systems to understand what soldiers know. Signal Magazine (March 2008), `http://www.afcea.org/signal/articles/templates/Signal_Article_Template.asp?articleid=1528&zoneid=228`
3. Department of the Army, Washington, D.C.: U.S. Army Field Manual 21-18: Foot Marches (1990)
4. van Gerven, M., Farquhar, J., Schaefer, R., Vlek, R., Geuze, J., Nijholt, A., Ramsey, N., Haselager, P., Vuurpijl, L., Gielen, S., Desain, P.: The brain-computer interface cycle. Journal of Neural Engineering 6(4), 41001 (2009)
5. Krusienski, D.J., Sellers, E.W., Cabestaing, F., Bayoudh, S., McFarland, D.J., Vaughan, T.M., Wolpaw, J.R.: A comparison of classification techniques for the P300 speller. Journal of Neural Engineering 3(4), 299 (2006)
6. Medvedev, A.V., Kainerstorfer, J., Borisov, S.V., Barbour, R.L., VanMeter, J.: Event-related fast optical signal in a rapid object recognition task: Improving detection by the independent component analysis. Brain Research 1236, 145–158 (2008)
7. Melzer, J.E., Brozoski, F.T., Letowski, T.R., Harding, T.H., Rash, C.E.: Guidelines for HMD design. In: Rash, C.E., Russo, M.B., Letowski, T.R., Schmeisser, E.T. (eds.) Helmet-Mounted Displays: Sensation, Perception and Cognition Issues, pp. 805–848. U.S. Army Aeromedical Research Laboratory, Fort Rucker (2009), `http://www.usaarl.army.mil/new/publications/HMD_Book09/`
8. U.S. Army PEO Soldier: Advanced combat helmet (ACH), `https://peosoldier.army.mil/factsheets/SEQ_SSV_ACH.pdf`

Interface Design Challenge for Brain-Computer Interaction

Jeremy Hill, Peter Brunner, and Theresa Vaughan

Laboratory of Neural Injury and Repair, Wadsworth Center,
New York State Department of Health, Albany, NY 12201, USA
jezhill@gmail.com, {pbrunner,Vaughan}@wadsworth.org

Abstract. Great things can be achieved even with very low bandwidth. Stephen Hawking has been able to break new ground in theoretical physics just by twitching his hand and cheek. Jean-Dominique Bauby was able to write a best-selling memoir by blinking one eyelid. By reading and decoding "brain-waves", the field of brain-computer interfacing (BCI) is poised to open up the possibility of such expression, even for people who can no longer move a single muscle. A BCI still requires an HCI front-end to be of practical use, but many currently-used HCIs do not adequately address limitations on the typical target user's input (e.g., limited eye movement leading to poor spatial vision) or output (e.g. variable delays, and false positives/negatives, in "pressing the button"). In this symposium, BCI experts will present their view of the challenges arising from these limitations. The HCI community is invited to participate in a competition to provide the best solutions.

Keywords: brain-computer interfacing (BCI), electroencephalography (EEG), human-computer interaction (HCI), human factors, spelling, augmentative and alternative communication (AAC), assistive technology, competition.

1 Introduction

Brain-computer interfacing (BCI) is a field of research which aims to develop the means for a person to communicate, or to issue a control signal, without using muscles or peripheral nerves [44]. Control signals are instead interpreted directly from activity in the cerebral cortex, measured via surface electrodes on the scalp (EEG), via electrodes implanted inside the skull on or in the cortex, or via some other brain imaging technique. The BCI user's communication or control intentions are then decoded from the signals. Bypassing the peripheral output channels makes the technology attractive as a potential tool for rehabilitation following stroke or brain injury, or for neuroprosthetics that replace lost function: for example for people who have amputated limbs, who have spinal cord injuries, or who are paralyzed as a result of disease [45, 42, 6]. Cases at the most extreme end of this spectrum of disability tend to attract the most attention in BCI: many studies focus on its applicability to people with advanced amyotrophic lateral sclerosis (ALS), a fast-progressing degenerative motor-neuron disease. Within a few years, ALS may lead to a "locked-in" state (LIS) [34] where only a very small number of individual muscles (typically

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 500–506, 2011.

eye muscles) can be moved voluntarily, and it may then progress beyond this to what has been termed the "totally locked-in" state (TLIS) [3, 21], in which no voluntary movement control remains at all (this latter state holds a particular interest for BCI researchers, since alternatives to BCI become unfeasible). In such a condition, restoration of efficient communication with other people and with one's environment becomes a high priority.

The locked-in state holds a fascination for the research community and the public alike. It is well-known that, with the appropriate method of augmentative and alternative communication (AAC), a surprising amount can be accomplished with low bandwidth. Nobel-prize-winning physicist Stephen Hawking, and best-selling author Jean-Dominique Bauby are two well-known cases in point [2]. BCI promises to go one step further, perhaps by improving further the range expression available to locked-in users, and perhaps by widening the category of people who can make use of AAC, even to include the totally-locked-in.

BCI strategies generally fall into two broad categories: self-actuated BCI and stimulus-driven BCI [24]. In self-actuated BCI, the user chooses when to start, and when to stop, performing a certain mental task or operation. Performing the mental task causes a measurable change in the strength of certain brain signals: usually the power of certain oscillations increases or decreases, at particular frequencies and spatial locations that are dependent on the task and on the individual. A common example is motor imagery: imagining making movements with one's hand, for example, leads to fluctuations in the power of signals that are measurable from the motor and pre-motor cortical areas of the brain, which can in turn be used as a continuous control signal [25, 33]. By contrast, a stimulus-driven BCI relies on the brain's response to a particular stimulus, delivered at a time decided by the computer [11, 28]. To use the interface, the user directs his or her attention to the stimulus he or she wishes to select: the BCI then depends on measuring the difference between the brain's response to relevant, attended stimuli and its response to irrelevant, unattended stimuli. Often the useful responses are in the form of event-related potentials (ERPs) which are brief deflections in the EEG lasting less than a second.

BCIs, whether they are stimulus-driven or self-actuated, elicit signals that have a low signal-to-noise ratio [30]. This effectively limits the communication performance of BCIs [35]. The current state-of-the-art in BCI control is still noisy enough that, generally, the utility of a BCI system to any one given user has not yet surpassed that of the conventional assistive technology available to the same user [4, 35]. The avenues that have tried to address this limitation all entail certain problems. For example, brain-signals recorded closer to the neural source, via invasive neurosurgical techniques [22, 18, 43, 9], are of higher signal-to-noise ratio, but obtaining them entails increased risk and cost [16]. Another avenue is to elicit signals of higher signal-to-noise ratio using better-designed stimuli and tasks [5, 13, 29]. However, these refinements (in common with the original design of the BCIs in question) are optimized specifically to elicit clearer brain signals, and do not explicitly address ease-of-use or obey the demands of conventional human factors. BCI performance may therefore also be limited by the error-prone behavior of the user in attempting to deal with a poorly-designed HCI in noisy conditions [27]. As an alternative to these approaches, BCIs might stand to improve their utility by improving the usability of their HCI front-ends.

2 Example: ERP-Based Spelling

The most widely-explored paradigm for BCI spelling is the stimulus-driven Donchin matrix speller [11, 10]. Letters are arranged in a grid on screen, and a subset of the letters (usually one row or one column) is highlighted at any one time. During a sequence of such highlighting events, the relevant, attended stimulus tends to cause larger ERPs than the unattended stimuli. The BCI can infer the desired character by determining the row and column that produced the largest ERP. The sequence of highlighting events will typically be repeated many times in order to overcome the inherently low signal-to-noise ratio. Although it does form the basis for the most comprehensive trials to date of BCI in target users' homes [37, 41, 31, 21, 38], its treatment in the BCI literature has been largely as a research tool rather than a practical AAC device. Its design was motivated more by the necessary conditions for eliciting and measuring changes in ERPs, and less by usability concerns. This holds also for many subsequent refinements [e.g. 17, 23, 19, 26, 39] which changing aspects of the arrangement of the stimuli in space and time, to increase the brain signals' signal-to-noise ratio or the robustness with which letters are encoded.

The usability gap is particularly clear when we consider the users who stand to benefit most from BCI, namely those people on the LIS/TLIS border and beyond, whose paralysis has progressed far enough to impact their ability to direct and focus their gaze, resulting in poor spatial vision. Two recent studies [8, 40] independently confirmed that the high performance reported in the standard laboratory settings with healthy users depends heavily on ERP components that are only generated when users fix their gaze precisely on the target letter. Both studies show, however, that the system can be used at a lesser level of performance without target fixation, since the P300 ERP itself does not depend on this (though it remains an open question how well a paralyzed person with poor spatial vision could deal with the practical demands of locating and attending to their desired letter within a dense visual array). One of the studies [40] showed that, by rearranging the stimuli spatially into a radial layout, and temporally into a two-step hierarchy, the system became more resistant to the limitations of poor gaze control. This is an excellent example of the impact of improved HCI design in BCI. Consideration of users' limited vision has also led to speller designs even less dependent on spatial vision [1] and entirely based on auditory stimuli [14, 20, 36].

3 Self-actuated BCI Spelling

The improved HCI mentioned in the previous section, called Hex-O-Spell, had previously made its debut in BCI as one of the few self-actuated spelling systems, driven by motor imagery [7, 29]. Other interesting user interfaces have also been used in conjunction with self-actuated BCIs, such as Dasher [12], and a similar information-theoretic approaches to spelling [32], although there has been little opportunity to assess their performance in the target user group. Another reason for the under-representation of self-actuated BCI in communication is that there is a larger variance in people's native ability to control them [15]. Stimulus-driven BCIs have a greater tendency to be usable "out of the box," where self-actuated BCI may require more user training to be accessible to all users. This is one area in which better HCI design may be valuable: to better reinforce a user's learning of the control signal. A related, and also valuable, design goal is to adapt better to variations in users' level of control. Adaptation to a range of proficiency levels is important in three contexts: on first use (to cope with variation in levels of control between individuals); during use (to cope with fatigue); and over longer timescales (to adapt to a given user's improvement in control as their repeated use of the interface allows them to learn).

4 Symposium and Competition

Our symposium at HCI 2011 will present the HCI challenges faced by the field of BCI, as seen through the eyes of BCI experts. It will also launch a competition into which the HCI community is invited to enter to meet these challenges: to design a better BCI-driven interface for communication. Provisionally, the competition it will consist of two independent streams, each with its own prize:

1. The design stream: entrants will submit two-page white-papers to us, provisionally by November 2011, detailing their design for a better HCI front-end to a brain-computer interface and explaining the design elements that best serve the target population. A prize will be awarded for the best idea, in the opinion of the judges.
2. The implementation stream: by the later provisional deadline of mid-February 2012, entrants will submit an actual implementation, built within a rapid-development software framework that we make available. This will allow us to test the effectiveness of the entrants' design in coping with a noisy input signal. The input signal will be triggered by the user making a mouse movement at the appropriate times, as dictated by the interface's stimuli. The resulting signal will emulate a slow ERP or a brief burst of self-actuated brain activity. Design of the interface's auditory and/or visual stimuli, and their arrangement in space and time, and the decoding of the sequences of input signals into arbitrary English text, will be the entrants' challenge. Error-correction and predictive spelling will be encouraged. Finalists will be chosen according to the judges' assessment of how well the designs meet the needs of the target user group. A time-limit will then be chosen for each finalist according to a "handicap"-like system, with the aim of leveling the playing-field between designs that meet different degrees of limitation on the user's sensory input. Naturally we expect auditory-only interfaces to be slower than those that require vision, but we nonetheless wish to ensure that a design that is useful to any subgroup of the target population has a chance of winning. A "spell-off' event will then be held for the finalists, at which the entrants will be required to use their interfaces to spell a given sentence in English: the winner will be finalist who has the largest number of (correct letters - incorrect letters) on the screen at the end of their allotted time period.

Details of the prize money, exact deadlines, and further rules will be posted on: http://bcimeeting.org/HCI2011Challenge/

5 Design Challenges for a BCI

In conclusion, we envisage that that the ideal HCI-for-BCI should:

1. cope with a noisy input signal (entailing many false positives and false negatives in selection) which may become noisier or more infrequent as the user tires;
2. allow efficient and robust error correction;
3. integrate prior knowledge about the task to be performed (for example predictive selection) in an intuitive and accessible way;
4. adapt to wide range of user proficiency levels between users, and within the same user from hour to hour and from day to day.
5. promote user behavior that leads to learning and improvement in BCI control;
6. be accessible to users whose vision is poor, or perhaps even to users who are functionally blind.

We look forward to the participation and the ideas of the HCI community. We expect that this competition will further enhance the non-muscular communication options available to people with severe motor impairments.

References

1. Acqualagna, L., Treder, M.S., Schreuder, M., Blankertz, B.: A novel brain-computer interface based on the rapid serial visual presentation paradigm. In: Conf Proc IEEE Eng Med Biol Soc, pp. 2686–2689 (2010)
2. Bauby, J.D.: The Diving Bell and the Butterfly. Knopf, New York (1997)
3. Bauer, G., Gerstenbrand, F., Rumpl, E.: Varieties of the locked-in syndrome. Journal of neurology 221(2), 77–91 (1979)
4. Berger, T.W., Glanzman, D.L.: Toward Replacement Parts for the Brain: Implantable Biomimetic Electronics as Neural Prostheses. MIT Press, Cambridge (2005)
5. Bin, G., Gao, X., Wang, Y., Hong, B., Gao, S.: VEP-based brain-computer interfaces: time, frequency, and code modulations [Research Frontier]. Computational Intelligence Magazine, IEEE 4(4), 22–26 (2009)
6. Birbaumer, N., Cohen, L.G.: Brain-computer interfaces: communication and restoration of movement in paralysis. J Physiol 579(Pt 3), 621–636 (2007)
7. Blankertz, B., Dornhege, G., Krauledat, M., Schroder, M., Williamson, J., Murray-Smith, R., Muller, K.R.: The berlin brain-computer interface presents the novel mental typewriter hex-o-spell. In: Proceedings of the 3rd International Brain-Computer Interface Workshop and Training Course (2006)
8. Brunner, P., Joshi, S., Briskin, S., Wolpaw, J.R., Bischof, H., Schalk, G.: Does the 'P300' speller depend on eye gaze? J Neural Eng 7(5), 56013 (2010)
9. Brunner, P., Ritaccio, A.L., Emrich, J.F., Bischof, H., Schalk, G.: Rapid communication with a "P300" matrix speller using electrocorticographic signals (ECoG). Front Neurosci 5(5) (2011)
10. Donchin, E., Spencer, K.M., Wijesinghe, R.: The mental prosthesis: assessing the speed of a P300-based brain-computer interface. IEEE Trans Rehabil Eng 8(2), 174–179 (2000)
11. Farwell, L.A., Donchin, E.: Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalogr Clin Neurophysiol 70(6), 510–523 (1988)

12. Felton, E., Lewis, N.L., Wills, S.A., Radwin, R.G., Williams, J.C.: Neural signal based control of the Dasher writing system. In: IEEE EMBS Conference on Neural Engineering, pp. 366–370 (2007)
13. Friedrich, E.V., McFarland, D.J., Neuper, C., Vaughan, T.M., Brunner, P., Wolpaw, J.R.: A scanning protocol for a sensorimotor rhythm-based brain-computer interface. Biol. Psychol (August 2008)
14. Furdea, A., Halder, S., Krusienski, D.J., Bross, D., Nijboer, F., Birbaumer, N., Kübler, A.: An auditory oddball (P300) spelling system for brain-computer interfaces. Psychophysiology 46(3), 617–625 (2009)
15. Guger, C., Daban, S., Sellers, E., Holzner, C., Krausz, G., Carabalona, R., Gramatica, F., Edlinger, G.: How many people are able to control a P300-based brain-computer interface (BCI)? Neurosci. Lett. 462(1), 94–98 (2009)
16. Hamer, H.M., Morris, H.H., Mascha, E.J., Karafa, M.T., Bingaman, W.E., Bej, M.D., Burgess, R.C., Dinner, D.S., Foldvary, N.R., Hahn, J.F., Kotagal, P., Najm, I., Wyllie, E., Lüders, H.O.: Complications of invasive video-EEG monitoring with subdural grid electrodes. Neurology 58(1), 97–103 (2002)
17. Hill, J., Farquhar, J., Martens, S.M.M., Biessmann, F., Schölkopf, B.: Effects of stimulus type and of error-correcting code design on bci speller performance. In: Koller, D., Schuurmans, D., Bengio, Y., Bottou, L. (eds.) Twenty-Second Annual Conference on Neural Information Processing Systems, Curran, Red Hook, NY, USA, pp. 665–672 (June 2009)
18. Hochberg, L.R., Serruya, M.D., Friehs, G.M., Mukand, J.A., Saleh, M., Caplan, A.H., Branner, A., Chen, D., Penn, R.D., Donoghue, J.P.: Neuronal ensemble control of prosthetic devices by a human with tetraplegia. Nature 442(7099), 164–171 (2006)
19. Hong, B., Guo, F., Liu, T., Gao, X., Gao, S.: N200-speller using motion-onset visual response. Clin Neurophysiol. 120(9), 1658–1666 (2009)
20. Klobassa, D.S., Vaughan, T.M., Brunner, P., Schwartz, N.E., Wolpaw, J.R., Neuper, C., Sellers, E.W.: Toward a high-throughput auditory P300-based brain-computer interface. Clin Neurophysiol. 120(7), 1252–1261 (2009)
21. Kübler, A., Birbaumer, N.: Brain-computer interfaces and communication in paralysis: extinction of goal directed thinking in completely paralysed patients? Clin Neurophysiol. 119(11), 2658–2666 (2008)
22. Leuthardt, E.C., Schalk, G., Wolpaw, J.R., Ojemann, J.G., Moran, D.W.: A brain-computer interface using electrocorticographic signals in humans. J. Neural Eng. 1(2), 63–71 (2004)
23. Martens, S.M., Hill, N.J., Farquhar, J., Schölkopf, B.: Overlap and refractory effects in a brain-computer interface speller based on the visual P300 event-related potential. J Neural Eng 6(2), 26003–26003 (2009)
24. Mason, S.G., Birch, G.E.: A general framework for brain-computer interface design. IEEE Trans Neur Syst Rehabil Eng 11(1), 70–85 (2003)
25. McFarland, D.J., Neat, G.W., Wolpaw, J.R.: An EEG-based method for graded cursor control. Psychobiology 21, 77–81 (1993)
26. McFarland, D.J., Sarnacki, W.A., Townsend, G., Vaughan, T.M., Wolpaw, J.R.: The P300-based brain–computer interface (BCI): Effects of stimulus rate. Clin Neurophysiol (in press, 2010)
27. McFarland, D.J., Sarnacki, W.A., Vaughan, T.M., Wolpaw, J.R.: Brain-computer interface (BCI) operation: signal and noise during early training sessions. Clin Neurophysiol. 116(1), 56–62 (2005)

28. Middendorf, M., McMillan, G., Calhoun, G., Jones, K.S.: Brain-computer interfaces based on the steady-state visual-evoked response. IEEE Trans. Rehabil Eng. 8(2), 211–214 (2000)
29. Müller, K.R., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., Blankertz, B.: Machine learning for real-time single-trial EEG-analysis: from brain-computer interfacing to mental state monitoring. J Neurosci. Methods 167(1), 82–90 (2008)
30. Niedermeyer, E., Lopes da Silva, F. (eds.): Electroencephalography. Basic Principles, Clinical Applications, and Related fields. Williams & Wilkins (1993)
31. Nijboer, F., Sellers, E.W., Mellinger, J., Jordan, M.A., Matuz, T., Furdea, A., Halder, S., Mochty, U., Krusienski, D.J., Vaughan, T.M., Wolpaw, J.R., Birbaumer, N., Kübler, A.: A P300-based brain-computer interface for people with amyotrophic lateral sclerosis. Clin Neurophysiol 119(8), 1909–1916 (2008)
32. Omar, C., Akce, A., Johnson, M., Bretl, T., Ma, R., Maclin, E., McCormick, M.P.C.T.: A feedback information-theoretic approach to the design of brain-computer interfaces. International Journal of Human-Computer Interaction 27(1), 5–23 (2010)
33. Pfurtscheller, G., Neuper, C.: Motor imagery activates primary sensorimotor area in humans. Neurosci Lett 239, 65–68 (1997)
34. Plum, F., Posner, J.: The diagnosis of stupor and coma, vol. 1. F. A. Davis Co., Philadelphia (1966)
35. Schalk, G.: Brain-computer symbiosis. J Neural Eng 5(1), 1–15 (2008)
36. Schreuder, M., Blankertz, B., Tangermann, M.: A new auditory multi-class brain-computer interface paradigm: spatial hearing as an informative cue. PLoS One 5(4) (2010)
37. Sellers, E.W., Kübler, A., Donchin, E.: Brain-computer interface research at the University of South Florida Cognitive Psychophysiology Laboratory: the P300 Speller. IEEE Trans Neural Syst Rehabil Eng 14(2), 221–224 (2006)
38. Sellers, E.W., Vaughan, T.M., Wolpaw, J.R.: A brain-computer interface for long-term independent home use. Amyotroph Lateral Scler 11(5), 449–455 (2010)
39. Townsend, G., LaPallo, B.K., Boulay, C.B., Krusienski, D.J., Frye, G.E., Hauser, C.K., Schwartz, N.E., Vaughan, T.M., Wolpaw, J.R., Sellers, E.W.: A novel P300-based brain-computer interface stimulus presentation paradigm: moving beyond rows and columns. Clin Neurophysiol 121(7), 1109–1120 (2010)
40. Treder, M.S., Blankertz, B.: C)overt attention and visual speller design in an ERP-based brain-computer interface. Behav Brain Funct 6(1), 28–28 (2010)
41. Vaughan, T.M., McFarland, D.J., Schalk, G., Sarnacki, W.A., Krusienski, D.J., Sellers, E.W., Wolpaw, J.R.: The Wadsworth BCI Research and Development Program: at home with BCI. IEEE Trans Neural Syst Rehabil Eng 14(2), 229–233 (2006)
42. Vaughan, T.M., Wolpaw, J.R.: The Third International Meeting on Brain-Computer Interface Technology: making a difference. IEEE Trans Neural Syst Rehabil Eng 14(2), 126–127 (2006)
43. Velliste, M., Perel, S., Spalding, M.C., Whitford, A.S., Schwartz, A.B.: Cortical control of a prosthetic arm for self-feeding. Nature 453(7198), 1098–1101 (2008)
44. Vidal, J.J.: Toward direct brain-computer communication. Annu Rev Biophys Bioeng 2, 157–180 (1973)
45. Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M.: Brain-computer interfaces for communication and control. Clin Neurophysiol 113(6), 767–791 (2002)

Trust in Human-Computer Interactions as Measured by Frustration, Surprise, and Workload

Leanne M. Hirshfield, Stuart H. Hirshfield, Samuel Hincks,
Matthew Russell, Rachel Ward, and Tom Williams

Department of Computer Science, Hamilton College, Clinton, NY 13323, USA
{lhirshfi,shirshfi,shincks,mprussel,rward,tewillia}@hamilton.edu

Abstract. We describe preliminary research that attempts to quantify the level of trust that exists in typical interactions between human users and their computer systems. We describe the cognitive and emotional states that are correlated to trust, and we present preliminary experiments using functional near infrared spectroscopy (fNIRS) and electroencephalography (EEG) to measure these user states. Our long term goal is to run experiments that manipulate users' level of trust in their interactions with the computer and to measure these effects via non-invasive brain measurement.

Keywords: fNIRS, EEG, electroencephalograph, near-infrared spectroscopy, workload, frustration, surprise, trust.

1 Introduction

We describe research that attempts to model and quantify the level of trust[1] that exists in typical interactions between human users and their computer systems. This can be useful for a number of reasons. For example, in order for users to interact with online websites, they must trust the security and validity of that site. If we can measure users' levels of trust during online interactions in usability studies, we can ensure that a given website is designed appropriately to maximize users' trust. Measuring trust would also be useful during usability studies of a number of applications and technologies, pointing designers to areas of the technology or interface that should be re-designed to maximize the user's comfort while working with the system. In addition, measuring trust can help to defer the wealth of money and time spent on training personnel to detect security breaches. If we can measure users' changing levels of trust while working with their computer systems, we can have a better understanding of the training needed to ensure that security personnel detect breaches quickly and accurately.

[1] There is a great deal of research from the management, economics, and recently, from the computer science disciplines, that focuses on building definitions and models that describe the concept of' trust'. While we describe some of this research in section 2, this work does *not* present new models or definitions of trust. We use the term in its most general, non-specific sense, as it is the word choice at this time that best describes the elements of the human-computer interactions that we are exploring.

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 507–516, 2011.

The novelty of our work stems primarily from the multi-modal approach of our analysis, which employs standard psychological testing techniques (both during and post-experiment surveys) to pinpoint emotional and cognitive components of trust as well as brain measuring technologies (EEG and functional near infrared spectroscopy, or fNIRS, recordings) to record physiological reactions. Specifically, our research focuses on the changing relationship from trust to distrust over time from the perspective of a person who is detecting deceitfulness, rather than being deceitful.

In our initial experiment we used a variation of "The Trust Game", a scenario which has been used by many trust researchers. In our version, we modified the level of trustworthiness of the computer agent with whom the subject was playing as the game progressed. Throughout the experiment, we measured, via subjective surveys, the cognitive and emotional changes that occurred while the user's level of trust toward the computer agent changed. The results indicated that the cognitive and emotional states of workload, frustration, and surprise are directly correlated to the users' changing level of trust with a computer agent, and that as users lost trust in the computer agent, they had increasing levels of workload, surprise, and frustration.

We hypothesize that these findings can be generalized and used to measure the amount of trust a user feels during a variety of human-computer interactions. Thus, if we can objectively measure users' levels of workload, surprise, and frustration throughout human-computer interactions, we can understand that user's current level of trust toward the computer, the computer agent, the website, or any other computer mediated communication with which the user is interacting. Our long term goal is to measure changing levels of trust during human-computer interactions. As a first step, we have been running preliminary experiments that isolate the cognitive components of workload, frustration, and surprise.

The rest of this paper is organized as follows: First, we provide a brief overview of research dealing with the user state of trust. Second, we describe the experiment that we conducted to manipulate trust and to understand the cognitive and emotional state changes that were correlated to trust. Third, we provide an overview of the past research on measuring the user states of workload, frustration, and surprise. Fourth, we describe our EEG and fNIRS devices, and we discuss the preliminary experiments that we ran with these devices as well as our experiment results. Fifth, we provide analysis of our results. Lastly, we describe avenues for future work in the measurement of trust during human-computer interactions.

2 Background and Relevant Literature

The topic of trust has sparked a wealth of research in the domains of management and economics. There are many working definitions of 'trust' which have been proposed in the literature[1-5]. The rapid evolution of the internet and the viruses, hackers, and malware that have surfaced in recent years allow for new, broader interpretations of previous notions of interpersonal trust [6]. How much do we trust our computers and the content being presented to us by our computers? What elements of the human-computer interaction affect our feelings of trust toward a computer, computer agent, or web site? Can we 'trust' a computer in the first place? We chose to use the term 'trust' throughout our paper because our research findings are based on the well known ‘Trust Game’ which has been used extensively in the trust literature, and because ‘trust’ is the word that is used most often in qualitative interviews that have been conducted in our lab while people talk about their interactions with computers.

2.1 Trust Experiment

We conducted an experiment that aimed to discover relationships between a computer user’s level of trust and that user’s changing cognitive and emotional states. To do so we asked participants to sit in front of a standard-size computer monitor and interact with a computer console on the computer screen. They then played a version of the “Trust Game,” developed by Berg [4], which has been used in many experiments dealing with trust, risk-taking, and money management [3]. In our version of the “Trust Game” both the computer and the user began with a fictional $10. The user and computer would take turns sending some amount of money, ranging from $0-$10, back and forth to one another. Each time some amount of money was sent between the user and computer, the amount sent was tripled while en route. In an ideal, high-trust scenario, the computer and the user would always send a the maximum amount of money back and forth to one another—maximizing the gain possible for each of the two Trust Game ‘players’. This process was repeated 23 times.

For the first eight transactions, the computer acted “trustworthy” in the sense that it typically returned a high amount of money back to the participant. For the next nine transactions, the computer acted with a mix of trustworthy and untrustworthy behavior, sometimes returning a high amount, and other times returning a low amount of money back to the participant. For the last six transactions the computer acted in a wholly untrustworthy manner, regularly returning a very low amount of money back to the participant.

We used data from Self Assessment Manikins which were administered after every six transactions to gauge the user’s cognitive and emotional state. Additionally, we collected the amount given and percentage returned for each transaction, as well as information regarding the subjects’ self-rated locus of control, trust, and computer familiarity.

From this information we gleaned that the median amount given increased during the computer’s trustworthy state, varied highly during the computer’s erratic state, and decreased during the computer’s untrustworthy state. Furthermore, we analyzed the data from the Self Assessment Manikins, turning users self reported measures of valence, arousal, and dominance into a set of discrete user states. We found a direct correlation between the trustworthiness of the computer agent and the user’s reported measures of workload, frustration, and surprise. Our results showed that overall workload increased throughout the experiment, as did frustration and surprise.

We were also interested in whether relationships existed between frustration, surprise, and workload. We found that significant correlations between frustration and workload in all four surveys existed. We also found significant correlations between workload and surprise in surveys 1, 3, and 4, as shown in Table 1 below.

Our results suggested that as computer users lost trust with the simulated agent they were interacting with in the Trust Game, the user states of workload, surprise, and frustration were directly correlated with the users’ changing level of trust.

Table 1. Significant Correlations between workload/frustration, and workload/surprise

	Survey 1	Survey 2	Survey 3	Survey 4
Frustration & Workload	$r(27)=.565$ $p<.01$	$r(25)=.733$ $p<.01$	$r(25)=.67$ $p<.01$	$r(26)=.739$ $p<.01$
Workload and Surprise	$r(27)=.559$ $p<.01$	$r(25)=.391$ $p<.06$	$r(25)=.638$ $p<.01$	$r(26)=.478$ $p<.02$

2.2 Linking Trust in Human-Computer Interactions to Surprise, Workload, and Frustration

While these results are restricted to the version of the Trust Game that users played, we hypothesize that the same user states will influence the level of trust in more realistic interactions between users and their computer systems. As an illustrative example, consider John Doe's interaction with his computer over the course of a year:

When John's computer was functioning properly, he had high trust in his interactions with and through his computer. During these high trust times, John had low levels of frustration, workload, and surprise—all interactions with his computer seemed to proceed as expected. However, one day John visited a new website and suddenly he noticed hundreds of pop ups infiltrating his screen (i.e., surprise and frustration). He later found out that his computer had a virus, which likely came from the site with all of the pop ups (frustration). Later on that year, John was Instant Messaging with his friend Alice. Alice was a classmate of John's and a user with her name as a username contacted John via IM. After a few minutes of messaging with who he presumed to be Alice, John began to become wary of the interactions. The IMer was not writing in a way that was consistent with Alice. John began to interact very cautiously with the IMer, hoping to determine if the person was indeed an imposter (workload, frustration, surprise). Also, over the course of time, John's computer became very slow because he downloaded too many programs and add-ons. He was frustrated while using it because it took a long time for him to get things done and he found it difficult to keep focused on the task at hand while waiting long intervals for his computer to catch up to his train of thought (frustration and workload).

All of these occurrences caused John's level of trust during his computer interactions to be lowered. We hypothesize that we can use measures of users' workload, frustration, and surprise to indicate that users' level of trust. In the next sections we describe cutting edge research that attempts to measure these user states objectively.

2.3 Measurement of Surprise, Workload, and Frustration

Acquiring quantitative data about computer users is a continual challenge for researchers in HCI. Although we can accurately measure task completion time and accuracy, measuring factors such as mental workload, frustration, and distraction are often done by qualitatively observing users or by administering subjective surveys to users. These surveys are often taken after the completion of a task, potentially missing valuable insight into the user's changing experiences throughout the task. They also fail to capture internal details of the operator's mental state. To address these evaluation issues, much current research focuses on developing objective techniques to measure, in real time, user states such as workload, frustration, and surprise [7-9]. Although this ongoing research has advanced user experience measurements in the HCI field, finding accurate and non-invasive tools to measure computer users' states in real working conditions remains a challenge. The user states that are addressed by this research are the states of workload, frustration, and surprise.

Surprise. Surprise has previously been measured in HCI studies using facial analysis software [10] as well as using Skin Conductivity, blood volume and heart rate [11]. Detecting surprise with electroencephalography (EEG) is a topic of much research in Psychophysiology. Surprise can be indicated by the presence of an Error-Related Potential (ErrP), in which EEG data contains error-related negativity (a sharp negative deflection around 80ms)[12], often followed by error-related positivity (a slow positive wave)[13]. ErrPs can be found both when a user makes an error and when a user notices an error made by the machine [14]. This makes the measurement of ErrPs useful in HCI-based interface analysis.

Frustration. Frustration is an important metric in HCI. Lazar, et al. studied frustration with computer interfaces by having employees that used computers in their workday keep journals that tracked their ongoing frustration as they used their routine computer programs [15]. The results showed that word processing and email were reported as the most frustrating activities, and that participants wasted an average of forty percent of their time trying to solve unnecessarily frustrating problems.

Biological methods of measuring frustration allow researchers to collect more reliable data. Scheier, et al., induced frustration in users with a mouse that sporadically froze and inhibited the user from winning a game [16]. A Hidden Markov Model analyzed skin conductivity, blood volume pressure and the state of the mouse, eventually learning the manifestations of frustration. The results indicate that a user's affective state can be automatically discriminated from events in their physiology [16]. Most recently, BCI devices enable automated detection of user frustration through discovering patterns in brain activity. Reuderink et al. developed an affective version of Pacman which places the user in a state of frustration [8]. They found significant differences in EEG activity during periods of frustration and the normal state.

Workload. The ability to acquire objective, real-time measures of a computer user's mental workload while (s)he works with a computer would be valuable to the field of HCI. Adaptive interfaces could adapt in real-time to a given user based on his or her current level of workload, keeping that user in *the flow* [17]. Also, measures of users' mental workload could be acquired during usability studies to help interface designers to pinpoint areas of the interface that may be un-intuitive for users [18, 19].Researchers have successfully used EEG or fNIRS to measure elements of mental workload such as working memory[20-23], response inhibition [21, 24], visual search [21, 25], as well as a myriad of other executive processes [26, 27].

3 Measurement Oriented Experiments

We conducted three preliminary experiments where we attempted to manipulate and measure the user states of surprise, frustration, and workload. In the future, we aim to acquire real time measures of these user states in order to predict one's level of trust during his or her computer interactions.

3.1 Surprise Experiment

An important component of trust is the moment of surprise; that is the moment when a person notices that something 'unexpected' has occurred in the computer system. This could be the moment users notice that a virus is on their computer, or the moment they realize that the person they are IMing with may be an imposter. To measure this, we exploited the oddball paradigm in order to elicit surprise. Three participants completed an experiment that was created using Eprime in which they pressed two different buttons depending on the position of an oval on the screen. The oval was in one of two positions, located either on the far left or the far right side of the screen. When the oval was on the left side of the screen the subjects were instructed to press the 'z' button, and when the oval was on the right side of the screen they were instructed to press the 'm' button.

Immediately following the subject response a feedback screen indicated whether or not the subject had pressed the correct key. Subjects completed 150 tasks where they simply hit the 'z' or 'm' keys to indicate the position of the oval on the screen. During the first 20 tasks, the feedback for the subjects was as expected. During the last 130 tasks, we randomly selected 15% of the tasks to provide incorrect, *or surprising feedback* to the user. In other words, 15% of the time, when subjects pressed the 'z' key, the feedback indicated that the 'm' key had been pressed, and vice versa.

The EEG used in the study was Advanced Brain Monitoring's b-alert wireless 10 channel EEG. Data was sampled at 256Hz (www.b-alert.com). The non-invasive EEG is an ideal brain monitoring device for use in human-computer interaction studies, where it may be important to keep participants comfortable while completing tasks in realistic working conditions.

The Eprime software sent markers to the EEG immediately before the subject saw the feedback screen. In this way, we planned to search for the presence of an ErrP that was caused when the surprising feedback occurred during 15% of the tasks.

Data Analysis and Results of Surprise Experiment. We used a similar procedure as Ferrez et al. [14] to preprocess our EEG data for classification. We took the data from the moment the feedback occurred through to 650ms after the feedback was shown for channels Cz and Fz. Like Ferrez et al., we chose these channels because ErrPs are usually found in a fronto-central distribution along the midline [14] Each temporal section of data was associated with one of two class labels: control or surprise, indicating whether or not the feedback the subject saw at that moment was the expected feedback or the surprising feedback.. We applied a 1-10 Hz bandpass filter as ErrPs have a relatively slow cortical potential. We downsampled our data from 256Hz to 128Hz and input our resulting timeseries data into a weighted K-nearest neighbor classifier (k = 3) with a Dynamic Time Warping distance measure. We ran our classification separately for each subject. Results are in Table 2. We were able to distinguish between the control and surprising feedback conditions with an average of 71% accuracy for our three subjects.

Table 2. Classifier accuracy distinguishing between the control and surprising feedback

	sub1	sub2	sub3	average
Classifier Accuracy	70%	74%	68%	71%

3.2 Frustration Experiment

In this section we report on an experiment that was completed in 2009 [28]. During the experiment six subjects completed a series of nback tasks [20, 21], which have been used in many experiments to manipulate working memory. In the 1-back task, depicted in Figure 1, subjects must indicate whether the current letter on their computer screen is a match ('m'), or not a match ('n') to the letter that was shown 1 screen previously.

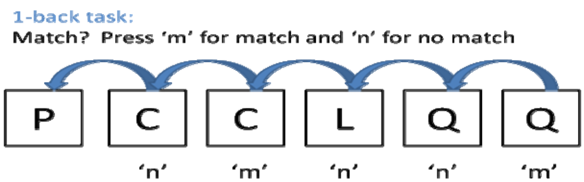


Fig. 1. Depiction of the 1 back task

Each task lasted 30 seconds with a rest time of 20 seconds between tasks. Half of the 1back tasks were completed by subjects as expected. However, during the other half of the 1back tasks, internet pop ups such as the one shown in Figure 2, were introduced into the computer systems. Subjects were told to finish the nback tasks as quickly as possible and with the highest accuracy possible. Six subjects (3 female, 3 male) completed the experiment. Subjects were all Tufts undergraduate students. A randomized block design with eight trials was used in this experiment.

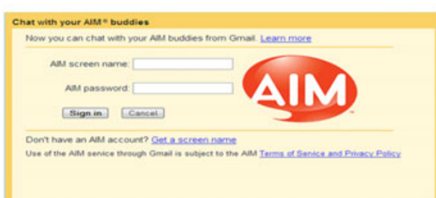


Fig. 2. An example of a pop up in the frustration experiment

In this experiment we used an OxyplexTS (ISS Inc. Champagne, IL) frequency-domain tissue spectrometer with two optical probes. Each probe has a detector and four light sources. Each light source emits near infrared light at two separate wavelengths (690nm and 830nm) which are pulsed intermittently in time. This results in 2 probes x 4 light sources x 2 wavelengths = 16 light readings at each timepoint (sampled at 6.25Hz).

Data Analysis and Results of Frustration Experiment. All subjects were interviewed following the experiment. All subjects indicated that the pop ups were a source of frustration throughout the experiment. We computed all machine learning analyses separately for each subject. For each subject, we recorded 16 channel readings throughout the experiment where we refer to the readings of one source detector pair at one wavelength, as one *channel*. We normalized the intensity data in each channel by their own baseline values. We then applied a moving average band pass filter to each channel (with values of .1 and .01 Hz) and we use the modified Beer-Lambert Law[12] to convert our light intensity data to measures of the relative changes in oxygenated (HbO) and deoxygenated hemoglobin (Hb) concentrations in the brain. This resulted in eight readings of HbO and eight readings of Hb data at each timepoint in the experiment. We then averaged together the channels from the left side of the head and the channels on the right side of the head, giving us 4 time series for each subject; 1) HbO on the left side of the head, 2) HbO on the right side of the head, 3) Hb on the left side of the head, and 4) Hb on the right side of the head. We then input these time series into a weighted KNN classifier (k = 3) with a distance measure computed via Symbolic Aggregate Approximation (SAX). For more information on SAX, see [29]. As shown in Table 3, we were able to distinguish between the control 1back tasks and the frustrating 1back tasks with an average of 73% accuracy across the six subjects.

Table 3. Classifier accuracy at distinguishing between the control (1back) and frustrating (1back with pop-ups) conditions

	sub1	sub2	sub3	sub4	sub5	sub6	average
Classifier Accuracy	69%	81%	63%	75%	75%	75%	73%

3.3 Workload Experiments

We have conducted several experiments, using the fNIRs device described above, to measure various aspects of mental workload. Using this device we have:

1. Used machine learning techniques to classify, on a single trial basis, the load placed on users visual search, working memory, and response inhibition resources [21].
2. Used machine learning techniques to classify various levels of working memory load in a simple counting and addition task [30].
3. Used machine learning techniques to distinguish between spatial and verbal working memory [19].

4 Conclusion and Future Work

We described preliminary research that attempts to quantify the level of trust that exists in typical interactions between human users and their computer systems. We described the cognitive and emotional states that we found to be correlated to trust, and we presented preliminary experiments using functional near infrared spectroscopy and electroencephalography to measure these user states. The experiments presented in this paper represent the beginning of our research on the measurement of trust during human-computer interactions. Ongoing work in our lab continues to manipulate, and measure, the user states of frustration, surprise, and workload. The longer term goal is to run experiments that manipulate users' level of trust in their interactions with the computer and to measure these effects via non-invasive brain measurement.

References

1. Mayer, R., et al.: An Integrative Model of Organizational Trust. The Academy of Management Review 20(3), 709–734 (1995)
2. Serva, M.A., Fuller, M.A., Mayer, R.: Trust in systems development: a model of management and developer interaction research in progress. In: Proceedings of the 2000 ACM SIGCPR Conference on Computer Personnel Research, ACM, Chicago (2000)
3. Lewicki, R., et al.: Trust and Distrust: New Relationships and Realities. The Academy of Management Review 23(3), 438–458 (1998)
4. Berg, J., Dickhaut, J., McCabe, K.: Trust, Reciprocity, and Social History. Games and Economic Behavior 10, 122–142 (1995)
5. McAllister, D.J.: Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. Academy of Management Journal 38(1) (1995)
6. Schneider, F.: Trust in Cyberspace. National Academy Press, Washington (1998)
7. Mandryk, R., Atkins, M., Inkpen, K.: A continuous and objective evaluation of emotional experience with interactive play environments. In: Proceedings of the SIGCHI conference, ACM Press, Montreal Canada (2006)
8. Reuderink, B., Nijholt, A., Poel, M.: Affective Pacman: A Frustrating Game for Brain-Computer Interface Experiments. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering (2009)
9. Savran, A., et al.: Emotion Detection in the Loop from Brain Signals and Facial Images. In: eNTERFACE 2006, Dubrovnik, Croatia (2006)
10. Ward, R.: An analysis of facial movement tracking in ordinary human-computer interaction. Physiological Computing 16(5), 879–889 (2004)
11. Ward, R., Marsden, P.: Physiological responses to different web page designs. International Journal of Human Computer Studies 59, 199–212 (2003)
12. Chavarriaga, R., Ferrez, P., Millán, J.: To Err is Human: Learning from Error Potentials in Brain-Computer Interfaces. Advances in Cognitive Neurodynamics, 777–782 (2008)
13. Nieuwenhuis, S., et al.: Psychophysiology, Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task: p. 752–760
14. Ferrez, P., Millán, J.: You Are Wrong!—Automatic Detection of Interaction Errors from Brain Waves. In: Proceedings of the 19th International Joint Conference on Artificial Intelligence (2005)
15. Lazar, J., Jones, A.S.: Workplace user frustration with computers: an exploratory investigation of the causes and severity. Behaviour & Information Technology, 239–251 (2006)

16. Scheirer, J., et al.: Frustrating the user on purpose: a step toward building an affective computer. Interacting with Computers, 93–118 (2002)
17. Csikszentmihalyi, M.: Flow: The Psychology of Optimal Experience, Harper Collins, 320 (1991)
18. Lee, J.C., Tan, D.S.: Using a low-cost electroencephalograph for task classification in HCI research. In: Proceedings of the 19th annual ACM symposium on User interface software and technology, ACM Press, Montreux (2006)
19. Hirshfield, L.M., et al.: Brain Measurement for Usability Testing and Adaptive Interfaces: An Example of Uncovering Syntactic Workload in the Brain Using Functional Near Infrared Spectroscopy. In: Conference on Human Factors in Computing Systems: Proceeding of the twenty-seventh annual SIGCHI conference on Human factors in computing systems (2009)
20. Grimes, D., et al.: Feasibility and Pragmatics of Classifying Working Memory Load with an Electroencephalograph. In: CHI 2008 Conference on Human Factors in Computing Systems, Florence, Italy (2008)
21. Hirshfield, L., et al.: This is your brain on interfaces: enhancing usability testing with functional near infrared spectroscopy. In: SIGCHI, ACM, New York (in press, 2011)
22. Sassaroli, A., et al.: Discrimination of mental workload levels in human subjects with functional near-infrared spectroscopy (2009); accepted in the Journal of Innovative Optical Health Sciences
23. Gevins, A., et al.: High-Resolution EEG Mapping of Cortical Activation Related to Working Memory: Effects of Task Difficulty, Type of Processing, and Practice. Cerebral Cortex (1997)
24. Schroeter, M.L., et al.: Near-Infrared Spectroscopy Can Detect Brain Activity During a Color-Word Matching Stroop Task in an Event-Related Design. Human Brain Mapping 17(1), 61–71 (2002)
25. Anderson, E.J., et al.: Involvement of prefrontal cortex in visual search. Experimental Brain Research 180(2), 289–302 (2007)
26. Tanida, M., et al.: Relation between asymmetry of prefrontal cortex activities and the autonomic nervous system during a mental arithmetic task: near infrared spectroscopy study. Neuroscience Letters 369(1), 69–74 (2004)
27. Joanette, Y., et al.: Neuroimaging investigation of executive functions: evidence from fNIRS. PSICO 39(3) (2008)
28. Hirshfield, L.M.: Enhancing Usabiltiy Testing with Functional Near Infrared Spectroscopy. In: Computer Science, Tufts University, Medford (2009)
29. Lin, J., et al.: A Symbolic Representation of Time Series, with Implications for Streaming Algorithms. In: Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, San Diego, CA (2003)
30. Hirshfield, L.M., et al.: Human-Computer Interaction and Brain Measurement Using Functional Near-Infrared Spectroscopy. In: Symposium on User Interface Software and Technology: Poster Paper, ACM Press, New York (2007)

Idea Visibility, Information Diversity, and Idea Integration in Electronic Brainstorming

Elahe Javadi and Wai-Tat Fu

University of Illinois, Urbana, IL 61801
{ejavadi2,wfu}@illinois.edu

Abstract. Despite the pervasive use of electronic media for idea generation and idea sharing, the extent and quality of idea integration and use is relatively understudied. Idea integration and use depends on information saliency but little is known about how idea integration may be facilitated by user interface features that influence information saliency. This paper examines the effect of idea visibility on idea integration and how that relationship is moderated by information diversity. Our laboratory experiment showed that although the basic level of idea integration, i.e. mere reference to partners' ideas increased when visibility increased, higher levels of idea integration decreased as visibility increased. Information diversity was found to be a significant moderator of the relationship between visibility and idea integration.

Keywords: Idea integration, visibility, information diversity, brainstorming.

1 Introduction

Research and practice shows that electronic brainstorming systems (EBSs) may have created an illusion of productivity as they seem to offer limited benefits in terms of quantity or quality of the ideas generated by individuals during brainstorming [4,10]. An underpinning thesis for losses during electronic brainstorming is associated with the lack of idea integration and use. Although many experimental studies have addressed individual's idea-sharing behavior in electronic settings [3] little research have been done to examine the extent to which individuals build on the ideas shared by others. To bridge this gap in the research literature, the current paper focuses on the effect of visibility on idea integration. Since individual's idea integration behavior depends on the extent and quality of attention allocated to the shared ideas and since user interface is the point of access to the shared ideas, we posit that channeling attention through manipulation of visibility of the ideas (i.e., information saliency) will influence idea integration behavior of the individuals.

1.1 Idea Integration in Electronic Brainstorming

Idea integration (also referred to as adoption, exploitation, combination or synthesis) is often considered the most fruitful phase of the creative process [9]. Integration occurs when dimensions of more than one individual's ideas are combined to create new ideas [5]. Assuming that no one individual has sufficient information to generate

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 517–524, 2011.

the best idea, idea integration becomes a key to realizing more fully the value of the individually generated ideas [11].

Idea is defined as a statement that consists of at least one testable proposition [12]. Idea dimensions, which are building blocks of idea integration, are defined as "unique testable propositions". Thus, an idea is called a multi-dimensional idea if it includes more than one unique testable proposition. An example of a one-dimensional idea is "I think some sort of tarp would be useful for shade and shelter". A multi-dimensional idea could be "some sort of outer shell jacket that is water proof, can be used to collect water if it rains, covers body at night". "We have to stick together though" is an example of a value statement which is not counted as an idea.

1.2 Idea visibility, Information Diversity and Idea Integration

For idea integration, individuals must attend to the ideas shared by others so as to discover new perspectives. Visibility of the ideas on user interface facilitates members' exposure to the different dimensions and is a predictor of the idea being used in an integration activity. With the shift from information scarcity to information richness in modern organizations, visibility of ideas becomes even more important [6]. Visibility identifies the extent to which ideas generated and shared by members of the group are exposed to other members.

Visibility is defined by the portion of the idea pool that is displayed on the screen at any given time. Increased visibility leads to an increased number of cues made available by visible ideas, which activates knowledge items in memory. Activation of more items in memory increases the possibility of the individuals' discovering and articulating connections among different ideas' dimensions. Thus, we derive the first proposition:

Proposition 1: idea integration is positively associated with idea visibility.

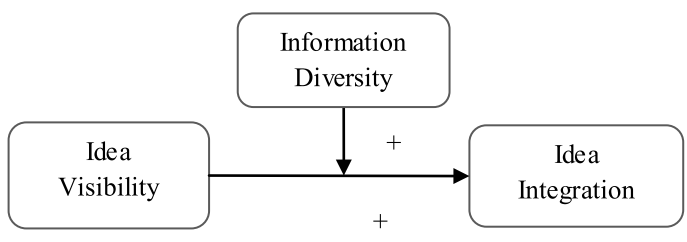


Fig. 1. Research Model

1.3 Information Diversity

When ideas that are attended to are more diverse, the potential for integration increases because information diversity will by itself stimulate integration [13]. Information diversity here represents variety of the ideas or more precisely difference in information contained in the ideas generated and shared by individuals within the group. Diversity of the ideas leads to increased diversity of cues, which in turn, facilitates knowledge activation and retrieval of more information from memory. As such, we propose that diversity moderates the relationship between visibility and idea integration:

> *Proposition 2: Information diversity moderates the relationship between idea visibility and idea integration. For higher the levels of information diversity, the extent to which visibility influences idea integration will be higher.*

2 Method

Our hypotheses were tested in the laboratory using an open idea generation task. Participants in our experiments discussed electronically within groups using an experimental software system that allows for manipulations of visibility. The software system presents users with a screen split horizontally with the posted ideas displayed across the top. The user types in an idea in the lower section of the screen and submit the idea when they are ready. The users can rank other ideas and also refer to other ideas as well. The software (at ideation-experiment.org) then generates the transcripts which will be used for measuring idea integration.

2.1 Participants

Participants were recruited from two upper-level business courses at a large Mid-Western university in the United States. Participants were awarded extra credits with an additional opportunity to win a lottery (for $50). Participants assigned to different experimental conditions randomly and all participants in a particular session participated in the same condition.

2.2 Task

We used a modified version of the desert survival task [7] which is an open-ended idea generation task was. The task posed a survival problem in a desert and participants were asked to discuss and generate ideas on the items they wished to take to help them survive. An idea could include a new item, a new use for an already proposed item, or follow-ups to and/or counter-arguments of ideas that had already been suggested. Participants were instructed that the suggested items (a) should be portable, and (b) participants should explain why the suggested items were important for surviving in the situation provided.

2.3 Procedures

Each session was approximately thirty minutes long. The experimenter briefed participants on the experiment for about five minutes. The participants then read four instructions pages on computers for about ten minutes and were allowed to ask questions of clarification. Participants then used the discussion forum of the system to exchange ideas and discuss the survival situation for fifteen minutes.

2.4 Treatments

This experiment used a three (Visibility low, medium, and high) by two (small groups, large group) factorial design. Visibility is varied by setting the number of ideas that are displayed on the screen at any given time. Users can view other posts by navigating through different pages. Pilots revealed that five posts on screen are considered low, twelve posts medium and twenty-five, high visibility.

Group Size. Like in most of the theoretical and empirical studies of electronic brainstorming, size of the group is included for possible moderation or predictive effect. Small groups in our experiments involved 2-3 participants and large groups involved 4-6 participants.

2.5 Scoring System

To motivate active participation during the experiment, each participant had a score that would increase for different activities that contributed to the group discussion including posting an idea, rating other participants' ideas and referring to other participants' ideas. The score of the individual then would influence individuals' chance of winning the prize.

3 Analysis and Results

3.1 Measurements

Idea integration measurement is described in the next section and information diversity is measured by latent semantic analysis (LSA) [8]. LSA have demonstrated its applicability in the measures of coherence among topics and its measures match well with actual human coding. Although not perfect, LSA is still valid measure of diversity. For each experimental session we computed LSA measure between any two posts using the system available at http://lsa.colorado.edu/. The average of all binary LSA measures in a session was used as information diversity measure. No conversion were performed on LSA measures, thus higher numbers represent lower diversity.

3.2 Identifying Unique Ideas and Idea Integration Levels

Two external coders who were blind to experimental conditions were trained and asked to code the transcripts of experimental sessions. The coders were asked to first read the entire transcript to understand how the discussion flowed among the individuals in the group. The coders then were asked to read each statement that was exchanged by individuals and code them as idea generation or integration, as shown in Table 1. The Cronbach's Alpha for inter-coder reliability was 0.78. For each experimental session we computed the total number of level 1, level 2 and level 3 idea integration. Examples of the three categories are available in Table 1. Our three levels coding of idea integration is a simplified version of the seven-level integrative complexity coding [1]. We normalized the total number of idea integration over the total number of posts exchanged during that experimental section.

3.3 Testing for Visibility and Group Size Effect and Combining Low and Medium Visibility Groups

Our 2-way ANOVA of sum of levels 1-3 idea integration on visibility (L,M,H) or group size (S,L) showed no significant effect for visibility or group size. Similarly, our two 2-way ANOVA of sum of levels 1-3 idea integration on visibility (L,M) and group size (S,L) showed no significant effect for visibility or group size. Therefore we decided to examine each level of idea integration separately.

Table 1. Levels of Idea Integration

Description	Definition	Example from Experimental Sessions
Idea Generation		
Item without reason	Suggesting an item without providing any reason	Flare gun
Item with reasons	Suggesting an item with reason(s).	We should bring thick gloves, because we will need to work with the cacti (they often have water in them)
Idea Integration		
Level 1		
Challenge without reason	Challenge of, query to someone else's idea without providing any reason	P1: Take a cooler P2: why?
Approve without additional reason	Approving somebody else's idea without providing any additional	P1: maybe some kind of solar powered flashlight to use with the compass for nighttime travel P2: I think the flashlight idea is good
Level 2		
Challenge with reason	Challenge of, query to someone else's idea: with reason but without	P1: Medical first aid kit from plan P2: but they said we weren't hurt
Approve with reason	Approving somebody else's idea and providing additional	P1: I think in the middle of nowhere map might be better P2: yes, especially if we are in a zone with no reception
Level 3		
Alternative	Alternative to or improvement of an existing idea	P1: How about a flashlight for when it gets dark? P2: maybe some kind of solar powered flashlight to use with the compass for nighttime travel

The next step was to conduct three 2-way ANOVA for different levels of idea integration (levels 1-3) integration on Visibility (L,M) and Size (S,L) and three 2-way ANOVA on, Visibility (M,H) and Size (S,L). The above 2-way analyses of variance showed that group size is not a statically significant predictor for any of the three levels of idea integration (Figure 2). Furthermore the first three 2-way ANOVA models showed that there was significant difference among three levels of idea integration between medium and high visibility groups but no significant difference between three integration levels for low visibility groups and medium visibility (Figure 2). As such for further analysis, we combined the low and medium visibility groups.

We then conducted three ANCOVA to examine the influence of idea visibility and group size on the three levels of idea integration when information diversity was included as a covariate. The three ANCOVA models consisted of two visibility levels (L+M, H) and two group sizes (S, L). Consistent with the findings of ANOVA (Figure 3), ANCOVA showed even lesser effect for group size after taking out the variance accounted by information diversity. The difference between idea integration level 1, level 2 and levels 3 was different with visibility levels (L+M) and (H) at 0.05. The direction of the difference is depicted in Figures 2 and Figure 3.

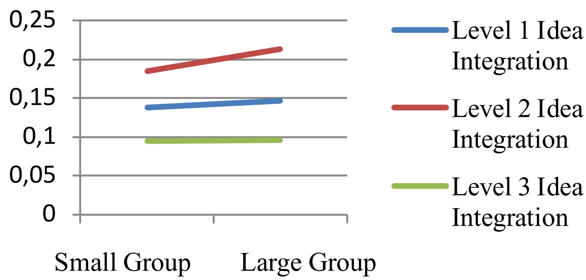


Fig. 2. Three Levels of Idea Integration for Small and Large Groups

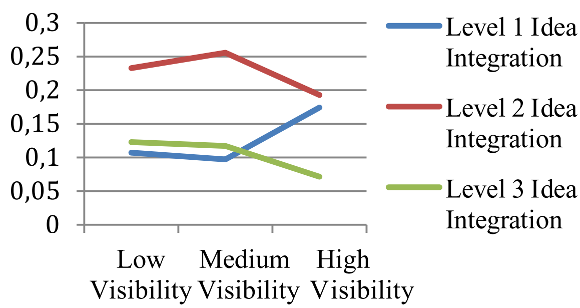


Fig. 3. Three Levels of Idea Integration for Different Visibility Levels

3.4 Visibility and the Moderating Effect of Information Diversity

Based on the findings from ANOVA and ANCOVA, we left out group size as a predictor and combined groups with low and medium visibility. We then conducted a regression analysis on the sum of level 2 and Level 3 idea integration to test for idea visibility effect with moderating effect of information diversity.

The coefficient for the interaction term was statistically significant. The negative coefficient indicates the favorability of smaller idea visibility and more information diversity for level 2 and level 3 idea integration The interaction coefficient is negative because the LSA score is lower for higher information diversity. Similarly our analyses of integration at level 1 showed positive relationship with visibility ($p<0.05$) with a marginally significant effect of diversity. Therefore as hypothesized earlier in this paper level 1 idea integration was found to be positively associated with idea visibility but because of cognitive overload, level 2 and level 3 idea integration were found to be negatively associated with visibility thus the coefficient is negative for idea visibility. Also information diversity was not expected to have any particular relationship with idea integration, and only the interaction of information diversity and idea visibility found to have an effect.

Table 2. Regression Model

Predictor	Model 1 Without interaction	Model 2 With interaction
Visibility	-0.074*	0-.80*
Information Diversity	.002	.473
Visibility * Information Diversity		-1.383**
Constant	0.017*	0.018*
N	65	65

* $p<0.05$ ** $p<0.01$

4 Discussion and Conclusion

In this paper we empirically examined the influence of idea visibility as a user interface feature on idea integration. In our experimental we found that the basic levels of idea integration when individuals only refer to each others' ideas, either approving or challenging, without any reason or justification, was higher for higher levels of idea visibility. We also found that higher levels of idea integration where more cognitive effort is required are diminished when idea visibility is high. We explained this diminishing effect by adverse effect of cognitive overload in higher visibility groups.

Since idea integration plays an important role in creativity, our findings have implications for the extent to which creativity support tools and systems alike expose individuals to the ideas generated in the group. Depending on the level of idea integration that is required for specific purposes, designers may adaptively expose participants to more or less visibility of ideas. Idea visibility, which was manifested in form of the number of visible ideas on the screen in our study, can take other forms. Our analysis on information diversity suggests that, for example, one could selectively present more diverse ideas (e.g., based on semantic analysis) to mitigate the effect of cognitive load. Future research can be directed to understand how cognitive load and semantic interpretation may interact in idea integration.

References

1. Baker-Brown, G., Ballard, E.J., Bluck, S., de Vries, B., Suedfeld, P., Tetlock, P.: The integrative complexity coding manual. In: Smith, C. (ed.) Handbook of Thematic Analysis, pp. 605–611. Cambridge University Press, Cambridge (1992)
2. Davidson, G., Dornburg, C., Stevens, S., Hendrickson, S., Bauer, T., Forsythe, C.: Sandia Research Team Studies Best Way to Solve Wicked Problems. Sandia National Laboratories (2007), `http://www.sandia.gov/news/resources/releases/-2007/wickedproblems.html`
3. Dennis, A.R.: Information Exchange and Use in Group Decision Making: You Can Lead a Group to Information, But You Can't Make it Think. MIS Quarterly 20(4), 433–457 (1996)
4. de Vreede, G.J., Davison, R.M., Briggs, R.O.: How a silver bullet lose its shine. Communications of the ACM 46(8), 96–101 (2003)
5. Gruenfeld, D.H., Hollingshead, A.B.: Sociocognition in Work Groups – The Evolution of Group Integrative Complexity and its Relations to Task-Performance. Small Group Research 24(3), 383–405 (1993)
6. Hansen, M.T., Haas, M.R.: Competing for attention in knowledge markets: Electronic document dissemination in a management consulting company. Administrative Science Quarterly 46(1), 1–28 (2001)
7. Homan, A.C., van Knippenberg, D., Van Kleef, G.A., De Dreu, C.: Bridging faultlines by valuing diversity: Diversity beliefs, information elaboration, and performance in diverse work groups. Journal of Applied Psychology 92(5), 1189–1199 (2007)
8. Landauer, T.K., Foltz, P., Laham, D.: Introduction to latent semantic analysis. Discourse Processes 25(2), 259–284 (1998)
9. Osborn, A.F.: Applied imagination; principles and procedures of creative thinking. Scribner, New York (1953)
10. Pinsonneault, A., Barki, H., Gallupe, R.B., Hoppen, N.: Electronic brainstorming: The illusion of productivity. Information Systems Research 10(2), 110–133 (1999)
11. Robert, L.P., Dennis, A.R., Ahuja, M.K.: Social capital and knowledge integration in digitally enabled teams. Information Systems Research 19(3), 314–334 (2008)
12. Simon, H.A.: Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization. Macmillan Co., New York (1947)
13. van Knippenberg, D., De Dreu, C.K.W., Homan, A.C.: Work group diversity and group performance: An integrative model and research agenda. Journal of Applied Psychology 89(6), 1008–1022 (2004)

The Challenges of Using Scalp-EEG Input Signals for Continuous Device Control

Garett Johnson, Nicholas Waytowich, and Dean J. Krusienski

Old Dominion University
Norfolk, VA 23529
dean.krusienski@ieee.org

Abstract. Whether aiming to control a computer cursor, a robotic arm, or a wheelchair, it remains a significant challenge to achieve responsive and reliable asynchronous control via EEG signals. The most promising scalp-recorded EEG signals for this task are sensorimotor rhythms and steady-state visual evoked potentials, which have both been demonstrated to be viable for continuous device operation in controlled laboratory settings. Several issues, such as handling signal nonstationarity and identifying reliable asynchronous modes of operation, must be addressed before these scalp-EEG signals can become practical for controlling devices outside of the laboratory.

1 Introduction

A brain-computer interface (BCI) provides a non-muscular channel for the brain to interact with the world, and is particularly useful for individuals with neuromuscular disabilities. Many such individuals, e.g., those with amyotrophic lateral sclerosis, still have normal cognitive capabilities but are 'locked in' and unable to communicate. These individuals are reliant upon this additional channel for basic communication, control, and a level of autonomy. Current BCIs have only recently been demonstrated for in-home use by disabled individuals [1]. These systems use BCI2000, a general framework capable of performing a variety of BCI paradigms, but have been found to be most practical for home use when operating either the P300 Speller or a sequential-menu driven system using sensorimotor rhythms (SMR). The P300 Speller is capable of performing discrete selections, but requires considerable trial averaging in order to provide accurate results, hence is not suitable for real-time continuous control. The sequential-menu driven SMR is based on 1 or 2-dimensional center-out tasks to make discrete selections from a set of menus. The appeal of these systems is partly that discrete capabilities allow the user to theoretically pause the system with a proper sequence of selections. This provides the user with a primitive form of asynchronous control, but still requires selections to be made in a given time frame.

In order to generate a more natural system for the user, continuous control needs to be implemented in an effective manner, such that truly asynchronous and reliable control can be achieved. Proven methods for continuous BCI control include steady-state visual evoked potentials (SSVEPs) and SMRs. SSVEPs are electrical potentials produced in the brain in response to a repetitive, periodic visual stimulus. SSVEPs

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 525–527, 2011.

have been demonstrated in laboratories to provide relatively high bitrates, some with communication rates exceeding 70 bits per minute [2]. Since this is an elicited signal, no user training is believed to be required. However, it is argued that gaze control is still required to obtain these results, which some disabled users may lack. Another possible BCI modality for continuous control includes SMRs, where imagined movement results in measurable power differences in certain EEG frequencies as compared to resting states. Currently, users are required to extensively train to gain adequate levels of control. Training occurs as the user adapts to the system, while the system is simultaneously adapted to the individual. Two of the most significant hurdles to overcome in order to achieve the level of reliable continuous control required for extended BCI home use are the identification of dependable asynchronous control modes and the proper treatment of the nonstationarities in EEG signals of interest. This nonstationarity hinders the training process for SMR control as well as the EEG feature translation and classification capabilities of most BCI modalities.

2 Asynchronous Modes of Operation

Asynchronous BCIs allow for the user to operate a device at his or her own pace, instead of being confined to time-constrained control intervals dictated by the BCI. Asynchronous BCIs are becoming an active area of research [3], with some recent efforts focusing on providing a switch to turn the control state of the device on or off. This method of control requires the user to perform some task, typically with a different BCI modality, to enable or disable the primary method of control. For instance, it has been demonstrated that an orthosis has been controlled by a combination of motor imagery and SSVEPs [4]. In this work, SMRs were used to enable the orthosis, while SSVEPs opened or closed the orthotic hand. Although the orthotic hand was not continuous in operation, these hybrid BCIs provides the first steps toward asynchronous control. A more natural method for continuous, asynchronous control would be to eliminate the need for a control switch. In this case, the no-control state is very important. When dealing within a physical environment (e.g., robotic arm control), minimizing the number of false positives are key to minimizing unwanted collisions within the environment. One key question remains, how to effectively determine the no-control state in a continuous asynchronous environment? Confidence levels, and thresholding of classifiers are possible methods, but can simultaneously hinder the speed of the system, e.g., causing too many false negatives. To further complicate the issue, the relevant signal nonstationarities must be identified, characterized, and effectively countered for such asynchronous approaches to successful.

3 Signal Nonstationarities

The characteristics that describe a user's EEG are continuously changing. These signal nonstationarities can result in significant differences within and across days, and can be attributed to a variety of factors such as: the user's physical and mental state, the development of new cortical activities, the user attempting to adapt to the system feedback while the system simultaneously adapts to the user (i.e., co-adaptation), changing recording conditions, etc. New and intelligent signal processing methods are required to effectively cope with these nonstationarites. Possible approaches to deal with the nonstationarity include 1) extracting the stationary signal components embedded within nonstationarities [5], or 2) implementing control schemes that adapt to the changing state of the EEG. An example of the first approach is that SMR phase information has been shown to carry additional directional information to supplement the traditional amplitude information, while being stable across several days for encoding hand movement direction [6]. For the second approach, several techniques for unsupervised adaptation have been proposed in [6], including 1) covariate shift adaptation / minimization, 2) feature adaptation, which focuses on adapting the parameters of the feature extraction method to account for subject learning, and 3) classifier adaptation for dealing with shifts in feature distributions in addition to the conditional distributions between features and classifiers.

References

[1] Sellers, E.W., et al.: A brain-computer interface for long-term independent home use. Amyotroph Lateral Scler 11(5), 449–455 (2010)
[2] Zhu, D., et al.: A survey of stimulation methods used in SSVEP-based BCIs. Comput Intell Neurosci., 702357 (2010)
[3] Bashashati, A., et al.: An Improved Asynchronous Brain Interface: Making Use of the Temporal History of the LF-ASD Feature Vectors. J. Neural Eng. 3(2), 87–94 (2006)
[4] Pfurtscheller, G., et al.: Self-paced operation of an SSVEP-Based orthosis with and without an imagery-based brain switch: a feasibility study towards a hybrid BCI. IEEE Trans. Neural Syst. Rehabil Eng. 18, 409–414 (2010)
[5] Bünau, P.V., et al.: Finding Stationary Subspaces in Multivariate Time Series. Physical Review Letters 103, 214101 (2009)
[6] Krusienski, D.J., et al.: Critical issues in state-of-the-art brain-computer interface signal processing. J. Neural Eng. 8 (2011)

Modeling Pharmacokinetics and Pharmacodynamics on a Mobile Device to Help Caffeine Users

Frank E. Ritter and Kuo-Chuan (Martin) Yeh

The Pennsylvania State University
{frank.ritter,martin.yeh}@psu.edu

Abstract. We introduce a mobile device application that displays key information about caffeine: the pharmacokinetics (time course of drug levels) and pharmacodynamics (the effects of caffeine level) visually on the iPhone, iPod Touch, and iPad. This application, Caffeine Zone, is based on an existing model of caffeine physiology using user inputs, including caffeine dose, start time, and consumption speed. It calculates the caffeine load in a user for the next twenty-four hours and displays it using a line chart. In addition, it shows whether the user is currently in the "cognitive alert zone" (the range of caffeine where a normal person might benefit most from caffeine) or the "possible sleep zone" (the range of caffeine where sleep is presumed not affected by caffeine level.) Understanding the pharmacokinetics and pharmacodynamics of caffeine can help people using caffeine to improve alertness, including in operational environments. Caffeine Zone may also help users create a mental model of caffeine levels when the device is not available. We argue that this app will both teach users the complex absorption/elimination process of caffeine and help monitor users' daily caffeine usage. The model, with additional validation, can be part of a system that predict cognitive state of users and provide assistances in critical conditions.

Keywords: pharmacokinetics, pharmacodynamics, caffeine, mobile app, modeling.

1 Introduction

Caffeine, the most widely used psychoactive substance [e.g., 1, 2], has long been regarded as an effective way to improve mental alertness and reactions, especially in critical operational environments like long-distance driving [e.g., 3], air traffic control, and nearly all operational military environments. Caffeine can be found in many different sources of foods, beverages, and medicines, including chewable gum in military rations. Some people take caffeine-contained substances, mainly coffee and tea, for well-being [2]; others for its pharmacological effects. Low to moderate doses of caffeine can indeed be very useful in military settings according to the National Academy's Institute of Medicine [4].

Overuse of caffeine, however, can impair cognition and health directly and indirectly. For example, higher levels of caffeine can lead to higher levels of cortisol

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 528–535, 2011.

[5]. Too much caffeine may disrupt sleep schedules[1] and may contribute to long-term chronic health issues such as agitation, anxiety [6], and insomnia. Users in stressful, high tempo situations might be particularly prone to using and overusing stimulants to maintain alertness (e.g., see a Naval Aerospace Medical Research Laboratory report [7]). Striking a balance between too much and too little is a challenging task because caffeine's use depends on understanding the pharmacokinetics of caffeine, because uptake and excretion are exponential processes, and because while the immediate benefit is during the task, the delayed response to eliminate caffeine may make users more sleep deprived later and over time. The computation and future impact of use may be beyond many of us to compute.

As a result, people who use caffeine for its pharmacological effects can end up with at least three possible problems. They can consume too little, and not be as alert as they need to be. They can consume too much at a single point in time and be jittery or have other side effects. Or they can consume a right amount but too close to when they would like to sleep and subsequently have trouble sleeping.

Before introducing an application to help users moderate their caffeine levels, we will briefly describe the caffeine model.

2 Understanding Non-linear Curves

We model two of the processes in the human body that modify the caffeine level: absorption and elimination. Absorption refers to caffeine being taken into bloodstream from its external form (liquid, tablet, gum, etc.). Elimination refers to caffeine being excreted from our body, mostly through urine. Both absorption and elimination rates are non-linear functions based on time. (We subsume distribution with absorption and metabolism with elimination.)

In our theory and in the software, we use the following equations taken from a review used for modeling caffeine in cognitive models and agents [8]:

$$\text{Caffeine absorption}_{t+1} = \text{Caffeine intake reservoir}_{t} * 2^{-(1/7\text{min})} \quad (1)$$

$$\text{Caffeine elimination}_{t+1} = \text{Caffeine in bloodstream}_{t} * 2^{-(1/240\text{min})} \quad (2)$$

That is, we have an absorption half-life of 7 minutes (eqn. 1) and an elimination half-life of 4 hours in (eqn. 2).

Soon after caffeine intake, the absorption and elimination processes start simultaneously: caffeine is being distributed into the bloodstream and excreted into urine. The exponential decay equations intertwined with one another may make it difficult for caffeine users to calculate the current caffeine intake in their bloodstream at any moment without external computation, and particularly difficult to predict the level in several hours when it will be time to sleep. This complexity is the major source of the aforementioned challenge—striking a balance of caffeine dose over time.

[1] The effect of caffeine on sleep varies. Some people are very sensitive to caffeine; some seem to have no sleep problems despite regular caffeine consumption in the evening [2].

3 Caffeine Zone

Caffeine Zone is an application that utilizes the ubiquity and computational power of modern mobile devices such that an inexpensive and portable real-time caffeine intake can be displayed graphically. The current version of Caffeine Zone works on the iOS operating system—provided on iPhone, iPod Touch, and iPad—version 3.1.2 and above. There is, however, no conceivable reason it cannot be ported to other mobile devices such as the Android, BlackBerry, and similar devices.

3.1 Software Architecture

The software architecture of Caffeine Zone is shown in Figure 1, providing also an overview of the functions in the app. The application consists of three major components: main, history, and settings. The main component is where the formula are calculated and displayed; the history component is in charge of recording/managing the consumption history; the settings component is the center where settings are achieved and retrieved. The data points of caffeine intake are calculated for each minute and stored in the SQLite database that comes with iOS to save calculation time whenever the line plot is to be displayed.

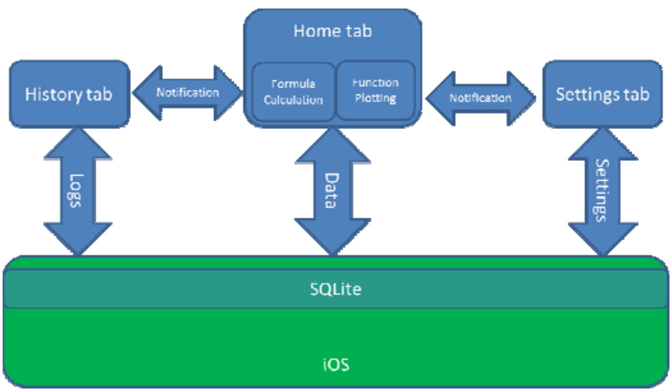


Fig. 1. The software architecture of Caffeine Zone

3.2 The Pharmacokinetics Equations and Assumptions

We provide several parameters as defaults. These numbers are extracted from previous research in the pharmacology effects of caffeine. They are the half-life of absorption and excretion (currently implemented in the calculation and are unchangeable), and the thresholds for minimum optimal caffeine for cognition and maximum optimal caffeine for cognition. We also include a threshold for sleep. Weight is used to calculate dosage (dose by weight) and units to display, doses (mg), or dosages (mg/kg).

The minimum and maximums for the cognitive range are based on our review [8], and assume that the users are typical, which not all users are. These two numbers represent the minimal does of caffeine that can keep humans alert or that helps cognitive performance and the maximum dose of caffeine that does not impair cognition.

The half-lives are taken from our review. We know that the half-life of elimination should be moderated for nicotine users [9]. In general, the half-life for nicotine users is about half as long as for non-nicotine users. We intend to add this effect in a future revision. The threshold for sleep is currently taken from anecdotal reports by the developers.

These parameters will vary from person to person. Therefore, these parameters are adjustable—users can change the settings based on their usage.

3.3 The Displays

Figure 2 shows two of the current displays, including the timeline of caffeine. By displaying the caffeine model users can understand the time course of their caffeine levels, and moderate their caffeine consumption more effectively. For example, they can switch to decaf coffee before they have consumed too much and go above the cognitive alert zone, or stop consuming caffeine before they will have trouble sleeping that evening, and obtain coffee before they get into traffic.

The screenshot on the left of Figure 2 shows the settings screen where users adjust some thresholds and parameters. Changing the Sleeping Level, Max Optimal, and

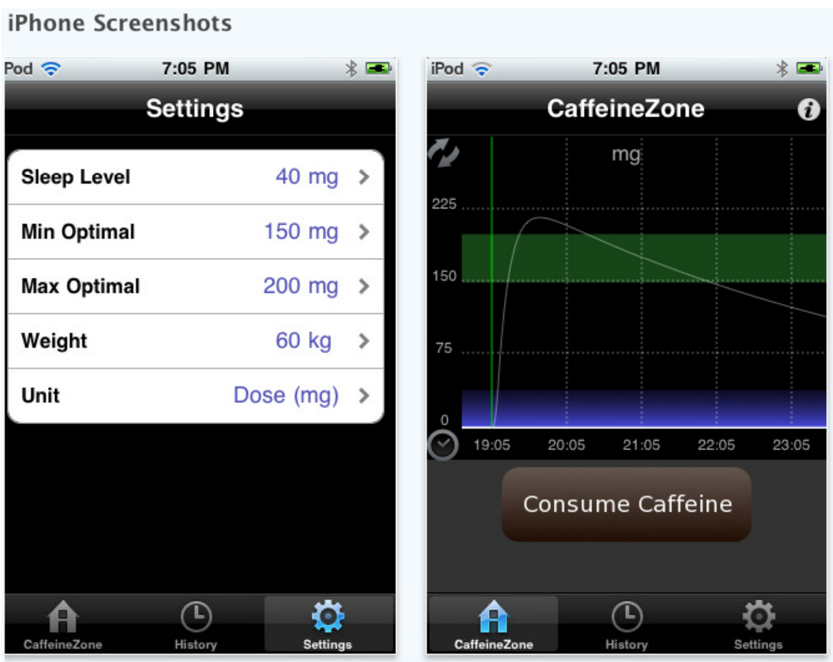


Fig. 2. The interface for Caffeine Zone, showing on the left some of the adjustable parameters, and on the right, the scrollable pharmacokinetics display

Min Optimal will change the range of the cognitive alert zone (green area on the right of Figure 2, 150 to 200 mg here) and the possible sleep zone (blue area on the right of Figure 2, approximately 0 to 40 mg here).

The screenshot on the right of Figure 2 shows the graphical plot of plasma caffeine concentration over a given time period on the main screen. In the screenshot, the user has consumed a dose of about 240 mg caffeine starting at 19:05. The plot area can be scrolled to left and right to cover forty hours of data points (twenty four hours before and after current point of time). The time on x-axis can be changed to clock time or canonical time.

3.4 Current State and Use

Currently, this application is distributed through the iTunes store as a free application. You can find it on iTunes by searching for 'Caffeine Zone'. Users can view more about it at http://www.caffeinezone.net, which includes information about the updates, features, and manual. We have had more than six hundred downloads from thirty-six different countries from August 2010 to January 2011.

3.5 Benefits of Caffeine Zone

This app may help operators create a mental model of caffeine and learn a better way to use caffeine as a psychoactive substance. Having a working knowledge of caffeine can reduce the risk of ineffective use of caffeine. Any operator wants to extend their cognitive attention or reduce reaction time may find this model of caffeine helpful.

In addition, an accurate model of caffeine can be used to predict when operators may loss their attention or become fatigue due to low caffeine load. Systems with the caffeine model will be able to predict users' cognitive state, or even emotional state, more accurately and provide assistance when needed.

4 Conclusions

We introduced a novel, yet readily accessible and inexpensive, way of visualizing caffeine intake for augmenting cognition through caffeine consumption for users. Because caffeine is already part of many people's daily diet, by using Caffeine Zone, users, including those in operational settings and safety critical systems, may be able to improve their vigilance and alertness and prolong cognitive attention more safely and with fewer side effects. There are many critical operational tasks, such as air traffic control, driving, and radar screen operators, that require constant alertness and heavy cognitive processes. Allowing these users to remain cognitively active in critical situations using and not over using caffeine may improve performance.

Caffeine Zone can be used as a caffeine monitoring device and a teaching tool for caffeine intake. Users who do not consume caffeine regularly can use it as a caffeine monitor—users with an iOS device now have a convenient and inexpensive real-time caffeine monitor to improve their cognitive alertness with the possibility of less over-use. Users who consume caffeine in certain pattern may build a mental model of caffeine that allows them to disengage from the tool but remain an effective caffeine consumer. Both types can learn from using this application and teach themselves without explicitly learning a complex model of caffeine. We suggest that this application will eventually help caffeine users to improve their cognitive performance by more explicitly self-monitoring their caffeine use, and to learn how and when to use caffeine to improve their alertness and sleep pattern.

Understanding the effect of substances in cognition, such as caffeine, is a precursor of two interesting applications. One will be a more complex and complete system to model more than caffeine, perhaps sleep and other moderators of cognition.

The other interesting application is creating more sophisticated computational models of humans in synthetic environments. Including effects like those of caffeine helps modelers realize the limitation in cognition and adds another dimension or layer to the traditional theory of human information processing that deals primarily with an internal, static, and completely rational model of cognition.

4.1 Limitations

There are several limitations to Caffeine Zone. These range from limitations to the theory to limitations of understanding the application's use and impact. Limitation with the equations: There is no consensus of a complete caffeine model to the best of our knowledge. Our model is based on our own search in this area, and we are aware that the parameters might vary for different populations and different environments (e.g., [10, 11]). The adjustable thresholds can mitigate part of this limitation, however.

Limitations with the interface: Our app does not include a rich set of caffeinated beverages and foods. More and more energy drinks include substantial amounts of caffeine. Including these would make the app easier to use. In addition, there are some inherent limitations in these numbers because there are variations in the amount of caffeine in the same type of beverage, for example, in coffees. Although there is a custom option when entering data to our app, using it may require more knowledge about caffeine levels in different substances than a novice user might have. For example, to know the amount of caffeine in a 30 oz. coffee will require the caffeine/ounce and then a multiplication. Providing an extensive set of predefined sources of caffeine will make data entry easier, faster, and less cognitively demanding for many users. To achieve this, we should include a more complete database of caffeine content in foods and drinks.

Limitations with impact: in this case, we do not know if and how it is used: As it stands today, there is not enough empirical data to support our conjuncture that users will often use and learn from our app as we expect they will. In fact, the history has taught us that people will use a technological tool differently than what the original designers expect. One local Caffeine Zone user told us he used it to "coach" him when to consume to remain alert at a specific time period in the future. We are planning to start collecting preliminary data to learn how such feedback works and to understand its impact to regular users.

Finally, it is not used enough to have a broad impact. There is only a small user base for this application, and we have not collected empirical data to see whether the current users do benefit from what we think they will. We need more users and feedback to analyze and maximize it impact and validate our absorption and elimination formulas.

4.2 Future Applications

Further work on this application is possible, including both near term and far term projects. In the near term, we will work on the area of extending the application to help coach users, that is, to generate a recommendation on when/how much to drink, given different goals such as to remain alert during a period of time, not being influenced by caffeine at certain time, etc. This will mean including a set of alarms to know when to start or stop consuming based on user-defined thresholds.

In particular, we are working on a revision that will include three alarms/guards to help users regulate caffeine consumption: (a) a Min alarm: this alarm generates a notification when the caffeine intake is about to drop out of the cognitive alert zone. Users can use this notification and decide whether additional caffeine consumption is appropriate; (b) Max guard: this guard pops up a warning when the app detects the consumption of the caffeine source will generate a caffeine intake that exceeds the maximal threshold. This function allows users to change the amount/way they are about to consume caffeine. (c) Sleep guard: this guard pops up a warning when the app detects that the consumption may interfere with sleep. Again, this function will allow users to change their behavior immediately.

Users might also be able to apply this application to different but similar drugs and substances. When the half-lives of absorption and eliminate are modifiable, other drugs can be modeled, and the platform can be used for teaching about other drugs, including nicotine, for example. This will extend the potential use of this app to a broader audience.

We are also starting a study to understand whether and how people change their use of caffeine based on this application. This may provide evidence about how Caffeine Zone alters the use of caffeine, and may provide additional feedback about how the app is used and can be improved. In addition, we would like to study the use of Caffeine Zone by different types of users. So far, we have studied users pretty much like ourselves. It would be useful to see how it is used by users on ships or in a desert, and in operational environments were ease of use might become acute, or where sleep cycles change, or where caffeine sources vary.

Longer term, we could extend this approach to include more aspects of behavior and the environment that affect alertness. These factors could include physical activity and previous rest, time of day, and other foods known to influence alertness. The app could also measure alertness directly and modify the equations accordingly for each user. If there are mobile sensors that can provide psychological information such as heart beats, blood pressure, and pupil dilation, incorporating these information with the caffeine model will provide accurate cognitive state measurement data for augmented cognitive systems.

Finally, this work may be useful inside other systems trying to predict and assist users. From a designer's point of view, a system should have a model of the user to assist understanding and predicting its users. If the users are tired or fatigued and use caffeine, the system could create visual or audio cues or offload cognitive intensive process to automation or other team members. This could help prevent errors. Our caffeine model can be used with other cognitive state measurements to provide a more accurate user model to help augment cognition and performance.

Acknowledgements. Background work on studying caffeine was supported by ONR grant N00014-03-1-0248. Discussions with Susan Chipman encouraged us to create this app. Discussions with Laura Klein helped us understand pharmacokinetics and pharmacodynamics. This presentation was improved by comments from Nathan Gerhart and Monique Beaudoin.

References

1. Kendler, K.S., Prescott, C.A.: Caffeine Intake, Tolerance, and Withdrawal in Women: A Population-Based Twin Study. American Journal of Psychiatry 156, 223–228 (1999)
2. Fredholm, B.B., Batting, K., Holmen, J., Nehlig, A., Zvartau, E.E.: Actions of Caffeine in the Brain with Special Reference to Factors that Contribute to its Widespread Use. Pharmacological Reviews 55(1), 83–133 (1999)
3. Horne, J.A., Reyner, L.A.: Counteracting Driver Sleepiness: Effects of Napping, Caffeine, and Placebo. Psychophysiology 33(3), 306–309 (1996)
4. Institute of Medicine: Caffeine for the Sustainment of Mental Task Performance. National Academy Press, Washington, DC (2001)
5. Klein, L.C., Bennett, J.M., Whetzel, C.A., Granger, D.A., Ritter, F.E.: Caffeine and Stress Alter Salivary α-Amylase Levels in Young Men. Human Psychopharmacology: Clinical and Experimental 25, 359–367 (2010)
6. Hughes, R.N.: Drugs which Induce Anxiety: Caffeine. New Zealand Journal of Psychology 25(1), 36–42 (1996)
7. Naval Aerospace Medical Research Laboratory: Performance Maintenance During Continuous Flight Operations: A Guide for Flight Surgeons, Vol. NAVMED P-6410. Naval Strike and Air Warfare Center (2000)
8. Morgan, G.P., Ritter, F.E., Stine, M.M., Klein, L.C.: The Cognitive Effects of Caffeine: Implications for Models of Users: unpublished mss (2006)
9. Seng, K.-Y., Fun, C.-Y., Law, Y.-L., Lim, W.-M., Fan, W., Lim, C.-L.: Population Pharmacokinetics of Caffeine in Healthy Male Adults Using Mixed-Effects Model. Journal of Clinical Pharmacy and Therapeutics 34, 103–114 (2009)
10. James, J.E.: Understanding Caffeine: A Biobehavioral Analysis. Sage, Thousand Oaks (1997)
11. Julien, R.M.: The Primer of Drug Action. Worth Publishers, New York (2001)

Designing Consumer Health Information Systems: What Do User-Generated Questions Tell Us?

Yan Zhang[1] and Wai-Tat Fu[2]

[1] School of Information, University of Texas at Austin, Austin, TX, 78701
[2] Department of Computer Science, University of Illinois, Urbana, IL 61801
yanz@ischool.utexas.edu, wfu@illinois.edu

Abstract. Searching for health information has become a prevalent activity on the web. The information found online has a significant impact on people's decisions on whether to seek medical care and what treatments to undergo. However, existing studies consistently suggest that general consumers have various difficulties in formulating search queries using existing search engines and the queries were often not effective in retrieving personal- and situational-relevant information. Understanding users' information needs is a gateway to designing effective information retrieval (IR) systems. In this study, we examined the types of information requested by users, the characteristics of consumers' expressions of their information needs, and their expectations for results by analyzing the questions that general users posted on Yahoo! Answers, a popular social Q&A site. Based on the results, we proposed design recommendations for facilitating users' ability to articulate their health information needs and recommendations for the presentation of information in health-related IR systems.

Keywords: Consumer health informatics, information retrieval, social Q&A, health information searching.

1 Introduction

Recent Pew studies show that more than 80% of the internet users in the U.S. seek health information online and this activity has become the third most popular activity across all age groups, after Email and search engine use [30]. Furthermore, the information found online significantly impacts people's decisions about their own health or the health of someone in their care [9].

In response to the high demand for health information and the significant impact of this information, a lot of efforts have been made, in the past decade, to make reliable and trustworthy health information sources available online. However, little attention has been paid to the usability of and the user's experience with such sources. Subsequently, there is still a significant lack of understanding of the factors that impact users' adoption of health information systems [14, 16]. Meanwhile, consumers consistently report difficulties in finding personally relevant information using existing sources [29]; and they frequently feel confused and frustrated by their searching experiences [1].

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 536–545, 2011.

In order to design not only useful but usable health information systems, it is necessary to understand how consumers conceptualize and articulate their health needs and how they use information systems to find relevant information [3, 14]. This study is an effort to improve the understanding of consumer health information needs. Specifically, we explore types of information requested by consumers, the ways in which they express their needs, and their expectations of the answers by analyzing health-related questions that general users posted on a social Q&A platform, Yahoo! Answers. The implications of the results to the design of health information systems are discussed as well.

2 Related Literature

The current research on consumer health information needs mainly takes two approaches. The first approach is from a clinical perspective using structured surveys to identify types of health information such as information about cure, spread of disease, and treatment, needed by patients or their caregivers. For example, several studies employed a structured information needs questionnaire (INQ) to examine the information needs of cancer patients or their care givers. The INQ includes nine items: information about cure, spread of disease, treatments, side effects of treatment, genetic risk, self care, sexual attractiveness, impact on social life, and impact on family [2, 7; 19]. Harrison, et al. [12] used the Toronto Informational Needs Questionnaire (TINQ) to survey women with breast cancer concerning their demand for specific information related to their disease, their treatment, and relevant investigative tests.

The second approach is analyzing transaction logs, particularly queries. In addition to identifying subject areas that people search for [22], query analyses can shed light on vocabularies that users use to search for health-related information and the difficulties that they have with the searching. These studies found that people submit short queries, with an average number of 2.2 to 3.3 terms per query, to search engines for health-related information [22, 27]. Short queries often lead to insufficient representation of health problems in search and unsatisfying results. At the same time, misspellings and abbreviations are common in queries, which tend to cause search failures [4, 20].

Query analyses also reveal conceptual difficulties that consumers have in searching for health information. First, their vocabularies do not match medical terminologies, such as those found in UMLA and used in the content or indexes of health information websites [15]. Second, they sometimes could not find proper terms to describe their actual needs. And third, they have different mental representations of their conditions from medical professionals and are likely to describe the conditions using simpler or more concrete terms [29].

Apparently, these two approaches provide valuable insight into what information is needed by general people and the difficulties that people have in retrieving health-related information. However, they fall short of revealing important elements of people's information search tasks that result from conscious information needs, such as how people conceptualize their problems, what their intentions are, and how they convey their conceptualizations to information systems [6, 21]. The research on interactive information retrieval (IIR) consistently suggests that users' perceptions of tasks, such as intended goals or purposes, task complexity, and task familiarity, have a significant impact on their mental models of the system that they are interacting with, their behavior of using the systems, and their experience with the systems [e.g., 24]. Therefore, an enhanced understanding of consumer health information needs, which give rise to information search tasks, will be a rich source for informing the design of better information systems.

In recent years, supported by the fast development of social media technologies, community-based online health forums, such as bulletin boards, blogs, wikis, social networks, and social Q&A sites, became widespread on the web. The emergence of these communities was considered critical in empowering patients to make decisions concerning their own health and promoting patient-centered healthcare [8]. These communities often consist of patients with a particular disease, such as breast cancer, or people who share a particular topic of interest, such as running, patients' caregivers, and sometimes, some health professionals. When faced with a problem, the most natural way for a person to seek information is to ask questions of someone with more expertise or who share similar experiences [28]. Therefore, the most popular activity on these social platforms is asking health-related questions or answering questions posted by peer users [10, 17]. Unlike in web search engines where users often type in short queries, these platforms allow the users to ask questions in natural language and in full sentences. Therefore, these questions can be a good source for learning a range of factors involved in consumers' health information needs and their related information search tasks, such as background of the askers, the askers' perception of the tasks, the nature and goals of the tasks, and the constraints around the tasks [23, 24].

This study presents a preliminary effort to explore the characteristics of consumer health information needs that have direct impact on information searching by analyzing health-related questions that general consumers posted on Yahoo! Answers, a major social Q&A site. Implications of the results for designing effective consumer health information systems are discussed, as well.

3 Research Method

A dataset including 77,903 questioning messages across 23 health-related categories, such as diabetes, heart disease, cancer, and diet and fitness were crawled from Yahoo! Answers [18]. At the time the data was crawled, this site had about 74% of the market share of U.S. visits among Q&A websites [13]. We randomly selected 276 questioning messages from the dataset and analyzed the messages using the qualitative content analysis method. A content analysis software, QSR, was employed to assist the data analysis. Each questioning message was a unit of analysis.

We first coded the number of questions contained in each message. A question is loosely defined as a request sent by an asker to one or more respondents to solicit knowledge on a certain subject, which the asker lacks but sincerely wants [26]. Then, the open coding method was employed to code the data. At the beginning of the coding, we read the text several times to gain an overview of the overall content. Then codes were derived inductively by closely reviewing themes that appeared in the messages. Whenever a theme related to health information needs appeared, such as for whom a question was asked, what information is needed, difficulties in expressing their needs, and expectations for answers, it was coded into a category. When coding a new text to a category, the text was compared to those already assigned to that category [11]. The constant comparison method allowed us to fully understand the theoretical properties of the category. We then examined the categories resulted from the open coding process to make sense of the properties and dimensions of the categories, identify relationships between them, and uncover patterns [5]. To improve the validity of the results, a second coder coded 20% of the questions. The inter-coder agreement reached 87.9%.

4 Results

Among the 276 questioning messages, nine were advertisements or messages eliciting comments and prayer from others. These messages were not intent to solicit knowledge that the asker lacked, thus did not meet our definition of questions. They were excluded from the subsequent analysis. In this section, we report three important aspects of consumer health information needs that are directly related to information searching and information system design: types of information needed, characteristics of users' expressions of the information needs, and users' expectations for results.

4.1 Types of Information Needed

Information systems, particular IR systems, are designed to help people find relevant information to solve particular tasks. Therefore, knowledge about types of information requested by users is important for system designers. The analysis of the questioning messages revealed the following major types of information needs. The number in the parentheses is the number of messages.

Information about a particular disorder/disease (123). These questions were asking about a specific aspect of a particular disease, mainly symptoms, causes, diagnoses, treatments, and prognoses. When asking about symptoms, users' questions followed two patterns: what are the symptoms of a disease or condition and whether it is normal for people with a particular disease or condition to have certain symptoms. When asking about causes, users were concerned about what factors or behaviors cause a disease or a condition. An example is: "*What raises people's blood pressure? Please give me list of stuff that raises your blood pressure and how it raises.*"

Questions concerning diagnoses had three different patterns. The first was that askers described certain symptoms, conditions, or behaviors and asked for possible diagnoses. For example, "*I'm really desperate to find out what the lump in the back of my throat is. I also have an inflamed gland on the side of my throat (neck). […] Do you think this sound like throat cancer?*" The second pattern was that askers were concerned about themselves having a particular condition and asked whether a symptom suggests that particular condition. The third pattern was that askers wanted to double check a doctor's diagnoses. For example, "*There is a hyper density 2mm in my liver. Can anybody tell me if it's something to worry about? My doctor said it's nothing but I need to be sure.*"

Questions concerning treatments were asking about effective treatments to certain conditions, the cost of a treatment, how long it takes, whether it is hard to get through, and whether there are other options. Askers asking for prognoses often had a diagnosis and wanted to find the prognosis for that condition. For example, "*Hubby has dilated cardiomyopathy, stage 4, what is his prognosis?*"

Some askers knew about diagnoses, but wanted to find out whether the disease or condition would affect their other life plans, such as carrying babies and traveling. An example is: "*What kind of risk [is] involved if I carry baby with heart disease?*" Some askers were asking about attributes of a disease or a condition, such as whether it is genetic, whether it is contagious, whether it is fatal, how serious it could be, how fast it spreads, whether it is rare, whether it will recur, and the age of getting the disease. A few askers also requested general information about a disease, for example, "*What is cancer?*" and "*What is pink eyes?*" A few people asked questions concerning the recovery of a disease, such as how long it will take to recover.

Information about drugs or supplements (23). Questions about drugs or supplements were mainly about the following themes: whether a particular drug is effective, what drugs to take, how much to take and when, whether there are side effects, what the ingredients are, whether it is safe, whether it interferes with other medical tests, whether the way in which a drug had been stored has impact on its effect. An exemplar question asking about the effectiveness of a drug is: "*Does relacore really work? I am exercising and I have been eating pretty well. I was thinking of taking relacore to help me lost weight, has anyone used it? Does it work?*"

Information about lifestyle, mainly diet and exercise (10). Questions concerning lifestyle often asked for recommendations for a healthy diet or exercise routine, given a specific weight, height, age, or health condition or checked with others whether their current exercise routine is reasonable, for example, "*Hi, I am 33 yrs old… My heart rate drops as I started exercising everyday (for about 2 months). I run 3-4 miles, sit-ups and play ping pong EVERYDAY.....I have a blood pressure monitor with which I measure my BP. and it seems to me it dropped down and heart rate dropped down from 72BPM to 55 BPM [...], I am also on Lexapro, (medicine for depression)… I am afraid I am over-exercising? I don't smoke or consume alcohol...*" Some of the questions in this category also asked for more information on a particular diet, such as calorie shifting diet, and a lifestyle to adopt after experiencing a disease. For example, "*What lifestyle to adopt after gastric ulcer disease? [..].*"

People with similar conditions (23). There were a number of askers looking for people with similar conditions on Yahoo! Answers, for general information or advice, success stories, treatment information, and drug information. An example is: "*Has anyone tried ortho-k lenses coupled with eye exercises for permanent vision improvement? How well does it work?*"

Information sources (12). Some askers asked about sources for information, such as quotes to cheer up friends, reliable websites to buy medications, animated diagrams of a heart attack, statistics about the causality of lung cancer and percentage of people smoking worldwide, or a yoga DVD for pregnant women.

Others (20). Askers also asked questions concerning medical professions, for example, "*I have a niece interested in becoming a Physical Therapy Assistant. What is the pay? We are in NC,*" body working mechanisms, for example, "W*hat would happen if the right side of heart pumped faster than the left side?*", interpretations of medical tests readings, such as blood pressure, information about coping with bad life situations and managing stress, and health insurance and policies. An example about insurance and policies is, "*Can I bring mum over to Australia for treatment for lung cancer from UK? Does NHS subsidize any costs (or medicare)? […]*"

It is worth noting that many askers were asking questions not for themselves, but for related others, such as sisters, brothers, parents, grandparents, and friends. Among the 267 messages, 25 explicitly stated that the questions asked were for someone other than themselves.

4.2 Characteristics of Users' Expressions of Health Information Needs

Characteristics of users' expressions of health information needs refer to linguistic features of the questioning messages. Queries or questions are a major means through which users interact with an information system. Therefore, features of users' expressions impact the quality of the results of user-system interaction. In this study, users' expressions of health information needs were analyzed at both the term level and the question level.

At the term level, inappropriate terms, particularly, misspellings (e.g., canser, diagnoised, suppliment, and gentically) and run-together phrases (forcancer, to'prevent'whatever, medication"zimvastatin 10mg", and cardiologygrizzln) were prevalent in the sample questioning messages. Sometimes, the misspellings were due to the lack of auto-correction mechanisms in Yahoo! Answer's question submit form. Sometimes, it was due to the fact that users had difficulties in spelling a medical term. For example, one asker stated that, "*I'm looking for a website on a rare skin disease, it sounds like dariase DARE E AZE? Uncertain of the correct spelling […].*" The second term level linguistic feature that could cause communication difficulties between user-system or user-user was the use of acronyms, such as B.P., AML, EKG, ECHO, CKMB, CK, and NHL. In this study, about 5% of the sample questioning messages contained such acronyms. The third term level difficulty was conceptual, specifically, users could not find terms to describe their conditions. One asker stated: *[…] I've inherited a condition, which gives me more of a chance of getting lung and liver cancer if I smoke and drink. My question is, what is the condition called? I want to know more about its history. […]*"

At the question level, in the sample questioning messages, about 75% of the messages contained only one question, 18% contained two questions, and the rest 7% contained 3 to 5 questions. Multiple questions in a questioning message were often asking about different aspects of a disease, drug, surgery or procedure. For example, an asker asked two related questions about a condition: "*Is clogging of the heart arteries reversible? If so how?*" Similarly, another asked "*What is a spleenic lesion and how is it treated?*" In many cases, one question in such messages was asking about whether there was someone with similar experience. For example, "*Who is taking hydrochlorothiazide 12.5 mg and lisnopril 10mg and does it help to lower your blood pressure? What is your blood pressure after taking these two medicines?*"

4.3 Expectations for the Answers

In asking questions, askers also expressed their expectations for the answers. The expectations had three dimensions, one concerning the quality of the information, one concerning the personal or situational relevance of the information, and the other concerning the social attitudes involved in the answers. For the quality of information, some pointed out in the messages that they wanted reliable information, for example, one asker asked, "*Where is a reliable site to buy medication online?*" Some askers indicated that they wanted answers from medical professionals, and some indicated that they wanted to see the source of the information.

As mentioned above, many users of Yahoo! Answers tried to seek information, advice, or personal experience from people with similar conditions. One reason for this phenomenon seems to be that users favor personally and contextually relevant information. For example, one asker tried to elicit personal stories "*Do you have a child with congenital heart disease? [Has he/she] undergone an open heart surgery? […] How's the operation? Please tell something about after the surgery.*"

Users of Yahoo! Answers also cared about the attitudes of the people who answered their questions. Askers wanted the answers to be genuine, candid, kind, and serious. For example, "*What's the best yoga/workout post natal DVD? I'm pregnant but due fairly soon, and what to start a collection of some workout dvds and need some input on the best for a post-natal body. ** Serious answers only … kids DON'T respond saying some dumb remarks and NO ignorant remarks […]*"

5 Discussion and Implications

Understanding the characteristics of consumer health information needs is essential to designing and constructing effective consumer health information systems [3]. In this study, we explored the types of information requested by users, the characteristics of their expressions of information needs, and their expectations for results. It is apparent that consumers need different types of health-related information, with the majority of the request focusing on different aspects of categories including diseases, drug and supplements, lifestyles, people with similar conditions, and information sources. This observation suggests that consumer health information systems should construct effective information architecture to support users' access to these different types of information. The first three categories were also identified in several other studies [e.g., 25, 26] and were made available in many consumer health websites, such as MedlinePlus, however, little attention has been paid to providing effective access to people with similar conditions and other relevant information sources (like websites and DVDs). Future design of consumer health information systems should make an effort to improve access to these two types of information.

In this study, it was also found that users sometimes have several different questions concerning a particular disease, medication, or surgery. This suggests that systems can provide cognitive assistance to help users find relevant information. Such assistance could be implemented by providing access to related topics. The selection of related topics could be based on data mining of user-generated questions.

Consistent with the existing literature [4, 20, 22], this study found that users have various linguistic and conceptual difficulties in expressing their health information needs. Misspellings and run-together phrases were common and these errors sometimes prevented systems or other users from understanding the messages correctly. Most of the misspellings and run-together phrases were due to users' cognitive slips and could be solved by providing a spelling check function in the question submit form. Sometimes, the misspellings were due to the fact that users do not know how to spell a medical term or they do not know what terms to use to describe their conditions. These difficulties suggest that the system should provide a function to assist users in finding appropriate medical terms to describe their thoughts. This help could significantly improve users' experience with health information searching.

Looking for health-related information online is a highly social activity. Many people asked questions for family and friends who they care about. People wanted information, advice, personal stories, and emotional support from those with similar conditions; and people wished the answers to be genuine, candid, and serious. These observations suggest that users want personally and situationally relevant information and they value the information from peers. Most health information systems provide systematic scientific knowledge on consumer health topics, however, due to the static nature, they often fall short in supporting people's highly idiosyncratic health information needs. One way to improve users' experience with health information seeking is to integrate social networks with traditional information retrieval system and leverage information searching using collective wisdom. Future research could investigate how to integrate social systems with traditional IR systems to best augment general users' cognition and achieve more effective health information searching.

This study also found that users are concerned about the quality and credibility of the results. They wanted to see the source for answers and they preferred answers from medical professionals. They also wanted the answers to be sincere and candid. This suggests that a consumer health information system should provide mechanisms to help users judge the quality of information. If a social network component is involved, it will be desirable to have means to prevent, block, or penalize malicious and ignorant remarks. In future studies, it will be worthwhile to investigate what criteria users employ to judge the quality and relevance of results.

Acknowledgments. The authors acknowledge the support of the Research Grant from the Office of the Vice President for Research at the University of Texas at Austin.

References

1. Arora, N.K., Hesse, B.W., Rimer, B.K., Viswanath, K., Clayman, M.L., Croyle, R.T.: Frustrated and confused: the American public rates its cancer-related information-seeking experiences. Journal of General Internal Medicine 23(3), 223–228 (2007)
2. Beaver, K., Witham, G.: Information needs of the informal carers of women treated for breast cancer. European Journal of Oncology Nursing 11, 16–25 (2007)
3. Belkin, N.J., Oddy, R.N., Brooks, H.M.: Ask for information retrieval: Part I. background and theory. Journal of Documentation 38, 61–71 (1982)

4. Boden, C.: Overcoming the linguistic divide: a barrier to consumer health information. Journal of the Canadian Health Libraries Association 30, 75–80 (2009)
5. Bradley, J.: Methodological issues and practices in qualitative research. Library Quarterly 63(4), 431–449 (1993)
6. Bystrom, K., Hansen, P.: Conceptual framework for tasks in information studies. Journal of the American Society for Information Science and Technology 56(10), 1050–1061 (2005)
7. Degner, L.F., Kristjanson, L.J., Bowman, D., et al.: Information needs and decisional preferences in women with breast cancer. Journal of the American Medical Association 277, 1485–1492 (1997)
8. Eysenbach, G., Powell, J., Englesakis, M., Rizo, C., Stern, A.: Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. BMJ 328(7449), 1166–1171 (2004)
9. Fox, S., Jones, S.: The Social Life of Health Information. Pew Internet & American Life project (2009) `http://www.pewinternet.org/~/media//Files/Reports/2009/PIP_Health_2009.pdf` (retrieved March 3, 2010)
10. Ginossar, T.: Online participation: A content analysis of differences in utilization of two online cancer communities by men and women, patients and family members. Health Communication 23, 1–12 (2008)
11. Glaser, B.G., Strauss, A.L.: The Discovery of Grounded Theory: Strategies for Qualitative Research. Aldine, New York (1967)
12. Harrison, D.E., Galloway, S., Graydon, J.E., Palmer-Wickham, S., van der Bij, L.: Information needs and preference for information of women with breast cancer over a first course of radiation therapy. Patient Education and Counseling 38, 217–225 (1999)
13. Jasra, M.: Yahoo Answers down but dominating Q&A websites (March 20, 2008), `http://www.webanalyticsworld.net/2008/03/yahoo-answers-dominates-q-websites.html`(accessed on May 3, 2010)
14. Keselman, A., Logan, R., Smith, C.A., Leroy, G., Zeng, Q.: Developing informatics tools and strategies for consumer-centered health communication. Journal of the American Medical Informatics Association 15(4), 473–483 (2008)
15. Keselman, A., Smith, G.A., Divita, G., Kim, H., Browne, A.C., Leroy, G., Zeng-Treitler, Q.: Consumer health concepts that do not map to the UMLS: Where do they fit? Journal of American Informatics Association 15, 496–505 (2008)
16. Kim, D., Chang, H.: Key functional characteristics in designing and operating health information websites for user satisfaction: An application of the extended technology acceptance model. International Journal of Medical Informatics 76, 790–800 (2007)
17. Klemm, P., Reppert, K., Visich, L.: A nontraditional cancer support group: the Internet. Computer in Nursing 16(1), 31–36 (1998)
18. Liu, Y., Bian, J., Agichtein, E.: Predicting information seeker satisfaction in community question answering. In: Proceedings of SIGIR (2008)
19. Luker, K.A., Beaver, K., Leinster, S.J., Owens, R.G.: The information needs of women with breast cancer: a follow up study. Journal of Advanced Nursing 23, 487–495 (1996)
20. McCray, A.T., Tse, T.: Understanding search failures in consumer health information systems. Proceedings of AMIA Symposium, 430–434 (2003)
21. Rose, D.E., Levinson, D.: Understanding user goals in web search. In: Proceedings of WWW 2004, pp. 13–19 (2004)

22. Spink, A., Yang, Y., Jansen, J., Nykanen, P., Lorence, D.P., Ozmutlu, S., Ozmutlu, H.C.: A study of medical and health queries to web search engines. Health Information and Libraries Journal 21, 44–51 (2004)
23. Taylor, R.S.: Question-negotiation and information seeking in libraries. College & Research Libraries 29, 178–194 (1968)
24. Toms, E.G., O'Brien, H., Mackenzie, T., Jordan, C., Freund, L., Toze, S., Dawe, E., MacNutt, A.: Task effects on interactive search: the query factor. In: Fuhr, N., Kamps, J., Lalmas, M., Trotman, A. (eds.) INEX 2007. LNCS, vol. 4862, pp. 359–372. Springer, Heidelberg (2008)
25. Warner, D., Procaccino, J.D.: Toward wellness: Women seeking health information. Journal of American Society for Information Science and Technology 55(8), 709–730 (2004)
26. White, M.D.: Questioning behavior on a consumer health electronic list. Library Quarterly 70(3), 302–334 (2000)
27. White, R.W., Dumais, S., Teevan, J.: How medical expertise influences web search interaction. In: Proceedings of SIGIR 2008, pp. 791–792 (2008)
28. Wildemuth, B.M., Bliek, R., Friedman, C.P., Miya, T.S.: Information-seeking behaviors of medical students: a classification of questions asked of librarians and physicians. Bulletin of Medical Library Association 82(3), 295–304 (1994)
29. Zeng, Q.T., Crowell, J., Plovnick, R.M., Kim, E., Ngo, L., Dibble, E.: Assisting consumer health information retrieval with query recommendations. Journal of American Medical Informatics Association 13, 80–90 (2006)
30. Zickuhr, K.: Generations 2010. Pew Internet & American Life project (2010), http://www.pewinternet.org/~/media//Files/Reports/2010/PIP_Generations_and_Tech10.pdf (retrieved December 21)

Part VI

Augmented Cognition in Complex Environments

Estimation of Cognitive Workload during Simulated Air Traffic Control Using Optical Brain Imaging Sensors

Hasan Ayaz[1], Ben Willems[2], Scott Bunce[3], Patricia A. Shewokis[1,4], Kurtulus Izzetoglu[1], Sehchang Hah[2], Atul Deshmukh[2], and Banu Onaral[1]

[1] School of Biomedical Engineering, Science & Health Systems, Drexel University
[2] Atlantic City International Airport: Federal Aviation Administration William J. Hughes Technical Center
[3] Penn State Hershey Neuroscience Institute, Penn State University
[4] College of Nursing and Health Professions, Drexel University
{ayaz,shewokis,ki25,banu.onaral}@drexel.edu,
{ben.willems,sehchang.hah,atul.ctr.deshmukh}@faa.gov,
sbunce@hmc.psu.edu

Abstract. Deployment of portable neuroimaging technologies to operating settings could help assess cognitive states of personnel assigned to perform critical tasks and thus help improve efficiency and safety of human machine systems. Functional Near Infrared Spectroscopy (fNIR) is an emerging non-invasive brain imaging technology that relies on optical techniques to detect brain hemodynamics within the prefrontal cortex in response to sensory, motor, or cognitive activation. Collaborating with the FAA William J. Hughes Technical Center, fNIR has been used to monitor twenty four certified professional controllers as they manage realistic Air Traffic Control (ATC) scenarios under typical and emergent conditions. We have implemented a normalization procedure to estimate cognitive workload levels from fNIR signals during ATC by developing linear regression models that were informed by the respective participants' prior n-back data. This normalization can account for oxygenation variance due to inter-personal physiological differences. Results indicate that fNIR is sensitive task loads during ATC.

Keywords: Optical Brain Imaging, Air Traffic Control, Cognitive Workload, Functional Near Infrared Spectroscopy, fNIR.

1 Introduction

Increasing air traffic in the United States has placed the current air traffic control (ATC) system under stress. To alleviate this stress, the Next Generation Air Transportation System (NextGen, Joint Planning and Development Office) has outlined a series of transformations designed to increase the capacity, safety, and security of air traffic operations [1]. Towards achieving this vision, sophisticated new features and systems are designed or will be implemented for future air-traffic management. Brain based measures of operator's cognitive workload could help

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 549–558, 2011.

assess utility of interfaces in human machine systems operating in complex environments and provide objective measures in addition to behavioral performance and subjective self-reported feedback [2-4]. The use of human brain imaging sensors and their capacity to integrate with behavioral and physiological measures, position them to play a key role in the development and operation of futuristic 'brain-in-the-loop' systems.

Significant progress has occurred over the last decade in understanding the physiological and neural bases of cognitive processes and behavior. New and improved brain imaging tools are used to monitor brain activity noninvasively and assess neurophysiological markers of human performance in various settings. Functional magnetic resonance imaging (fMRI), positron emission tomography (PET) and Magneto-encephalography (MEG) have enabled studying localized brain activity in humans and carry out studies for better understanding the neural basis of mental states [5-8]. However, these techniques are not amenable to ecologically valid, regular operating settings as they are expensive, highly sensitive to motion artifact, require participants in confined positions and may expose individuals to potentially harmful materials or loud noise. More recently, functional near-infrared (fNIR) spectroscopy has been used as a noninvasive tool to monitor concentration changes of oxygenated hemoglobin (oxy-Hb) and deoxygenated hemoglobin (deoxy-Hb) at the cortex [9-11]. fNIR technology allows the design of ambulatory, safe and affordable monitoring systems. These qualities pose fNIR as an ideal candidate for monitoring cognitive activity related hemodynamic changes not only in laboratory settings but also under working conditions [12-15].

In this study, we have incorporated fNIR into ongoing studies at the FAA's William J. Hughes Technical Center (WJHTC) where certified controllers were monitored with fNIR while they managed realistic ATC scenarios under typical and emergent conditions. A critical transition defined in NextGen involves augmenting the current auditory-based communications between ATC and the flight deck with text-based messaging, or DataComm systems. DataComm systems are expected to allow ATC to manage more air traffic at a lower level of cognitive load, thereby increasing both the capacity of national airspace system and safety of passengers. Self-report measures of workload suggest that DataComm systems require less cognitive effort than voice-based systems to manage the same amount of traffic [16, 17], and this has been recently corroborated with objective brain based measures as the first part of the current study [14].

The objective of this experiment was to use physiological measures to predict changes in cognitive workload during a complex cognitive task: ATC. The hypothesis derived from a standardized working memory task (n-back) [14] was that blood oxygenation in the dorsolateral prefrontal cortex (DLPFC), as measured by fNIR would increase with increasing task difficulty and sustained cognitive effort.

2 Cognitive Workload

There is no singular definition of cognitive workload [18]. There are at least two major theoretical approaches to the construct: 1) mental workload may be defined such that a given task's requirements are viewed as an independent, external variable with which the working subjects have to cope more or less efficiently; or 2) mental workload may be defined in terms of an interaction between task requirements and human capabilities or resources [19, 20]. In either of these paradigms, the definition of workload involves the "objective" effects of task difficulty on the participant, and the participant's effort involved in maintaining performance. Workload is an intervening variable between task and environmental demands and the operator's performance, defined by the relationship between task demand and the participants resource supply, i.e., the portion of operator's limited mental capacities actually required to perform a particular task. That is, workload can be defined in terms of some "objective" criteria for task difficulty (e.g., managing 6 aircrafts versus 12 aircrafts), or in terms of the participant's capacities to perform the identified task. As such, workload can be dissociated from performance. Two people performing the same task can have identical performance, yet one operator may have significant cognitive resources free to allocate to concurrent tasks, whereas the second operator may be just on the brink of performance failure. The difference between the required capacity and the available capacity of an individual is referred to as the mental or cognitive reserve.

There are four basic methodologies for the assessment of workload. The first category includes subjective assessment that use self-reported rating scores such as Modified Cooper-Harper Scale [21], Subjective Workload Assessment Technique (SWAT) [22], NASA Task Load Index [23] and self -reported mental effort [24]. The second category of cognitive workload assessment methods compares behavior and primary task performance measures of the participant, recorded during the task, to identify any workload effects within them. Accuracy and speed of response are widely in use [25]. The third category includes secondary task measures where participant's mental workload is evaluated based on the performance on the additional secondary task [26]. The final group of cognitive workload assessment methods is based on physiological measures such as eye movements [27], eye blinks [28], pupil dilation [29], skin temperature [30], galvanic skin response [31], heart rate [32], blood pressure [28], and respiration rate [33].

Primary task performance measures are critical in that they define the task that needs to be accomplished. However, they are only sensitive to changes in workload at the limits of mental capacity, which is not very useful for ATC. If operators can compensate for increased workload by increasing effort, then the task performance measure is insensitive. Finally, task performance measures do not provide any direct evidence about the participant's degree of effort or level of arousal. Subjective ratings provide an important measure of the operator's perceived load, and have good face validity. However, they are intrusive to collect while on task, and are often dissociated from actual performance and potential failure. Secondary tasks techniques can give some sense of cognitive or mental reserve, but they are often intrusive, and cannot reasonably be employed during actual ATC activities. Physiological measures can be unobtrusive, and provide an "objective" measure of workload. However, many of the typical physiological systems measure autonomic responses, i.e., the fight or flight system, which assesses stress and emotion, whereas some of the difference in workload may more appropriately be measured using a measure of cognitive workload, rather than stress.

On the other hand, technologies that assess Central Nervous System functioning can provide measures of cognitive functioning as well as measures of stress or emotion. Commonly employed techniques such as electroencephalography (EEG), event-related brain potentials (ERPs), MEG, PET and fMRI have dramatically increased our understanding of a broad range of cognitive and emotional states. EEG have been applied for mental workload assessment of air traffic controllers [34] and operators [35]. Methods that directly measure the summation of neural function, such as EEG, ERP and MEG allow researchers to observe neuronal activation related electromagnetic signals with temporal resolution on the order of milliseconds and have been utilized for workload assessment. However, these technologies also have limited spatial resolution [11] and are susceptible to electromagnetic field artifacts. In contrast, PET and fMRI monitor the hemodynamic and metabolic changes associated with neural activity with impressive spatial resolution, but are limited in temporal resolution and are associated with neuronal activity through a poorly-understood neurovascular coupling function [11]. In addition, PET do not allow for continuous or repeated measurements because they require the use of radioactive isotopes, which also limits their use in children. fMRI is currently considered the "gold standard" for measuring functional brain activation because it offers safe, noninvasive neuroimaging with high spatial resolution. However, fMRI is expensive to operate, requires massive installations and a large team of technicians to operate. In addition, fMRI is highly sensitive to motion artifact and confines participants to restricted positions, and exposes participants to loud noises.

fNIR is an emerging optical brain imaging modality that measures hemodynamic response, similar to fMRI, by using near infrared light [36, 37]. fNIR has been demonstrated to be sensitive to cognitive workload [12-14, 36, 37]. fNIR technology allows devices to be miniaturized, portable and battery-operated, making them field deployable. These qualities make fNIR suitable for the study of cognitive- and affect-related hemodynamic changes under field conditions.

3 Method

3.1 Participants

Twenty-four certified professional controllers (CPC) between the ages of 24 to 55 volunteered. All participants were non-supervisory CPC with a current medical certificate and had actively controlled traffic in an Air Route Traffic Control Center between 3 to 30 years. Prior to the study, all participants signed informed consent forms.

3.2 Experiment Protocol

All participants were asked to complete two types of tasks: n-back and ATC tasks. The n-back task is a standardized working memory and attention task with four incremental levels of difficulty. Participants were asked to monitor stimuli (single letters) presented on a screen serially and click a button when a target stimulus arrives. Four conditions were used to incrementally vary working memory load from zero to three items. In the 0-back condition, participants responded to a single target letter (e.g., “X”) with their dominant hand (pressing a button to identify the stimulus). In the 1-back condition, the target was defined as any letter identical to the one immediately preceding it (i.e., one trial back). In the 2-back and 3-back conditions, the targets were defined as any letter that was identical to the one presented two or three trials back, respectively. The total test included seven sessions of each of the four n-back conditions (hence, a total of 28 n-back blocks) presented in a pseudo-random order. The task was designed and presented in E-prime (Psychology Software Tools).

For the ATC task, each CPC controlled traffic on workstations with a high-resolution (2,048 x 2,048), 29” radarscope, keyboard, trackball, and Direct Access Keypad for 10 minutes. To display the air traffic, the DESIREE ATC simulator and the TGF systems that were developed by software engineers at the WJHTC were used.

Six simulation pilots were used within scenarios by supporting one sector or two sectors and entering data at their workstations to maneuver aircraft, all based on controller clearances. Two types of communications, either voice (VoiceComm) or data (DataComm) communications were used in separate sessions in a pseudo-random order. For each communication type, task difficulty was varied by the number of aircraft in each sector, containing 6, 12 or 18 aircraft.

3.3 fNIR Device

The continuous wave fNIR system (fNIR Devices LLC; www.fnirdevices.com) used in this study is connected to a flexible sensor pad that contains 4 light sources with built in peak wavelengths at 730 nm and 850 nm and 10 detectors designed to sample cortical areas underlying the forehead. With a fixed source-detector separation of 2.5 cm, this configuration generates a total of 16 measurement locations (voxels) [38]. For data acquisition and visualization, COBI Studio software (Drexel University) was used. The sampling rate of the system is 2Hz. During the n-back task, serial cable between the fNIR data acquisition computer and E-prime stimulus presentation computer was used to transfer time synchronization signals (markers) that indicate the start of sessions and onset of stimuli.

4 Data Analysis

For each participant, raw fNIR data (16 voxels x 3 wavelengths) was low-pass filtered with a finite impulse response, linear phase filter with order of 20 and cut-off frequency of 0.1Hz to attenuate the high frequency noise. Saturated channels (if any), in which light intensity at the detector was higher than the analog-to-digital converter limit were excluded.

Using time synchronization markers, fNIR data segments for rest periods and task periods (28 sessions per participant for n-back task and 6 sessions per participant for ATC task) were extracted. Blood oxygenation changes within dorsolateral prefrontal cortex for 16 voxels were calculated using the Modified Beer Lambert Law (MBLL) for task periods with respect to rest periods at beginning of each task [13, 14]. Average oxygenation change for each session was used as the dependent measure. For statistical analysis, 2 (Communication: Data-based, Voice-based) X 3 (Task Difficulty: 6, 12, 18 aircraft) ANOVA with repeated measures on both factors for average self-reported ratings and oxygenation changes were applied. Geisser-Greenhouse correction was used with Tukey's post hoc tests to determine the locus of main effects. The significance criterion was 0.05.

4.1 Normalization by Linear Modeling

A separate first order polynomial regression model was established for each participant and the model parameters were estimated by using min and max of the respective participant's n-back data. This individualized model, mapped oxygenation (of ATC task) through an affine transformation that kept the collinearity, in other words it applied the same scaling and translation for all conditions for that subject. This model transforms ATC oxy values on a standardized n-back conditions (0 to 3) axis by using the min and max nback oxy values as scale. The general model:

$$Y_i = \beta_0 + \beta_1 X_i \tag{1}$$

where X is the oxygenation value and Y is the normalized output response. β_0 and β_1 are scalar model parameters (of a participant) estimated by using two (nback-oxygenation, nback-condition) coordinates, where nback-condition is either 0, 1, 2 or 3. Using minimum and maximum oxygenation (and respective condition), parameters can be solved from the following equation set:

$$\begin{bmatrix} n_{min} \\ n_{max} \end{bmatrix} = \beta_0 + \beta_1 \begin{bmatrix} oxy_{min} \\ oxy_{max} \end{bmatrix} \tag{2}$$

Finally, using this model, ATC oxygenation values for all 6 (2 communication x 3 task difficulty) conditions of the participants were transformed.

5 Results

N-back behavioral and fNIR results were reported in [14]. For the self-assessment data of ATC tasks, there were two significant main effects, Task Difficulty denoted by number of aircraft [$F_{2,46} = 6.79$, $p < 0.05$] and Communication [$F_{1,23} = 4.53$, $p < 0.05$] which is depicted in Figure 1. Tukey post hoc tests for Task Difficulty ($q_{0.05/2, 46} = 3.43$, $p < 0.05$) showed that the 18 aircraft condition had significantly higher oxygenation change than the 6 and 12 aircraft conditions. Also, the interaction between aircraft number and communication type was significant [$F_{2,46} = 4.66$, $p < 0.05$].

For the fNIR data of ATC tasks, two subjects were excluded from the analyses because of high motion artifact and low signal-to-noise ratios. The most significant measurement location was voxel 8 in the medial prefrontal cortex, and there were two significant main effects, Task Difficulty [$F_{2,42} = 4.52$, $p < 0.05$] and Communication [$F_{1,21} = 5.03$, $p < 0.05$] which are depicted in Figure 1. Tukey post hoc tests for Task Difficulty ($q_{0.05/2, 42} = 3.44$, $p < 0.05$) showed that the 18 aircraft condition had significantly higher oxygenation change than the 6 aircraft condition.

The individualized linear model trained by nback data used to normalize oxygenation changes (See Figure 2). Fitted ATC responses indicate that this normalization improved the spatial localization of activity pattern in DLPFC and represent a monotonic increase with increasing task difficulty within anterior medial prefrontal cortex (PFC) [$F_{2,21} = 11.26$, $p < 0.05$].

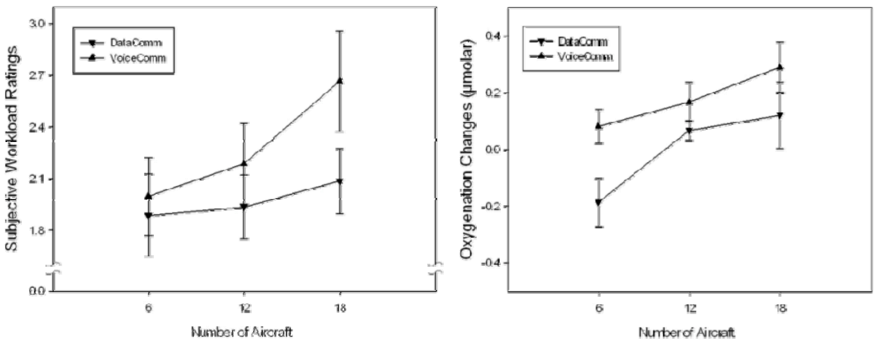


Fig. 1. Self-reported difficulty ratings (left) and average oxygenation changes (right) for each condition (6, 12 and 18 aircraft). Error bars are standard error of the mean.

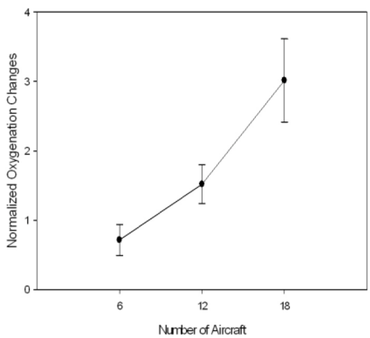


Fig. 2. Average Oxygenation changes normalized by the individualized linear model

6 Discussion

In this study, ATC brain responses under highly regulated and well-known experimental conditions were assessed for comparison with simulated air traffic control tasks. Changes in blood oxygenation in dorsolateral prefrontal cortex, as measured by fNIR, were shown to be associated with increasing cognitive workload. The results further indicate that an individualized linear model improved contrast for comparing responses across task difficulty.

The hypothesis was that, for increasing task difficulty, there would be increasing oxygenation as measured by fNIR. ATC results supported the hypothesis as n-back and earlier results [12, 14]. Average oxygenation in the anterior medial prefrontal cortex, imaged at voxel 8, increased monotonically with our manipulation of increasing task difficulty (number of aircrafts) as illustrated in Figure 1, suggesting this cortical area could provide a valid measure of workload for the air traffic control task. Furthermore, applying normalization resulted in response across larger spatial area by including voxels 7, 9 and 10. For voxel locations please see [38]. The distinction of focal regions for n-back (working memory) and ATC (planning/decision making) are also in parallel with the fMRI findings of functional dissociation of lateral and medial PFC [39-41]. The fNIR results from the main effect of communication type ($p < 0.05$), confirms that VoiceComm condition results in higher oxygenation compared to DataComm with a small to moderate effect size ($d = 0.28$ for voxel 8, and $d=0.25$ for voxel 6). These results are consistent with the idea that, given the same cognitive workload (identical scenarios), DataComm required fewer cognitive resources.

These results must be interpreted with caution, as not all areas of the brain that are involved in ATC could be measured with fNIR technology. It is possible that workload was shifted to other areas of the brain that were not being monitored under the DataComm condition, rather than a reduction in working memory resources. This hypothesis requires further evaluation. However, combined with performance evaluations, the fNIR data provides preliminary evidence that DataComm may require fewer cognitive resources than VoiceComm by lowering the demand on working memory and attentional resources.

In summary, fNIR is a portable, safe, affordable and negligibly intrusive optical brain monitoring technology that can be used to measure hemodynamic changes in the prefrontal cortex. Changes in blood oxygenation in the dorsolateral prefrontal cortex, as measured by fNIR, were shown to be associated with increasing cognitive workload. The results further indicate that text-based communications required less brain activation of the operator than legacy voice based communication systems. These fNIR results corroborate with subjective assessments of operators and earlier studies.

Acknowledgments. This work was supported by the U.S. Federal Aviation Administration through BAE Systems Technology Solutions Services Inc. under Primary Contract, DTFA01-00-C-00068 and Subcontract Number, 31-5029862.

References

1. FAA: FAA's NextGen Implementation Plan. NextGen Integration and Implementation Office (2010)
2. Schmorrow, D., Kruse, A.A.: DARPA's Augmented Cognition Program-tomorrow's human computer interaction from vision to reality: building cognitively aware computational systems. In: Proceedings of the IEEE 7th Human Factors and Power Plants, p. 7, 1-4. IEEE, Scottsdale (2002)
3. Reeves, L., Schmorrow, D., Stanney, K.: Augmented Cognition and Cognitive State Assessment Technology – Near-Term, Mid-Term, and Long-Term Research Objectives. In: Schmorrow, D.D., Reeves, L.M. (eds.) HCII 2007 and FAC 2007. LNCS (LNAI), vol. 4565, pp. 220–228. Springer, Heidelberg (2007)
4. Parasuraman, R., Wilson, G.: Putting the brain to work: Neuroergonomics past, present, and future. Human factors 50, 468 (2008)
5. Cabeza, R., Nyberg, L.: Imaging cognition II: An empirical review of 275 PET and fMRI studies. Journal of Cognitive Neuroscience 12, 1–47 (2000)
6. Wood, J., Grafman, J.: Human prefrontal cortex: processing and representational perspectives. Nature Reviews Neuroscience 4, 139–147 (2003)
7. Ramnani, N., Owen, A.: Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. Nature Reviews Neuroscience 5, 184–194 (2004)

8. Osaka, N., Osaka, M., Kondo, H., Morishita, M., Fukuyama, H., Shibasaki, H.: The neural basis of executive function in working memory: an fMRI study based on individual differences. Neuroimage 21, 623–631 (2004)
9. Chance, B., Zhuang, Z., UnAh, C., Alter, C., Lipton, L.: Cognition-activated low-frequency modulation of light absorption in human brain. Proceedings of the National Academy of Sciences of the United States of America 90, 3770–3774 (1993)
10. Villringer, A., Chance, B.: Non-invasive optical spectroscopy and imaging of human brain function. Trends in neurosciences 20, 435–442 (1997)
11. Strangman, G., Boas, D.A., Sutton, J.P.: Non-invasive neuroimaging using near-infrared light. Biological psychiatry 52, 679–693 (2002)
12. Izzetoglu, K., Bunce, S., Onaral, B., Pourrezaei, K., Chance, B.: Functional Optical Brain Imaging Using Near-Infrared During Cognitive Tasks. International Journal of Human-Computer Interaction 17, 211–227 (2004)
13. Izzetoglu, M., Izzetoglu, K., Bunce, S., Ayaz, H., Devaraj, A., Onaral, B., Pourrezaei, K.: Functional near-infrared neuroimaging. IEEE Trans Neural Syst Rehabil Eng 13, 153–159 (2005)
14. Ayaz, H., Willems, B., Bunce, B., Shewokis, P.A., Izzetoglu, K., Hah, S., Deshmukh, A., Onaral, B.: Cognitive Workload Assessment of Air Traffic Controllers Using Optical Brain Imaging Sensors. In: Marek, T., Karwowski, W., Rice, V. (eds.) Advances in Understanding Human Performance: Neuroergonomics, Human Factors Design, and Special Populations, pp. 21–31. CRC Press, Boca Raton (2010)
15. Menda, J., Hing, J.T., Ayaz, H., Shewokis, P.A., Izzetoglu, K., Onaral, B., Oh, P.: Optical Brain Imaging to Enhance UAV Operator Training, Evaluation, and Interface Development. Journal of Intelligent & Robotic Systems 61, 423–443 (2010)
16. Willems, B., Hah, S., Phillips, R.: The effect of data link on en route controller workload. FAA William J. Hughes Technical Center (2006)
17. Hah, S., Willems, B., Phillips, R.: The effect of air traffic increase on controller workload. Human Factors and Ergonomics Society Annual Meeting 50, 50–54 (2006)
18. Cain, B.: A Review of the Mental Workload Literature. Defence Research and Development, Toronto, Canada, Published online (2007), `http://handle.dtic.mil/100.2/ADA474193`
19. Welford, A.: Forty years of experimental psychology in relation to age: retrospect and prospect. Experimental gerontology 21, 469–481 (1986)
20. Hancock, P.A., Chignell, M.H.: Mental workload dynamics in adaptive interface design. IEEE Transactions on Systems, Man and Cybernetics 18, 647–658 (1988)
21. Cooper, G., Harper, R.: The use of pilot rating in the evaluation of aircraft handling qualities. NASA, Washington (1969)
22. Sheridan, T., Simpson, R.: Toward the definition and measurement of the mental workload of transport pilots (FTL Rept. R 79-4). Cambridge, MA: Massachusetts Institute of Technology, Flight Transportation Laboratory (1979)
23. Hart, S., Staveland, L.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Human mental workload 1, 139–183 (1988)
24. Paas, F.G.W.C., Van Merriënboer, J.J.G.: The efficiency of instructional conditions: An approach to combine mental effort and performance measures. Human Factors: The Journal of the Human Factors and Ergonomics Society 35, 737–743 (1993)
25. Embrey, D., Blackett, C., Marsden, P., Peachey, M.: Development of a Human Cognitive Workload Assessment Tool. MCA Final Report, Lancashire (2006)
26. Meshkati, N., Hancock, P.A., Rahimi, M., Dawes, S.M.: Techniques of mental workload assessment. In: Wilson, J., Corlett, E.N. (eds.) Evaluation of human work: A practical ergonomics methodology, pp. 749–782. Taylor & Francis, London (1995)

27. Ahlstrom, U., Friedman-Berg, F.J.: Using eye movement activity as a correlate of cognitive workload. International Journal of Industrial Ergonomics 36, 623–636 (2006)
28. Veltman, J., Gaillard, A.: Physiological indices of workload in a simulated flight task. Biological Psychology 42, 323–342 (1996)
29. Neumann, D.L., Lipp, O.V.: Spontaneous and reflexive eye activity measures of mental workload. Australian Journal of Psychology 54, 174–179 (2002)
30. Wang, L.-m., Duffy, V.G., Du, Y.: A Composite Measure for the Evaluation of Mental Workload. In: Duffy, V.G. (ed.) HCII 2007 and DHM 2007. LNCS, vol. 4561, pp. 460–466. Springer, Heidelberg (2007)
31. Helander, M.: Applicability of drivers' electrodermal response to the design of the traffic environment. Journal of Applied Psychology 63, 481–488 (1978)
32. Bedny, G., Karwowski, W., Seglin, M.: A heart rate evaluation approach to determine cost-effectiveness an ergonomics intervention. International Journal of Occupational Safety and Ergonomics: JOSE 7, 121–133 (2001)
33. Roscoe, A.: Assessing pilot workload: Why measure heart rate, HRV and respiration? Biological Psychology 34, 259–287 (1992)
34. Brookings, J., Wilson, G., Swain, C.: Psychophysiological responses to changes in workload during simulated air traffic control. Biological Psychology 42, 361–377 (1996)
35. Berka, C., Levendowski, D.J., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T., Olmstead, R.E., Tremoulet, P.D., Craven, P.L.: EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning, and Memory Tasks. Aviation, space, and environmental medicine 78, B231–B244 (2007)
36. Bunce, S.C., Izzetoglu, M., Izzetoglu, K., Onaral, B., Pourrezaei, K.: Functional near-infrared spectroscopy: An Emerging Neuroimaging Modality. IEEE Eng Med Biol Mag 25, 54–62 (2006)
37. Ayaz, H., Shewokis, P., Bunce, S., Schultheis, M., Onaral, B.: Assessment of Cognitive Neural Correlates for a Functional Near Infrared-Based Brain Computer Interface System. In: Schmorrow, D., Estabrooke, I., Grootjen, M. (eds.) FAC 2009. LNCS, vol. 5638, pp. 699–708. Springer, Heidelberg (2009)
38. Ayaz, H., Izzetoglu, M., Platek, S.M., Bunce, S., Izzetoglu, K., Pourrezaei, K., Onaral, B.: Registering fNIR data to brain surface image using MRI templates. In: Conf Proc IEEE Eng Med Biol Soc, pp. 2671–2674 (2006)
39. Bechara, A., Damasio, H., Tranel, D., Anderson, S.: Dissociation of working memory from decision making within the human prefrontal cortex. Journal of Neuroscience 18, 428–437 (1998)
40. Koechlin, E., Corrado, G., Pietrini, P., Grafman, J.: Dissociating the role of the medial and lateral anterior prefrontal cortex in human planning. Proceedings of the National Academy of Sciences of the United States of America 97, 7651–7656 (2000)
41. Simons, J., Gilbert, S., Owen, A., Fletcher, P., Burgess, P.: Distinct roles for lateral and medial anterior prefrontal cortex in contextual recollection. Journal of Neurophysiology 94, 813–820 (2005)

Distributed Logging and Synchronization of Physiological and Performance Measures to Support Adaptive Automation Strategies

Daniel Barber[1] and Irwin Hudson[2]

[1] University of Central Florida Institute for Simulation and Training Applied Cognition and Training in Immersive Virtual Environments Laboratory,
3100 Technology Parkway, Orlando, FL 32826
[2] U.S. Army Research Laboratory, SFC Paul Ray Smith Simulation & Training Technology Center (STTC), Orlando, FL
dbarber@ist.ucf.edu

Abstract. As advances in physiological sensors make them more minimally intrusive and easier to use, there is a clear desire by researchers in the fields of Augmented Cognition and Neuroergonomics to incorporate them as much as possible. To best support use of multiple measures, the data from each sensor must be accurately synchronized across all devices and tied to performance and environment events. However, each sensor provides different sampling frequencies, local timing information, and timing accuracy making data synchronization in logs or real time systems difficult. In this paper, a modular architecture is presented to address the issue of how to synchronize data to support analysis of physiological and performance measures. Specific design requirements are presented to ensure the ability to accurately measure raw sensor data and compute metrics in a distributed computing environment to support adaptive automation strategies in a research environment. Finally, an example system is described which combines multiple minimally invasive physiological sensors.

Keywords: Adaptive Automation, Closed-Loop Training System, Data Synchronization.

1 Introduction

Current advances and capabilities of physiological sensing technologies have made them minimally intrusive and portable enough for use in research efforts for subjective measurement of a user's state [1]. When developing a closed-loop adaptive system or one in which the human acts as a sensor, the ability to measure the workload and potential cognitive state accurately using multiple sensors and performance measures is critical [2]. However, one of the major issues with these devices is that each provides different sampling frequencies, timing resolution, and local time stamps making it difficult to synchronize data across disparate devices and other measurements (e.g. task performance). These problems become even more

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 559–566, 2011.

pronounced when working within a distributed computing environment where multiple machines are used for different devices, tasks, and team members as shown in Fig. 1.

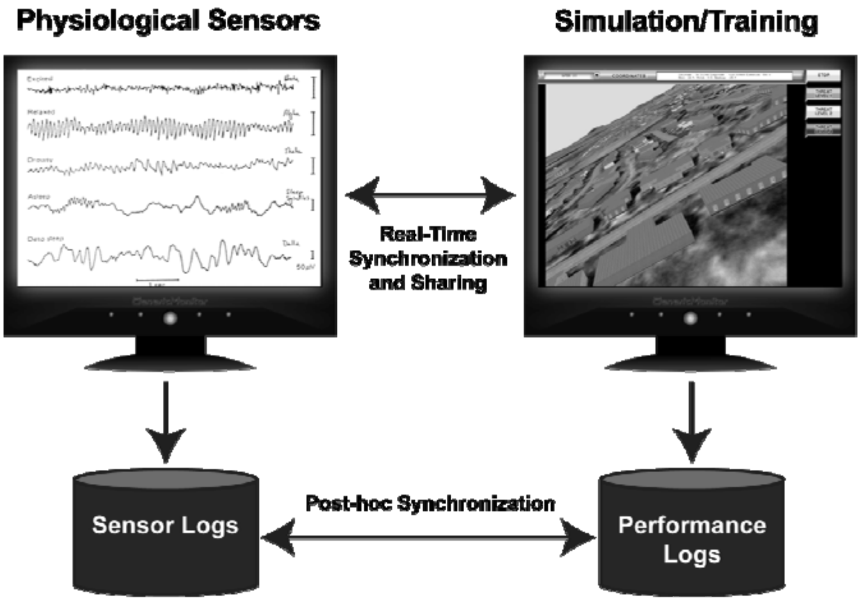


Fig. 1. Example distributed simulation and training environment using physiological sensors

As shown in the above figure, collected data can be shared and synchronized post-hoc, but depending on the timing information recorded this can be a tedious process, and usually performed manually by the researcher. If an appropriate minimum set of timing data, as described in Section 0, is collected it is possible to automate the data synchronization process and provide real time sharing of metrics to support closed-loop training and feedback as described in [1][2][3]. Moreover, the inclusion of unique identifiers for a participant and their group/team makes it possible to design larger systems that can adapt based on the physiological and thus cognitive state of a team. For example, a team of four tasked with supervising multiple unmanned ground systems in a reconnaissance and surveillance mission. During this mission each team member is required to respond to audio and text communications, monitor video feeds, and re-route vehicles based on changes in environment. Using physiological measures from an eye tracker (e.g. eye blink rate (EBR), pupil dilation (PD)) and electroencephalography (EEG), the system may identify that a member of the team is under a high workload condition and based on past measures is likely to enter a period of decreased performance. In this event, the system can decide to re-assign an unmanned system to another member of the team with the lowest workload level, ensuring overall team workload and performance stays within optimal levels.

In the following sections, this paper describes an architecture framework for building systems capable of synchronizing physiological and other measures for post-hoc and real-time analysis of an individual or team state. The required minimum set of data for synchronization of disparate data is described in addition to the design of a system developed at the University of Central Florida's Institute for Simulation and Trainings ACTIVE Laboratory to support real-time adaptation in a closed-loop experimental environment.

2 Synchronization and Correlation of Data

As mentioned previously, sensors and other system modules provide time information for data in different formats, typically relative to when the device or data collection was activated. Using this information alone makes it impossible to match up the times of different data types on a single machine and across a network. Therefore, a time stamp that is from a global reference frame like Coordinated Universal Time (UTC) is required. UTC is a time standard based on International Atomic Time and includes the following information: Day, Hour, Minute, Second, Milliseconds. UTC is also used as the primary method for system time on most modern operating systems making it an ideal choice as a global time reference for tagging data. By recording the current UTC time stamp in a log or in software, when data collection begins for any sensor, it is possible to convert the local sensor time stamps to UTC time for data synchronization.

In addition to recording UTC time stamps for physiological sensors, simulation and operational environments must also be designed to use UTC timing in recorded data to correlate user state with environmental events. An example of a simulation environment that does this is the Mixed Initiative Experimental (MIX) Testbed [4]. The MIX Testbed is a simulation environment developed for research in supervisory control of unmanned systems. It provides log files necessary to support the correlation of physiological data to simulation events [3]. One such log file is called the "Simulation Time Log." This log provides UTC, Local Time (e.g. Eastern Time), and Simulation Time as events occur within the simulation (e.g. Simulation Start, Pause, Resume, Stop). Using this log file it is a simple to extract physiological sensor data from log files that correspond to when a user was performing a mission within the simulation environment post-hoc.

With the use of UTC times associated to data, the next step is to tag sensor and performance records with appropriate identifiers (e.g. User ID, Group ID). With unique identifiers an operational environment or log analyzer can pull sensor data for specific users and teams for analysis or triggering changes in the environment (e.g. re-assign tasking, prompt user). The minimum amount of data that must be associated with recorded data is presented in Table 1. Building off of this minimum set it is possible to create a datagram (i.e. byte array structure) format for transmission of physiological and environment data between computers in a distributed network. Also, with the inclusion of a time stamp representing when the data point was produced, any latency effects in sharing the data across a network will be minimal on filters which consume data to produce new metrics of user state.

Table 1. Record Data Requirements

Field	Name	Description/Interpretation
1	Time (UTC)	UTC Time when this data point was generated or captured by source.
2	Participant/User Number	Unique ID for the participant or user the data point is associated with
3	Group/Team Number	Unique ID for the group or team the participant or user the data point is associated with

3 Operational Psychophysiological Sensor Suite (OPSS)

The design of metrics to determine the real-time changes in cognitive state and workload of a human performing a task using psychophysiological measures is an ongoing field of research [1][5]. The Operational Psychophysiological Sensor Suite (OPSS) is a suite of state-of-the art wearable psychophysiological sensing devices and software tools assembled by the ACTIVE Laboratory at UCF-IST for measuring participants physiological response to cognitive state in real-time. This suite includes devices for eye tracking, Electro Dermal Response (EDR), Heart Rate Variability (HRV), and EEG measurement. To support the synchronization and correlation of data generated by these devices, a software architecture framework has been developed which expands upon the data requirements described in section 2 to support enhanced data logging, filtering, and real-time sharing over a network connection.

As shown in Fig. 2, this architecture design introduces the concept of a Sensor and Filter which both have similar capabilities that include data logging and network communication. A Sensor is an interface to an actual physical device attached to a computer (e.g. Eye Tracker). A Filter is a software module which can be “connected” to a single or multiple sensors and other Filters, and is designed to produce new metrics or cleanup of raw data. Connections between a Sensor and Filters can be made within the software application or through subscriptions over a network connection (e.g. UDP/IP, TCP/IP) producing a “Filter Graph.” Although possible to transmit raw data from sensors, the Filter Graph concept provides the ability to reduce network bandwidth usage by calculating and transmitting only the metrics required by a 3rd party data subscriber. Data logging is an option that can be enabled or disabled for any Sensor/Filter and is used for record keeping or post-hoc analysis through batch processing.

An example of a Filter Graph based on the OPSS System Architecture for an Eye Tracker is described in Fig 3. In this example, there are two filters in the graph: Eye Tracker Filter and Areas of Interest (AOI) Filter. The Eye Tracker Filter takes the raw eye gaze data it receives the Eye Tracker Sensors and produces classifications such as fixations and workload (e.g. Nearest-Neighbor Index) [6]. The downstream AOI Filter uses this data to identify what areas of the screen a user is focusing on for a given point in time. Based on the metrics of the AOI Filter, a program can trigger system events to notify the user that they need to pay more attention to a specific area of the screen.

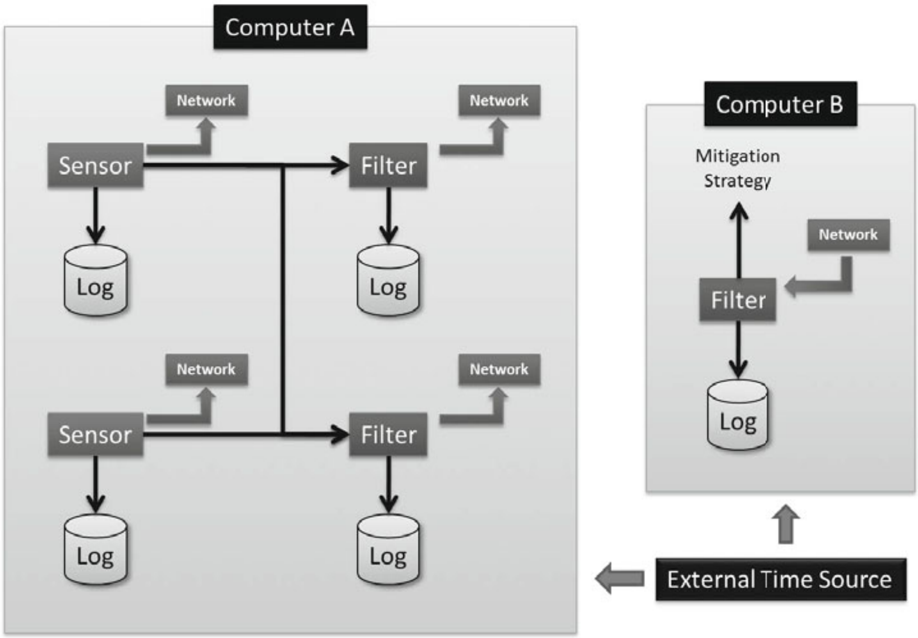


Fig. 2. System Architecture

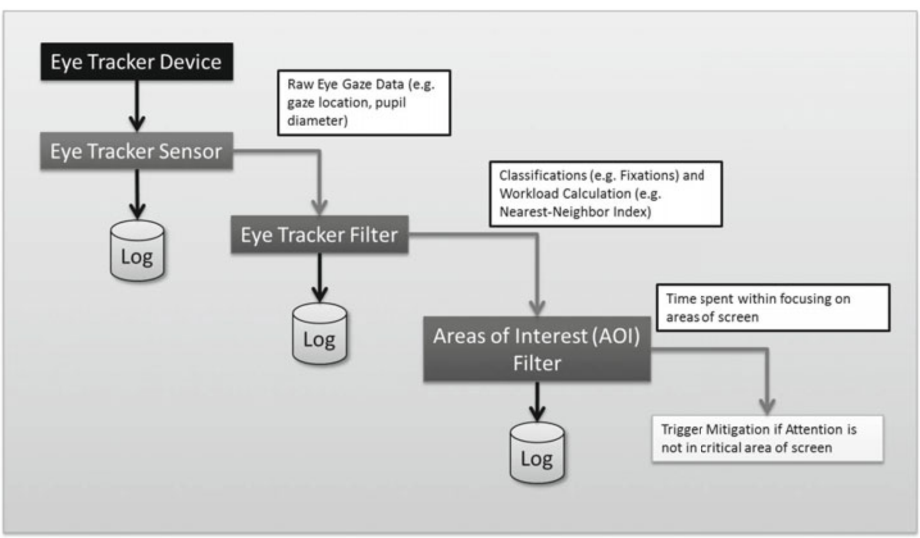


Fig. 3. Example Filter Graph for Eye Tracking Data

Another aspect of physiological sensors is that many provide large amounts of data at high sampling frequencies with timing resolution of 1ms. However, many of the devices provided run only on Windows operating systems, which provide UTC time resolution around 10ms [7]. To work around this issue, an external time source can be integrated into the system as shown in Fig. 2. The OPSS provides software tools to create external time sources which can be connected over a network connection to provide 1ms timing resolution in data collected.

Finally, for network sharing of data, the dataset from Table 1. is expanded upon (Table 2) to provide additional information necessary for distinguishing between types of data being transmitted so that subscribers can select only the data they require for processing. Within the OPSS, any record that a Sensor or Filter produces that can be logged can be modified to support conversion to the datagram format in Table 2 for transmission to subscribers.

Table 2. Message format for network streaming of data

Field	Name	Description/Interpretation
1	Message Start	4 Byte value containing unique identifiers for parsing message data and bit flags for priority and transmission confirmation
2	Time (UTC)	4 Byte UTC Time when this data point was generated or captured by source.
3	Participant/User Number	1 Byte Unique ID for the participant or user the data point is associated with
4	Group/Team Number	1 Byte Unique ID for the group or team the participant or user the data point is associated with
5	Sequence Number	2 Byte values representing the sequence number for a multi-packet sequence of data. First value always starts at 0 and increases for each packet in multi-packet stream. For standalone packets value is always 1.
6	Message Type	2 Byte Unique ID for the type of message data included in payload
7	Message Payload Size	2 Byte integer representing size of data in bytes

Table 2. (*continued*)

Message Start Format		
Bits	**Name**	**Description**
0-7	Start Byte	Byte representing the start of a message. Value is always equal to 0x02
8-15	Source Node ID	Unique Identifier representing the node source of the message. A node is defined as a PC or device with physical interface. Valid values range from [0,255]
16-23	Source Component ID	Unique Identifier representing the source application on the node. This is used to distinguish between multiple applications on the same physical source. Valid values range from [0,255]
24	Multi-Sequence Message packet	If value is 1, than this packet is part of a multi-packet sequence which is needed for transmitting messages greater than maximum bytes per packet of transport medium being used. Use the Sequence Number field for ordering of data packets. If 0, then the pack is stand alone and contains the full message.
25-32	Reserved	Reserved bits for future expansion

4 Conclusion

The synchronization process described in this paper provides a solution to the issues related to using multiple disparate physiological sensors. Specifically, how timing information is represented using these devices and methods for synchronizing the information so it can be correlated to external system events was discussed. An architecture framework for defining how sensors can be paired together and combined to produce new metrics for use in measurement of cognitive state and operator workload was described. The architecture implemented by the OPSS defines modules (e.g. Sensor, Filter) that make up a Filter Graph. If system developers choose to follow this framework in describing their own designs, facilitation of software and hardware re-use is achievable between researchers. Finally, a message structure format is proposed for the transmission of physiological data over network connections. It is the authors belief that the development of standards for sharing of physiological data between different hardware/software implementations, similar to the Joint Architecture for Unmanned Systems (JAUS) [8], would be advantageous to the larger research community. Through the use of standards, and a desire by customers of sensors for their use, the cost of adding new sensors to adaptive systems will decrease and research results more easily shared.

Acknowledgements. This work was supported by the US Army RDECOM (W91CRB08D0015). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of RDECOM or the US Government. The US Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

References

1. Vartak, A.: Cognative State Estimation for Adaptive Learning Systems Using Wearable Physiological Sensors. In: 1st International Conference on Biomedical Electronics and Devices (2008)
2. Sciarini, L.W.: Assessing Cognitive State with Multiple Phyiological Measures: A Modular Approach. In: Human-Computer Interaction International, pp. 533–542 (2009)
3. Fidopiastis, C.M.: Impact of Automation and Task Load on Unmanned System Operator's Eye Movement Patterns. In: Human-Computer Interaction Interntional, pp. 229–238 (2009)
4. Barber, D.: The Mixed Initiative Experimental (MIX) Testbed for Human Robot Interactions with Varied Levels of Automation. In: 26th Army Science Conference, Olrando (2008)
5. Nicholson, D.: An Adaptive System for Improving and Augmenting Human Performance. In: Foundations of Augmented Cognition, pp. 215–222. Strategic Analysis, Inc., Arlington (2006)
6. Camilli, M.: ASTEF: A simple tool for examining fixations. Behavior Research Methods, 373–382 (2008)
7. Nilsson, J.: Implementing a Continuously Updating High-Resolution Time Provider for Windows. In: Microsoft, `http://msdn.microsoft.com/en-us/magazine/cc163996.aspx` (accessed March 2004)
8. Committee, A.-4.: JAUS Core Service Set. In: Society for Automotive Enginers. HYPERLINK, `http://www.sae.org/technical/standards/AS5710` (accessed December 2008)

Augmenting Robot Behaviors Using Physiological Measures

Daniel Barber[1], Lauren Reinerman-Jones[1],
Stephanie Lackey[1], and Irwin Hudson[2]

[1] University of Central Florida Institute for Simulation and
Training Applied Cognition and
Training in Immersive Virtual Environments Laboratory,
3100 Technology Parkway, Orlando, FL 32826
[2] U.S. Army Research Laboratory, SFC Paul Ray Smith
Simulation & Training Technology Center (STTC), Orlando, FL

Abstract. In recent years, advancements in Unmanned Systems have allowed Human Robot Interaction (HRI) to transition from direct remote control to autonomous systems capable of self-navigation. However, these new technologies do not yet support true mixed-initiative solider-robot teaming where soldiers work with another agent as if it were another human being. In order to achieve this goal, researchers must explore new types of multi-modal and natural communication strategies and methods to provide robots improved understanding of their human counterparts' thought process. Physiological sensors are continuously becoming more portable and affordable leading to the possibility of providing new insight of team member state to a robot team member. However, steps need to be taken to improve how affective and cognitive states are measured and how these new metrics can be used to augment the decision making process for a robot team member. This paper describes current state of the art and next steps needed for accurate profile creation for improved human robot team performance.

Keywords: Multi-Modal Communion, Implicit Communication, Human Robot Interaction, Physiological Measures for State Measurement.

1 Introduction

Unmanned Systems have been developed for waypoint navigation with obstacle avoidance to operate in complex environments. Such system capabilities have been driven by events like the DARPA Grand Challenge [1] and recent advances in sensors and other technologies. Although this is a major technological achievement, these capabilities do not support true mixed-initiative teams [2] where Soldiers work with robot agents as team members, not as directly controlled assets. Reaching this goal will require the exploration of new methods for Human Robot Interaction (HRI), starting with incorporating multiple communication modalities to provide human and robot team members understanding of each other's actions, intent, and needs.

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 567–572, 2011.

2 Background

Traditional HRI in military operations has involved a human operator explicitly controlling or supervising an unmanned asset using a Human Computer Interface (HCI) [3]. With teleoperation remaining the contemporary standard, humans do not interact with the unmanned asset as another team member when co-located in a real-world dynamic operational environment. Moreover, having a remotely located commander supervising human and robot team members can affect the human team member's confidence in decision making and reduce ability to fully understand the commander's true intent [4]. In addition to reduced team confidence, adding a robot team member further increases the complexities of communication by adding an additional step in the process [5]. That is, instead of being able to communicate directly to the team members, the commanders must communicate through a robot operator first.

In order to decrease the cognitive load required from the commanders and subsequently also reduce the time it takes to issue a command, natural multimodal communications must be implemented to replace some of the traditional control features [6]. Through the combination of explicit and implicit communications over different modalities (i.e. audio, visual, tactile), robot systems can be developed and successfully integrated within human-robot teams. Physiological sensors have often been studied for applications involving adaptive automation and on the Warfighter for improving performance. Along those lines, a logical extension is to test sensor effectiveness for communicating cognitive and emotional state from a human to a robot. For example, studies have found an effective measure of sleep onset, which includes fatigue, provided by Electroencephalogram (EEG) characteristics [7]. Numerous sensors are indicative of emotional and cognitive state including pupil diameter [8][9], EEG characteristics [10], eye tracking patterns [11], and heart rate [12]. The direct information attained from sensors concerning emotional and cognitive state can be relayed from human to robot for one or more robots to act upon.

2.1 Augmenting Robot Behaviors: Example Vignette

In order for a robot to truly behave appropriately in a mixed-initiative team, it must be able to respond to a dynamic environment as quickly as possible, reducing the burden of human teammates to explicitly direct actions. In the following dismounted example, a Soldier and a robot teammate are performing a reconnaissance and surveillance mission in an urban environment (*Figure 1*). At some point while navigating down a street, a point of interest is identified and the Soldier commands the robot teammate to investigate. The Soldier is equipped with multiple physiological sensors (e.g. EEG, Electrocardiogram (EKG)) with the specific metrics and trends being continuously transmitted to the robot for monitoring its team members' stress levels and cognitive demand. While the robot team member navigates toward the point of interest, the Soldier moves to a non-line-of-site location, awaiting feedback from the investigation.

At some point while approaching the point of interest, the robot detects a change in the physiological sensor data indicating that the Soldier's stress and workload have increased by an abnormal amount. Based on this new information, the robot queries the soldier to see if it should change its current orders and return back to the non-line-of-site location to assist. The Soldier responds back that he/she is in danger and needs the robot to return so that the team can move to a more secure location. Through the use of implicit physiologic data, the robot is able to react to a changing situation without the burden of the Soldier needing to explicitly alert the robot teammate.

Fig. 1. – Point of Interest Vignette

3 Affective and Cognitive State for Direct Communication

Based on the example vignette in 0, the case for reduced cognitive demand and improved communication efficiency using physiological measures is made. Specific measures that have been explored for augmenting robot behavior is Heart Rate Variability (HRV), Electromyography (EMG), and Galvanic Skin Response (GSR) [13][14][12]. In a series of experiments, Rani and Sarkar (2005) developed methods of classifying operator engagement and affective state as a means of implicit communication to a robot team member [13]. In this study, an online classification system was developed to measure five affective states: engagement, anxiety, boredom, frustration, and anger. The system was trained using cognitive tasks (e.g. pong, anagrams) designed to elicit affective states through emotional responses from participants. Once trained over a period of multiple sessions, participants then performed a teleoperation task with a mobile robot. During this mission, the robot would modify its behavior based on the level of engagement by the operator, slowing down or stopping as engagement decreased. The results of the study found that the indices chosen showed significant correlation with engagement with Interbeat Interval (IBI) being the highest. However, indices were different across participants making it difficult to identify common patterns across measures [13].

Based on these efforts, it has been shown that it is possible to use physiological measures to modify robot behaviors. However, these methods only identify the anxiety and engagement levels of a team member while performing a given task, which does not necessarily correlate with the current level of workload. For example, a soldier can be engaged in the mission, but without knowing their current cognitive demand it is difficult to make any judgment to modify current behaviors to be more supportive. Also, the tasks described do not include a dismounted operation with an autonomous robot like the vignette described in section 0 where a Solider is moving around in the environment. In an operational context, a Soldier will at times be extremely physically active and current classification systems do not take this into account. In a true mixed-initiative team, a flexible interaction strategy is used in which each agent (human or robot) contributes the capability it is best suited for at the most appropriate time. Without the ability to acquire measures of workload and physical activity level directly for describing a team members' state (see *Figure 2)*, it is not as effective or efficient for a robot team member to make appropriate changes to its behavior. Thus, optimal support for the team is not attained in an operational environment.

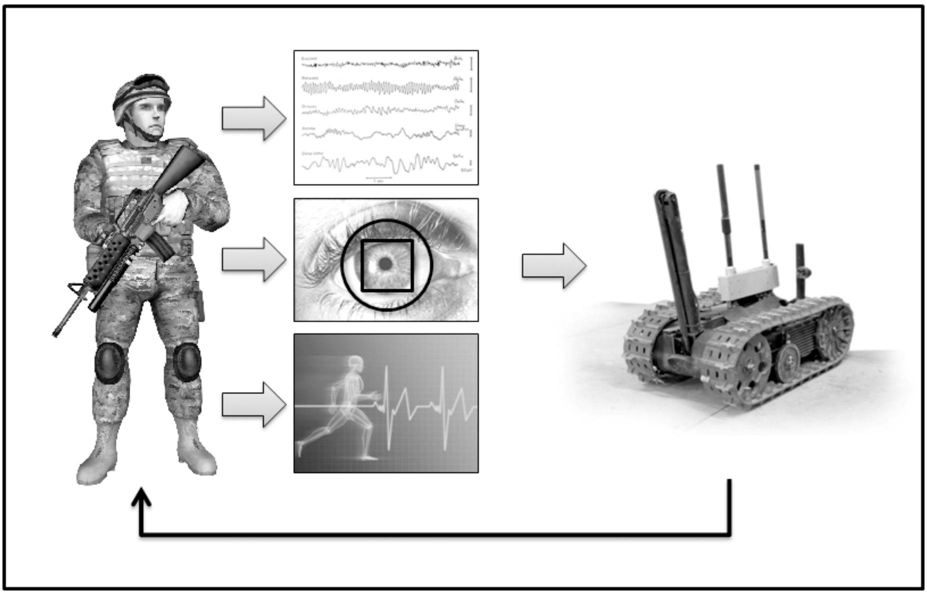


Fig. 2. Physiologically Driven Robot Behavior

4 Measuring Workload Using Physiological Sensors

In order to accomplish the goal for direct physiological communication from a human to a robot team member, the states Soldiers experience that are mission critical need to be identified. The impact of workload on performance has been investigated extensively and is often associated with changes in physiological activity. Workload is a transactional process with regard to the human's evaluation of a given task and more specifically is the cost accrued by engaging in task requirements [15]. Specifically, as workload increases, performance suffers [16].

As previously mentioned, physiological measures have been used to index workload. Physiological correlates of workload include changes in EEG patterns, heart rate, and eye scanning. Previous research reported limited success with measures of HRV, GSR, pupil dilation, and event-related potentials (ERP) [17][18]. More recently, investigations into workload measured via eye-movements using sophisticated eye tracking technology occurred. McCarley and Kramer, [19], provided a comprehensive review of the role of eye tracking in predicting and describing workload on various tasks such as driving, baggage inspection, and cockpit instrumentation. Nearest Neighbor Index (NNI) is a distance statistic that expresses proximity of each point (e.g. object or instrument) on a plane of space relative to all other surrounding points [20]. When applied to eye movements, NNI measures the randomness of fixation patterns as a ratio of the average of the observed minimum distances between fixation points and the mean distance once would expect if the distribution was random. "This ratio is equal to 1 for a distribution that is random. Values lower than 1 suggest grouping, whereas values higher than 1 suggest regularity (i.e. the point pattern is dispersed in a non-random way). Theoretically, the NNI lies between 0 (maximum clustering) and 2.1491 (strictly regular hexagonal pattern) [21]." Another method for using eye tracking to assess workload is wavelet analysis of pupil diameter [9]. The Index of Cognitive Activity (ICA) measures abrupt discontinuities in the signal created from continuous recording of pupil diameter [9]." Unlike NNI, ICA is independent of eye gaze, and has the added benefit of running at nearly real-time over signals of any length [9]. Based upon this literature, workload appears to be a grounded starting point to begin investigating physiologically driven robot behavior.

5 Conclusion

Although some initial work has been performed using EKG data to characterize stress and affective state of an individual for HRI, additional information needs to be provided to a robot teammate for it to make decisions that will improve human-robot team performance. With the inclusion of EEG and pupil diameter, which can be used for measures of workload in near real-time, advanced online classification frameworks can be developed to form a closed-loop system for augmenting robot behavior. Through additional research into combinations of multiple sensors like EKG, EEG, and eye tracking, these classification frameworks will support dynamic operational environments and improve the performance of human-robot teams.

Acknowledgements. This work was supported by the US Army RDECOM (W91CRB08D0015). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of RDECOM or the US Government. The US Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

References

1. DARPA: URBAN Challenge. In: DARPA. HYPERLINK, `http://www.darpa.mil/grandchallenge/index.asp` (accessed February 3, 2011)
2. Hearst, M., Allen, J., Guinn, C., Horvitz, E.: Mixed-Initiative Interaction: Trends & Controversies. IEEE Intelligent Systems, 14–23 (1999)

3. Barnes, M., Jentsch, F. (eds.): Human-Robot Interactions in Future Military Operations. Ashgate (2010)
4. Grey, A., Redden, E., Coovert, M., Elliot, L.: Empowering followers in virtual teams: Guiding principles from theory and practice. Computers in Human Behavior 24, 1884–1906 (2008)
5. Cosenzo, K., Capstick, E., Pomranky, R., Dungrani, S., Johnson, T.: Soldier Machine Interface for Vehicle Formations: Interface Design and an Approach Evaluation and Experimentation. Technical Report ATRL-TR-4678, Aberdeen Proving Ground (2009)
6. Pettitt, R.A., Carsten, E.S.R., Scalablity, C.B.: of Robotic Controllers: Speech-based Robotic Controller Evaluation (ARL-TR-4858). Aberdeen Proving Ground, MD: US Army Research Laboratory, 1-46 (2009)
7. Kruegar, G.P.: Sustained work, fatigue, sleep loss, and performance: A review of the issues. Work & Stress, 129–141 (1989)
8. Bradley, M., Miccoli, L., Escrig, M., Lang, P.: The pupil diameter as a measure of emotional arousal and autonomic aviation. Psychophysiology 45, 602–607 (2008)
9. Marshall, S.: The index of measuring cognitive workload. IEEE 7th Human Factors Meeting, Scottsdale, AZ, 7.5–7.9(2002)
10. Berka, C., Levendowski, D., Lumicao, M.A.Y., Davis, G., Zivkovic, V., Olmstead, R., Tremoulet, P., Craven, P.: EEG correlates of engagement and mental workload in vigilance, learning, and memory tasks. Aviation, Space, and Environmental Medicine 78, B231–B244 (2007)
11. Camilli, M.: ASTEF: A simple tool for examining fixations. Behavior Research Methods, 373–382 (2008)
12. Rani, P., Sims, J., Brackin, R., Nilanjan, S.: Online stress detection using psychophysiological signals for implicit human-robot coorperation. Robotica, 673–685 (2002)
13. Rani, P., Sarkar, N.: Operator Engagement Detection and Robot Behavior Adaptation in Human-Robot Interaction. In: IEEE International Conference on Robotics and Automation, pp. 2051–2056 (2005)
14. Rani, P., Sarkar, N., Smith, C., Kirby, L.: Anxiety Detecting Robotic System-Towards Implicit Human-Robot Collaboration. Robotica 22, 85–95 (2004)
15. Hart, S., Wickens, C.: Workload assessment and prediction. In: Booher, H.R. (ed.) MANPRINT: An approach to systems integration. Van Nostrand Reinhold, New York (1990)
16. Wickens, C.D.: Multiple resources and performance prediction, pp. 159–177 (2002)
17. Kramer, A.: Physiological measures of workload: A review of recent progress. In: Damos, D. (ed.) Multiple Task Performance. Taylor and Francis, London (1991)
18. Parasuraman, R.: Event-related brain potentials and human factors research. In: Rohrbaugh, J., Parasuraman, R., Johnson, R. (eds.) Event-related Brain Potentials: Basic and Applied Issues. Oxford University Press, New York (1990)
19. McCarley, J.S., Kramer, A.F.: Cerebral hemodynamics and vigilance performance. In: Parasuraman, R., Rizzo, A.M. (eds.) Neuroergonimcs: The Brain at Work, pp. 95–112. MIT Press, Cambridge (2007)
20. Clark, P.J., Evans, F.C.: Distance to nearest neighbor as a measure of spatial relationships in populations. Ecology 35, 445–453 (1954)
21. Di Nocera, F., Camilli, M., Terenzi, M., Nacchia, R.: Cognitive aspects and behavioral effects of transitions between levels of automation. Technical Report FA8655-05-1-3021, EOARD (2007)

Operational Neuroscience: Neuroscience Research and Tool Development to Support the Warfighter

Monique E. Beaudoin[1] and Dylan D. Schmorrow[2]

[1] Strategic Analysis, Inc., 4075 Wilson Blvd,
Suite 200, Arlington, VA 22203
[2] Office of the Secretary of Defense, 1777 N. Kent St., Arlington VA 22209
mbeaudoin@sainc.com, dylan.schmorrow@osd.mil

Abstract. This paper provides a summary of the presentations presented in the Operational Neuroscience session during Augmented Cognition International 2011 at Human Computer Interaction International 2011 in Orlando, Florida, July, 2011.

Keywords: Neuroscience, military operations, warfighter support, Cognitive readiness, Neurotechnology, hemodynamics, pharmacokinetics, brain bio-markers.

1 Introduction

The concept of applying basic neuroscience research and tool development to the improvement of cognitive performance of warfighters and to provide individualized operational cognitive support in the military setting is not a new one. Early influential work especially by NASA [1] showed that workload could be quantified through subjective measurement of several variables (mental, physical, and temporal demand; performance; effort; frustration) related to cognition and performance, and thereby monitored. NASA has become one of the largest governmental funding agencies for research regarding performance under various stressful conditions and scenarios [2]. The principle of quantifying cognitive workload and functioning for optimization of cognitive performance was then explored simultaneously by multiple groups, and as neuroscience methodology and tools progressed in technical capabilities, measures of cognitive state and work load were combined with neurodetection technologies and user interfaces to optimize performance in various fields. This is essentially the foundation of Augmented Cognition (a term resulting from a DARPA-funded research program, 2002-2008), and for research in Neuroergonomics (see [3, 4]) and Human Computer Interaction. When the term “operational neuroscience” was coined [5], it reflected a growing interest in defense research and development in the area of neurotechnologies for the operational military setting [6]- especially by DARPA [7]. Around the same time, the National Research Council published reports on the importance of neuroscience to defense research and military applications [2, 8], which further increased public interest in Neuroscience for these purposes. Neuroscience and cognitive science research/development efforts are currently ongoing in various organizations, including military, academia, and industry. As investment and interest

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 573–577, 2011.

has increased, so has the need for scientific rigor in development of diagnostic capabilities as well as innovative brain-in-the-loop tool development, while maintaining a constant focus on the needs of the warfighter in the field as well as throughout training and in rehabilitation. This session overview paper provides a sampling of the various research and development efforts from across the spectrum of the field of neuroscience and cognitive science, for application to the operational setting.

2 Neuroscience and Cognitive Science - Basic Science Research for AugCog

The basic technological capabilities of psychophysiological measurement technologies (EEG, fNIR, eyetracking, etc) for integration and application to Augmented Cognition technologies have been the subject of some recent debate- this issue has been termed a "technology chasm"[9]. This chasm describes the difficulties involved in the leap from obtaining accurate psychophysiologic measurements in the field to development of tools, all while under conditions of uncertainty during both measurement and analysis. "Without critical advances in EEG and other neurologic and physiologic technologies, the AugCog effort cannot make significant progress or be operationalized."[9] This comment highlights the critical importance of the solid basic research foundation to the burgeoning field of operational neuroscience for quality tool development.

2.1 Exploring New Methodologies for the Analysis of Functional Magnetic Resonance Imaging (fMRI) Following Closed-Head Injuries

Our current and basic understanding and assumptions about the functioning of the brain under stressful conditions are based on currently-available scanning technologies, and these assumptions are often taken as solid fact when applied for use in neurotechnologies and cognitive modeling. In their paper, Walker and Davidson [10] challenge the idea that fMRI scanning has reached its maximum technical potential. They contend that Blood Oxygen Level Dependence (BOLD) fMRI is a relative measurement tool, and cite growing concerns over the reliability of findings, thus bringing doubt to prior cognitive assumptions based on BOLD-based interpretations of scanning results. Such concerns include poor temporal resolution, spatial resolution and lack of dimensionality. For the purpose of cognitive rehabilitation following brain injury, these limitations are particularly troublesome. The paper describes a new data analytic approach to fMRI which uses a low-rank tensor approximation to overcome these obstacles, providing the potential for more accurate modeling, analysis, and translation to multiple platforms and applications.

2.2 Non-invasive Functional Brain Biomarkers for Cognitive-Motor Performance Assessment: Towards New Brain Monitoring Applications

As the basis for assessing level and skill of performance on sensorimotor and cognitive tasks, various biomarkers from behavioral/kinetic, physiological, or brain states can be utilized. These biomarkers are best applied to the operational setting through detection with non-invasive techniques and newly emerging mobile technologies for detection of cognitive state. In his work, Gentili [11] presents EEG brain biomarkers which can track changes in performance, and some preliminary work using fNIRS for the same purpose. The work discussed has the potential to be applied to assessment of sensori/cognitive motor performance in bioengineering and augmented cognition neurotechnologies, for increased field-based performance, training effectiveness, and efficiency.

3 Human-in-the-Loop for Brain-in-the-Loop Tool Development

Basic principles of pharmacology and known drug pharmacodynamics, in combination with basic science research about alertness and cognition, are powerful tools for operational neuroscience researchers and developers. While much research aims to remove the "user" from the loop and rely solely upon the brain for direct feedback, some have argued for the importance for the human to be directly involved in inputting of physical data to the suite of tools which aim to measure cognitive function. The principle lies in the fact that some biological data- critical as it may be to accurate neurophysiological modeling- simply cannot be determined using external non-invasive neurodetection technology (such as EEG). Use of pharmaceuticals or stimulants may prove to be essential to the accurate interpretation of data derived from the use of direct measurements using neurotechnologies. Without appropriate incorporation of current and varied pharmaceutical use data, poor modeling adaptation and data analysis may result for a single user over time, due to a gradual alteration in the baseline measurements. With extreme demands on the time of field operators and trainers, it is not often possible for field technicians/scientists to individually interview a large number of warfighters on a daily, let alone hourly, basis, as they use fielded neurofeedback technologies. More effective, individualized cognitive modeling and optimal usage/development of field-based neurotechnologies may be achievable in the near-term through the incorporation of human input- via pharmacokinetic modeling software on a handheld platform- to suites of real-time neurodetection technologies.

3.1 A Mobile Tool to Help Users Moderate Caffeine Intake by Displaying Caffeine Pharmacokinetics and Pharmacodynamics

As described above, human-in-the-loop elements of brain-in-the-loop augmented cognition technologies may prove to be important to the development of robust operational neuroscience applications. In their paper, Ritter and Yeh [12] discuss the development of a handheld device which they call Caffeine Zone, which utilizes known pharmacokinetics of caffeine, along with caffeine modeling in cognitive models and agents [13]. The tool is inexpensive and portable, and could be rapidly applied to other integrated suites of neurotechnologies for operational use. The device is a software program (iOS operating system compatible, usable also on smartphones such as Blackberry, Android) which uses input by the user of his/her dietary or supplemental caffeinated substance intake, along with kinetics of caffeine absorption and elimination to display real-time optimal times and doses of caffeine to assist in the maintenance of alertness and cognitive readiness. Because of caffeine's ability to both enhance cognition and also negatively affect sleep cycles and stress hormone levels, the added data including its stimulant use could prove quite relevant to the analysis of real-time cognitive data obtained with purely brain-in-the-loop tool suites.

4 Enlisting Operational Neuroscience Research

To reach the shared end goal of integrated real-time tool development using non-invasive methods and hardware for continual monitoring in the field, basic research must eventually make the leap into the applied setting. The concept of using individual neurofeedback tools operationally has become more widely accepted in the past decade- whether the tools are developed for direct use by the warfighter, the analyst, or teams of warfighters- to measure cognitive load, cognitive readiness, and team dynamics, among other metrics.

4.1 Brain Dynamics of Coordinated Teams

Team coordination is highly complicated phenomenon that social and cognitive science researchers are beginning to better understand in this decade. To address the intricacies of team dynamics, Kovacs et al. [14] present results in their paper from an ecologically valid laboratory task assessing cognitive, social, perceptual, behavioral, and attentional processes. They enlist dual-EEGs to study team dynamics in a virtual environment, which the goal of capturing neural dynamics continuously over time. Temporal distribution of neural markers during well-characterized and unplanned team challenges were used to diagnose team cooperation, using a novel framework that captures overt behavioral choices as well as the simultaneous underlying neural activity.

5 Conclusions

The important neuroscience research and development work which is being carried out for and applied to the operational military setting is as diverse and multifaceted as the warfighters who will benefit from it. Under the auspices of Augmented Cognition and Neuroergonomics, Operational Neuroscience research and tool development provides an excellent example of the range of basic and applied scientific research that can be inspired through an accurate understanding of the end user (or, beneficiary of such research and development). Operational Neuroscience continues to expand as a field, which is a credit to the researchers who have exhibited the flexibility necessary in their collaborations and development of expertise to encompass a variety of techniques and platforms, while maintaining critical ties to the warfighter.

Acknowledgements. The opinions expressed here are those of the authors and do not necessarily reflect official United States Department of Defense views.

References

1. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock, A., Meshkati, N. (eds.) Human Mental Workload. North Holland Press, Amsterdam (1988)
2. Opportunities in Neuroscience for Future Army Applications. Committee on Opportunities in Neuroscience for Future Army Applications; National Research Council. The National Academies Press (2009)
3. Parasuraman, R., Wilson, G.F.: Putting the brain to work: neuroergonomics past, present, and future. Hum. Factors 50(3), 468–474 (2008)
4. Parasuraman, R.: Neuroergonomics: research and practice. Theoretical Issues in Ergonomics Science 4(1-2), 5–20 (2003)
5. Kruse, A.A.: Operational neuroscience: neurophysiological measures in applied environments. Aviat Space Environ Med. 78(5), 4–191 (2007)
6. Beaudoin, M.E., Cohn, J., Schmorrow, D.: The Convergence of Neuroscience and Cognitive Science Research in the Department of Defense. In: Annual Conference on Association of Military Surgeons of the United States (AMSUS 2010), Phoenix, AZ (2010)
7. Areas of Interest in DARPA DSO website (2011)
8. Emerging Cognitive Neuroscience and Related Technologies; National Research Council. Committee on Military and Intelligence Methodology for Emergent Neruophysiological and Cognitive/Neural Research in the Next Two Decades. The National Academies Press (2008)
9. Cummings, M.L.: Technology Impedances to Augmented Cognition. Ergonomics In Design, 26–27 (Spring 2010)
10. Walker, P., Davidson, I.: Exploring New Methodologies for the Analysis of Functional Magnetic Resonance Imaging (fMRI) Following Closed-Head Injuries. In: Human Computer Interaction International. Springer, Orlando (2011)
11. Gentili, R.: Non-Invasive Functional Brain Biomarkers for Cognitive-Motor Performance Assessment: Towards New Brain Monitoring Applications. In: Human Computer Interaction International. Springer, Orlando (2011)
12. Ritter, F., Yeh, K.: A Mobile Tool to Help Users Moderate Caffeine Intake by Displaying Caffeine Pharmacokinetics and Pharmacodynamics. In: Human Computer Interaction International. Springer, Orlando (2011)
13. Morgan, G., Ritter, F., Stine, M., Klein, L.: The Cognitive Effects of Caffeine: Implications for models of users. The Pennsylvania State University (2006)
14. Kovacs, A., Tognoli, E., Afergan, D., Coyne, J., Gibson, G., Stripling, J., Keso, J.A.S.: Brain Dynamics of Coordinated Teams. In: Human Computer Interaction International, Orlando, FL. Springer, Heidelberg (2011)

Performance Measures to Enable Agent-Based Support in Demanding Circumstances

Fiemke Both[1], Mark Hoogendoorn[1], Rianne M. van Lambalgen[1], Rogier Oorburg[2], and Michael de Vos[2]

[1] Vrije Universiteit Amsterdam, Department of Artificial Intelligence
De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands
{fboth,mhoogen,rm.van.lambalgen}@few.vu.nl
[2] Defence Materiel Organization, CAMS-Force Vision
P.O. Box 10000, 1780 CA Den Helder, The Netherlands
{r.oorburg,m.de.vos}@forcevision.nl

Abstract. In this paper, an evaluation of measurements that can be used by a personal support agent to measure the quality of human task performance is addressed. Such measurements are important in order for a support agent to give effective and personalized support during the performance of demanding tasks. Hereby, the performance quality measurement is addressed from two perspectives, namely the human's perspective as well as the task perspective. The former represents the idea the human has about the current task performance, whereas the latter measures the actual task performance compared to the goals set for the task at hand. Criteria have been identified to compare the various measurements, and an experiment has been conducted for evaluation. Based on these evaluation results, the most useful measurements are identified to be adopted within personal support agents.

1 Introduction

When humans perform demanding tasks, it is known that their performance can severely degrade over time when their available resources are being exceeded (see e.g. [1]). Such degrading performance is highly undesired, especially in critical domains. Within the research field of augmented cognition, one of the goals is to develop systems that take such limitations of a human's capacity to process information into account and avoid performance degradation by intervening (e.g. [2]). For this purpose, personal assistant agents (e.g. [3], [4]) can be designed where agents interact with sensors in the environment to monitor the human's performance quality and contribute support in case it is needed. Hereby, having information on the performance quality of the human is of essence in order to give appropriate support.

The measurement of how well a human is performing is however not trivial. The quality can be measured from different perspectives, namely the human's perspective (the judgment of performance the human has) as well as the task perspective (depending on the actual task performance). In the field of augmented cognition, both are useful. In order to accept the help of a personal assistant, the human needs to have the idea that the system "understands" the human, hence the agent needs to contain a

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 578–587, 2011.

model of the human's experience of performance. Furthermore, discrepancies between the human's idea of performance and the actual task performance can also be a basis for an intervention. Of course, the actual task performance is important from the perspective of the eventual outcome of the task.

A variety of measurements that have been proposed in the past as indicators for performance quality can potentially be utilized by an agent applied in a system that is aware of the human state. Indicators for the human's performance are for instance measured using the NASA-TLX [5], or using physiological measurements such as ECG (to measure heart rate). For measurements from the task perspective, agents can use workflow oriented approaches to measure how well the workflow has been followed (see e.g. [6]). In this paper, the measurements are compared to see how suitable they are for usage in a personal assistant agent. Hereby, criteria are identified to score the various measurements, and an experiment has been conducted using a simulation based training environment to evaluate the measurements for their use in agent-based support.

This paper is organized as follows. First, an overview is given of existing performance measurements in Section 2. Thereafter, the criteria for evaluation of measurements are identified in Section 3. Section 4 presents the simulation based training environment used to conduct the experiments, and an evaluation using the data from the experiment is shown in Section 5. Finally, Section 6 is a discussion.

2 Performance Measurements

First, the performance measurements from the human's perspective are described in this section, followed by the measurements from the task perspective.

2.1 Human's Perspective

When looking at performance from a human perspective, the focus is on performance measurements that can be defined by looking at the human. Such measurements can be subjective (e.g. the agent could ask the human to fill in a questionnaire) or psychophysical (the agent could communicate with measurement devices that measure the heart rate). Also, in previous literature human performance is described by looking at the mental effort someone has put in a task. Hockey [7] states that when looking at task performance it is important to take the efficiency of behavior into account. Instead of only looking at a specific task output, it is important to also look at the costs of achieving such an output (i.e. a person's mental effort).

A **subjective measurement** of performance gives information to the agent on how the human is observing the performance. In order to perform these measurements, the subjective scales NASA-Task Load Index (NASA-TLX, [5]) and Subjective Workload Assessment Technique (SWAT, [8]) can be used. Both scales consist of subscales where aspects of mental workload are rated by the human performing a task. In addition, one of the subscales of the NASA-TLX is a performance measure and asks humans to indicate their own performance. The major disadvantage of the subjective performance measurement is that the person performing a task needs to be interrupted by the agent. This can easily be done in an experimental setting, but is not practical in a real world setting.

Physiological measurements provide the personal assistant agent with information about bodily responses to task execution. Examples are EEG (brain activity), Eye Blink Activity and ECG to measure heart rate (HR). HR is known to increase with increasing task demands and decreasing performance ([9]). Concerning Eye Blink Activity, research shows that the time between two successive blinks increases when visual load increases, but decreases when mental (non-visual) load increases ([10]). Both HR and eye blink activity can be very useful as an indicator for task performance for the personal assistant agent. A disadvantage of psychophysiology is that the measurements required can be intrusive and therefore not very desired to use in real world settings (however, less intrusive measurements are also being developed, see e.g. [11]). In addition, when considering HR, other factors should be taken into account. For instance, HR can be influenced by physical exercise, sleep or coffee as well. This should be taken into account by a support agent that uses psychophysical input to reason about a human's state.

2.2 Task Perspective

Some of the approaches described in the previous section are difficult to measure, especially in applications in the real world. Measurements from a task perspective are less intrusive and provide a different type of information about the performance. Depending on the precise reason for which the personal assistant wants to use the task performance for its support actions, one or more of the approaches described below can be used. For example, in a stressful situation it may not be important at all whether the correct procedure is followed, only the outcome matters. Three types of performance measurements from a task perspective are considered here. Section 5.1 gives a detailed description on these measurements applied to the case study.

Effectiveness. The correctness of handling a task based on the set goal, is referred to as *effectiveness*. In this paper, two different perspectives on effectiveness are taken into account. The first is an *absolute* perspective by looking at the outcome regardless of the process leading to that outcome. As the absolute correct outcome is not always available during task execution, it is difficult for an agent to use this information to measure real-time performance. The second perspective is a more realistic perspective on effectiveness, called *realistic* effectiveness. Here, the correctness of a response depends on the workflow that is followed (the correctness of the individual steps that are taken to achieve the outcome).

Productivity. Performance can also be viewed by taking into account the *productivity*. Productivity is often seen as the ratio between the output of a task and the input of a task [12]; the faster the input of a specific task is processed, the more output is generated within a time unit and the higher the productivity. In this paper, two different productivity measurements are taken into account: *average completion time*, and *percentage of cases handled*. These measurements both evaluate the amount of data that is processed within the task.

Efficiency. In addition to productivity and effectiveness, performance from a task-based perspective can be measured by looking into *efficiency*. In this paper, efficiency is defined as the costs of performing a specific task relative to the minimal amount of costs that are necessary to perform the task. Costs are represented by the resources spent on a task, for example money or material.

3 Performance Measurement Evaluation Criteria

In order to compare the measurements for task performance to see how suitable they are for usage in personal assistant agents, a number of criteria have been identified. Hereby, first of all inspiration has been drawn from the work done by [13] in which criteria have been identified to evaluate workload assessment techniques. These can be reused for evaluating performance measurements to be utilized by an agent and are listed below. Note that only the relevant subset of the criteria is taken.

Sensitivity. The sensitivity refers to the capability of the measurement to detect differences in the performance of the human. Some measurements might be relatively coarse grained whereas other can measure on a fine granularity. In this case, two types of sensitivity are involved, namely the sensitivity for the human's perception of performance as well as the sensitivity for the actual task performance. Both are important as argued in the introduction already. As a baseline for actual task performance, the precise performance of the human upon the task at hand is used. Hereby the performance measurement (absolute effectiveness) is directly linked to the goal as provided to the human in the beginning of the task. As a golden standard for the human's perception of performance, the NASA-TLX is used, as this is known to be very reliable for measuring subjective performance. In order to precisely measure how accurate a measurement *m* is for the human's perceived performance and the actual task performance, a linear regression method is used namely *simple linear regression*. The parameters of the linear regression model were a curve of the form $\hat{y} = b_0 + b_1 \cdot x$. Hereby, the x-scale denotes the observed value of measurement *m*, whereas the y-scale indicates the value of the golden standard at the same time point (in this case the precise performance upon the task at hand). $\hat{y}$ denotes the predicted y value for measurement x, and b_0 and b_1 are estimated by means of the *ordinary least square* method and are calculated as follows:

$$b_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad (1)$$

and:

$$b_0 = \bar{y} - b_1 \cdot \bar{x} \qquad (2)$$

In the equation, (x_i, y_i) are pairs of observed measurements (*n* in total). The suitability of the measurement is now defined by taking the average squared error:

$$error = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} \qquad (3)$$

Intrusiveness. One of the criteria related to the sensing devices themselves is the *intrusiveness* of the sensor to perform measurements. In case the sensors are very intrusive, this might lead to the human feeling uncomfortable, as there is a continuous awareness of everything being measured. The sensors are scored by taking into account how much the human is disturbed during the task itself (e.g. freezing the computer screen to allow the assistant agent to pose a question), and how visible the sensors are. A five point scale is used to score this criterion, ranging from '--' for very intrusive, to '++' for highly non-intrusive ('o' is for neutral).

Reliability. When measurements are performed another important criteria is how robust the measurements are. Some measurements might only be robust when they are performed under laboratory conditions whereas the developed assistant agent might be meant for more demanding conditions. There is often a trade-off between the intrusiveness of sensors, and their robustness. Measuring heart rate using electrodes is more robust compared to measuring it via a watch. However, the latter is less intrusive compared to having electrodes attached to your body. Again, a five point scale is used, whereby '--' stands for not reliable and '++' stands for very reliable.

Implementation requirements. Another criterion includes the requirements of the measurement to be performed, and how difficult it is for an assistant agent to interpret the sensing data. Some data is very easily understandable (e.g. the heart rate), whereas other measurements require the assistant agent to have a more thorough knowledge on how to use the information (e.g. an EEG). In this case, the five point scale ranges from '--' representing heavy implementation requirements to '++' for hardly any requirements.

Task dependence. In the design of personal assistant agent, the goal is often not only to investigate support for a single task, but for multiple tasks to allow for more generic support. Therefore, it is important that the agent does not entirely depend on measurements that highly depend on the characteristics of the tasks being performed. Therefore, the portability of the measurements to other tasks is also included as a criterion. Also here, the same five point scale is used, whereby '--' indicates highly task dependent whereas '++' stands for task independent.

Cost. The last factor is cost. Some sensors are relatively cheap, whereas others can be quite expensive. Again, the five point scale '--' to '++' is used for very high costs to very low costs respectively.

4 Experiment

This section briefly describes the setup of the experiment that has been conducted to evaluate the various measurements. First, the task environment is discussed, followed by the concrete measurements that were performed. Finally, the setup of the experiment is described.

4.1 Simulation-Based Training Environment

The main task that was used in this study consists of identifying incoming contacts on a computer screen and, based on the outcome of identification, deciding to eliminate the contact (by shooting) or allowing it to land (by not shooting). Contacts appear at a random location on the top of the screen and fall down to a random location at the bottom. Shooting is performed by means of a stationary weapon placed on the bottom of the screen. Before a contact can be identified, it has to be perceived. This is done by a mouse click at the contact, which reveals a mathematical equation underneath the contact (e.g. 12*3=36). The identification task is to check the correctness of the mathematical equation (which is less difficult in less demanding situations). A correct equation means that the contact is an ally; an incorrect equation indicates that the contact is an enemy. Identification is done by pressing either the left or right arrow for respectively an ally or enemy. When a contact is identified a green (for an ally) or a red (for an enemy) circle appears around the contact. The contacts that have been identified as an enemy have to be shot before they land. A missile is shot by executing a mouse click at a specific location; the missile will move from the weapon to that location and explode exactly at the location of the mouse click. When a contact is within a radius of 50 pixels of the exploding missile, it is destroyed. The scenario can in the future easily be extended with a personal assistant agent that measures progress, and takes care of some missiles in cases the human is becoming overloaded. A preliminary study addressing a personal assistant agent for this task environment can be seen in [14], note that the proposed performance measurements in this paper have not been incorporated in the personal assistant agent yet.

4.2 Performance Measurements for the Task

As already stated before, the performance from a *human perspective* was measured with use of a subscale of the NASA-TLX (taken as the golden standard). Each 2.5 minutes participants were asked to rate their performance. In order to conduct the sensitivity evaluation described in Section 3, the participants' ratings were scaled to a number between 0 and 100. For physiological measurements, ECG was measured throughout the entire experiment to calculate the heart rate. Eye blinks were measured using a Tobii x.60 tracker.

For all performance measurements from the *task perspective*, a moving average with a time window of 86 seconds was calculated. To calculate the *absolute effectiveness*, a contact that was handled correctly (e.g. a friend was landed and an enemy was shot) was given an acceptance of 1, a contact that was not handled correctly was given an acceptance of 0. In case of the *realistic effectiveness* acceptance depended on the participants' identification of a contact: an acceptance of 1 was given when a contact identified as friend landed or a contact identified as ally was shot; an acceptance of 0 was given when a contact identified as friend was shot and a contact identified as enemy landed. When a contact was missed, realistic effectiveness was 0. As stated in Section 2, *productivity* was separated in two measurements. First, the average handle time was calculated from the average completion time (time from the time point a contact was instantiated to the time point a contact was handled) and the average reactivity time (time from the time point a contact was instantiated to the time point a contact was perceived:

$$avg\_handle\_time = avg\_completion\_time - avg\_reactivity\_time \quad (4)$$

In addition, the percentage of handled cases was calculated:

$$perc\_handled\_cases = handled\_cases \,/\, (handled\_cases + expired\_cases) \quad (5)$$

Finally, the *efficiency* was calculated by dividing the amount of bullets by the amount of handled contacts.

4.3 Participants and Procedure

In this study, 2 female participants and 3 male participants with a mean age of 22.8 took part. All participants already had some experience with the experimental environment.

The experiment consisted of 4 blocks of 20 minutes of the simulation-based training environment. In the first 10 minutes of one block, task demands were low (contacts appear every 10 to 20 seconds) and in the second 10 minutes of one block, task demands were high (contacts appear every 2.25 to 4.5 seconds). In the first and third block, the environment froze after every 2.5 minutes, in the second and fourth block no freezes appeared. The purpose of the freezes was to put the experiment on hold and ask the participants questions about the participants' perceived performance quality. The following sentence was shown: "Gameplay frozen. After this message, a computer version of the NASA-TLX was shown, where participants had to indicate their performance and mental effort. In the future this would be a task that performed by the personal assistant agent.

At the start of the experiment, onscreen instructions were given on the task environment and freezes. The instructions were followed by a practice block of two minutes medium task demands to get familiar with the environment. After practice, participants started with the first block. After each block, the participant was given a three minute break before continuing with the next block.

5 Results

In Table 1 the scores of the various measurements upon the criteria are shown that have been identified to measure the suitability of a measurement for the personal assistant. For calculation of sensitivity, mean values were obtained for each performance measurement and regression analysis was performed. The mean squared error (MSE) (as explained in Section 3) was calculated and averaged over participants. The sensitivity is determined by 1-MSE and scores are presented in Table 1. For the sensitivity with respect to the human's perceived performance, 8 data points were taken from each measurement, 1 for each NASA-TLX measurement in one stage. For the sensitivity with respect to the actual task performance, one data point represented an interval of 20 seconds, the first data points of each part were taken out as no objective data was present yet.

The sensitivity values in Table 1 show that absolute effectiveness (golden standard for task performance) is highly sensitive to the human's perceived performance. The relationship suggests that humans are good in rating their own task performance. Realistic effectiveness is highly sensitive to both task as well as perceived performance. In addition, the completion time is also very sensitive to both types of performances. This could be due to a speed-accuracy trade off: when a case is handled faster, there is more chance of making an error which causes a decrease in performance. The sensitivity scores do not reveal much difference between perceived and actual task performance, except that the measurement eye blink has a relatively high sensitivity for perceived performance compared to task performance. When looking into the data it can be seen that human's perceived performance increases as the time between blinks increases. This effect could be indirectly caused by task demands: as task demands increase, both performance and time between blinks increases.

The rationale for the score on the evaluation criteria apart from the sensitivity criterion is as follows. The NASA-TLX scores negative on intrusiveness as answering

Table 1. Performance Measures Evaluation

Measurement	Task. Sens	Human Sens.	Intru-siveness	Reliability	Implementation requirements	Task Depen-dence	Cost
NASA-TLX	0.977	1.0	--	++	++	++	++
Heart rate	0.900	0.834	o	o	++	++	++
Eyeblink	0.887	0.916	+	--	++	++	++
Absolute effectiveness	1.0	0.957	++	++	o	--	o
Realistic effectiveness	0.976	0.915	++	++	o	--	o
Efficiency	0.936	0.894	++	++	o	--	o
%handled cases	0.926	0.811	++	++	o	--	o
Completion Time	0.959	0.872	++	++	o	--	o

the NASA-TLX questions means that the personal assistant agent would have to interrupt the execution of the current task. The NASA-TLX scores well on reliability, implementation requirements, task dependence, and cost [15]. The heart rate measurement scores neutral on the intrusiveness as well as on the reliability. This is because heart rate can be measured non-intrusive and less reliable (e.g. sensors in clothes), or more reliable and more intrusive (e.g. via ECG using electrodes on the chest). The measurement scores well on implementation requirements, task dependence, and cost. Regarding the eye blinks, the sensor scores well on the intrusiveness, implementation requirements, task dependence, and cost. It does however score relatively bad on reliability, as other environmental aspects can affect the amount of eye blinks (e.g. the amount of sun, tiredness).

Finally, absolute and realistic effectiveness, efficiency and both productivity measurements all score low on task dependence. This is because each time these measurements are used in a different task, a new metric needs to be adopted by the agent. Furthermore, they score mediocre on the implementation requirements as well as cost, as often the software environment in which the task is performed needs to be accessed and possibly extended to allow for a precise measurement to be available for usage by the personal assistant agent. The measurements score high on reliability, because they are highly task dependent. The custom made measurement, that has to be designed for each task, does allow for a reliable representation of performance within the specific task.

6 Discussion

For a personal assistant agent in dynamic circumstances it is useful to have access to different measures of task performance to know the current performance of the human, allowing for such an agent to give dedicated support. This support could for instance avoid degradation of performance, which is important in the field of augmented cognition. This paper describes several performance measurements that were measured in an experimental setting, all aiming at a different aspect of human performance. The measurements were scored based on a number of criteria and evaluated for their use within a personal assistant agent. The paper shows that especially realistic effectiveness can be perfectly used to substitute both subjective and objective performance as the sensitivity to both measurements is very high. With respect to the psychophysiological measurements, especially eye blink was more predictive of subjective performance compared to objective performance. A possible explanation is that the rating of subjective performance is based upon the responses of the body observed by the human. However, it could also be that both the body and the subjective performance respond to the demands of the task. More research has to be done on the causal nature of this relationship.

The relatively high sensitivity score of all measurements shows that they all can be used to replace either the very intrusive NASA-TLX or the absolute effectiveness that is often not measurable in a real world setting. Depending on the purpose of the support system and the task environment, different approaches can be more or less useful. The advantages of the task-based approaches are the low intrusiveness and high reliability. However, they are very task dependent. In other words, the human does not need to be disturbed at all, but for every new task a new measurement needs to be adopted by the agent. The NASA-TLX questionnaire is also very sensitive, but the very low score for intrusiveness can make it difficult to apply in a real world situation such as an operator working in Air Traffic Control. Here, interruption of the operator can have disastrous consequences.

This research shows that there are several different, very useful performance measurements possible for an agent to use in the example simulation-based training environment. For future research, the idea is to incorporate the most promising performance measurements in a personal assistant agent, and see how well this support agent is able to support the human. Note that a preliminary study concerning this has already been performed (see [14]), however in that setting not all promising measurements have been utilized by the personal assistant agent yet.

References

1. Posner, M.I., Boies, S.J.: Components of attention. Psychological Bulletin. 78, 391–408 (1971)
2. Fuchs, S., Kelly, S., Stanney, K.M., Juhnke, J.S., Dylan, D.: Enhancing mitigation in augmented cognition. Journal of Cognitive Engineering and Decision Making 1(3), 309–326 (2007)
3. Modi, P.J., Veloso, M.M., Smith, S.F., Oh, J.: CMRadar: A Personal Assistant Agent for Calendar Management. In: Bresciani, P., Giorgini, P., Henderson-Sellers, B., Low, G., Winikoff, M. (eds.) AOIS 2004. LNCS (LNAI), vol. 3508, pp. 169–181. Springer, Heidelberg (2005)
4. Myers, K., Berry, P., Blythe, J., Conley, K., Gervasio, M., McGuinness, D.L., Morley, D., Pfeffer, A., Pollack, M., Tambe, M.: An Intelligent Personal Assistant for Task and Time Management. AI Magazine Summer, 47–61 (2007)

5. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Resulsts of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload, pp. 139–183. North-Holland, Amsterdam (1988)
6. Muehlen, M.: Workflow-based Process Controlling – Or: What You Can Measure You Can Control. In: Fischer, L. (ed.) Workflow Handbook, pp. 61–77 (2001)
7. Hockey, G.R.J.: Compensatory control in the regulation of human performance under stress and high workload: a cognitive energetical framework. Biological Psychology 45, 73–93 (1997)
8. Reid, G.B., Nygren, T.E.: The Subjective Workload Assessment Technique: a scaling procedure for measuring mental workload. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload, pp. 185–218. North Holland, Amsterdam (1988)
9. Both, F., Hoogendoorn, M., Lambalgen, R., van, O.R., Vos, M.: de, Relating Personality and Physiological Measurements to Task Performance Quality. In: Proc. of the 31th Annual Conference of the Cognitive Science Society (CogSci 2009), Austin, TX. Cognitive Science Society (2009)) (to appear)
10. Veltman, J.A., Gaillard, A.W.K.: Physiological workload reactions to increasing levels of task difficulty. Ergonomics 41(5), 656–669 (1998)
11. Pandian, P.S., Mohanavelu, K., Safeer, K.P., Kotresh, T.M., Shakunthala, D.T., Gopal, P., Padaki, V.C.: Smart vest: wearable multi-parameter remote physiological monitoring system. Medical Engineering & Physics 30(4), 466–477 (2008)
12. Coelli, T., Prasada Rao, D.S., O'Donnell, C.J., Battese, G.E.: An introduction to efficiency and productivity analysis. Springer, Heidelberg (2005)
13. Eggemeier, F.T., Wilson, G.F., Kramer, A.F., Damos, D.L.: Workload assessment in multi-task environments. In: Damos, D.L. (ed.) Multi-Task Performance, pp. 206–216. CRC Press, Boca Raton (1991)
14. Bosse, T., Both, F., Duell, R., Hoogendoorn, M., Klein, M.C.A., Lambalgen, R., van Mee, A., van der Oorburg, R., Sharpanskykh, A., Treur, J., de Vos, M.: An Ambient Agent System Assisting Humans in Complex Tasks by Analysis of a Human's State and Performance. In: Proceedings of the Second IEEE International Conference on Intelligent Human Computer Interaction (IHCI 2010). Springer, Heidelberg (2010) (to appear)
15. Rubio, S., Diaz, E., Martin, J., Puente, J.M.: Evaluation of subjective mental workload: a comparison of SWAT, NASA-TLX, and workload profile methods. Applied Psychology: an International Review 53(1), 61–86 (2004)

Cognitive Adaptive Man Machine Interfaces for the Firefighter Commander: Design Framework and Research Methodology

Maurits de Graaf[2], Michel Varkevisser[1,*], Masja Kempen[1], and Nicolas Jourden[2]

[1] Human Factors & Cognition Laboratory, Thales Research & Technology, Delft, The Netherlands
[2] Thales Nederland, Land Forces & Security, Huizen, The Netherlands
michel.varkevisser@d-cis.nl

Abstract. The ARTEMIS CAMMI project aims at developing a joint-cognitive system to optimise human operator's performance under demanding labour conditions. The CAMMI domain applications concern avionics, automotive, and civil emergencies. In this paper we address the development of a joint-cognitive system for firefighter commanders to optimise situational and team awareness by reducing the workload through mitigation strategies and an adaptive HMI. A general framework and a research methodology are presented to explore the possibilities of applying the CAMMI building blocks in the development of systems to support the handling of firefighter emergencies.

Keywords: HMI, Situational Awareness, Team Awareness, Mental Load, Mitigation Strategies.

1 Introduction

Firefighter commanders in the control-loop of a complex system are typically exposed to high task demands and under continuous time pressure, which usually leads to decreased situational and team awareness (Bass, Zenyuh, Small & Fortin, 1996), and consequently in performance decrement and safety risks in the overall mission. The CAMMI[1] (Cognitive Adaptive Man Machine Interface) project, is anchored in the concept of a closed-loop architecture where the Human Machine Interface (HMI) supports adaptive mitigation strategies based on the measured workload of the operator. Thus, a workload exceeding the operator's capability will lead to off-loading of non-critical, time-consuming tasks through optimised automation strategies. CAMMI addresses various domains, namely flight-management systems for aircraft, ground control stations for UAVs, automotive domain and the crisis management domain. This paper addresses automation strategies allowing the firefighter commander to focus on his critical tasks (Conway & Hockey, 2007). In the domain of

---

* Corresponding author.

[1] CAMMI is funded by the European Technology Platform ARTEMIS (Advanced Research & Technology for EMbedded Intelligence and Systems), Grant agreement no.: 100008.

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 588–597, 2011.

civil emergency management it is expected that workload measurement and optimised mitigation strategies can improve the level of situational and team awareness and thereby increase operational performances and safety (Endsley, 2000; Perry, Wiggins, Childs & Fogarty, 2009).

Due to the nature of their tasks and responsibilities, firefighter commanders typically experience high cognitive load. The commander is responsible for the processing of a vast amount of incoming data to maintain situational awareness and to coordinate his teams (team awareness). Yet, little systematic research has been conducted on mental load assessment of firefighter commanders. There are some indications of mental overload in a recent study (Roja, Kalkis, Kalkis & Pencis, 2009). Offloading secondary tasks might be very beneficial for a commander to accomplish his primary tasks (Conway & Hockey (2007). Assuming that advanced technology will be used to support team commanders (Kontogiannis & Kossiavelou, 1999), we propose a design framework and a research methodology on how a firefighter commander can be assisted through use of technology, including mental load reduction strategies.

2 Work Domain

The command structure for firefighters consists of a fire truck with a team of 4 to 6 fire men and a driver under the command of one team captain. When two or more trucks are cooperating, a commander is added, who commands the team captains. It is of vital importance for the commander to obtain an accurate and consistent picture of the situation to prevent further escalation, minimise the number of casualties, and restrict the damage.

The incident management typically consists of a disorganised phase and a relatively controlled phase. The stages in the disorganized phase can be separated into a) the approach, b) the arrival on-scene, and c) the partitioning of the area. The approach is expected to involve relative high cognitive work load. The main task of the firefighter commander in the approach stage is to build up and maintain situational awareness (SA). From research it is known that an experienced firefighter engages in anticipatory thinking to build situation awareness by means of pattern matching, trajectory tracking, and convergence (Klein et al., 2007; McClellan et al., 2009). To maintain SA it is important for a commander to continually find, update and interpret and share the pieces of information he needs at any given time (Endsley, 2000; Kaber & Endsley, 1998; Weisband, 2002). As in emergency situations information is typically scattered and sometimes contradictory and contact with the teams oftentimes troublesome, it is presently cognitively demanding for a commander to maintain SA.

In the ‘arrival-on-scene’ and ‘partitioning of area’, SA consists of sending firefighters into the building to gather as much information as possible, e.g. estimation of people inside, the scope of the fire, and exit options (Fig. 1). Note that the commander does not enter the building. Here workload is high for the commander: the teams continually communicate information from inside the building and its perimeter to the commander. In parallel, for Team Awareness (TA) it is vital to monitor the time firemen have been inside the building, possible signs of heat exertion, and their exact location. Presently the commander is forced to make a mental and paper map. Mental load is expected to be very high under these conditions and needs to be assessed in a systematic way (Section 4).

Fig. 1. Impression of the situation, with in the middle picture the commander in communication with his teams through a portophone. Updates are marked with pen and paper.

We propose that our support system should, through optimised mitigation strategies, assist the commander in building and maintaining SA and TA. According to the 2008 National (US) Fire Fighter Near-Miss Annual Report "loss of situational awareness is a significant factor in firefighter near-miss events: unintentional unsafe occurrences that could have resulted in an injury, a fatality, or property damage." This observation that SA and TA are key, holds also for (near miss) incidents in The Netherlands[2]. An integrated, intelligent system that can seamlessly acquire, fuse, reason about, distribute, and protect information to provide enhanced decision support and situational understanding as well as foster effective collaboration will have a great advantage compared to traditional methods.

3 CAMMI Architecture

The CAMMI architectural approach is represented by an augmented HMI design: the traditional HMI model is enriched with an additional (active) component performing continuous monitoring of the operator's workload levels which serve as a basis for the decision of triggering automation/adaptation (mitigation) strategies in those cases where the assessed workload levels seem critical or potentially dangerous to the overall safety and performance of the mission. The main blocks of this architectural design are depicted in Fig. 2, and are summarized briefly. For the current paper we will mainly focus on the mental load control loop (Fig. 3).

3.1 Cognitive Monitor Subsystem

The Cognitive Monitor combines psychophysiological measures, thereby taking into account task activity, time pressure and potential environmental stressors to determine the Firefighter Commander's mental load in real-time. The concept of workload needs

[2] Public governmental reports analysing major incidents in the Netherlands are written in the Dutch language, and are available upon request.

to be explored in more detail in the domain of firefighting management to determine how a mental load index can be calculated. The difficulty with the concept of workload is that it is multidimensional and can only be measured indirectly, through measurement of physiological, psychological, and/or behavioral processes (Cain, 2007). For the firefighter commander, in the current project workload is operationalised as a physiological index and embedded in a mental load loop (Fig. 3). From recent research it has also been shown that physiology measurement could be used for task adaptation (Boucsein et al., 2007; Mulder et al. 2008).

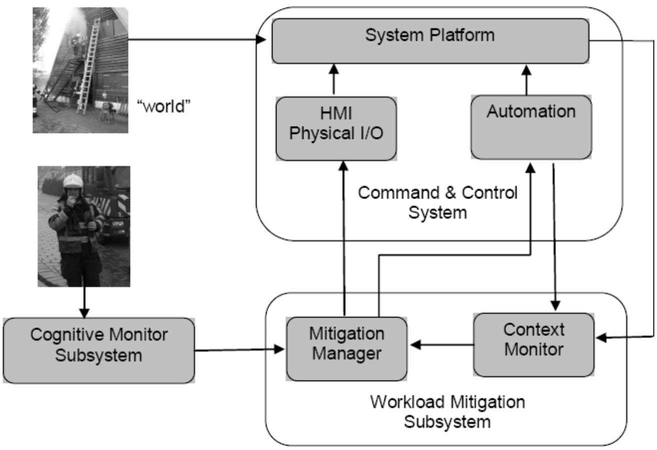


Fig. 2. High level view of CAMMI framework

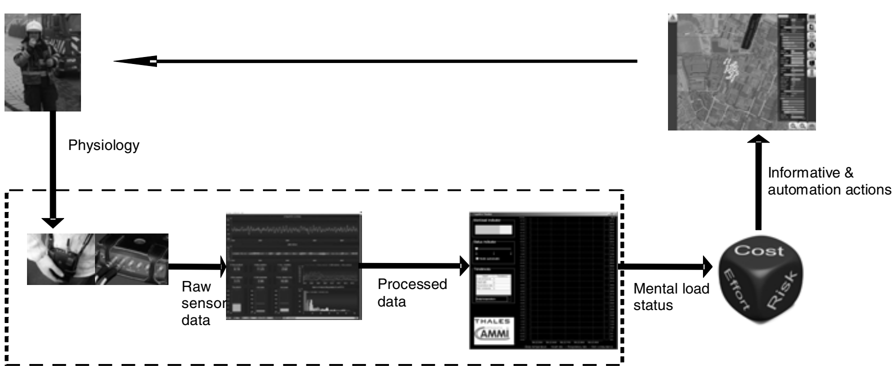


Fig. 3. The mental load feedback loop. The dashed box indicates the components within the cognitive monitor, which sends a mental load status to the mitigation server. The mitigation server decides to send a message to the HMI, based on mental load and context parameters.

Mental load is known to be influenced mainly by the type of tasks, time pressure, and human cognitive capabilities. A number of task characteristics determine the level of experienced mental load. Neerincx et al. (2004, 2009) developed a model for cognitive task load in complex environments. Time pressure is evidently a major contributor to the experience of mental load (Bass, Zenyuh, Small & Fortin, 1996).

Several anticipatory, task-related and general/non-specific psychophysiological responses are to be expected from operations in the field, ranging from acute elevations in physiological arousal to chronic changes in physiological patterns. For the operational assessment the primary elements will be physiological arousal and emotional changes as a function of stress and time. When the tasks are too demanding (mental overload), peaks are to be expected in several physiological indicators (Al'Absi et al., 1997; Kelsey et al., 2004, Varkevisser & Keyson, 2007). If this prolongs over a longer period of times, energy will be depleted, resulting in fatigue or exhaustion and serious performance decrement (Neri et al., 1997).

The Cognitive Monitor generates a mental load index varying between three levels: baseline Workload (WL 0), medium Workload (WL 1) and high Workload (WL 2). These values will heavily depend on the above-mentioned influences of tasks, time, and cognitive capabilities at any given time and should be carefully considered in an experimental setup (Section 4). The participants will thus be measured under different conditions, ranging from routine tasks to working-memory to multiple parallel tasks embedded in operational scenarios in a simulated environment. This approach will provide a means to understand the dynamics of workload and its influence on SA and TA.

3.2 Workload Mitigation Subsystem

The level of automation in a joint human-automation system can vary from completely manual, where the entire task is performed by the operator, to fully automatic, where the system does everything autonomously (Sheridan and Verplank, 1978). Adaptive automation can either provide adaptive aiding, which makes a certain component of a task simpler, or can provide adaptive task allocation, which shifts an entire task from a larger multitask context to automation (Parasuraman, Mouloua, & Hilburn, 1999). Adaptive mitigation strategies can include task management, optimising information presentation via modality management, task sharing, and task loading. For example, an air traffic controller might be presented with decision aids for conflict detection and resolution by the automated system when it detects a rapid increase in traffic density or complexity (Hilburn, Jorna, Byrne & Parasuraman, 1997). Ultimately, the goals of adaptive automation are similar to those of automation in general, such as avoiding operator out of the loop conflicts or mistrust in the automation.

The Mitigation Manager has two broad areas of adaptation: at the level of the HMI component and at the level of the Automation component. Based on the workload index, the Mitigation Manager decides what action of the HMI or Automation component is necessary in order to optimize the joint human-automation, whilst taking into account the information from the Context Monitor. This information is important as in certain contexts high cognitive workload is appropriate and no mitigation needs to be triggered, while in other contexts high cognitive workload would indicate a need to mitigate. So the Mitigation Manager is in fact ‘aware’ of the possible entry points for automation and HMI adaptation. Triggered by the Mitigation Manager, the Automation component ensures that tasks that are being performed by the commander are taken over entirely or partially (e.g. simplify tasks).

Based on observations and interviews with firefighter commanders it turned out that (sub-)tasks related to TA, being the commander’s primary responsibility, are most eligible for mitigation. TA includes: 3D location of the actors (movement), monitoring of (body) temperature, monitoring of vital signs of firemen (heart rate, air tank level, breathing air consumption rate, and other vital signs information. In the CAMMI-project, we focus on mitigation strategies for `alarming situations’, defined as: personnel entering a forbidden area, an extreme temperature, or a dangerous gas.

3.3 HMI

The Human-Machine Interface (HMI) is the primary way the commander will interact with the systems, including the platform and the automation components. In the current project, the commander will use a command and control (C2) application that shows a map with the locations of the firemen, their vital signs, their problems and possible solutions. The C2 application’s HMI supports voice command interactions for both the commander and the firemen, and allows the commander to send routes, places and information to the crew. We have developed a tool named Cerberus to build and maintain SA and TA (Fig. 4).

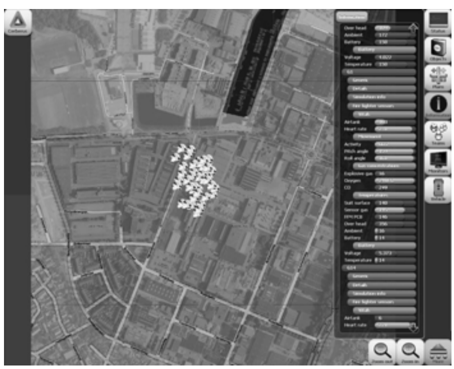


Fig. 4. Cerberus screen appearance. The application displays the relevant parameters of each firemen, such as location (represented by helmets), reported incidents, fire development, and risk of collapse.

3.4 Remaining Components

The joint cognitive system operates in an operational context (represented by the globe in Fig.1). Events and states of the operational context may impact the operation of the information platform or affect the mission, e.g. weather condition, smoke development, presence of people, size of the site. Firefighters provide the commander with the relevant information to maintain TA (firefighter silhouettes in Fig.1). The

Platform consists of the physical systems controlled or operated by the human, including sensors in which it can detect aspects of the outside world, actuators to control its movement or affect the state of the outside world, and communication systems with which to transmit information.

4 Research Methodology

4.1 Development of the Cognitive Monitor

In a first series of experiments, the focus will be on the measurement and interpretation of physiological indices of mental load (Collet et al., 2009) and in what way mental load potentially mediates SA and TA. Among the most commonly used parameters are heart rate, heart rate variability, skin conductance, respiration, and muscle tension. The match between an individual's baseline state and his operational state is known to give an accurate assessment of the level of psychophysiological reactivity towards different stressors (Varkevisser & Keyson, 2007). Also important is the individual's adaptability, i.e. the ease with which he or she is able to move between state levels without experiencing mental overload or strain (Wilson et al., 2003).

As a proof of concept, our main aim is to perform tests in a simulated emergency environment (on a computer) to systematically explore mental load issues involved in emergency management by firefighter commanders. For the acquisition of the signals the Nexus-10 polygraph will be used (Fig. 3). It converts the raw analogue values from the sensors to processed digital data, such as heart rate average, which are fed into the monitoring platform (Fig. 3). Algorithms need to be developed to combine the physiological parameters in order to generate a mental load index. We will possibly build on earlier algorithms as proposed by Healey & Picard (2005) and Ohsuga et al. (2001). The mental load state algorithms should be able to reliably discriminate between the three different levels (Section 3) of load on a moment to moment basis.

Through an iterative experimental process a specific subset of sensor data will be chosen that correlate with high cognitive load. It is important to stress that currently no existing fully working models exist in the firefighting domain that directly translate physiological signals into indices for workload. The challenge lies in finding the right combinations of physiological signals in relation to domain-specific user activity data. Moreover, based on the work domain analysis we have to develop a technique where the accuracy is high and the learning time of the system is low. The power of physiological signals can be increased by incorporating/adding subjective assessment, yet these are typically more intrusive.

In the first series of experiments, participants will be evaluated in a simulated environment in which different levels of stress will be induced. In the simulation a participant (representing a commander) has to complete a certain assignment by managing a number of virtual people (representing firefighters) through a maze (representing an abstract building). The 'commander' will manage the 'firefighters' by means of a digital assistant. For this purpose he has to carry out abstract tasks which guide the virtual firefighters through a 'maze-like' world, who –through scripting– report on certain events they encounter The ‘commander’ has to respond to those events, e.g. by placing objects on a digital map, by re-routing the characters, etc. The tasks will be scaled-up in difficulty during the scenario, to ensure an increase of mental load. This will be validated by means of subjective mental effort and performance measures linked to mental load, such as response latencies, objects missed, and strategies used. The main goal is to extract the common characteristics from all the studied physiological values to produce a global rule that will be adapted to fit in a fuzzy logic controller within the Cognitive Monitor. At this time we are defining the test procedure to gather the data and extract a global rule for one environment. The participants will use a simplified version of the Cerberus HMI developed at Thales Nederland B.V., located in Huizen, The Netherlands (Section 3).

In a pilot experiment we have already explored the development of a look-up table based on the physiological dataset (4 parameters: temperature, skin conductance, respiratory rate and heart rate variation) of an experimental participant, in order to qualify each of the individual physiological curves as a three state value: increase, decrease and constant. A lookup table provides values matched to a referenced table and is used at run time. Since retrieving a value from memory is often faster than undergoing an 'expensive' computation, the use of a look-up table saves processing time and space in terms of transformations. Although, a general look-up table would be difficult to develop this solution is robust and easy to configure. An approach using neural network based classification algorithms may be more viable.

4.2 Development of the Mitigation Manager

The mitigation strategies for alarming situations (defined as: personnel entering forbidden area, an extreme temperature, or a dangerous gas), as proposed in Section 3 will be implemented in a simulator and evaluated together with domain experts to determine whether these strategies are valuable in assisting the commander in building and maintaining SA and TA. The same test-bed as for the development of the Cognitive Monitor may be used, but now with a mental load index and a system mitigator (pro)actively supporting the operator. Finally, in an iterative research cycle, the design of the Cerberus HMI will be optimised through assessment of SA and TA parameters.

References

[1] Al’Absi, M., Bongard, S., Buchanan, T., Pincomb, G.A., Licinio, J., Lovallo, W.R.: Cardiovascular and neuroendocrine adjustment to public speaking and mental arithmetic stressors. Psychophysiology 34, 266–275 (1997)

[2] Bass, E.J., Zenyuh, J.P., Small, R.L., Fortin, S.T.: Context-based approach to training situation awareness. In: Proceedings of the IEEE Symposium on Human Interaction with Complex Systems, Dayton, OH, USA, pp. 89–95 (1996)

[3] Boucsein, W., Haarmann, A., Schaefer, F.: Combining Skin Conductance and Heart Rate Variability for Adaptive Automation During Simulated IFR Flight. In: Harris, D. (ed.) HCII 2007 and EPCE 2007. LNCS (LNAI), vol. 4562, pp. 639–647. Springer, Heidelberg (2007)

[4] Cain, B.: A review of the mental workload literature. In: NATO report RTO-TR-HFM-1211-Part-II, Toronto, Ontario, Canada (2007)
[5] Conway, G.E., Hockey, G.R.J.: Effects of cyclic loading on complex performance and operator functional state. In: de Waard, D., Hockey, G.R.J., Nickel, P., Brookhuis, K.A. (eds.) Human Factors Issues in Complex System Performance, pp. 339–344. Shaker Publishing, Maastricht (2007)
[6] Collet, C., Averty, P., Dittmar, A.: Autonomic nervous system and subjective ratings of strain in air-traffic control. Applied Ergonomics 40, 23–32 (2009)
[7] Endsley, M.R.: Theoretical underpinnings of situation awareness: a critical review. In: Endsley, M.R., Garland, D.J. (eds.) Situation Awareness Analysis and Measurement, pp. 1–24. Lawrence Erlbaum Associates, Mahwah (2000)
[8] Healey, J.A., Picard, R.W.: Detecting Stress During Real-World Driving Tasks Using Physiological Sensors. IEEE Transactions on Intelligent Transportation Systems 6, 156–166 (2005)
[9] Hilburn, B., Jorna, P.G., Byrne, E.A., Parasuraman, R.: The Effect of Adaptive Air Traffic Control (ATC) Decision Aiding on Controller Mental Workload. In: Mouloua, M., Koonce, J. (eds.) Human-Automation Interaction: Research and Practice, pp. 84–91. Lawrence Erlbaum, Mahwah (1997)
[10] Kaber, D.B., Endsley, M.R.: Team Situation Awareness for Process Control Safety and Performance. Process Safety Progress 17, 43–48 (1998)
[11] Kelsey, R.M., Soderlund, K., Arthur, C.M.: Cardiovascular reactivity and adaptation to recurrent psychological stress: Replication and extension. Psychophysiology 41, 924–934 (2004)
[12] Klein, G., Snowden, D., Chew, L.P.: Anticipatory thinking. In: Proceedings of the Conference on Naturalistic Decision Making, Pacific Grove, CA, USA (2007)
[13] Kontogiannis, T., Kossiavelou, Z.: Stress and team performance: principles and challenges for intelligent decision aids. Safety Science 22, 103–128 (1999)
[14] Mclennan, J., Elliot, G., Holgate, A.M.: Anticipatory Thinking And Managing Complex Tasks: Wildfire Fighting Safety And Effectiveness. In: Proceedings of the Industrial & Organisational Psychology Conference, Sydney, Australia, pp. 90–95 (2009)
[15] Mulder, L.J.M., de Waard, D., Hoogeboom, P., Quispel, L., Stuiver, A.: Using physiological measures for task adaptation: towards a companion. In: Westerink, J.H.D.M., Ouwerkerk, M., Overbeek, T.J.M. (eds.) Probing Experience: From Assessment of User Emotions and Behaviour to Development of Products, pp. 221–234. Springer, Heidelberg (2008)
[16] Neerincx, M.A., Grootjen, M., Veltman, J.A.: How to manage cognitive task load during supervision and damage control in an all-electric ship. IASME Transactions 2, 253–258 (2004)
[17] Neerincx, M.A., Kennedie, S., Grootjen, M., Grootjen, F.: Modeling the Cognitive Task Load and Performance of Naval Operators. In: Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience, pp. 973–978. Springer, Heidelberg (2009) ISBN 978-3-642-02811-3, no. 5638
[18] Neri, D.F., Dinges, D.F., Rosekind, M.R.: Sustained carrier operations: sleep loss, performance, and fatigue countermeasures. Internal Report, Fatigue Countermeasures Program Flight Management and Human Factors Division, NASA Ames Research Center(1997)
[19] Ohsuga, M., Shimono, F., Genno, H.: Assessment of phasic work stress using autonomic indices. International Journal of Psychophysiology 40, 211–220 (2001)
[20] Parasuraman, Mouloua, & Hilburn (1999) (To Be Done)

[21] Perry, N., Wiggins, M., Childs, M., Fogarty, G.: Decision Support For Competent And Expert Firefighters. In: Proceedings of the Industrial and Organisational Psychology Conference, Sydney, Australia, pp. 101–105 (2009)
[22] Roja, Z., Kalkis, V., Kalkis, H., Pencis, I.: Assessment of firefighters-rescuers' work severity in relation with interaction between physical and mental load. In: Proceedings of the Latvian Academy of Sciences. Section B. Natural, Exact, and Applied Sciences, vol. 63, pp. 264–270 (2009)
[23] Sheridan, T.B., Verplank, W.L.: Human and Computer Control of Undersea Teleoperators. MIT Man-Machine Systems Laboratory, Cambridge, MA, Tech. Rep. (1978)
[24] Varkevisser, M., Keyson, D.V.: The impact of VDU tasks and continuous feedback on arousal and well-being: Preliminary findings. In: Dainoff, M.J. (ed.) HCII 2007 and EHAWC 2007. LNCS, vol. 4566, pp. 151–156. Springer, Heidelberg (2007)
[25] Weisband, S.: Maintaining awareness in distributed team collaboration: implications for leadership and performance. In: Hinds, P.J., Kiesler, S. (eds.) Distributed Work, pp. 311–334. Institute of Technology, Massachusetts (2002)
[26] Wilson, G.F., Russell, C.A.: Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. Human Factors 45, 635–643 (2003)

An Intelligent Infrastructure for In-Flight Situation Awareness of Aviation Pilots

Alessandro G. Di Nuovo[1,2], Rosario Bruno Cannavò[1], and Santo Di Nuovo[2]

[1] Cognitive Technologies and Services S.r.l., Catania, Italy
[2] Università degli Studi di Catania, Catania, Italy

Abstract. This paper presents an infrastructure that integrates intelligent agents in order to monitor, in real time, the attention of aviation pilots during training/operative flight missions. The primary goal of this infrastructure is to make the decision process easier and increase Situation Awareness, thus to increase flight safety pro-actively. The proposed hardware/software platform could be able to anticipate the onset of problems which can lead to incidents, and to make easier the decision making process toward a positive solution of the problem. To attain the goal, a multi-agent system is designed using the most recent technology in the field of artificial vision and of the measurement of psychophysical parameters, starting from the most recent knowledge of visual attention to arrive at the development of an original and innovative model of Augmented Reality. Finally it is provided a case study based on an event actually occurred to prove effectiveness of the proposed platform.

Keywords: Situation Awareness, Intelligent Agents, Augmented Reality.

1 Introduction

With the doubling of air traffic predicted for 2020 ([1]) the total number of incidents will rise even though the ratio of incidents/flight hours will stay low, and this will impact on the passengers' perception of the safety of air transport, as they take more notice of the number of incidents than their ratio. Since it has been widely documented that at least the 70% of commercial aviation incidents, in the last 15 years, are connected with human errors [2,3],. This is linked with the increasing difficulties to interact with ever more complex planes, as was found by the different studies [4] and [5], that highlight how on the one hand there has been an enormous increase in avionics and on-board systems which, taken one by one, should increase safety (FMS, Narrow spacing for VHF frequency, TCAS); on the other hand, budget requirements, the need to retrofit numerous old-generation aircraft (forced to suffer invasive technological upgrading to be able to use airspaces efficiently and economically), have produced the installation on board of low cost apparatuses and with often in-existent considerations of the ergonomics of the interface used, leading to an overall increase in the workload and a compromise of the global situation awareness.

Therefore the complexity of modern flight deck of commercial and / or military aircraft puts the pilot on the sidelines of a huge flow of information that helps to

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 598–607, 2011.

create a safe and efficient flight profile. The various aircraft systems, however, offer only a limited set of data for pilot analysis, e.g. only the information that are deemed useful to maintain the pilot within the attention loop needed to exercise appropriate control over the system and to take the right actions to maintain the level of safety and efficiency expected. This leads both the marginalization of the situation awareness of the pilot, that in routine operations has to perform simple and repetitive actions through the instrumentation and to monitor the environment; on the other hand relevant data are processed by the aircraft systems only in order to do not interfere with the perception of the pilot and to avoid overloading of the cognitive process of the subject. In case of detection of an anomaly in the profile of flight the pilot must: (i) Take all the information; (ii) Elaborate them all, focusing the problem; (iii) Develop one or more coping strategies; (iv) implementing a decision of containment of the event. Usually the time available to the pilot for problem solving is short, in which he must seize as much information as possible and process them all starting from the already established patterns, using its system expertise, in order to arrive at an answer to the original problem. In the case of high workload, however, the capacity to collect and analyze information and matching the current case with past experience is severely penalized and often the decision making process arises as a result of an inadequate analysis of information and a substantial revival of behaviors already encoded is applied, and it is not adapted to the new situation. The heterogeneous composition of aircraft and avionics makes even more critical the decision making process and even more important to capture the largest possible number of appropriate information for a correct interpretation of the actual situation.

To help the pilot we propose an integrated infrastructure, based on intelligent agents, to monitor in real time the attention given by aviation pilots during flight operations, to facilitate decision-making and to enhance situation awareness (i.e. awareness of the events in surrounding environment). The infrastructure is based on pilot's eye and gaze tracking, as well as on the monitoring of attention arousal in flight deck, using real-time hardware systems that process signals from opto-electric and psychophysiological sensors. The output from the tracking system will be send as input to a comparison agent that determines the level of attention and its allocation in the cockpit. If the infrastructure detects a form of inattention it shall issue a warning signal with the aim to restore the level and / or the appropriate allocation of attention, according to the criticality detected.

The proposed infrastructure works comparing the behaviors shown by the pilot himself in similar situations to the behaviors measured in real time during the flight. To this end the infrastructure include a virtual cockpit system, that records in real-time both the telemetry data of on-board avionics system and the behavior of the subject, and compares them all with previous records in telemetry and behaviors databases. If computing these elements it is identified an inappropriate attention process, either for quantitative or for allocation, the infrastructure will produce a warning signal that will stimulate the situation awareness and, providing useful information to the analysis of the problem, the process of decision-making.

From a methodological point of view, we will propose the development of new closed loop technologies by means of the computerization of the environment in which pilots find themselves operating, so that the system can identify attention lapses, analyze potentially dangerous situations and produce alert signals which allow correct decision making and/or full situation awareness. In this sense the proposed multi-agent infrastructure will be integrated with the on-board instruments. The infrastructure is composed of optoelectronic sensors, which produce useful information about the physical environment in which pilots operate. In addition the proposed system will allow us, for the first time, to model the attention process based on real individual vision. The computerized psycho-physiological study will thus allow us to integrate into a group behavior model the specific, individual, results of the subject, and definitively, to have both a complete panorama of the information present in the operative context, and feedback based on the correlation of actual data and perceived data and/or those taken into consideration, as well as a comparative analysis between the current parameters and previous ones (sampled in other experiences), with the relative decision results. The computerization of behaviors measured thus will allow us to set up a database aiming at reducing the work load brought by knowledge based happenings relative to routine and/or non-routine situations through the development of an appropriate alert protocol which informs the pilot of a difference between an effective operating situation and the perceived one.

But what will place this study strategically at the edges of current technology will be the integration of the outcomes with the concept of Augmented Cognition. The Augmented Cognition program, developed by the Defense Advance Research Projects Agency (DARPA) [7], has as its core elements the focus on the real time measurement of the subject's cognitive state, measured by modern neuroscientific/psycho-physiological instruments, and the concept of "closed loop", in which the cognitive state of the operator is identified in real time and produces an appropriate adaptation of the system around him [8,9]. The current state of the art sees augmented cognition still to be implemented in the field of the measurement of visual parameters which, in the aeronautical field, are rendered even more complex by the dynamism of the environment. The most recent studies have underlined this gap, highlighting the inadequacies of psycho-physiological (visual) parameters which did little to help the measurement of inattention [10].

The rest of the paper is organized as follows: Section 2 shows the model of the proposed infrastructure; in Section 3 we will study the effectiveness of our system model using a real world event actually occurred. Finally Section 4 reports our conclusion and future works.

2 SAMI: A Multi-agent Infrastructure for Situation Awareness Monitoring

It has been demonstrated in various fields that productivity can be greatly enhanced through the use of automation technologies. In order to be fully accepted in aviation, automation not only needs to increase productivity but do so cooperatively with the pilot. That is, the pilot must know what the automation is currently doing and what it will do in the future. Rather than replacing what the pilot currently does, we aim to augment his capabilities with signals that back up his internal knowledge and enhance his SA. The main aims of our infrastructure is to increase flight safety pro-actively through early detection of problems that can lead to incidents, and to make easier the decision making process which will lead to a positive solution of the problem. To attain the goal, in this section we will introduce a multi-agent system, which exploits the most recent technology in the field of artificial vision and of the measurement of psychophysical parameters, starting from the most recent knowledge of visual attention to arrive at the development of an original and innovative model of Augmented Reality.

The Agent-based Model of SAMI. In our model the information flows in a feed-forward fashion, from flight deck, where two interface modules collect and select relevant data, toward the reasoning module of SAMI that is in charge to take decision if the attention is adequate or not and, if not, to send a signal to pilot. Therefore, the output of SAMI is to alert the pilot either if he loses quantitative attention during the routine of the flight or if he lacks of qualitative attention in non-routine operations. In other words SAMI will send a warning signal to restore the attention level when the pilot shows a low level of attention during routine or will send an alert signal to focus the pilot's attention to a rising problem.

As the attention level during the flight operations has a gradual degradation, and, meanwhile, the attention focus can be more or less near to the points which need constant supervision, the adequacy of the attention must be computed as a fuzzy value [11]. For this reason they are modeled using fuzzy logic, that is able to deal with reasoning that is robust and approximate rather than brittle and exact [12]. Therefore fuzzy logic variables may have a truth value that ranges in degree between 0 and 1. Furthermore, when linguistic variables are used, these degrees may be managed by specific functions.

As shown in Figure 1 three main components form the system, they are listed with a brief description of their functions in the following.

HMI: Human Machine Interaction Recorder. This component records all the interactions of pilot with the cockpit. It uses the available technologies to track psychometric, neural-physio-logical parameters, e.g. eye and gaze movements, heartbeat, blood pressure, skin conductance, EEG.

CAV: Cockpit Artificial Vision. This component simulates a reference pilot that is able to acquire and to process all the information from the aircraft. In fact, it derives telemetry data from all on-board control systems, like for example the Flight Management Guidance System (FMGS) and the Enhanced Ground Proximity Warning System (EGPWS). Collected data is computed and aggregated in real-time, then relevant parameters are sent to SAMS for processing.

SAMS: Situation Awareness Monitoring System. This component is the core of the infrastructure. It is modeled as a fuzzy inference system, which is the process of formulating the mapping from a given input to an output using fuzzy logic. SAMS is divided into two elements: the ***Pilot Attention Monitor*** (PAM) and the ***Situation* Awareness Monitor (SAM).** PAM has three modules: the Attention Level Evaluator (ALE), the Attention Focus Evaluator (AFE) and the System Database (SD). In a preliminary phase PAM identifies the current pilot and does an initial assessment of the attention level. Next, during flight operations, all three modules receive raw input information from HMI and CAV. ALE and AFE evaluate two indexes that indicate respectively the pilot's (quantitative) attention level and the pilot's (qualitative) attention focus. The SD stores the input data from flight deck and the two indexes evaluated by ALE and AFE, meanwhile SD compare new inputs with the reference data and select the fuzzy rule and data bases according to the current pilot and flight situation. Personalized fuzzy rule and databases are sent, along with ALE and AFE indexes, to the SAM. PAM uses the information in SD to personalize the computation to the specific individual that is recognized by the preliminary identification as the current pilot. Feedback information received from SAM is used to update the personalized fuzzy rule and data bases. Data for individual information database can be collected during simulated training.

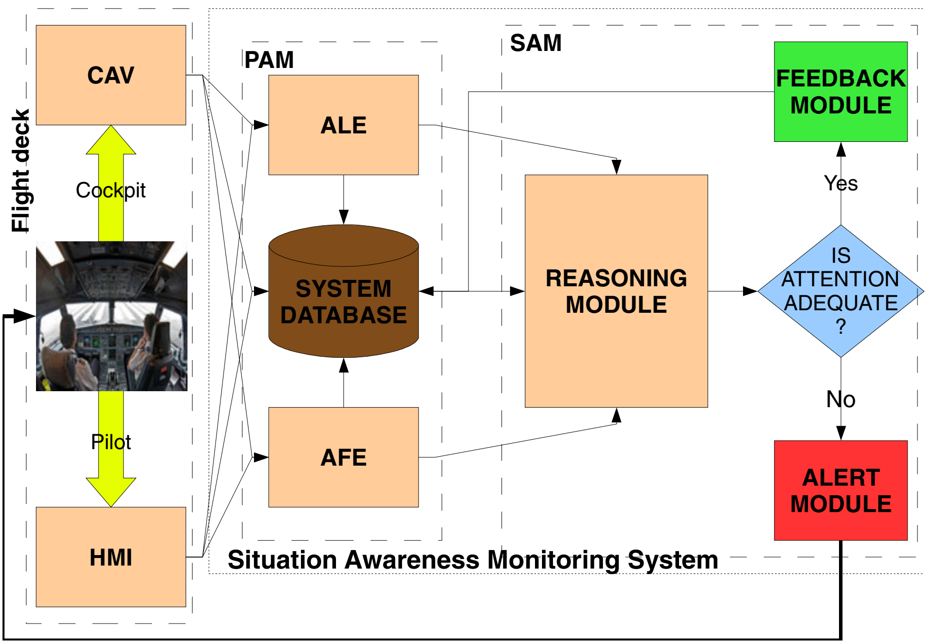


Fig. 1. Model for an Agent-based Situation Awareness Monitoring Infrastructure

SAM main module is the Reasoning Module (RM) that computes the fuzzy inference system to evaluate if the attention of the current pilot is adequate for current flight situation. The three input variables of the fuzzy inference system of SAM are the artificial indexes that PAM evaluates from raw data collected by HMI and CAV. If the answers are positive SAM send a feedback to the SD to update the information about the current pilot. If the answer is negative SAM activates the Alert Unit, which will send a signal to the pilot. Signals could be of three main categories: "*wake up*" signals, that are sent in case of loss of quantitative attention (i.e. SAMS output is *low attention*) with the aim to restore the correct level; “*warning*” signals, that are sent in order to catch pilot's attention in case of critical situations, when the pilot attention focus is in a wrong position (i.e. SAMS output is "*wrong-focus*"); "*alarm*" signals, that, in case of "*not adequate*" attention and "*dangerous*" situations, intervene with on-board instrumentation to cut off the unnecessary information and thus increase situational awareness. At the end of its computation SAM send feedback information to PAM to update the SD with the new data for system customization.

Realization Remarks. For all devices present in an aircraft the main rule to follow is "*do not interfere with the other on-board systems*". After this the proposed infrastructure can be classified as a "add-on", in other words it is not critical for the correct function of the other on-board systems and if it has problems it can be simply switched-off without any loss in the standard safety of the aircraft. However it is better to include in its design a self-check system able to automatically switch off the SAMS infrastructure in case of malfunctioning. Moreover it is safer to avoid false signals from it, for this reason even if a single module of the system is not functioning the whole system must be shut down. It is preferable that the pilot wears sensors that should not hinder the movement. For commercial pilots, in order to increase eye and gaze tracking performance without using invasive sensors, a camera can be integrated in glasses that the pilot should wear during the flight. This is not needed for military pilots that wear a full helmet in which it is yet possible to integrate a lot of devices, including a camera.

3 Case Study

In this section we will demonstrate the effectiveness of our system model using case study, based on an event actually occurred. The case under consideration is a Controlled Flight Into Terrain identified as CFIT. This category of accidents is one that aims to reduce the SAMI system. Implement the CFIT category represents worldwide the second leading cause of serious accidents on commercial aviation aircraft, behind the loss of control of the aircraft caused by damage to facilities or systems.

In this section we will report the actual development of the incident step by step and, in parallel, we will describe how the events could have been carried out if our SAMI model was operational.

The Incident with and without SAMI. On 11 March 2005, an Airbus A321-200 operated by British Mediterranean Airways, executed two unstable approaches below applicable minima in a dust storm to land in Khartoum Airport, Sudan. The crew were attempting a third approach when they received information from ATC that visibility was below the minimum required for the approach and they decided to divert to Port Sudan where the A320 landed without further incident.

In the following the record of the incident is divided into 5 steps, for each one we highlight the expected actions of SAMI and how they would be useful to improve situation awareness and thus flight safety.

Step	Official Report of the serious incident	SAMI actions
1	*Runway 36 was in use but the ILS on this runway was out of service. The commander assessed the weather conditions passed to him by ATC and believed that he was permitted, under his company's operations policy, to carry out a Managed Non-Precision Approach (MNPA) to Runway 36. This type of approach requires the autopilot to follow an approach path defined by parameters stored in the aircraft's commercially supplied Flight Management and Guidance System (FMGC) navigation database*	At this step there is no criticality on this scenario, the proposed SAMI model would not found anything unusual, but thanks to the CAV module it would record a non-precision approach on FMGC, loading the corresponding baseline parameters for this procedure from the system database as benchmark.
2	*On the pilot's approach chart, which was also commercially supplied but from a different supplier, the final descent point was depicted at 5 nm from the threshold of Runway 36 whereas the FMGC's navigational database had been correctly updated with a recent change to this position published by the Sudanese CAA which placed it at 4.4 nm from the threshold. The discrepancy amounted to a difference in descent point of 0.6 nm from the Khartoum VOR/DME beacon, the primary navigation aid for the non-precision approach.* *The pilots commenced the approach with the autopilot engaged in managed modes (i.e. the approach profile being determined by the FMGC instead of pilot selections). The aircraft began its final descent 0.6 nm later than the pilots were expecting. Believing the aircraft was high on the approach, the handling pilot changed the autopilot mode in order to select an increased rate of descent.*	Also at this stage our SAMI model records but does not intervene. It monitors the trajectory of the aircraft (altitude and position readings from FMGC) and the increase in the descent rate, meanwhile the attention level is monitored and compared with the benchmark. Outputs of SAMI at this step represent a situation compatible with standard safety operations : a strategic management of FMGS and in addition a subsequent tactical change by the use of selected vertical speed to handle the vertical trajectory. Furthermore, the sequence of actions recorded in the activation of pilot vision and psychophysical parameters when there is a change between strategic and tactical management are still part of the routine of the flight.
3	*The approach became unstable and the aircraft descended through 1,000 ft agl at an abnormally high rate.*	Now CAV detects the abnormal descent rate in relation to the position of the aircraft. According to our model CAV output shifts *operative situation* from n*ormal* to *critical* and AFE moves a*ttention focus* towards *far*, due to the analysis of the pilot being off the target parameter (high descent rate below 1000 ft agl). SAMI output is now *wrong-focus* and a warning signal is triggered in order to lead the pilot attention toward the target parameter (descent rate)

4	*The aircraft then passed through its Minimum Descent Altitude (equivalent to a height of 390 ft agl) with neither pilot having established the required visual references for landing. Instead each pilot believed, mistakenly, that the other pilot was in visual contact with the runway approach lights.*	At this step, SAMI output turns to l*ow-attention*, therefore monitoring pilot attention level, ALE detects the difference between the problem reported by the CAV and the current pilot vision. In this case activates a second signal, an alert to instruct pilots to stop the descent and follow the missed approach path.
5	*When the confusion between the two pilots became apparent, the aircraft had descended to approximately 180 ft agl and the handling pilot commenced a go-around. Between 3.4 and 5.1 seconds later, with the aircraft at a radio altitude of approximately 125 ft agl, in a position approximately 1.5 nm short of the runway, the Enhanced Ground Proximity Warning System (EGPWS) "TERRAIN AHEAD, PULL UP" audio warning was triggered. The correct emergency pull-up procedure was not followed in full, partly because the handling pilot had already initiated a go-around. The minimum recorded terrain clearance achieved during the recovery maneuver was 121 ft. One further non-precision approach to Runway 36 was attempted using selected autopilot modes*	SAMI detects a dangerous situation, because the confusion between the pilots is evaluated by PAM as *low-attention* and *far* focus. But the two warning signals from our SAMI model provided an important advance in time and avoid confusion between the two pilots. For this reason this step simply would not happen if a SAMI was present in the aircraft.

What made possible to interrupt the chain of events during the case is clearly the decision of the crew to stop the descent at 180ft (about 60 meters from the ground). In this case the only support system was identifiable in the EGPWS, which generates an alert at 125ft, with the AM aircraft located about 2km from the runway. The airplane was down about 55ft in roughly 4.5 seconds, and then considering the projection of a similar rate the aircraft would have been found to impact with the ground after about 13.5 seconds from the time when the alert signal is activated from EGPWS. Consider also that the go-around maneuver resulted in an initial decrease of altitude of 60ft and was started at 180 ft, if it was initiated activation signal EGPWS you can assume a minimum of 65ft and thus a time latency of the decision more than 4 seconds could lead to the impact with the ground even in the presence of a signal EGPWS correct and appropriate avoidance maneuver.

In this case the intervention of the SAMI instead would rise an alert prior to the signal from EGPWS and in strict adherence with the company's Standard Operating Procedures, thus providing also the correct, safe, execution of the missed approach procedure (go-around). SAMI provides an important advance in time (even if only a few seconds) but also acts as a disruptive element in chain of events, providing an important structural support to the process of situation awareness and decision-making. Furthermore we want to underline a second safety improvement related to

SAMI: the ability to discriminate the operational environment in which the aircraft is operating, and to relate it with the pilot's S.A. this can be evaluated in step 3, where the system shifts from *normal* to *critical* operational situation; the comparison between the situation detected and the current pilot focus or level of attention could lead to an alert well ahead the reaching of MDA, thus reducing in a significant manner the stress and the workload caused by the interruption of the descent path and of the subsequent missed approach procedure.

4 Conclusion

In this paper we presented the model of an infrastructure which integrates intelligent agents in order to monitor in real time the attention paid by aviation pilots during training/operative flight missions, to make the decision process easier and increase Situation Awareness (SA). To achieve this goal in our work we proceeded, from the methodological point of view, reversing the terms of the problem. In other words, we used the most advanced technology to build an agents-based infrastructure to interpret the reality in which the pilot is set. The model of the infrastructure, we called Situation Awareness Monitoring Infrastructure (SAMI), is based on intelligent agents, that cooperating with each other act like a virtual co-pilot in order to augment capabilities of the real pilot, enhancing his SA, and to maintain and recover pro-actively its attention. Starting from an event actually occurred, a case study scenario was given to prove the enhancement given by SAMI in pilot's SA and thus in flight safety.

References

1. Flight Safety Foundation: Flight safety digest, `http://flightsafety.org/aerosafety-world-magazine/`
2. Harris, D., Helen, C.: Muir: Contemporary issues in human factors and aviation safety. Ashgate Publishing, Ltd. (2005)
3. CAA, P-NPA 25-310, Issue 1 Human Centered Design Requirements. Gatwick, England, UK: UK Civil Aviation Authority (April 2000)
4. AIA/AECMA: Project Report, Propulsion system malfunction plus inappropriate crew response (PSM+ICR). Tech. Rep. 1 (1998)
5. Singer, G., Dekker, S.: The ergonomics of flight management systems: Fixing holes in the certification net. Applied Ergonomics 32(3), 247–254 (2001)
6. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., MacIntyre, B.: Recent advances in augmented reality. IEEE Computer Graphics and Applications 21(6), 34–47 (2001)
7. Kobus, D.A., St John, M., Morrison, J.G., Schmorrow, D.: Overview of the darpa augmented cognition technical integration experiment. International Journal of Human-Computer Interaction 17(2), 131–149 (2004)
8. Schmorrow, D.D.: Foundations of Augmented Cognition. Lawrence Erlbaum, Mahwah (2005)

9. Schmorrow, D.D., Kruse, A.: Improving Human Performance Through Advanced Cognitive System Technology. In: A1 LCDR MSC USN, Defense Advanced Research Projects Agency. A2 Strategic Analysis Inc, Arlington, VA (2005)
10. D. A. Kobus, M. St. John, M. R. Risser.: A real-time closed-loop system for predicting and counteracting lapses of attention. Pacific Science And Engineering Group Inc., San Diego CA, USA, Final Tech. Rep. (2008)
11. Zadeh, L.A.: Fuzzy sets. Information and Control 8, 338–353 (1965)
12. Zadeh, L.A.: Fuzzy Sets, Fuzzy Logic, Fuzzy Systems. World Scientific, Singapore (1996)

Applications of Functional Near Infrared Imaging: Case Study on UAV Ground Controller

Kurtulus Izzetoglu[1], Hasan Ayaz[1], Justin Menda[1], Meltem Izzetoglu[1], Anna Merzagora[1], Patricia A.Shewokis[1,2], Kambiz Pourrezaei[1], and Banu Onaral[1]

[1] School of Biomedical Engineering, Science & Health Systems
[2] College of Nursing and Health Professions, Drexel University
Drexel University, Philadelphia, PA, U.S.A
{kurtulus.izzetoglu,ayaz,jm973,meltem,am396,shewokis, kambiz,banu.onaral}@drexel.edu

Abstract. Functional Near-Infrared (fNIR) spectroscopy is an emerging optical brain imaging technology that enables assessment of brain activity through the intact skull in human subjects. fNIR systems developed during the last decade allow for a rapid, non-invasive method of measuring the brain activity of a subject while conducting tasks in realistic environments. This paper introduces underlying principles and various fNIR designs currently applied to real-time settings, such as monitoring Unmanned Aerial Vehicle (UAV) operator's expertise development and cognitive workload during simulated missions.

Keywords: Near-infrared spectroscopy, optical brain imaging, fNIR, human performance assessment.

1 Introduction

Near infrared spectroscopy (NIRS) has been increasingly applied for the noninvasive measurement of changes in the relative ratios of oxygenated hemoglobin (oxy-Hb) and deoxygenated hemoglobin (deoxy-Hb) during brain activation. In the late 1980s, Delpy designed and tested an NIRS instrument on newborn heads in neonatal intensive care [1]. In the late 1980s and early 1990s, Dr. Britton Chance and his colleagues, using pico-second long laser pulses, spearheaded the development of time-resolved spectroscopy techniques in an effort to obtain quantitative information about the optical characteristics of the tissue [2]. These efforts by Chance, Delpy [3] and others [4], expedited the translation of NIRS based techniques into a neuroimaging modality for various cognitive studies [5-9]. Based on the NIRS technique, the Drexel Optical Brain Imaging team has developed a functional brain monitoring prototype, called fNIR. The portable fNIR system enables the study of cortical cognition-related hemodynamic changes in various field conditions.

Neural activity has a direct relation with hemodynamic changes in the brain [10]. Research on brain-energy metabolism has elucidated the close link between hemodynamic and neural activity [11]. Traditional neuroimaging techniques, such as fMRI cannot be used to measure these hemodynamics for a variety of real-life applications that could yield important discoveries and lead to novel uses. By

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 608–617, 2011.

contrast, the fNIR system can be deployed to assess hemodynamic responses and help understand human brain activation by providing neurophysiological markers derived from neural responses to different experimental settings under field conditions. However, research should be conducted to establish the validity of the fNIR signal as well as to demonstrate the acquisition of accurate and viable signals under real-life conditions. Hence, in addition to general review of underlying principles and various fNIR designs currently applied to real-time settings, this paper also introduces the deployment of this emerging fNIR device to human performance assessment, such as monitoring the changes in UAV operator's level of expertise during simulated missions.

1.1 Physiological Principles of fNIR in Brain Activity Assessment

Understanding the brain energy metabolism and associated neural activity is important for realizing principles of fNIR spectroscopy in assessing brain activity. The brain has small energy reserves and the great majority of the energy used by brain cells is for processes that sustain physiological functioning [12]. Ames III [12] reviewed the studies on brain energy metabolism as related to function and reported that the oxygen (O_2) consumption of the rabbit vagus nerve increased 3.4-fold when it was stimulated at 10 Hz and O_2 consumption in rabbit sympathetic ganglia increased 40% with stimulation at 15 Hz. Furthermore, glucose utilization by various brain regions increased several fold in response to physiological stimulation or in response to pharmacological agents that affect physiological activity [12]. These studies provide clear evidence that large changes occur in brain energy metabolism in response to changes in activity. Moreover, based on this brain energy metabolism, methods and imaging modalities that measure deoxy-Hb and/or oxy-Hb, such as fNIR and fMRI, are implemented to provide correlates of brain activity through oxygen consumption by neurons. Because oxy-Hb and deoxy-Hb have characteristic optical properties in the visible and near-infrared light range, the change in concentration of these molecules during increase in brain activation can be measured using optical methods.

1.2 Physical Principles of fNIR in Brain Activity Assessment

Most biological tissues are relatively transparent to light in the near infrared range between 700-900 nm, largely because water, a major component of most tissues, absorbs very little energy at these wavelengths (Fig. 1). Within this window the spectra of oxy- and deoxy-hemoglobin are distinct enough to allow spectroscopy and measures of separate concentrations of both oxy-Hb and deoxy-Hb molecules [13]. This spectral band is often referred to as the 'optical window' for the non-invasive assessment of brain activation [14].

If wavelengths are chosen to maximize the amount of absorption by oxy-Hb and deoxy-Hb, changes in these chromophore concentrations cause alterations in the number of absorbed photons as well as in the number of scattered photons that leave the scalp. These changes in light intensity measured at the surface of the scalp are quantified using a modified Beer–Lambert law, which is an empirical description of optical attenuation in a highly scattering medium [3]. By measuring absorbance/scattering changes at two (or more) wavelengths, one of which is more sensitive to oxy-Hb and the other to deoxy-Hb, changes in the relative concentration of these chromophores can be calculated. Using these principles, researchers have demonstrated that it is possible to assess hemodynamic changes in response to brain activity through the intact skull in adult human subjects [15-19].

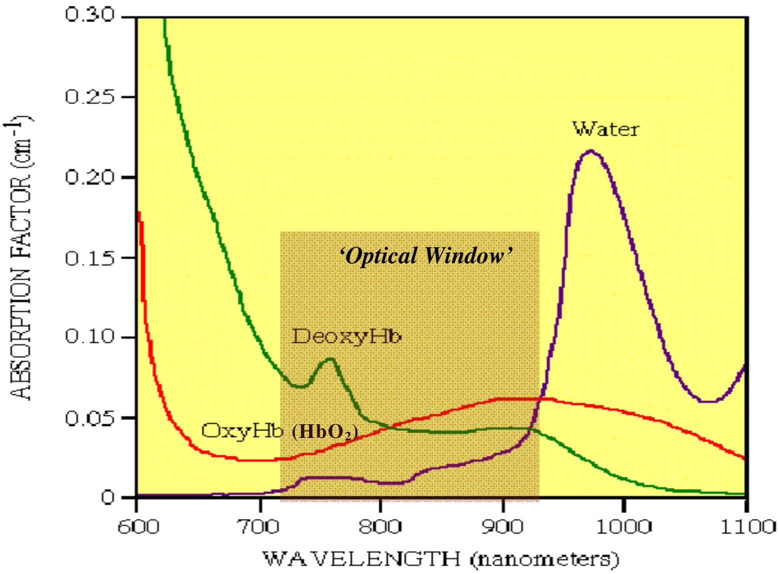


Fig. 1. Absorption spectrum in NIR window: spectra of oxy-Hb and deoxy-Hb in the range of 700 to 900 nm allow spectroscopy methods to assess oxy-Hb and deoxy-Hb concentrations, whereas water absorption becomes substantial above 900 nm, and thus majority of photons are mainly absorbed by water [13].

Typically, an optical apparatus consists of a light source by which the tissue is radiated and a light detector that receives light after it has interacted with the tissue. Photons that enter tissue undergo two different types of interaction, namely absorption and scattering. According to the modified Beer-Lambert Law [13], the light intensity after absorption and scattering of the biological tissue is expressed by the equation:

$$I=GI_oe^{-(\alpha_{HB}C_{HB}+\alpha_{HBO2}C_{HBO2})*L} \tag{1}$$

where G is a factor that accounts for the measurement geometry and is assumed constant when concentration changes. I_o is input light intensity, α_{HB} and α_{HBO2} are the molar extinction coefficients of deoxy-Hb and oxy-Hb, C_{HB} and C_{HBO2} are the concentrations of chromophores, deoxy-Hb and oxy-Hb respectively, and L is the photon path which is a function of absorption and scattering coefficients μ_a and μ_b.

By measuring optical density (OD) changes at two wavelengths, the relative change of oxy- and deoxy-hemoglobin versus time can be obtained. If the intensity measurement at an initial time is I_b (baseline), and at another time is I, the OD change due to variation in C_{HB} and C_{HBO_2} during that period is:

$$\Delta OD = \log_{10}\frac{I_b}{I} = (\alpha_{HB}\Delta C_{HB} + \alpha_{HBO_2}\Delta C_{HBO_2})L \tag{2}$$

Measurements performed at two different wavelengths allow the calculation of ΔC_{HB} and ΔC_{HBO_2}. Change in oxygenation and blood volume or total hemoglobin (Hbt) can then be deduced:

$$Oxygenation = \Delta C_{HBO_2} - \Delta C_{HB} \tag{3}$$

$$BloodVolume = \Delta C_{HBO_2} - \Delta C_{HB} \tag{4}$$

1.3 Near-Infrared Spectroscopy Based Brain Imaging Systems

The combined efforts of the researchers [3, 4, 15] led to the development of three distinct NIRS implementations, namely, time resolved spectroscopy (TRS), frequency domain and continuous wave (CW) spectroscopy [20]. In TRS systems, extremely short incident pulses of light are applied to tissue and the temporal distribution of photons that carry the information about tissue scattering and absorption is measured. In frequency domain systems, the light source is amplitude modulated with frequencies in the order of tens to hundreds of megahertz. The amplitude decay and phase shift of the detected signal with respect to the incident are measured to characterize the optical properties of tissue. In CW systems, light is continuously applied to tissue at constant amplitude. The CW systems are limited to measuring the amplitude attenuation of the incident light[20].

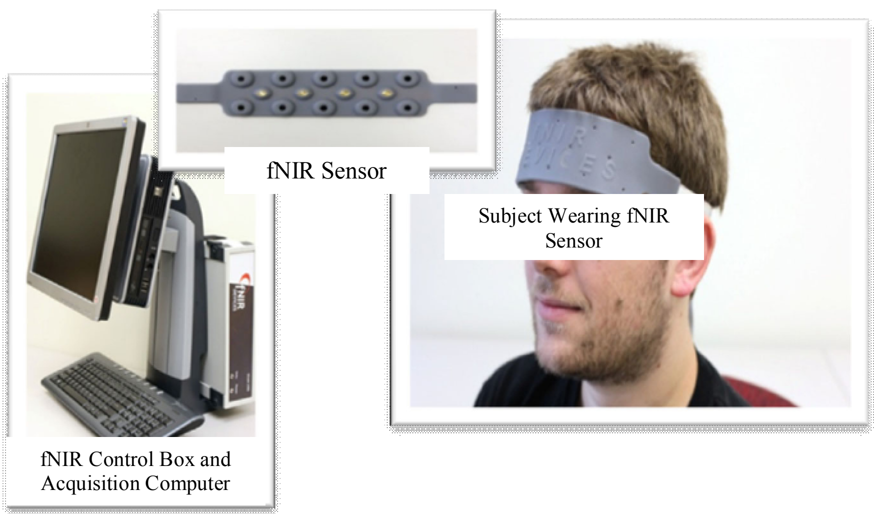


Fig. 2. Overview of the continuous wave 16-channel fNIR system

CW systems have a number of advantageous properties that have resulted in wide use by researchers interested in brain imaging relative to other near-infrared systems; it is minimally intrusive and portable, affordable, and easy to engineer relative to frequency and time domain systems [21]. These CW systems hold enormous potential for research studies and clinical applications that require the quantitative measurements of hemodynamic changes during brain activation under ambulant conditions in natural environments.

The continuous wave fNIR system used in this study was originally described by Chance et al. [15]. The current generation, flexible headband sensor developed in the Drexel's Optical Brain Imaging laboratory, consists of 4 LED light sources and 10 detectors (Figure 2).

The fNIR sensor, illustrated in Fig. 2, reveals information in localizing brain activity, particularly in dorsolateral prefrontal cortex. Fig. 3 shows the spatial map of the 16-channel fNIR sensor on the curved brain surface, frontal lobe [22].

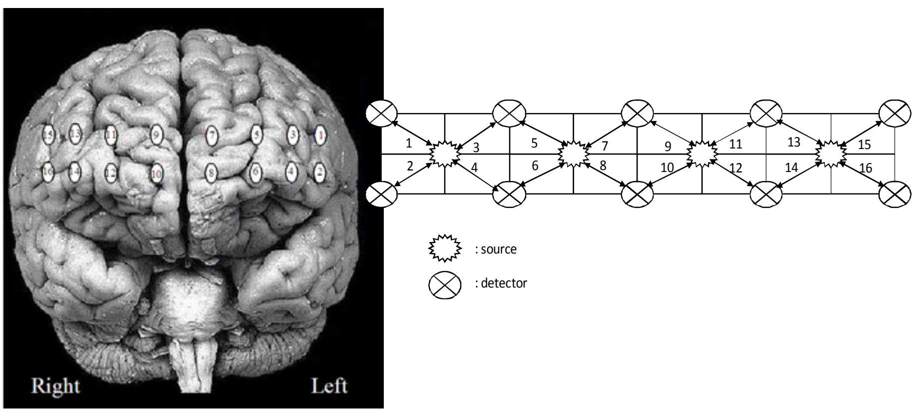


Fig. 3. Spatial map of the 16-channel fNIR sensor on the curved brain surface, frontal lobe

2 Method: Brain Activity Monitor during UAV Simulations

A 16-channel CW- fNIR system (Fig. 2) has been utilized to monitor the prefrontal cortex during simulated missions. An integrated simulation environment is constructed to allow novice participants to operate a simulated MQ-1 Predator UAV (Fig. 4). Missions and scenarios have been developed to represent a variety of tasks typical of UAV training and Predator operations, such as visual search/target categorization tasks and flight maneuvers [23].

To run the simulation reliably and with a high degree of realism, high performance hardware is specified, including an Intel Core i7 925 CPU and an nVidia GeForce GTX 280 graphics processor. The simulation is presented on a triple-display system by Digital Tigers, using 19" LCD monitors with 4:3 aspect ratios in a horizontal configuration (Fig. 5).

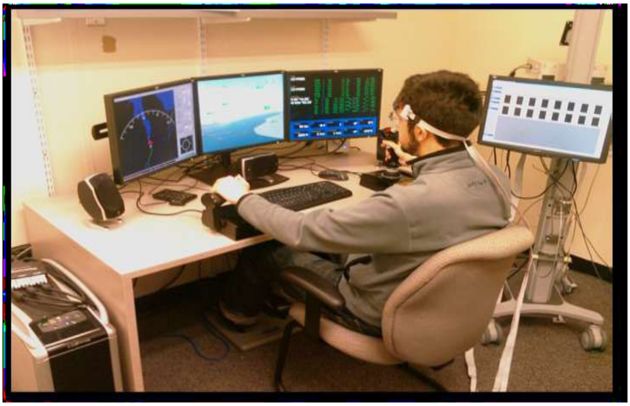


Fig. 4. Subject operating the Predator UAV simulator with fNIR sensor attached and data acquisition apparatus on far right

Fig. 5. Screenshot of flight simulation interface with Predator UAV add-on

Subjects control the simulated Predator UAV using a Thrustmaster HOTAS Cougar joystick-and-throttle system and a CH Pro Pedals rudder pedal system. FS Recorder, an add-on for Flight Simulator X, is implemented to record behavioural data during the simulated flights [23]. The time synchronization between fNIR recording and task events is facilitated by a custom application implemented to send event markers to the fNIR data acquisition computer via RS232.

2.1 Experimental Procedure

Prior to the study, all participants signed informed consent statements approved by the Human Subjects Institutional Review Board at Drexel University and by the U.S. Army Medical Research and Materiel Command (USAMRMC), Office of Research Protections (ORP), Human Research Protection Office (HRPO). The flight scenarios have been designed to represent a variety of tasks required of UAV operators (coordinate-based navigation, landing, visual search/target categorization, etc.) and to incorporate workload factors (e.g. crosswinds, cloud cover, fuel constraints, etc.). After an "introduction" session for the purpose of familiarization with the protocol and simulation, subjects fly these scenarios during eight subsequent flight sessions. Each subject performs one flight session per day, each lasting approximately two hours, for a total of 18 hours over 9 days per subject.

In the first session, after being given an overview of the experiment and providing informed consent, each subject completes the Edinburgh Handedness Inventory and a brief questionnaire regarding previous flight and video game experience. Then, the fNIR sensor is attached and subjects perform an intro flight of up to 1.5 hours, during which they are introduced to the UAV simulation. In sessions 2 through 9, each subject attempts one or more of the flight scenarios, with the fNIR sensor attached to the forehead and gathering data during the flight. At the end of each session, a confidence survey and the NASA-TLX are administered to allow subjects to self-rate overall performance.

2.2 Data Acquisition

Throughout the entire sessions, the following physiological and behavioral data were collected; i. fNIR sensor recordings acquired at every half second; ii. events including position, orientation, and velocity of the simulated aircraft in all three axes, and positions of all flight controllers (joystick, rudder pedals, throttle) recorded at 1/8 seconds intervals. A flexible fNIR sensor pad (Figs. 2 & 3) hosting 4 light sources with built in peak wavelengths at 730 nm and 850 nm is placed over subject's forehead to scan cortical areas. With a fixed source-detector separation of 2.5 cm, this configuration generates a total of 16 measurement locations per wavelength. For data acquisition and visualization, COBI Studio software (©2010, Drexel University) was used. Raw light intensity measures were low-pass filtered with a finite impulse response, linear phase filter with order of 20 and cut-off frequency of 0.1Hz to attenuate the high frequency noise. Using filtered raw fNIR measures, we calculated oxy-Hb, deoxy-Hb and blood volume (totalHb) using the formulas 1,2,3, and 4.

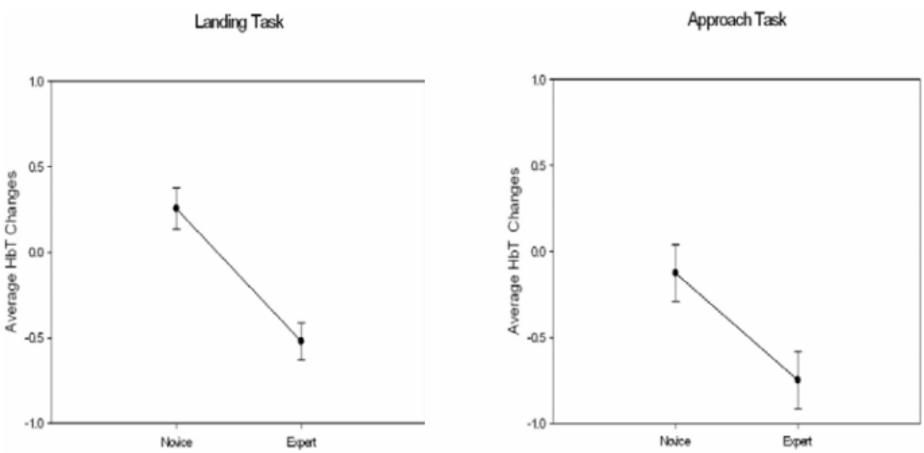


Fig. 6. Average blood volume (totalHb) changes in channel 2 during the transition from novice to expert. Left: Flight landing, and Right: Flight approach tasks. Error bars represent the standard error of the mean.

3 Results

For statistical analysis, repeated-measures ANOVA was used for the fNIR data to compare within subject factor of two expertise levels: novice (sessions 2, 3 & 4) versus expert (sessions 7, 8 & 9). The significance criterion for the tests was $\alpha = 0.05$. The same statistical analyses were performed for the flight approach and landing tasks.

There is a significant decrease in totalHb during the transitioning from novice to expert for both flight landing task ($F(1,4) = 13.00$; $p<0.005$) and flight approach task ($F(1,4) = 9.22$; $p<0.005$) (Fig. 6). The results also reveal that channel 2 (see Fig.3 for the spatial mapping) is the significant activation location. This area, left inferior frontal gyrus, was also reported to be sensitive to working memory by Ayaz et al [24] in the cognitive workload monitoring study for the air traffic controllers.

4 Discussion

This paper introduces a case study with a preliminary finding that fNIR, a portable optical brain imaging system, can monitor changes in level of expertise by measuring activation in the prefrontal areas relative to task performance. Decrease in the fNIR measures, shown in Fig. 6, is significant and a valid hypothesis can be derived from the evidence that expertise tends to be associated with overall lower brain activity relative to novices, particularly in prefrontal areas [25]. Both practice and the development of expertise typically involve decreased activation across attentional and control areas, freeing these neural resources to attend to other incoming stimuli or task demands. As such, measuring activation in these attentional and control areas relative to task performance can provide an index of level of expertise and illustrate how task-specific practice influences the learning of tasks. The differences in activation of the attentional and control regions of the prefrontal cortex may also indicate neural plasticity as a function of task-specific practice [26].

In summary, a field deployable optical brain imaging (fNIR) holds enormous potential for research studies and clinical applications that require the quantitative measurements of hemodynamic changes during brain activation under ambulant conditions in natural environments. As such, fNIR has been already deployed in many field settings for objective measurements of cognitive state and expertise development which will, among other advantages, allow for dynamic interventions in the training process, and helping to assure robust performance under adverse circumstances. Other fNIR application areas include, but are not limited to, brain computer interface for cognitive enhancement, neurological and gaming applications, pediatric solutions, education and training and cognitive aging.

Acknowledgments. The U.S. Army Medical Research Acquisition Activity, 820 Chandler Street, Fort Detrick, MD 21702-5014 is the awarding and administering acquisition office. This investigation was funded under a U.S. Army Medical Research Acquisition Activity; Cooperative Agreement W81XWH-08-2-0573. The content of the information herein does not necessarily reflect the position or the policy of the U.S. Government or the U.S. Army and no official endorsement should be inferred.

References

1. Delpy, D.T., Cope, M.C., Cady, E.B., Wyatt, J.S., Hamilton, P.A., Hope, P.L., Wray, S., Reynolds, E.O.: Cerebral monitoring in newborn infants by magnetic resonance and near infrared spectroscopy. Scand J. Clin. Lab. Invest Suppl. 17, 9–17 (1987)
2. Patterson, M.S., Chance, B., Wilson, B.C.: Time resolved reflectance and transmittance for the non-invasive measurement of tissue optical properties. Appl. Opt. 28, 2331–2336 (1989)
3. Cope, M., Delpy, D.T.: System for long-term measurement of cerebral blood and tissue oxygenation on newborn infants by near infra-red transillumination. Medical & Biological Engineering & Computing 26, 289–294 (1988)
4. Fishkin, J.B., Gratton, E.: Propagation of photon-density waves in strongly scattering media containing an absorbing semi-infinite plane bounded by a straight edge. J. Opt. Soc. Am. A 10, 127–140 (1993)
5. Okada, F., Takahashi, N., Tokumitsu, Y.: Dominance of the nondominant hemisphere in depression. J. Affect Disord. 37, 13–21 (1996)
6. Villringer, A., Chance, B.: Non-invasive optical spectroscopy and imaging of human brain function. Trends in Neurosciences 20, 435–442 (1997)
7. Obrig, H., Villringer, A.: Near-infrared spectroscopy in functional activation studies. Can NIRS demonstrate cortical activation? Advances in Experimental Medicine and Biology 413, 113–127 (1997)
8. Boas, D.A., Gaudette, T., Strangman, G., Cheng, X., Marota, J.J.A., Mandeville, J.B.: The accuracy of near infrared spectroscopy and imaging during focal changes in cerebral hemodynamics. NeuroImage 13, 76–90 (2001)
9. Izzetoglu, K., Bunce, S., Onaral, B., Pourrezaei, K., Chance, B.: Functional optical brain imaging using near-infrared during cognitive tasks. International Journal of Human-Computer Interaction 17, 211–227 (2004)
10. Kruggel, F., von Cramon, D.Y.: Temporal properties of the hemodynamic response in functional MRI. Hum Brain Mapp 8, 259–271 (1999)
11. Magistretti, P.J.: Cellular bases of functional brain imaging: insights from neuron-glia metabolic coupling. Brain Res. 112, 108–112 (2000)
12. Ames, A.: CNS energy metabolism as related to function. Brain Res. Rev. 34, 42–68 (2000)
13. Cope, M.: The application of near infrared spectroscopy to non-invasive monitoring of cerebral oxygenation in the newborn infant. vol. Ph.D. University of London (1991)
14. Jobsis, F.F.: Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. Science 198, 1264–1267 (1977)
15. Chance, B., Zhuang, Z., Unah, C., Alter, C., Lipton, L.: Cognition-Activated Low-Frequency Modulation of Light-Absorption in Human Brain. P. Natl. Acad. Sci. 90, 3770–3774 (1993)
16. Gratton, G., Corballis, P.M., Cho, E., Fabiani, M., Hood, D.C.: Shades of gray matter: noninvasive optical images of human brain responses during visual stimulation. Psychophysiology 32, 505–509 (1995)
17. Hoshi, Y., Tamura, M.: Dynamic multichannel near-infrared optical imaging of human brain activity. J. Appl. Physiol. 75, 1842–1846 (1993)
18. Kato, T., Kamei, A., Takashima, S., Ozaki, T.: Human visual cortical function during photic stimulation monitoring by means of near-infrared spectroscopy. J. Cereb. Blood Flow Metab. 13, 516–520 (1993)

19. Villringer, A., Planck, J., Hock, C., Schleinkofer, L., Dirnagl, U.: Near infrared spectroscopy (NIRS): a new tool to study hemodynamic changes during activation of brain function in human adults. Neurosci. Lett. 154, 101–104 (1993)
20. Strangman, G., Boas, D.A., Sutton, J.P.: Non-invasive neuroimaging using near-infrared light. Biol. Psychiatry 52, 679–693 (2002)
21. Chance, B., Anday, E., Nioka, S., Zhou, S., Hong, L., Worden, K., Li, C., Murray, T., Ovetsky, Y., Pidikiti, D., Thomas, R.: A novel method for fast imaging of brain function, non-invasively, with light. Opt. Express 2, 411–423 (1998)
22. Ayaz, H., Izzetoglu, M., Platek, S.M., Bunce, S., Izzetoglu, K., Pourrezaei, K., Onaral, B.: Registering fNIR data to brain surface image using MRI templates. Conf. Proc. IEEE Eng. Med. Biol. Soc. 1, 2671–2674 (2006)
23. Menda, J., Hing, J.T., Ayaz, H., Shewokis, P.A., Izzetoglu, K., Onaral, B., Oh, P.: Optical Brain Imaging to Enhance UAV Operator Training, Evaluation, and Interface Development. J. Intell. Robotics Syst. 61, 423–443 (2011)
24. Shewokis, P.A., Izzetoglu, K., Hah, S., Deshmukh, A., Onaral, B.: Cognitive Workload Assessment of Air Traffic Controllers Using Optical Brain Imaging Sensors. In: Marek, T., Karwowski, W., Rice, V. (eds.) Advances in Understanding Human Performance: Neuroergonomics, Human Factors Design, and Special Populations, pp. 21–32. CRC Press, Boca Raton (2010)
25. Milton, J.G., Small, S.S., Solodkin, A.: On the road to automatic: dynamic aspects in the development of expertise. J. Clin. Neurophysiol. 21, 134–143 (2004)
26. Kelly, A.M., Garavan, H.: Human functional neuroimaging of brain changes associated with practice. Cereb Cortex 15, 1089–1102 (2005)

Augmented Phonocardiogram Acquisition and Analysis

Nancy E. Reed[1] and Todd R. Reed[2]

[1] Department of Information and Computer Sciences
[2] Department of Electrical Engineering
University of Hawaii at Manoa
Honolulu, Hawaii, 96822, USA
{nreed,trreed}@hawaii.edu

Abstract. Heart auscultation (the interpretation of heart sounds by a physician) is a widely used screening method for heart disease. It is well documented, however, that with the exception of expert cardiologists, physicians' auscultation skills are limited. It has also been shown that standard training methods do little to improve these skills. In this paper, we propose an architecture for a phonocardiogram analysis system that can augment a physician's auscultation abilities and serve as a training aid to improve those abilities.

1 Introduction

This paper presents a system to augment a physician's evaluation of a patient for heart disease, or to serve as a stand-alone screening system in countries where access to physicians is limited.

Analyzing heart sound recordings (phonocardiograms or PCGs) requires the detection of various sounds produced by the heart, while removing or ignoring the sounds caused by other factors (patient motion, speech, etc). Most physicians and other health care workers are not able to distinguish normal sounds from ones indicating disease. The less than desirable auscultatory performance of non-specialists (those that are not cardiologists) has been well documented [6,9,13,15,25].

Our goal is to provide both diagnostic advice and a training aid. Our system has the potential for use in developing countries due to simplicity of operation and low cost.

Previous work has focused on phonocardiograms taken from a single site (location on the chest), and often acquired under carefully controlled, non-clinical conditions. This is not how cardiologists perform auscultation. Because different locations provide cues to different pathologies, multiple sites are examined. Clinically acquired sounds often include noise and other artifacts that make analysis challenging. These must be dealt with by any system intended for clinical use. For these reasons, we base this work on simultaneously acquired multi-site phonocardiograms, acquired under clinical conditions.

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 618–627, 2011.

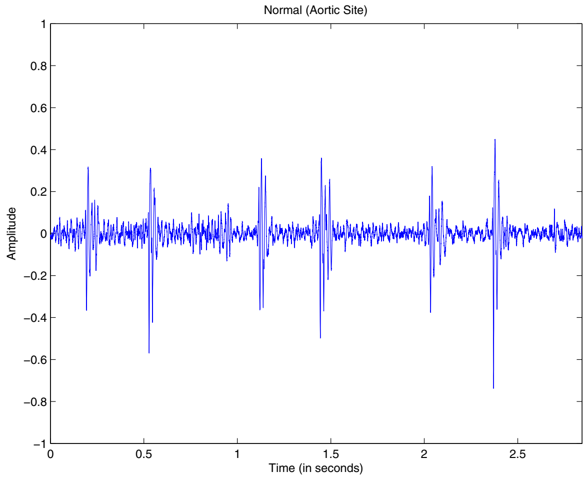


Fig. 1. The acoustic signal from a normal heart over three cardiac cycles. The peaks show the first and second heart sounds, respectively.

The rest of this paper is organized as follows. The next section describes the heart sound signal, and the signal processing and analysis methods that form the basis of symptom detection. The next section describes the design of software to complement the signal processing and analysis in a decision-support system. The last section contains a summary of this work.

2 Signal Processing, Analysis and Symptom Detection

The *cardiac cycle* refers to the events related to the flow of blood that occurs from the beginning of one heartbeat to the beginning of the next [26]. Every 'beat' of the heart involves two major phases, systole (ventricular contraction) and diastole (ventricular relaxation). Heart sounds S_1 and S_2 mark the beginning of systole and diastole respectively (Figure 1). The phonocardiogram shown was acquired from a patient in the supine position, from the aortic location on the chest. The sampling rate is 8kHz, with 16 bits/sample.

S_1 is associated with the closure of the mitral and tricuspid valves. In the vast majority of cases, the valves close nearly simultaneously, so that the individual closures are not discernible in the S_1. S_2 is primarily due to the closure of the aortic and pulmonary valves. These valves may close at significantly different times, such that the individual events may be audible. The resulting S_2 is referred to as "split", with subcomponents A_2 (aortic) and P_2 (pulmonary). The time between these subcomponents and their relative amplitudes can be diagnostically important.

Figure 2 shows a computer generated figure representing two (non consecutive) heart cycles. S_1 is the same in both sections of the figure. The distance

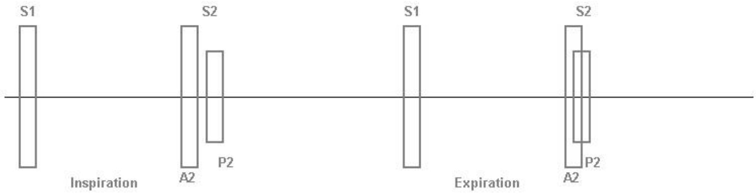


Fig. 2. A computer-generated idealized image of a normal person's heart sounds displaying 2 heart cycles, inspiration on the left and expiration on the right. Vertical bars indicate S_1 and S_2.

between the components of S_2, A_2 and P_2 (associated with the aortic and pulmonary valves respectively) change. On the left, A_2 and P_2 are separated by a small amount, while on the right, they are overlapping. This is normal, and called a physiological split. The change in S_2 is produced by a difference in pressure on the two sides of the heart that occur due to breathing. If the width of the split doesn't change over time, if it is wider than shown, or if the two components do not have the relative difference in volume indicated, it is a symptom of disease.

Two additional heart sound components are referred to as S_3 and S_4. The S_3 is associated with rapid ventricular filling. When present, it appears shortly after the S_2. Often associated with innocent murmurs in children, when heard in patients over the age of 35, it is a sign of pathology. The S_4, associated with late diastolic filling, occurs shortly before the S_1 (when present). An audible S_4 is always an indicator of abnormality.

Murmurs are another category of sounds with diagnostic utility. Murmurs may appear in systole, diastole or both (in which case they are referred to as continuous) They appear as segments of increased (noise-like) activity over extended periods of time. There are also innocent murmurs (that do not indicate pathology) that must be identified as distinct from the murmurs which indicate disease.

Finally, very short, impulsive sounds sometimes called *clicks* and *snaps* may be heard. Clicks are associated with mitral (and possibly tricuspid) valve closures and occur shortly after S_1. An opening snap sometimes accompanies mitral stenosis and is heard after S_2. When audible they are always a sign of underlying abnormality.

Physicians listen primarily to six locations on the torso, as shown by the large and small circles in Figure 3. Using multi-site auscultation, all six sites are acquired simultaneously from pediatric patients under clinical conditions. Diagnostically important information is contained in the differences among heart sounds when the patient is inhaling compared to exhaling. Trying to align breathing as well as cardiac cycles in multiple signals recorded at different times on the same patient has proven exceedingly difficult.

The multi-site signals are recorded simultaneously so that (within propagation time differences) they are time-aligned (Figure 4). The signals are acquired at the locations indicated in Figure 3 and are titled accordingly. Pulse and ECG

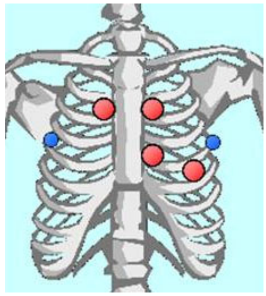


Fig. 3. The four primary sites used in auscultation are shown in large circles in the figure. From left to right and top to bottom they are the Aortic, Pulmonic, Tricuspid, and Mitral. The two smaller circles identify the right and left axilla.

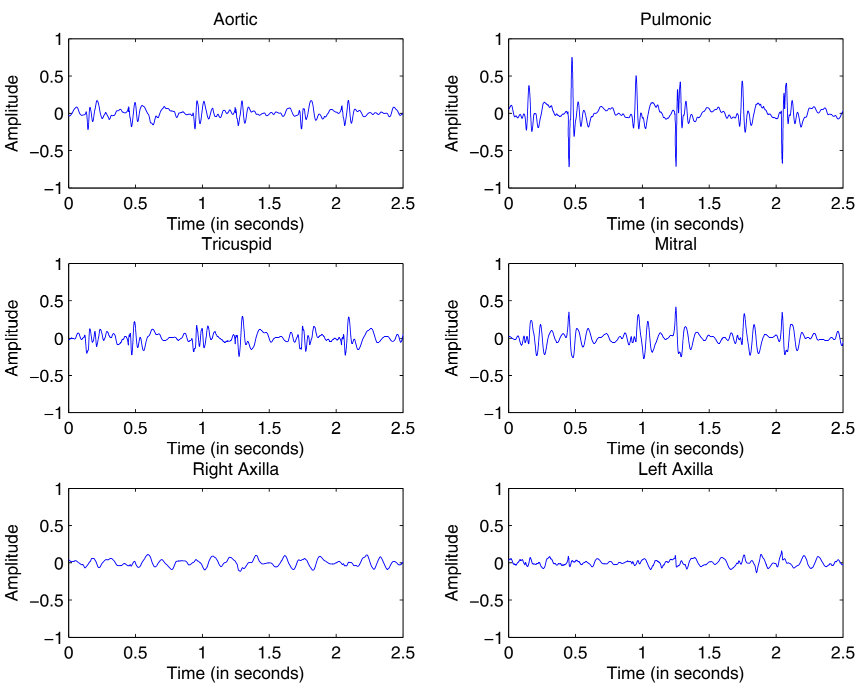


Fig. 4. Heartsounds collected (simultaneously) from the six locations indicated in Figure 3. Note that the signals have distinctively different characteristics, indicating differences in the information emphasized at each site.

signals (not shown) are simultaneously recorded as well. The signals shown are from a patient in the sitting position. The sampling rate is 10 kHz, with 16 bits/sample. The phonocardiograms exhibit a small atrial septal defect (ASD), a breach between the upper two chambers of the heart.

As described in the following sections, the signal processing system performs noise reduction, heart rate estimation, feature extraction, segmentation, labeling and symptom detection on the multi-site signals. The resulting symptoms are passed to the decision support system.

2.1 Noise Reduction

The reduction of noise is very important when designing a system intended for use in a clinical or otherwise difficult to constrain setting (e.g., in the home). A variety of approaches have been considered [16,17,21].

In signal processing, it is often assumed that noise is random (typically with a uniform or Gaussian statistical distribution), with a relatively wideband spectrum. It is also assumed that the desired signal has a spectrum that is largely disjoint from the noise spectrum, and concentrated at low frequencies. When these assumptions are met, noise reduction can be accomplished by simple lowpass filtering.

Phonocardiograms violate both of these assumptions. In addition to the typical random component due, e.g., to electronic noise, there may be deterministic components due to speech, footsteps, and patient motion. We will refer to these noise sources as environmental noise. The spectra of these components may overlap significantly with the desired signal. The desired signal may also contain random components due to turbulent blood flow (referred to as "murmurs"). These components are important for diagnosis, so should be retained.

In addition, the analysis required to establish symptoms has different (and conflicting) requirements for noise reduction. Identifying the heart sound events (S_1 , S_2 and S_3 and S_4 if they exist) is made difficult by the conventional noise, environmental noise, and murmurs. For this task, then, it is desirable to remove all three. In identifying symptoms, however, as mentioned previously, murmurs are important. There are also events with high frequency components (valve closures) that can be diagnostically relevant.

We therefore propose two parallel approaches to noise reduction: one that removes electronic and environmental noise and murmurs; and a second that removes noise to the degree possible while preserving murmurs and valve closure signatures. In the simplest form, these could be implemented using lowpass filters with different cutoff frequencies. A wavelet based approach, which also facilitates feature selection (Section 2.3) is used currently. More sophisticated methods (mostly nonlinear) are under investigation.

2.2 Heart Rate Estimation

The estimation of heart rate (both average and instantaneous) [7,10] is important in determining the parameters of feature extraction, segmentation and labeling (below). For example, expected relative timing between the S_1 and S_2 and S_2 and the following S_1 is often used to label the S_1 and S_2 components in a phonocardiogram. The relative timing changes with heart rate, and is less reliable at high heart rates (the two intervals approach equality).

Instantaneous heart rate is also important in determining the degree of heart rate variability. This is an important symptom for diagnosing, e.g., arrhythmia.

One approach to finding the average heart rate is to identify maxima in the phonocardiogram spectrum. Time-frequency approaches, discussed briefly in the next section, are also a useful mechanism for estimating instantaneous heart rate.

2.3 Feature Extraction

The features derived from the phonocardiogram are important for identifying components of the heartsound (S_1, S_2, S_3, S_4 and murmurs), the characteristics of the components (e.g., the murmur is rising or falling in amplitude) and for detecting impulsive events (clicks, snaps, and valve closures).

To accomplish these tasks, features must be localized in time. It is also desirable that the features be concise. That is, characteristics of interest should be described in as few features as possible. To facilitate this, the signal representation from which the features are derived should itself be concise (in mathematical terms, "sparse").

Two classes of representations that have been of considerable interest in the signal processing community are the time-frequency [2,4] and time-scale (wavelet) [12,19] representations. For the current application, both approaches have merit. Wavelet representations have proven especially suitable for detecting impulsive events (e.g., valve closures). Time-frequency approaches are particularly suitable for determining, e.g., heart rate variability. Both have promise for the overall segmentation task (Section 2.4).

An additional process may be applied to make the signal description even more concise. Principle component analysis (PCA) is widely used. Independent component analysis (ICA) may also be used.

In our current investigations, we use a wavelet decomposition (based on the Daubechies 15 wavelet) followed by principle component analysis.

2.4 Segmentation and Labeling

Different heartsound components occur at different times in the cardiac cycle. S_1 and S_2 occur periodically, at intervals predicted by the heart rate. This is also true of S_3 and S_4 if they exist. Murmurs may occur throughout the cardiac cycle. The expected interval between murmurs in successive cycles is also often indicated by the heart rate.

Segmentation [1,3,8,12,19] involves the detection (but not identification) of heartsound components. This typically involves examining features (as described above) to look for time intervals with high activity. This task is complicated by the fact that murmurs may overlap other components. It is desirable to extract the murmurs, so that they can be analyzed separately.

Labeling consists in identifying the heart sound components (e.g., as S_1, S_2, murmurs, etc.) The multisite approach proposed in this work is of significant benefit in this effort, since different locations have different characteristics for certain components (such as S_1 and S_2). This allows labeling in cases of high heart rate, where relative timing is not reliable. Once heart sound components are identified and labeled, component-specific symptoms can be identified. An example is the time between the aortic and pulmonic components of the S_2, referred to as the S_2 "split".

2.5 Symptom Detection

Common structural heart diseases include aortic stenosis (AS), atrial septal defect (ASD), pulmonary stenosis (PS), ventricular septal defect (VSD), mitral insufficiency (MI), tricuspid insufficiency (TI) and tetralogy of Fallot (TF). In ASD and VSD, there is a communication (hole) between the upper and lower chambers of the heart, respectively. AS and PS are abnormal constrictions near the aortic and pulmonary valves, respectively. MI and TI indicate reverse (backward) blood flow through the mitral and tricuspid valves. TF consists of PS and VSD.

Identifying which (if any) of the diseases described above is the focus of the Decision Support system. Detecting symptoms that can lead to a diagnosis is the ultimate goal of the Signal Processing, Analysis and Symptom Detection system.

For example, for the case of an ASD (illustrated in Figure 4), symptoms include wide, fixed splitting of the S_2. A mid-diastolic murmur is also common [18].

3 Decision Support

Decision-support for cardiac auscultation can focus on one or more tasks [14]. He describes two primary tasks, screening innocent from pathologic heart sounds, and differentiating among diseases and sub-types. The latter involves a detailed examination of case data with respect to expectations for specific diseases. Other tasks may be the focus, for example monitoring the severity of congestive heart failure or monitoring replaced valves.

The information acquired via cardiac auscultation is crucial to determining whether a patient has heart disease. Information from a physical exam, ECG, X-ray, or other tests may also be diagnostically useful. For example, blood pressure, external pulses and whether or not the patient has cyanosis (a bluish look to the skin from low oxygen levels) can suggest some diseases and rule others out.

A prototype system providing decision-support in identifying congenital heart disease is described in [22,23]. This program integrates data from multiple sources. In addition to single diseases, more than one disease may be present in the same patient. In cardiology there may be up to three co-occurring diseases. Multiple diseases may interact, which means that the cues from a case of X and Y are not the union of cues produced by X alone and Y alone. Cues may be missing, added, or altered.

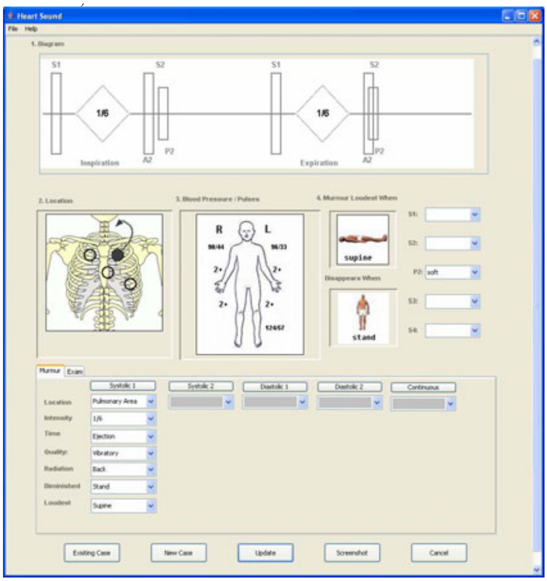


Fig. 5. A screen shot displaying auscultation data in both text and graphic formats

Decision support for an example case, number 0212 follows. The patient was referred to a cardiologist because a murmur was detected during a routine exam. The heart auscultation reveals a normal S_1 and a wide, fixed split S_2. There is a grade I/VI mid-systolic (ejection) murmur heard best in the pulmonic area, and radiating to the back. The murmur does not increase in loudness in the supine position, and does not disappear in a sitting or supine position. There are no clicks or snaps. No diastolic murmur is detected. The ECG shows normal rhythm and no hypertrophy.

The soft systolic ejection murmur detected is characteristic of an innocent (Stills) murmur, an Atrial Septal Defect or Pulmonary Stenosis. A normal S_1 and normal ECG are consistent with all three. The wide, fixed split of S_2 is characteristic of both ASD and PS, but not an innocent murmur. With PS, a systolic ejection click in the pulmonary area is also expected, but one was not observed.

A decision-support system using the explanation points metric [23,22] would report the following on this case. ASD explained all symptoms (the murmur and S_2), e.g. 2/2 abnormal cues or 100%. A Stills murmur could explain the systolic ejection murmur, but could not explain the abnormal S_2, resulting in only 1 of 2 or 50% of the abnormal cues explained. The expected abnormal click for PS is not present, meaning PS can explain only 2 of 3 or 66% of the abnormal cues. Echocardiography, the "gold standard", revealed a small ASD.

It is highly desirable that the results of auscultation be displayed in an *idealized* graphical format in the electronic medical record. Nie, et al. [20,24]

developed a graphical format that is similar to those in medical textbooks such as [5,11,18]. It is thus recognizable by people at various skill levels without having to listen to 12 or more recordings. An example is shown in Figure 5.

The top part of the figure shows the timing and intensity of heart sounds in case 1105. The figure of the chest in the second row shows the location and radiation direction of the murmur. In this case, the murmur is in the Pulmonary area and radiates to the back (indicated with an arrow). The graphics are generated from the text shown in the lower left side of the figure when the case is viewed.

4 Summary

Heart auscultation is a widely used screening method for heart disease. It is well documented, however, that with the exception of expert cardiologists, physicians' auscultation skills are limited and that standard training methods do little to improve these skills. In this paper, we propose an architecture for a phonocardiogram analysis system that can augment a physician's auscultation abilities and serve as a training aid to improve those abilities.

Acknowledgments

This work was supported in part by the Telemedicine and Advanced Technology Research Center's Army Advanced Medical Technology Initiative (AAMTI) program. We would like to thank Dr. C. B. Mahnke for generously providing the multisite phonocardiograms, for Figure 3 and for helpful discussions. Dr. Mahnke is with Tripler Army Medical Center, Pediatrics (Cardiology), Honolulu, HI, USA.

References

1. Ari, S., Saha, G.: On a robust algorithm for heart sound segmentation. Journal of Mechanics in Medicine and Biology 7(2), 129–150 (2007)
2. Boutana, D., Djeddi, M., Benidir, M.: Identification of aortic stenosis and mitral regurgitation by heart sound segmentation on time-frequency domain. In: Proc. 5th Int. Symp. on Image and Signal Proc. and Analysis, Istanbul, pp. 1–6 (2007)
3. Choi, S., Jiang, Z.: Comparison of envelope extraction algorithms for cardiac sound signal segmentation. Expert Systems With Applications 34(2), 1056–1069 (2008)
4. Debbal, S.M., Bereksi-Reguig, F.: Computerized heart sounds analysis. Computers in Biology and Medicine 38(2), 263–280 (2008)
5. Erickson, B.: Heart Sounds and Murmurs: A Practical Guide, Second Edition. Mosby-Year Book Inc. St. Louis, MO (1991)
6. Gaskin, P., Owens, S., Talner, N., Sanders, S., Li, J.: Clinical auscultation skills in pediatric residents. Pediatrics 105(6), 1184–1187 (2000)
7. Godinez, M., Jimenez, A., Ortiz, R., Pena, M.: On-line fetal heart rate monitor by phonocardiography. In: Proc. 25th Int. Conf. of IEEE EMB Soc., vol. 4, pp. 3141–3144 (2003)
8. Gupta, C.N., Palaniappan, R., Swaminathan, S., Krishnan, S.M.: Neural network classification of homomorphic segmented heart sounds. Applied Soft Computing 7(1), 286–297 (2007)

9. Iversen, K., Teisner, A.S., Dalsgaard, M., Greibe, R., Timm, H.R., Skovgaard, L.T., Hrobjartsson, A.: Effect of teaching and type of stethoscope on cardiac auscultatory performance. American Heart Journal 152(1) (2006)
10. Kovács, F., Torok, M., Habermajer, I.: A rule-based phonocardiographic method for long-term fetal heart rate monitoring. IEEE Trans. on Biomedical Engineering 47(15), 124–130 (2000)
11. Lehrer, S.: Understanding pediatric heart sounds, 2nd edn. Saunders (1992)
12. Lin, T.C., Reed, T.R.: Heart sound segmentation for computer-aided auscultation. In: Proc. 7th IASTED Int. Conf. on Signal and Image Proc., Honolulu, pp. 122–127 (2005)
13. Mahnke, C.B., Norwalk, A., Hofkosh, D., Zuberbuhler, J.R., Law, Y.M.: Comparison of two educational interventions on pediatric resident auscultation skills. Pediatrics 113(5), 1331–1335 (2004)
14. Mahnke, C.B.: Automated heartsound analysis/computer-aided auscultation: A cardiologist's perspective and suggestions for future development. In: Proc. 31st Int. Conf. of the IEEE EMB Soc. pp. 3115–3118 (2009)
15. Mangione, S.: Cardiac auscultatory skills of physicians-in-training: a comparison of three English-speaking countries. The Amer. J. of Medicine 110(3), 210–216 (2001)
16. Messer, S.R., Agzarian, J., Abbott, D.: Optimal wavelet denoising for phonocardiograms. Microelectronics Journal 32, 931–941 (2001)
17. Mittra, A.K., Shukla, A., Zadgaonkar, A.S.: System simulation and comparative analysis of foetal heart sound de-noising techniques for advanced phonocardiography. Int. J. of Biomedical Eng. and Tech. 1(1), 73–85 (2007)
18. Moller, J.H.: Essentials of Pediatric Cardiology, 2nd edn. A. Davis Company (1978)
19. Nazeran, H.: Wavelet-based segmentation and feature extraction of heart sounds for intelligent PDA-based phonocardiography. Methods of Inf. in Medicine 46, 135–141 (2007)
20. Nie, Y.: Heart Sounds and Murmurs in an Electronic Medical Record. Master's report, Computer Science Dept, University of Hawaii (2007)
21. Paul, A.S., Wan, E.A., Nelson, A.T.: Noise reduction for heart sounds using a modified minimum-mean squared error estimator with ECG gating. In: Proc. 28th Int. Conf. of the IEEE EMB Soc., pp. 3385–3390 (2006)
22. Reed, N.E.: Constructing the correct diagnosis when symptoms disappear. In: Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI 1998), pp. 151–156 (1998)
23. Reed, N.E., Gini, M., Johnson, P.E., Moller, J.H.: Diagnosing congenital heart defects using the Fallot computational model. Artificial Intelligence in Medicine 10(1), 25–40 (1997)
24. Reed, N., Nie, Y., Mahnke, C.: A portable graphical representation tool for phonocardiograms. In: Proc. 31st Int. Conf. of the IEEE Engineering in Medicine and Biology Soc., pp. 3111–3114 (2009)
25. Roy, D.L.: The paediatrician and cardiac auscultation. Paediatrics and Child Health 8(9), 561–563 (2003)
26. Shaver, J.A., Leonard, J.J., Leon, D.F.: Auscultation of the Heart: Part 4: Examination of the Heart. American Heart Association (1990)

Today's Competitive Objective: Augmenting Human Performance

Kay M. Stanney and Kelly S. Hale

Design Interactive, Inc.
1221 E. Broadway, Suite 110
Oviedo, FL 32765 USA
{Kay.Stanney,Kelly.Hale}@DesignInteractive.net

Abstract. Gaining competitive advantage requires acquiring or developing a capability that allows an organization or individual to outperform its competitors. In today's technology-driven environment, where human capabilities are struggling to keep up with technology offerings, techniques for augmenting human performance are becoming the critical gap that is precluding realizing the full benefits that these technology advances have to offer. The challenge is thus to develop tools and techniques that augment the human potential in order to best couple it to advancing complex interactive systems. In this void, those who are developing the capability to support real-time measurement, diagnosis, and augmentation of human performance may be the first to gain the competitive edge.

Keywords: Augmented cognition, Adaptive systems, human performance.

As natural human capacities become increasingly mismatched to data volumes, processing capabilities, and decision speeds, augmenting human performance will become essential for gaining the benefits that other technology advances can offer.

Technology Horizons: A Vision for Air Force Science & Technology During 2010-2030
Werner J.A. Dahm, United States Air Force Chief Scientist
May 15, 2010 (p. 58)

1 Introduction

The concept of human performance augmentation is not new. Efforts to enhance human performance have a long evolutionary history (Spohrer, 2002), which commenced millions of years ago with the human brain evolving its reasoning abilities, followed approximately one hundred thousand generations ago with the introduction of widespread spoken language and then the introduction of tools ten thousand generations ago; creative arts and written language appeared about 500

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 628–635, 2011.

generations ago; over the past few hundred generations we have seen libraries and universities appear, along with printing presses, clocks, telephones, assembly lines, radios, televisions, computers, and the Internet. All of these advances have enhanced the human's ability to perform their intended tasks - allowing humans to reach beyond their innate abilities - by taking an "outside-in" tactic (i.e., providing external aiding devices) (see Figure 1). With the advent of computers in general, and the information age specifically, the realities of exceeding the limits of human capabilities – even given the performance augmentations that evolved up to the information age – have become manifest.

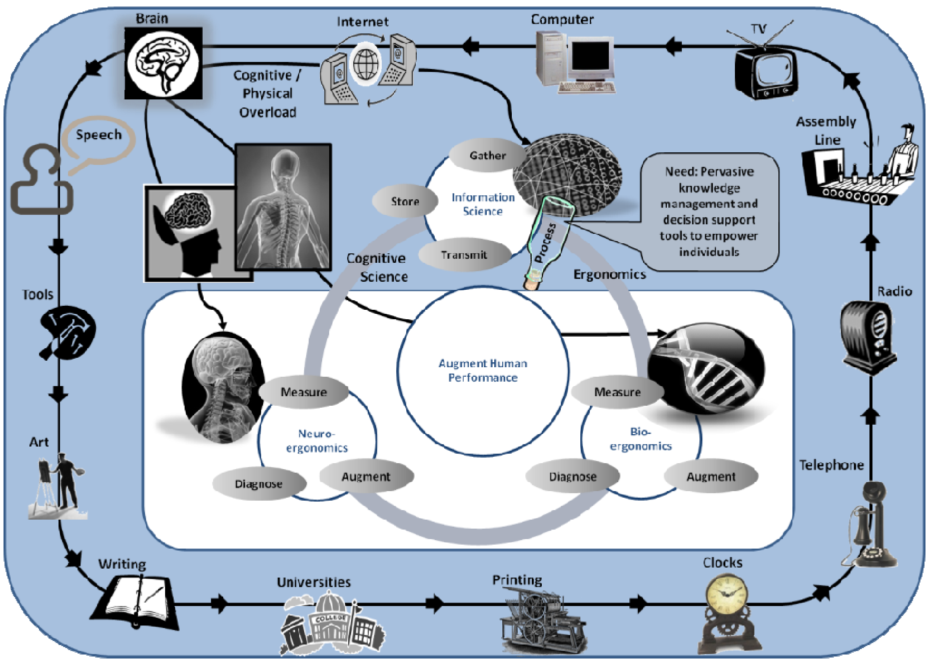


Fig. 1. Evolutionary history of human performance augmentation

2 The Augmentation Challenge

> *…there are many examples of systems that have either failed entirely or have been adopted despite their inadequacies because of the need for their capabilities. Often the reasons these adopted systems were considered unsuccessful are because they failed to meet the requirements of the human users—they required unreasonable workload, induced psychological and physical stress, or resulted in costly human error. They failed because their developers had inadequate understanding of, or overlooked consideration of, the unique capacities and limitations of people.*
>
> Pew & Mavor, 2007 (p. 12)

Initially, the mismatch between system and human was addressed in the conventional manner of performance augmentation – by providing external aids. Specifically, the field of information science emerged and provided technologies that directed how best to gather, store, transmit, and process information (Roco & Bainbridge, 2002). The resulting bottleneck to performance augmentation resulting from the information sciences was the "process" stage – while the information sciences brought the capabilities to inundate operators with massive data and work long, extended hours as one can work anytime, anyplace, these new capabilities are exceeding the human's capabilities to process (both cognitively and physically) at these higher levels. To address this bottleneck, the field of ergonomics first rose to the challenge and provided design principles that focused on redesigning information systems so that the physical stress on the operator was reduced; followed by the field of cognitive science, which introduced design principles that focused on redesigning the user interface to enhance information throughput, thereby reducing cognitive load (Stanney, 2010).

Yet, persistent challenges to human capabilities - such as sustained operations, environmental ambiguity, and information overload - require a new type of human performance augmentation– one set apart from its predecessors by the fact that it works from the "inside-out." Rather than the external performance aids of the past few hundred generations that store and make information accessible when needed to plan, inform, and guide (Rossett & Schafer, 2006), or the more recent solutions that support the physical and cognitive aspects of interaction with information systems(Stanney et al, 2001), contemporary forms of augmentation are delving inward - peering into the brain and body to measure and model human states that can be augmented in real-time so that the human can handle larger physical and cognitive loads. To realize this objective, two fields have emerged – neuroergonomics, which seeks to develop means of translating neural signals into computational cognitive models that trigger cognitive performance enhancements, and bioergonomics, which seeks to develop means to real-time monitor an individual's health in terms of physical stress and physiological condition and trigger administration of cognitive performance enhancing drugs or other such supplements (Stanney, 2010). Taken together these fields aim to overcome the nettlesome challenge of the information age - to provide pervasive knowledge management and decision support tools that truly empower individuals (see Figure 1).

Licklider (1960, pp. 4–5) envisioned such internally-driven performance enhancement a half century ago, when he mused:

> *that, in not too many years, human brains and computing machines will be coupled together very tightly, and that the resulting partnership will think as no human brain has ever thought and process data in a way not approached by the information-handling machines we know today.*

Today industry is seeking to advance the level of maturity of such coupled systems. Whether it is to support stock market analysts aiming to achieve competitive intelligence to maximize gains, physicians mining massive data to uncover important and reliable relationships between symptoms and diagnoses, or Commanders uncovering their adversary's intent through mining daily message traffic, the ability to augment natural human capacities that are being outstripped by increasing data volumes, processing capabilities, and decision speeds is becoming essential for gaining a competitive advantage. In this regard, a number of advances have been made over the past decade that have brought us closer to operational systems that can support real-time measurement, diagnosis, and augmentation of human performance based on physiological data.

3 Physiologically-Based Human Performance Augmentation

Table 1 provides an overview of the current technology maturity levels of neuroergonomic and bioergonomic technologies that aim to augment human performance in terms of Technology Readiness Levels (TRLs), which measure the maturity of evolving technologies (DOD, 2009). TRLs range from level one, where basic principles are being observed and reported, to mid-levels four and five, where component validation in laboratory and applied relevant environments occurs, to a high level of nine, where a full-scale system is proven successful in an operational setting.

Table 1. Technology Readiness Level of Neuroergonomic and Bioergonomic Technologies

	TRL 2-3	TRL 3-5	TRL 5-6	TRL 6-8
Neuroergonomics				
Measure		fNIR, Posture, Pupilometry	EDR, EEG, EMG, GSR, Respiration	ECG, Eye/ Gaze Tracking
Diagnosis		Neuro-physiologically Informed ICA		
Augment		Presentation, Schedule, System Autonomy		
Bioergonomics				
Measure	Real-time Blood and Tissue Analysis		Salivary Cortisol, SpO_2 Oximetry	
Diagnosis		Generalized Biomathematical Models		
Augment		Auto-dispense Performance-Enhancing Drugs and Supplements		

3.1 Measurement Component

The current maturity of neuroergonomic measurements tools ranges from TRL 3-5 (technology development; e.g., functional near infrared Imaging [fNIR], posture tracking, pupilometry) to TRL 5-6 (technology demonstration; e.g., electrodermal response [EDR], EEG, electromyograph [EMG], galvanic skin response [GSR]) to TRL 6-8 (system/subsystem development; e.g., electrocardiogram [ECG], eye/gaze tracking) (Schnell et al., 2008). The maturity of bioergonomic measurement tools ranges from TRL 3-5 (e.g., real-time blood and tissue chemistry analysis; Merrill, 2009) to TRL 5-6 (e.g., real-time salivary cortisol). Several efforts have demonstrated real-time cognitive and biological state measures over the past decade that meet the requirements of (1) sensitivity to different brain states and/or processes, (2) reliability, and (3) practicality in fielded use (Stanney, Schmorrow, & Hale, 2010). The current challenge with cognitive state measurement is in developing means to substantially advance real-time data fusion and classifier construction techniques (Fidopiastis & Wiederhold, 2009; Hale, Stanney, & Schmorrow, 2010). The current challenge for biological state measurement is in advancing real-time blood and tissue chemistry analysis for monitoring physical stress and physiological condition (Merrill, 2009).

3.2 Diagnostic Component

The current maturity of both neuroergonomic and bioergonomic diagnostic tools is a TRL 3-5 level, that of technology development. Most research into automated human performance augmentation has been conducted in the context of intelligent tutoring systems (see reviews by Murray, 1999 and Corbett et al., 2001), but these systems are often associated with high costs for knowledge engineering and implementation (Stevens et al., 2009). From a cognitive state perspective, the current challenge is to use machine learning to automatically produce models of expert behavior for automated performance assessment (Abbott, 2006; Santarelli et al., 2009; Stevens et al., 2009, 2010), as well as to combine multiple component models of cognitive performance into an integrative computational model (Santarelli et al., 2009), while ensuring the resulting models can provide accurate assessments on complex real-world tasks. From a biological state perspective, most research to date has resulted in sophisticated biomathematical models that can be used to create and assess work–rest schedules (Dean et al., 2007; Hursh et al., 2006). These models are limited to making general predictions about relative alertness in a post hoc manner (Van Dongen, 2004) or to making specific a priori predictions regarding the impact of fatigue (e.g., cognitive slowing; cognitive lapses) on cognitive performance (Gunzelmann, et al., 2009). Thus, the challenge to biological state diagnosis is to develop generative computational models that can make real-time predictions about the consequences of physical fatigue and other physiological stressors to cognitive performance in real-world settings.

3.3 Augmentation Component

The current maturity of both neuroergonomic and bioergonomic augmentation tools is a TRL 3-5 level, that of technology development, as much work has been done to identify “when, what, and how” to augment human performance in terms of adaptations of information presentation and schedule, level of system autonomy, and use of performance enhancing supplements or drugs (Fuchs et al., 2007; Stanney, 2010). The current challenges are to empirically identify the most effective individually tailored performance augmentations, as well as to develop trusted, adaptable, and flexible task allocation schemes and seamless and elegant interruption management strategies that not only consider the cognitive and biological states of the operator or trainee, but also individual learning/operating styles, preferences, and physiological predispositions that may develop over time as the system is being used.

4 Conclusions

While efforts to enhance human performance have a long evolutionary history from tools to libraries to clocks and gadgets to computers and the Internet, the quest to realize the human potential from the "outside-in" has been ever-present. This review of neuroergonomic and bioergonomic technologies that aim to support a paradigm shift in human performance augmentation by taking an "inside-out" approach, demonstrates that while substantial progress has been made in developing and validating these technologies, considerable work is still needed to realize full-scale systems that prove successful in operational, real-world settings. Those who propel these advancements are likely to gain a competitive advantage and be in position to outthink their competitors, outlearn their colleagues, and outsmart their adversaries.

References

1. Abbott, R.G.: Automated expert modeling for automated student evaluation. Intelligent Tutoring Systems 4053, 1–10 (2006)
2. Corbett, A.T., Koedinger, K.R., Hadley, W.H.: Cognitive tutors: From the research classroom to all classrooms. In: Goodman, P.S. (ed.) Technology enhanced learning: Opportunities for change, pp. 235–263. Lawrence Erlbaum, Mahwah (2001)
3. Dean II, D.A., Fletcher, A., Hursh, S.R., Klerman, E.B.: Developing mathematical models of neurobehavioral performance for the real world. Journal of Biological Rhythms 22, 246–258 (2007)
4. Department of Defense (DOD;, Technology Readiness Assessment (TRA) Deskbook. (July 2009), `http://www.dod.mil/ddre/doc/DoD_TRA_July_2009_Read_Version.pdf`
5. Fidopiastis, C.M., Weiderhold, M.: Mindscape retuning and brain reorganization with hybrid universes: The future of virtual rehabilitation. In: Schmorrow, D., Cohn, J., Nicholson, D. (eds.) The PSI Handbook of Virtual Environments for Training & Education: Developments for the Military and Beyond, pp. 427–434. Praeger Security International, Westport (2008)
6. Fuchs, S., Hale, K.S., Stanney, K.M., Juhnke, J., Schmorrow, D.: Enhancing mitigation in augmented cognition. Journal of Cognitive Engineering and Decision Making 1(3), 309–326 (2007)
7. Gunzelmann, G., Byrne, M.D., Gluck, K.A., Moore, L.R.: Using computational cognitive modeling to predict dual-task performance with sleep deprivation. Human Factors 51(2), 251–260 (2009)

8. Hale, K.S., Stanney, K.M., Schmorrow, D.D.: Augmenting cognition in HCI: 21st century adaptive system science and technology. In: Jacko, J., Sears, A. (eds.) Handbook of Human-Computer Interaction, 3rd edn. CRC Press, Boca Raton (2010)
9. Hursh, S.R., Raslear, T.G., Kaye, A.S., Fanzone, J.F. (2006). Validation and Calibration of a Fatigue Assessment Tool for Railroad Work Schedules, Summary Report. U.S. Department of Transportation Report Number: DOT/FRA/ORD-06/21, October 31 (2006), http://www.fra.dot.gov/downloads/Research/ord0621.pdf.
10. Licklider, L.C.R.: Man–computer symbiosis. IRE Transactions on Human Factors in Electronics 1, 4–11 (1960), http://www.memex.org/licklider.pdf
11. Merrill, M.: Monitoring technology being developed for astronauts could benefit patients on Earth. Healthcare IT News (April 30, 2009), http://www.healthcareitnews.com/news/monitoring-technology-being-developed-astronauts-could-benefit-patients-earth.(Accessed February 22, 2010)
12. Murray, T.: Authoring intelligent tutoring systems: An analysis of the state of the art. International Journal of Artificial Intelligence in Education 10, 98–129 (1999)
13. Pew, R.W., Mavor, A.S.: Human-system integration in the system development process: A new look., 11893th edn. The National Academies Press, Washington, DC (2007), http://www.nap.edu/catalog.php?record_id=11893 (accessed January 6, 2010)
14. Roco, M.C., Bainbridge, W.S. (eds.): Converging technologies for improving human performance (2010), http://www.wtec.org/ConvergingTechnologies/Report/NBIC_report.pdf (accessed September 28, 2010)
15. Rossett, A., Schafer, L.: Job aids and performance support: moving from knowledge in the classroom to knowledge everywhere. Pfeiffer, CA (2006)
16. Santarelli, T., Maulitz, R., Zachary, W., Barnieu, J., O'Connor, B.: Training Healthcare Providers to Confront Diversity in Clinical Settings. In: The Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC 2009). National Training Systems Association, VA (2009)
17. Schnell, T., Keller, M., Poolman, P.: Quality of training effectiveness assessment (QTEA): A neurophysiologically based method to enhance flight training. In: IEEE/AIAA 27th Digital Avionics Systems Conference, St. Paul, MN, October 26-30 (2008)
18. Spohrer, J.: NBICs (nano-bio-info-cogno-socio) convergence to improve human performance: opportunities and challenges. In: Roco, M.C., Bainbridge, W.S. (eds.) Converging Technologies for Improving Human Performance, Chapter B, pp. 101–117 (2002); NSF/DOC-sponsored report, http://www.wtec.org/ConvergingTechnologies/Report/NBIC_report.pdf (accessed September 28, 2010)
19. Stanney, K.M.: Augmenting human performance. In: Fafrowicz, M., Marek, T., Karwowski, W., Schmorrow, D. (eds.) Neuroadaptive Systems: Research, Theory, and Applications. Taylor & Francis/CRC Press (2010) (in press)
20. Stanney, K.M., Schmorrow, D., Hale, K.: Neuroergonomics and augmented cognition. In: Salvendy, G. (ed.) Handbook of human factors and ergonomics, 4th edn. John Wiley, New York (2010) (in press)
21. Stanney, K.M., Smith, M.J., Carayon, P., Salvendy, G.: Human-computer interaction. In: Salvendy, G. (ed.) Handbook of Industrial Engineering, 3rd edn., pp. 1192–1236. John Wiley, New York (2001)

22. Stevens, S.M., Forsythe, J.C., Abbott, R.G., Gieseler, C.J.: Experimental assessment of accuracy of automated knowledge capture. In: Schmorrow, D.D., Estabrooke, I.V., Grootjen, M. (eds.) FAC 2009. LNCS, vol. 5638, pp. 212–216. Springer, Heidelberg (2009)
23. U.S. Air Force Chief Scientist (AF/ST) (2010), Technology Horizons: A Vision for Air Force Science & Technology During 2010-2030. AF/ST-TR-10-01-PR (May 15, 2010), http://www.af.mil/shared/media/document/AFD-100727-053.pdf
24. Van Dongen, H.P.A.: Comparison of mathematical model predictions to experimental data of fatigue and performance. Aviation, Space, and Environmental Medicine 75(3), A15–A36 (2004)

Measuring the Effectiveness of Stress Prevention Programs in Military Personnel

Andrea H. Taylor and Sae Schatz

Institute for Simulation and Training, University of Central Florida.
3100 Technology Parkway, Orlando, FL 32826
{ATaylor,SSchatz}@ist.ucf.edu

Abstract. The effects of stress on military personnel are a pervasive concern. To mitigate stress's negative impacts, Defense agencies employ stress inoculation training and, more recently, have begun to provide stress resilience instruction. However, such pre-deployment programs suffer from measurement limitations, rendering their assessment difficult. Novel application of objective, individual, repeated measures, conducted under realistically stressful settings, may help address this gap. Towards that end, we reviewed common neurophysiological techniques and examined their usefulness for measuring stress reactions. These techniques include: 1) cortisol in the blood or saliva, 2) adrenaline in the blood or urine, 3) skin conductivity, 4) EEG, 5) Skin conductance, and 6) Heart rate.

Keywords: Stress, Training, Resilience, Inoculation, Physiological Measurement.

1 Introduction

Stress in the military population is a pervasive concern that grows dramatically in times of combat. Mental health disorders are the second leading cause for hospital admissions in military members [1], and exposure to extreme stressors can lead to any number of mental health disorders, including Post Traumatic Stress Disorder (PTSD). For example, a recent Mental Health Advisory Team study [2] revealed that approximately 23% of Soldiers in Iraq reported moderate or severe levels of stress, with a total of 7.3% of Soldiers testing positive for anxiety, 6.9% for depression, and 15.2% for PTSD.

In addition to potential chronic health effects (such as anxiety, depression, and PTSD), stress can also significantly decrease operational performance. Following the traditional inverted-U model of arousal, at low levels of physiological arousal, stimulation generally facilitates performance and, complementarily, at higher levels of physiological arousal, stimulation degrades performance. Extreme physiological arousal causes the suppression of cognitive and physiological resources [3], which greatly affects verbal, perceptual, and motor performance [4].

The military currently utilizes several programs to evaluate and treat stress-related mental health issues. The majority of these efforts entail post-deployment treatment rather than pre-deployment prevention. However, post-deployment treatment has not been shown to substantially reduce chronic disorders (e.g., [5], [6]) or improve future

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 636–646, 2011.

operational performance [7] although they do demonstrate limited positive effects regarding self-reported openness to, and perceptions of, mental health treatment (e.g., [8]). To address these limitations, pre-deployment prevention programs may be a viable addition to post-deployment treatment approaches.

Military personnel already received some pre-deployment stress inoculation training, and in the last decade, the Services have more formally emphasized both inoculation and resilience training. However, due to measurement limitations, assessment of these pre-deployment efforts is difficult. To address this, we suggest that novel application of objective, individual, repeated measures, conducted under realistically stressful settings, are necessary to validate contemporary military stress prevention efforts.

2 Stress Prevention Techniques

The two primary stress-prevention efforts are resilience and inoculation training. Stress resilience training aims to teach stress management techniques, and inoculation aims to build stress tolerance through exposure. In this paper, we detail the strengths and weaknesses of available measurement techniques for these two approaches, and we offer suggestions for improving their use in military stress-prevention training programs.

2.1 Resilience Training

Resilience training (i.e., the instruction of stress coping mechanisms while in a non-stressful setting) has been shown to reduce subjective stress assessments and increase performance of participants—at least in highly controlled settings [9], [10].

Over the last decade, the Services have expanded their efforts to support stress training [11]. For example, the Army's Resilience Training Program (formerly "Battlemind" training) is a standard, one hour pre-deployment briefing that covers basic signs and symptoms of post-deployment mental illness and encourages family members to support treatment [12]. Additionally, the U.S. Military Academy at West Point incorporates the Performance Enhancement Program. This year-long program integrates systematic psychological training into regular coursework to build mental and emotional strength, in order to maximize combat skills performance [13]. The program expanded in 2007, having now established sites at eight U.S. Army bases.

To date, military resilience programs have demonstrated limited effectiveness (e.g., [14], [15]). While the programs themselves may truly be effective, current measurement approaches have difficulty quantifying operational outcomes for several reasons. First, most studies focus on a single performance evaluation or post-training assessment to determine outcome effectiveness. Second, the majority of resilience training studies fail to screen participants for potential mediating variables such as previous exposure to stress reduction skills or clinical anxiety symptoms [16], [17]. Finally, it is difficult to measure the ecological validity toward military efforts, as the established laboratory measurement techniques show questionable effectiveness in operational contexts.

2.2 Stress Inoculation

Inoculation training involves the presentation of high-intensity situations to trainees in order to increase their stress tolerance through exposure [18]. The Navy's SERE exemplifies this type of tolerance-building training. For instance, the military requires military personnel who may be captured to undergo SERE training, which provides a uniquely realistic training environment. SERE includes classroom instruction, field training, and a period of confinement in a Resistance Training Laboratory. This captivity scenario is intended to accurately portray an environment of extreme stress, and its ultimate goal is to inoculate war fighters to against negative stress effects and improve their performance in live field situations [19], [20].

However, as with resilience training, several problems exist with traditional inoculation training assessment, including lack of evaluation in non-clinical environments, lack of utilization of objective stress measures, use of non-standardized training guidelines, and lack of repeated acute stress measurement. In addition, the vast majority of traditional stress inoculation studies are performed using clinically anxious participants and one-on-one training with mental health professional throughout multiple sessions [21], [22]. Even when these studies are performed in the field, inoculation training effectiveness is tested under minimally stressful conditions such as test anxiety [23], [24], dental procedures [25], [26], and speech anxiety [27], [28].

3 Stress Measurement Tools

As discussed in the previous section, although some data have been collected, measurements used to assess stress tolerance training have produced ambiguous results. Thus, quantifying the impact of this training has proven challenging. To address this gap in quantification, we suggest that objective, individual, repeated measures should be integrated into current assessment efforts and that these evaluations must be conducted under realistically stressful, complex conditions (i.e., not in clinical laboratories). More specifically, we suggest that the use of neurophysiological sensors may improve the measurement of this training.

Toward that end, we evaluated six common neurophysiological techniques and examined their usefulness for measuring stress reactions. These techniques include: 1) cortisol in the blood or saliva, 2) adrenaline in the blood or urine, 3) skin conductivity, 4) EEG, 5) eye tracking, and 6) heart-rate monitoring. Self-report questionnaires are also briefly described as complementary tool (as well as comparative foil) that may be used in conjunction with objective stress measures.

3.1 Cortisol

The physiological relationship between cortisol (or more formally, *hydrocortisone*) levels and stress has been tested extensively, and cortisol is the most commonly collected biological stress marker in experimental setting. Cortisol has been significantly related to self-reported stress symptoms, in settings such as military survival school, induced social stress, and surgical operations [29], [30], [31]. Plus, cortisol is highly correlated with various general stressors, such as test-taking and public speaking [32], [33], [34], [35], [36]. In general, cortisol measurements show a correlation (R^2) of .35–.50 with self-report measures of stress and a correlation of .55–.60 with heart rate. In a recent study, salivary and self-report data were collected from 25 Navy survival school students at baseline, during two stress exposure time points, and at recovery. Cortisol levels also show moderate correlations with self-reported stress responses at all time points (Adjusted R^2 = .46): from baseline, through exposure, and into recovery time [37].

The popularity of cortisol measurements continues to increase as new techniques facilitate its collection. For instance, cortisol levels can be monitored via saliva swabs, a minimally intrusive way to repeatedly monitor levels. Also, because swabs can be collected systematically, data can be assessed from specific time points within the exposure period [38], [39]. Further, saliva is not considered a biohazard, enabling researchers to collect samples with minimal training and limited administrative or safety concerns [40].

This method of measurement is not without its limitations, though. Peripheral collection techniques, such as through saliva, are subject to ongoing debate regarding their external validity [41], [42], [43]. Also, the more consistently valid collection method, i.e., intravenous sampling, is logistically challenging to employ, especially in the field and blood draws, themselves, may induce stress, causing confound [44]. Finally, cortisol measures are subject to circadian fluctuations unrelated to stress, so researchers must collect several baseline samples at various times of day [45], [46] and account for this variance during analyses.

3.2 Adrenaline

A less popular, but still validated, indicator of stress is adrenaline. Adrenaline measurements are about as reliable as cortisol swabs for monitoring stress [47], [48], [49]. A study investigating stress responses induced by strenuous air-to-air combat maneuver training in Japanese Air Self-Defense Force students showed significantly increased adrenaline levels post-flight, compared to pre-flight [50].

Unlike cortisol, adrenaline cannot be measured via saliva collection, and instead must be collected via urine or blood [51]. While urine sampling is less invasive than intravenous sampling, it still poses difficulties for operational use, in terms of storage, transport, and participant unease. Additionally, researchers can only obtain average levels of adrenaline from urine, because levels accumulate and average-out between urinations [52]. Since adrenaline levels increase and decrease very rapidly, collecting average levels significantly weakens urine-collection's reliability.

3.3 Skin Conductivity

Another popular measure of stress is skin conductance, or galvanic skin response (GSR). GSR has the ability to show real-time measurements throughout the entirely of stress exposure periods, and it has been shown as positively correlate with changes in other stress indicators, such as blood pressure, heart rate, norepinephrine and epinephrine levels [53], [54], [55], [56]. Also, at baseline or at rest, GSR is not influenced by normal circulatory changes (blood pressure, heart rate) [57], [58].

However, GSR measurement may prove less appropriate for field testing than other available objective measures. GSR shows high variability based on uncontrolled external factors, such as weather or room temperature [56]. Also, GSR electrodes must be placed on the hands or fingers, and are therefore continuously invasive to tasks that involve physical elements.

Possibly due to these constraints, field studies utilizing skin conductivity are virtually nonexistent. A laboratory study with Swedish military personnel showed significant differences in skin conductance between a combat-experienced group and a comparison group while viewing combat-related photos [59]. These results indicate a strong correlation between stress and skin conductance in military settings, but reiterates the difficulties of use outside a controlled lab environment.

3.4 EEG

An electroencephalogram (EEG) records electrical activity produced by neurons firing in the brain. Data are collected from electrodes placed on the scalp and these signals are commonly divided into three spectrums: Alpha, Beta, and Theta. The Alpha channel is an effective measure for minimal stress conditions (e.g., [60]); however, Alpha does not seem to strongly correlate with self-report measures of stress when higher levels of arousal are induced (e.g., [61], [62]). In general, the use of EEGs in stress measurement is still under debate, and EEG results often fail to significantly correlate with other autonomic functions (skin conductance, ECG, heart rate, etc.; [63], [64], which makes it difficult to draw reliability conclusions when monitoring multiple measures.

Additionally, although technology advances in EEG collection hardware (cap, electrodes, encoder box, and wiring) have improved, EEGs are still quite invasive to participants, take approximately 30 minutes to apply, and can only tolerate being worn for short periods of time [65]. The overall use of EEG equipment is cumbersome, as the data collection hardware is limited in portability. Once data is collected, it entails expert extrapolation of raw data before analysis can begin. Finally, a fully-equipped EEG instrument costs around $80,000, which may be too expensive for many research labs [66].

Acknowledging some of these limitations, especially the need for more portable methods of data collection in field studies, one EEG research and design company assessed the utility of a mobile EEG product during Marine Corps operational training. This testing showed external validity with heart rate variability, subjective self-assessments, and performance measures [67]. Therefore, although this improved technology is not yet widely available to the research community, it appears to be rapidly approaching.

3.5 Eye Tracking

Use of eye tracking to monitor stress has also grown in popularity as this technology continues to become more user-friendly [68]. Eye tracking data can capture reliable (R^2 around .70, compared to self-report), real-time measures of stress, since stress causes significant, immediate changes in pupil dilation [69].

However, use of eye-tracking for stress monitoring has been limited to the laboratory, since pupil dilation is only a reliable/valid measure of stress under very controlled conditions [69]. Also, eye-tracking hardware and software must be precisely calibrated and any movement between the participant and the eye-tracking device requires recalibration. Some researchers and eye tracking manufacturers are attempting to solve these challenges. For example, the Department of Homeland Security is utilizing mobile eye tracking technology in assessments of TSA vigilance tasks performed in virtual environments [70]. However, until eye tracking hardware technology becomes more mobile, the use of this measurement tool is impractical for field testing.

3.6 Heart-Rate Monitoring

Much like eye tracking methods, heart rate monitoring has been utilized for decades as a reliable, real-time stress indicator [71]. Heart rate is very easy and inexpensive to monitor, and is positively correlated (adjusted R^2 around .25) with self-reported stress levels [72]. Again, much as eye tracking, heart rate monitoring is inappropriate for field research, as stress can only be measured while the participant is immobile. As field research incorporates rigorous physical activity, heart rate monitoring for stress indicators is impossible.

3.7 Complementary Self-report Questionnaires

Numerous self-report surveys purport to reliably measure individual stress levels. Popular examples include the *State-Trait Anxiety Inventory* [73], *Beck Anxiety Inventory* [74], *Derogatis Stress Profile* [75], and *Perceived Stress Scale* [76]. These apparatus are generally comprehensive, show adequate validity and reliability, and correlate with a variety of psychological and somatic health outcomes. Additionally, self-report questionnaires can be as brief or as robust as necessary, offer anonymity to respondents, and are relatively easy to administer in most settings.

However convenient, subjective questionnaires are not necessarily useful as the sole measurement technique of resilience or inoculation training. As Shaw and colleagues eloquently explain: “Memory is fallible: when people are asked to assess the occurrence of stressful life events retrospectively, the accuracy of their recall diminishes with longer periods of assessment, so that some events may be unreported, forgotten, or denied” [77]. Additionally, individuals’ recollection of their historic stress levels is greatly influenced by their arousal at the time of questioning [78].

Nonetheless, when paired with objective measurement approaches, self-report apparatus can contribute useful insights. A study performed with 109 U.S. Army survival training students assessed self-reported stress and cortisol levels before and after training [29]. Although both measures indicated an increase in stress, the self-report data was not significantly correlated to the cortisol levels. As is a common occurrence in mental health studies in the military, participants self-reported lower stress levels than the cortisol measure suggested. Therefore, the use of two measurement techniques provided insight into participant stress that would not have been as complete with one measure.

4 Conclusion

By recognizing effective measurement tools, researchers may be able to provide war fighters with improved training programs to build stress tolerance and teach stress management techniques that reduce the potential for stress-related mental illness while maintaining operational performance. Based on the review of resilience and inoculation measurement techniques, practices can be recommended in terms of evaluating stress-prevention programs (see Table 1). It appears that a combination of urinary cortisol, multiple self-report measures, and EEG would provide researchers with a much more complete picture when assessing the effectiveness of pre-deployment stress-training programs. These measurement tools are capable of providing objective, acute data at multiple time points, and can be utilized in a field-training environment. These recommendations could be used to make up a systemized measurement technique to be integrated into military stress-prevention training.

Table 1. Summary of measures' key characteristics

Measure	Ease of Use (in field)	Reliability* (R^2)	Intrusiveness	Expense
Cortisol in saliva	+ + +	~.43	☹	$$
Cortisol in blood	+	~.51	☹ ☹ ☹	$$$
Adrenaline in urine	+ +	~.38	☹ ☹	$$
Adrenaline in blood	+	~.47	☹ ☹ ☹	$$$
Skin Conductivity	+	~.65	☹ ☹	$$
EEG	+ +	*too low to determine*	☹ ☹	$$$
Eye tracking	+	~.70	☹ ☹ ☹	$$$
Heart Rate	+	~.25 *(Adj. R^2)*	☹ ☹	$
Self-Report Apparatus	+ + +	~.80	☹ ☹	$

A heuristic summary based upon the authors' interpretation of common contemporary stress measures. Under "ease of use," + indicates difficult of use, ++ indicates neutral, +++ indicates relatively convenient to use. Under "invasiveness" ☹ indicates limited invasiveness, ☹☹ indicates moderate invasiveness, ☹☹☹ indicates extreme invasiveness. Under "expense," $ indicates relatively inexpensive approaches, $$ indicates more expensive tools, $$$ indicates especially monetarily expensive tools to acquire/employ.
* Reliability estimates based upon comparisons with self-report apparatus; see text above for specific details and citations.

References

1. Hoge, C., Lesikar, S., Guevara, R., Lange, J., Brundage, J., Engel, C., Messer, S., Orman, D.: Mental Disorders Among U.S. Military Personnel in the 1990s: Association With High Levels of Health Care Utilization and Early Military Attrition. American Journal of Psychiatry 159 (2002)
2. Mental Health Advisory Team (MHAT) Report, U.S. Army Surgeon General (2003)
3. Hardy, L., Jones, G.: Current issues and future directions for performance related research in sport psychology. Journal of Sports Sciences 12 (1994)
4. Lazarus, R.S., Deese, J., Osler, S.F.: The effects of psychological stress upon performance. Psychological Bulletin. 49, 293–317 (1952)
5. Deahl, M., Gillham, A., Thomas, J., Searle, M., Scrinivasan, M.: Psychological sequelae following the Gulf War: Factors associated with subsequent morbidity and the effectiveness of psychological debriefing. British Journal of Psychiatry 165, 60–65 (1994)

6. Shalev, A., Freedman, S., Peri, T., Brandes, D., Sahar, T., Orr, S., Pitman, R.: Prospective Study of PTSD and Depression Following Trauma. Am. Jour. Psych. 155 (1998)
7. Stecker, Fortney, Hamilton, Ajzen: An Assessment of Beliefs About Mental Health Care Among Veterans Who Served in Iraq. Psychiatric Services 58, 1358–1361 (2007)
8. Milliken, C.S., Auchterlonie, J.L., Hoge, C.W.: Longitudinal assessment of mental health problems among Active and Reserve Component soldiers returning from the Iraq War. Journal of the American Medical Association 298(18), 2141–2148 (2007)
9. Driskell, J., Johnston, J.: Stress Exposure Training. In: Cannon-Bowers, J., Salas, E. (eds.) Making Decisions Under Stress-Implications for Individual & Team Training, Washington, DC (1998)
10. Gaillard, A.: Concentration, stress and performance. In: Hancock, P.A., Szalma, J.L. (eds.) Performance Under Stress. Asghate Publishing, Aldershot (2008)
11. Zinsser, N., Perkins, L., Gervais, P., Burbelo, G.: Military application of performance-enhancement psychology. Military Review, 62–65 (2004)
12. Walter Reed Army Institute of Research (2008)
13. `http://www.acep.army.mil` (2010)
14. Bekker, Nijssen, Hens: Stress prevention training: sex differences in types of stressors, coping, and training effects. Stress and Health 17, 207 – 218 (2001)
15. Steinhardt, Dolbier: Evaluation of a resilience intervention to enhance coping strategies and protective factors and decrease symptomatology. J. Am. Coll. Health 56 (2008)
16. Johnston, J., Cannon-Bowers, J.: Stress and human performance. Mahwah, NJ (1996)
17. Jay, S., Elliot, C.: A stress inoculation program for parents whose children are undergoing painful medical procedures. Journal of Consulting and Clinical Psychology 58 (1990)
18. Meichenbaum, D.: Stress inoculation training. Pergamon Press, New York (1985)
19. Thompson, M., McCreary, D.: Enhancing Mental Readiness in Military Personnel. In: Human Dimensions in Military Operations- Military Leaders' Strategies for Addressing Stress and Psychological Support, Meeting proceedings RTO-MP-HFM-134, Paper 4, pp. 4-1 – 4-12 (2006)
20. Steffian, G., Bluestein, B., Ogrisseg, J., Doran, A., Morgan, C.: Code of conduct and the psychology of captivity: training, coping, and reintegration. In: Britt, T., Adler, A., Castro, C. (eds.) Military Life: The Psychology of Serving in Peace & Combat, Westport, CT, vol. 2 (2006)
21. Meichenbaum, D., Deffenbacher, J.L.: Stress inoculation training. Couns. Psychol. 16 (1988)
22. Swezey, R.W., Salas, E. (eds.): Teams: their Training and Performance, Norwood, NJ (1992)
23. Sharp, J.J., Forman, S.G.: A comparison of two approaches to anxiety management for teachers. Behavior Therapy 16, 370–383 (1985)
24. Zeidner, M., Klingman, A., Papko, O.: Enhancing students' test coping skills: Report of a psychological health education program. Journal of Educational Psychology 80, 95–101 (1988)
25. Getka, E.J., Glass, C.R.: Behavioral and cognitive-behavioral approaches to the reduction of dental anxiety. Behavior Therapy 23(3), 433–433 (1992)
26. Kvale, G., Berggren, U., Milgrom, P.: Dental fear in adults: A meta-analysis of behavioral interventions. Community Dentistry in Oral Epidemiology 32, 250–264 (2004)
27. Altmaier, E.M., Leary, M.R., Halpern, S., Sellers, J.E.: Effects of stress inoculation and participant modeling on confidence and anxiety: Testing predictions of self-efficacy theory. Journal of Social and Clinical Psychology 3, 500–505 (1985)
28. Cradock, C., Cotler, S., Jason, L.A.: Primary prevention: Immunization of children for speech anxiety. Cognitive Therapy and Research 2, 389–396 (1978)

29. Morgan III, C., Wang, S., Mason, J., Southwick, S., Fox, P., Hazlett, G., Charney, D., Greenfield, G.: Hormone profiles in humans experiencing military survival training. Biol. Psych. 47 (2000)
30. Gaab, J., Blattler, N., Menzi, T., Pabst, B., Stoyer, S., Ehlert, U.: Randomized controlled evaluation of the effects of cognitive-behavioral stress management on cortisol responses to acute stress in health subjects. Psychoneuroendocrinology 28, 767–779 (2003)
31. Arora, S., Tierney, T., Sevdalis, N., Aggarwal, R., Nestel, D., Woloshynowych, M., Darzi, A., Kneebone, R.: The Imperial Stress Assessment Tool (ISAT): A Feasible, Reliable and Valid Approach to Measuring Stress in the Operating Room. World Journal of Surgery 34 (2010)
32. Mason, J.W.: A Review of Psychoendocrine Research on the Pituitary-Adrenal Cortical System. Psychosomatic Medicine 30, 576–607 (1968a)
33. Mason, J.W.: The scope of psychoendocrine research. Psychosomatic Medicine 30(5) (1968b)
34. Raven, P.B., Davis, T.O., Shafer, C.L., Linnebur, A.C.: Maximal Stress Test Performance While Wearing a Self-Contained Breathing Apparatus. Journal of Occupational Medicine 19(12) (1977)
35. Sapolsky, R.: The adrenocortical axis. In: Schneider, E., Rowe, J. (eds.) Handbook of the Biology of Aging, 3rd edn., pp. 330–348. Academic, New York (1990)
36. Schedlowski, M., Flüge, T., Richter, S., Tewes, U., Schmidt, R., Wagner, T.: β- Endorphin, but not substance-P, is increased by acute stress in humans. Psychoneuroendocrinology (1995)
37. Morgan III, C.A., Rasmusson, A.M., Wang, S., Hoyt, G., Hauger, R.L., Hazlett, G.: Neuropeptide-Y, cortisol, and subjective distress in humans exposed to acute stress: Replication and extension of previous report. Biological Psychiatry 52(2), 136–142 (2002)
38. Smyth, J., Ockenfels, M., Porter, L., Kirschbaum, C., Hellhammer, D., Stone, A.: Stressors and mood measured on a momentary basis are associated with salivary cortisol secretion. Psychoneuroendocrinology 23(4), 353–370 (1998)
39. Bassett, J., Marshall, P., Spillane, R.: The physiological measurement of acute stress (public speaking) in bank employees. International Journal of Psychophysiology 5(4) (1987)
40. Center, U.S.: for Disease Control (2010)
41. Takai, N., Yamaguchi, M., Aragaki, T., Eto, K., Uchihashi, K., Nishikawa, Y.: Effect of psychological stress on the salivary cortisol and amylase levels in healthy young adults. Archives of Oral Biology 49, 963–968 (2004)
42. Noto, Y., Sato, T., Kudo, M., Kurata, K., Hirota, K.: The relationship between salivary biomarkers and state-trait anxiety inventory score under mental arithmetic stress: A pilot study. Anesthesia & Analgesia 101, 1873–1876 (2005)
43. Tessner, K.D., Walker, E.F., Hochman, K., Hamann, S.: Cortisol responses of healthy volunteers undergoing magnetic resonance imaging. Human Brain Mapping 27, 889–895 (2006)
44. Pollard, T.: Use of cortisol as a stress marker: Practical and Theoretical Problems. American Journal of Human Biology 7(2) (1995)
45. Weitzman, E.D., Fukushima, D., Nogiere, C., Roffwarg, H., Gallagher, T.F., Hellman, L.: Twenty-four hour pattern of the episodic secretion of cortisol in normal subjects. Journal of Clinical Endocrinology & Metabolism 33, 14–22 (1971)
46. Brantley, P.I., Dietz, L.S., McKnight, G.T., Jones, G.N., Tulley, R.: Convergence between the daily stress inventory and endocrine measures of stress. J. Consult. Clin. Psych. 56 (1988)

47. Hansen, M., Dyreborg, A., Larsen, Rugulies, R., Garde, A., Knudsen, L.: A Review of the Effect of the Psychosocial Working Environment on Physiological Changes in Blood and Urine. Basic & Clinical Pharmacology & Toxicology 105, 73–83 (2009)
48. Baum, A., Grunberg, N.: Measurement of stress hormones. In: Cohen, S., Kessler, R., Underwood Gordon, G. (eds.) Measuring Stress: A Guide for Health & Social Scientists. Oxford Press (1997)
49. Sluiter, J., van der Beek, A., Frings-Dresen, M.: Work stress and recovery measured by urinary catecholamines and cortisol excretion in long distance coach drivers. Occup. Environ. Med. (1998)
50. Otsuka, Y., Onozawa, A., Kikukawa, A., Miyamoto, Y.: Effects of flight workload on urinary catecholamine responses in experienced military pilots. Perceptual and Motor Skills 105(2) (2007)
51. Krapp, K. (ed.): The Gale Encyclopedia of Nursing & Allied Health. Farmington Hills, MI (2002)
52. Bassett, J.R., Marshall, P.M., Spillane, R.: The physiological measurement of acute stress (public speaking) in bank employees. International Journal of Psychophysiogy (1987)
53. Storm, H., Myre, K., Rostrup, M., Stockland, O., Lien, M., Raeder, J.: Skin conductance correlates with perioperative stress. Acta Anaesthesiologica Scandinavica 46, 887–895 (2002)
54. Edleberg, R.: Electrical properties of the skin. In: Brown, C.C. (ed.) Methods in Psychophysiology, pp. 1–53. Williams and Wilkins, Baltimore (1967)
55. Lindberg, L., Wallin, B.G.: Sympathetic skin nerve discharges in relation to amplitude of skin resistance responses. Psychophysiology 18, 268–270 (1981)
56. Bini, G., Hagbarth, K., Hynninin, P., Wallin, B.: Thermoregulatory & rhythm-generating mechanisms governing sudomotor & vasoconstrictor outflow in human cutaneous nerves. J. Physio. (1980)
57. Hagbarth, K.E.: Hallin. R.G., Hongel, A., Torebjork, H.E., & Wallin, B.G, General Characteristics of Sympathetic Activity in Human Skin Nerves. Acta Physiologica Scandinavica (1972)
58. Wallin, B., Sundlöf, G., Delius, W.: The effect of carotid sinus nerve stimulation on muscle and skin nerve sympathetic activity in man. Pflügers Archiv 358 (1975)
59. Najström, M., Högman, L.: Skin conductance response in Swedish United Nations soldiers exposed to fear-relevant stimuli. Cognitive Behavior Therapy 32(4) (2003)
60. Tyson, P.: Task-related stress & EEG alpha biofeedback. Biofeedback & Self-Regulation 12 (1987)
61. Frost, R.O., Burish, T.G., Holmes, D.S.: Stress and EEG-alpha. Psychophysiology 15 (1978)
62. Plotkin, W.B.: The alpha experience revisited: Biofeedback in the transformation of psychological state. Psychological Bulletin 86(5) (1979)
63. Hankins, T., Wilson, G.: A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight. Aviation, Space and Environmental Medicine 69 (1998)
64. Storm, H.: Changes in skin conductance as a tool to monitor nociceptive stimulation and pain. Current Opinion in Anaesthesiology 21, 796–804 (2008)
65. Nikulin, V., Kegeles, J., Curio, G.: Miniaturized electroencephalographic scalp electrode for optimal wearing comfort. Clinical Neurophysiology 121(7) (2010)
66. Jacquin, A., Causevic, E.: Method and Device for Probabilistic Objective Assessment of Brain Function. Patent #:US 2010/0191139 A1. Washington, DC: U.S. Patent & Trademark Office (2010)

67. Berka, C., Davis, G., Johnson, R., Levendowski, D., Whitmoyer, M., Fatch, R., Ensign, W., Yanagi, M., Olmstead, R.: Psychophysiological Profits of Sleep Deprivation and Stress during Marine Corps Training (2007)
68. Taylor, M., Mujica-Parodi, L., Padilla, G., Markham, A., Potterat, E., Momen, N., Sander, T., Larson, G.: Behavioral predictors of acute stress symptoms during intense military training. Journal of Traumatic Stress 22(3) (2009)
69. Shirtcliff, E., Granger, D., Schwartz, E., Curran, M.: Use of salivary biomarkers in biobehavioral research: cotton-based sample collection methods can interfere with salivary immunoassay results. Psychoneuroendocrinology 26(2) (2001)
70. Auto-Diagnostic Adaptive Precision Training for Baggage Screeners, Screen-ADAPT (2010), `http://www.designinteractive.net/projects`
71. Lazarus, R., Speisman, J., Mordkoff, A.: The Relationship Between Autonomic Indicators of Psychological Stress: Heart Rate and Skin Conductance. Psychosomatic Medicine 25 (1963)
72. Vrijkotte, T., van Doornen, L., de Geus, E.: Effects of Work Stress on Ambulatory Blood Pressure, Heart Rate, and Heart Rate Variability. Hypertension 35 (2000)
73. Spielberger, C.: Assessment of state and trait anxiety: Conceptual and methodological issues. The Southern Psychologist 2, 6–16 (1985)
74. Beck, A.T., Steer, R.A.: Manual for the Beck Anxiety Inventory. San Antonio, TX (1990)
75. Derogatis, L.: The Derogatis Stress Profile (DSP): Quantification of psychological stress. Advances in Psychosomatic Medicine 30, 30–54 (1987)
76. Cohen, S., Kamark, T., Mermelstein, R.: A global measure of perceived stress. J. Heal. Soc. Beh. (1983)
77. Shaw, W.S., Patterson, T.L., Dimsdale, J.: Stress and life events measures. Handbook of Psychiatric Measures. American Psychiatric Association, Washington (2000)
78. Lazarus, R.S., Cohen, J.B.: Environmental stress. Human Behavior and Environment 2 (1977)

Adaptive Attention Allocation Support: Effects of System Conservativeness and Human Competence

Peter-Paul van Maanen[1,2], Teun Lucassen[3], and Kees van Dongen[1]

[1] TNO Human Factors, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands
{peter-paul.vanmaanen,kees.vandongen}@tno.nl
[2] Vrije Universiteit Amsterdam, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands
[3] Department of Cognitive Psychology and Ergonomics, University of Twente
P.O. Box 215, 7500 AE Enschede, The Netherlands
t.lucassen@gw.utwente.nl

Abstract. Naval tactical picture compilation is a task for which allocation of attention to the right information at the right time is crucial. Performance on this task can be improved if a support system assists the human operator. However, there is evidence that benefits of support systems are highly dependent upon the systems' tendency to support. This paper presents a study into the effects of different levels of support conservativeness (i.e., tendency to support) and human competence on performance and on the human's trust in the support system. Three types of support are distinguished: fixed, liberal and conservative support. In fixed support, the system calculates an estimated optimal decision and suggests this to the human. In the liberal and conservative support types, the system estimated the important information in the problem space in order to make a correct decision and directs the human's attention to this information. In liberal support, the system attempts to direct the human's attention using only the assessed task requirements, whereas in conservative support, the this attempt is done provided that it has been estimated that the human is not already paying attention (more conservative). Overall results do not confirm our hypothesis that adaptive conservative support leads to the best performances. Furthermore, especially high-competent humans showed more trust in a system when delivered support was adapted to their specific needs.

1 Introduction

In the domain of naval warfare, information volumes for navigation, system monitoring and tactical tasks will increase while the complexity of the internal and external environment also increases [6]. For tactical picture compilation tasks also the dynamics in behavior and ambiguity of threat is expected to increase. Furthermore, the trend of reduced manning is expected to continue as a result of economic pressures and humans will be responsible for more and more demanding tasks. Although attention can be divided between tasks, problems with attention allocation and task performance are expected since attentional resources are limited [13,8]. Experience, training and better interfaces can lift these limitations, but only to a certain level. Even with experienced users, attentional problems are still likely to occur [11].

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 647–656, 2011.

Automation can assist humans by directing attention to critical events [14]. It can heuristically identify and prioritize objects of interest by highlighting high priority objects and dimming low priority objects. This helps humans to focus on the right subset of objects and thereby effectively reduces the number of objects that must be monitored. The downside is that this form of cuing can impede the detection of important objects that are mistakenly left unhighlighted when the automation is imperfect or when the situation is uncertain [12]. In these situations problems with inappropriate trust in the automation, resulting in over-reliance on the automation, are expected. This paper suggests to use different types of adaptive automation that adapts the support to the human need of support: only support when really needed, and in this way limit the above mentioned downside of automation as much as possible.

In [7] adaptive automation is defined as follows: "adaptive automation refers to a system capable of dynamic, workload-triggered reallocations of task responsibility between human and machine". There are multiple reasons for applying adaptive automation. How well people perform a certain task is affected by the allocation of their attention. People that are more experienced will be better at dividing attention between different sources of information. Research on the effects of playing video games has shown that visual attention abilities often improve with training. Experienced players of video games required less attentional resources for a given target [5]. In the case of a tactical picture compilation task, experts will be able to track more contacts. Experts will also be able to determine more quickly whether a contact is a possible threat. As opposed to poor performers, good performers will apply the rules correctly. Adaptive automation can help by assisting (the less well performing) humans in their allocation of attention through estimating their current allocation of attention and intervene when the human should reallocate his attention.

In [10] it is proposed to use adaptive automation to prevent 'complacency' and to increase the changes of detecting errors. By shifting task responsibility between humans and machines, humans will be more involved in tasks, which causes more errors to be detected and therefore the performance with respect to monitoring automation will increase.

In this paper the effect is investigated of different types of adaptive automation (decision support) with respect to system conservativeness (high and low) and human competence (good and poor) in terms of task performance, trust, understandability and responsibility. The paper is composed of the following sections. In Section 2 the proposed support types are described and several hypotheses are given and motivated about these support types. Then, in Section 3 a experiment is described in which the support types are evaluated. The results of this experiment are given in Section 4 and the paper ends with a discussion and conclusions in Section 5.

2 Attention Allocation Support

2.1 Generic Support Model

One way of implementing adaptive automation is to use computational cognitive models of attention as a basis for triggering change in automation. A cognitive model of attention is a model which estimates a human's focus of attention at each moment in time for a given task (see e.g., [2,1]). Together with a normative model, which estimates where attention should be optimally allocated for that same moment in time and task, a decision support system can aid the user in distributing limited attentional resources when there is a large difference between the two. In [2,1] it is shown, for instance, that in this way it is possible to support humans in their allocation of attention.

The support evaluated in this study has three variants, namely the *fixed*, *liberal* and *conservative* support type. The fixed support is defined as support that advices a human user what decision to make, without taking into account whether it is needed to support the human at that moment. The outcome of the task is shown to the human who can then decide whether to comply with the advice or to rely on his own judgment. As stated earlier, a potential risk of fixed support is inappropriate reliance. The fixed support system always gives its advice. So the easiest way to perform the task is to follow the advice as given by the system, which can lead to problems with complacency. This means that if the fixed system occasionally gives incorrect advice, it is more likely to be (incorrectly) taken over by the human, compared to an adaptive support system.

The alternative for the fixed support system is to direct the attention of the human to areas that are estimated to need human attention, instead of suggesting a specific decision. This way, the human is supported during an earlier stage of information processing, namely information acquisition, and hence leaving information interpretation and decision making to the human. The result is that the human can no longer completely rely on the support with respect to deciding what to do. Errors in the support are thus likely to be less influential on the decisions of the human. Wrong advice of the support system is also expected to be detected more easily by humans, because the advice is checked more thoroughly due to the fact that it needs to be processed more before a decision is made. This basic idea of bringing the human back 'in the loop' is also the underlying property of the last two support variants.

The liberal and conservative support type are different with respect to system conservativeness. *System conservativeness* is defined as the inverse tendency of the system to provide support to the human. It can be varied through adaptation to the behavior of the human. Examples of this behavior are mouse clicks, reaction times to events, and point of gaze. The models used for liberal adaptive support will use less behavioral data than those for conservative adaptive support. For instance, in the context of the tactical picture compilation task described in [2,1], liberal can be defined as support that adapts only to the current selection of threatening contacts. In this case, it is estimated (by a mathematical model) whether support is needed through adaptation to the clicking behavior of a human operator. For the conservative support, next to adaptation to the clicking behavior, an estimation of the current human attention allocation is also incorporated. Overall this means that the liberal support is less adapted to the user than the conservative support.

2.2 Hypotheses

For the above mentioned three support types, the effects on 1) task performance, 2) trust, 3) understandability and 4) responsibility are studied, which are discussed in sections 2.2, 2.2, 2.2 and 2.2, respectively. In sections 2.2 and 2.2 overall effects of system conservativeness and human competence are discussed. The above mentioned discussions lead to a total of 6 hypotheses.

Task Performance. When support is fixed, humans are expected to be more prone to over- and under-rely on the automation, whereas with adaptive support less problems with inappropriate reliance or complacency are expected. This can be explained by the fact that fixed support allows humans to rely entirely on the support (i.e., just take over the computer's advice). Adaptivity can stimulate the human's involvement in the task by automatically applying support, only where and when needed. It is expected that higher levels of such adaptivity to the human also results in higher task performance. This basically boils down to the following hypothesis:

Hypothesis 1. *The proposed adaptive support results in higher task performance than fixed (non-adaptive) support.*

Trust. Another important factor is trust in the support: Do participants trust adaptive decision support more than fixed decision support? Errors will inevitably occur in the support. However, these errors are likely to be much more salient in when applying fixed support. This boils down to the following hypothesis:

Hypothesis 2. *The proposed adaptive support results in more trust in the support system compared to fixed support.*

Understandability. A potential problem in adaptive support is understandability. Adaptive support systems are likely to be more complex than fixed support systems (or in any case, no support). This leads to the following hypothesis:

Hypothesis 3. *The proposed adaptive support results in a poorer understanding of the support compared to fixed support.*

Responsibility. Since we expect the human to be more involved in the task when applying the proposed adaptive support, we also expect that the responsibility for a good result as felt by the human is higher.

Hypothesis 4. *The proposed adaptive support results in a greater feeling of responsibility for the eventual outcome compared to fixed support.*

System Conservativeness. As has been mentioned, conservativeness can be varied through adaptation to the behavior of the human. This adaptivity can come in various degrees: A more conservative adaptive support system depends to a higher degree on the behavior of the human. When the system is uncertain about information, conservative support will withhold information longer than liberal support. This is expected to result in a stronger effect with respect to task performance, trust and understandability. This boils down to the following hypothesis:

Hypothesis 5. *The claimed effects in Hypotheses 1 to 4 are stronger for conservative adaptive support than for liberal adaptive support.*

Human Competence. We expect that, since adaptive support takes actions of the human into account, the task performance of the human (with or without support) contributes to the performance of the support. When the actions of the human are in line with the task model of the support, the support will be more appropriate. The hypothesis:

Hypothesis 6. *The claimed effects in Hypotheses 1 to 4 are stronger for good performers than for poor performers.*

3 Method

3.1 Participants

Forty college students (17 male, 23 female) with an average age of 23.9 years ($SD = 2.6$) participated in the experiment as paid volunteers.

3.2 Apparatus

Participants had to perform a (simplified) naval tactical picture compilation task as performed in naval warfare. The goal of this task is to build up awareness of possible threats surrounding the own ship (contacts). A screenshot of the task environment is shown in Figure 1.

Participants had to mark the five most threatening contacts by clicking on them. To determine if a contact is a possible threat the following criteria were used: speed, heading, distance to own ship and position in or out a sea-lane. All criteria were equally important. The five contacts scoring highest on these criteria had to be selected as most threatening. The behavior of the contacts was such that the threat varied over time. For instance, a contact could get out of a sea lane, speedup, or change its heading toward the own ship. Contacts that were mistakenly identified as threats (false alarms) or contacts that were mistakenly not identified as threats (misses) resulted in a lower task performance.

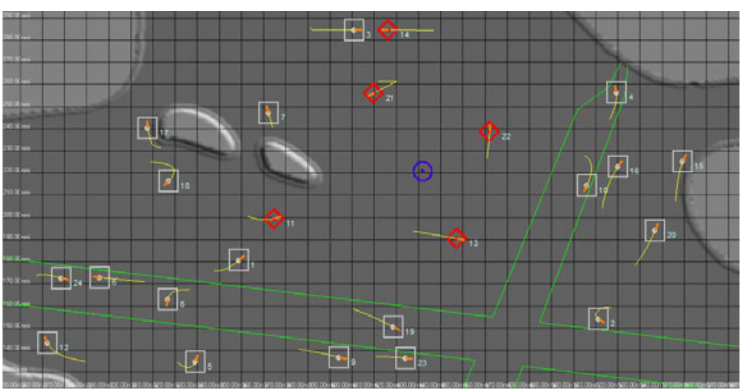


Fig. 1. Screenshot of the task environment

All participants were exposed to the same task complexity. The complexity was determined by the ambiguity and the dynamics of the behavior of the contacts. Concerning ambiguity, small differences in the threat level of contacts made it more difficult to identify the five most threatening contacts. Dynamics were determined by the varying number of threat level changes of contacts over time. Changes in threat levels were such that the number of times that each contact needed to be re-evaluated was high.

The task, including the different developed adaptive support conditions (see Section 3.4), was implemented using the game implementation software Gamemaker.[1]

3.3 Design

A 4 (system conservativeness) $\times$ 2 (human competence) mixed design was used. System conservativeness was a within-subjects independent variable and the order was balanced between the participants. Human competence was a between-subjects quasi-independent variable.

3.4 Independent Variables

Two independent variables were used: system conservativeness and human competence.

System Conservativeness. There were four levels of conservativeness for the used support system: no support, fixed support, liberal adaptive support and conservative adaptive support.

In the no support (NS) condition participants performed the tactical picture compilation task without any form of support.

The fixed support (FS) used a task model to determine the five most threatening contacts. Threat levels were determined on the same criteria as used by the participants. The five most threatening contacts were highlighted.

The liberal adaptive support (LAS) also takes the current selection of the user into account. Firstly, the two contacts with the lowest threat levels which are already selected by the user were highlighted. This was done because they are candidates for deselection and should therefore receive attention. Secondly, the three contacts with the highest threat levels which are not selected by the user were highlighted. This was done because they are candidates for selection and should receive attention for this reason.

The conservative adaptive support (CAS) basically highlighted the same contacts as the liberal support, but now only when the user paid little attention to these contacts. Attention values for all contacts were calculated using the cognitive model of attention proposed in [2,1].

The reliability of the task model in each support condition was manipulated by adding errors to the actual threatlevels. This manipulation was done in order to simulate an imperfect task model. If the task model was perfect, there was no use of comparing FS with the other support types, since it would always be a good decision to follow the advice of FS, resulting in a maximum task performance.

[1] For more information on Gamemaker, see `http://www.yoyogames.com/gamemaker`

Human Competence. After the experiment, a median split on the task performance in the NS condition was performed to distinguish a good and poor human competence group.

3.5 Dependent Variables

Four dependent variables were measured: task performance, trust, understandability and responsibility.

Task Performance. Task performance was determined by measuring the accuracy of the identification of the five most threatening contacts during the task by means of a penalty function. The severity of errors was also taken into account by the task performance measure.

Trust. After each trial participants estimated the reliability of the support system on a scale between 0 and 100% correct. Since trust is for an important part determined by perceptions of reliability [9,4,3], this was considered as a good measure of trust.

Understandability. Participants rated after each trial the degree to which they thought the decision making process of that of the support system was understandable on a 7-point Likert scale between -3 (not understandable) and 3 (understandable).

Responsibility. Participants rated after each trial the degree to which they felt responsible for the outcome of the task on a 7-point Likert scale between -3 (not responsible) and 3 (responsible).

4 Results

Two out of the 40 retrieved data sets have been removed due to unintended errors during the experiment. All participants passed the test on paper and were therefore expected to be able to correctly apply the classification criteria.

Task Performance. Lilliefors tests have shown that task performance in the NS, FS, LAS and CAS conditions were all normally distributed (i.e., null hypothesis that they are normally distributed could not be rejected).

To check whether the design of the fixed support system was a fair competitor for the adaptive variants, at least it should hold that FS results in higher performances than NS. This was indeed the case: Participants in condition FS ($M = 89.24, SD = 2.20$) performed significantly better compared to NS ($M = 87.62, SD = 3.27$), $t(74) = 2.53$, $p < .001$.

A repeated measures analysis of variance (ANOVA) showed a significant main effect ($F(2, 72) = 27.40, p < .001$) of system conservativeness on task performance.

A post-hoc Bonferroni test showed that there is a significant difference between conditions FS and LAS ($p < .001$), FS and CAS ($p < .001$), but not between LAS and CAS ($p = 1$). Hence participants performed worse in the LAS ($M = 87.34$, $SD = 1.73$) and CAS condition ($M = 87.16$, $SD = 1.87$) than in the FS condition ($M = 89.24$, $SD = 2.20$). Hypotheses 1 and 5 (for task performance) are therefore not accepted.

No interaction effect was found ($F(2, 72) = 0.22$, $p = .80$) between system conservativeness and human competence on task performance. Hence Hypothesis 6 (for task performance) is not accepted.

Trust. A repeated measures analysis of variance (ANOVA) did not show a significant main effect ($F(2, 72) = 0.47$, $p = .63$) of system conservativeness on trust. Hypotheses 2 and 5 (for trust) are therefore not accepted.

A significant interaction effect was found ($F(2, 72) = 3.17$, $p = .048$) between system conservativeness and human competence for trust.

A post-hoc Bonferroni test showed that there is a significant difference in trust between the good ($M = 1.11$, $SD = 0.29$) and poor ($M = -0.32$, $SD = 0.29$) competence group in the CAS condition ($p = .02$), but not in the LAS ($p = 1$) and FS ($p = 1$) condition. Hence the claimed effect in Hypothesis 2 is stronger for good performers than for poor performers in the case of CAS. For CAS, Hypothesis 6 (for trust) is therefore accepted, but not for LAS.

Understandability. A repeated measures analysis of variance (ANOVA) did not show a significant main effect ($F(2, 72) = 0.42$, $p = .66$) of system conservativeness on understandability. Hypotheses 3 and 5 (for understandability) are therefore not accepted.

No significant interaction effect was found ($F(2, 72) = 0.92$, $p = .40$) between system conservativeness and human competence for understandability. Hypothesis 6 (for understandability) is therefore not accepted.

Responsibility. A repeated measures analysis of variance (ANOVA) did not show a significant main effect ($F(2, 72) = 0.37$, $p = .69$) of system conservativeness on responsibility. Hypotheses 4 and 5 (for responsibility) are therefore not accepted.

No significant interaction effect was found ($F(2, 72) = 1.39$, $p = .26$) between system conservativeness and human competence for responsibility. Hypothesis 6 (for responsibility) is therefore not accepted.

5 Conclusion and Discussion

In this study we investigated the benefits of adaptive attention allocation support over fixed (non-adaptive) support in a tactical picture compilation task. We expected task performance using adaptive support to be higher than in fixed support. However, this first hypothesis was not accepted. Trust in adaptive and fixed support did not differ significantly, also rejecting the second hypothesis. In contrary to our third hypothesis, our participants did not report to have a poorer understanding of the more complicated adaptive support than of the fixed support. Also the fourth hypothesis, stating that the feeling of responsibility would be higher in the adaptive condition, could not be accepted based on the results in this study.

The influence of system conservativeness and human competence was also investigated on the first four hypotheses on task performance, trust, understandability and responsibility. The results did not show a significant effect of system conservativeness on any of these variables, so the fifth hypothesis could not be accepted.

For human competence, the effect was significant for trust, but only in the conservative adaptive support condition. This means that well performing participants had more trust in conservative adaptive support than poorly performing participants. This confirms the sixth hypothesis (for trust), but only for the conservative condition. The sixth hypothesis could not be accepted for task performance, understandability, and responsibility. The increase of trust in the conservative adaptive support for good performers can be explained by the effect that good performers are more likely to understand the task and the effect support systems have on task performance. In [9], for instance, it is shown that the use of automation decreases when the effect of automation to performance is not properly perceived.

There are several possible explanations for why an increase of task performance was not found in our experiment. Our implementation of adaptive support aimed at reducing inappropriate reliance on fixed support. However, this comes with a cost in the form of added complexity. Although participants did report a clear understanding of how both adaptive support systems worked, it is still possible that the disadvantage of added complexity is larger than the advantages of such a system. Working with complex (support) systems can raise the cognitive load on the human, leaving less capacity to focus on the actual monitoring of contacts. In this case, this resulted in a significantly higher task performance in the fixed support condition than in both adaptive support conditions. Future design of adaptive support systems should aim at keeping the system as simple as possible, though preserving the expected advantages of adaptivity.

For the adaptive support investigated in this study, it was not possible for the human to simply follow suggestions of the support system. This was because, instead of suggesting a possible answer to a problem, only areas of interest were indicated by the system. This meant that, in any case, the proposed adaptive support must have eliminated inappropriate reliance on the support. The found results in this study are not a reason for rejecting this principle and therefore more research on adaptive attention allocation support is suggested, focusing on the *requirements* in which such a system can help to gain task performance.

Acknowledgments

This research has partly been supported by the programme Cognitive Modeling" (V524), funded by the Dutch Ministry of Defense. The authors would also like to thank Karel van den Bosch, Tibor Bosse, Anja Langefeld and Jan-Willem Streefkerk for their helpful comments.

References

1. Bosse, T., van Lambalgen, R., van Maanen, P.P., Treur, J.: Attention manipulation for naval tactical picture compilation. In: Proceedings of the 2009 IEEE/WIC/ACM International Conference on Intelligent Agent Technology, IAT 2009 (2009)
2. Bosse, T., van Lambalgen, R., van Maanen, P.P., Treur, J.: Automated visual attention manipulation. In: Paletta, L., Tsotsos, J.K. (eds.) WAPCV 2008. LNCS, vol. 5395, pp. 257–272. Springer, Heidelberg (2009)
3. Dzindolet, M.T., Beck, H.P., Pierce, L.G., Dawe, L.A.: A framework of automation use. Technical Report ARL-TR-2412, Army Research Laboratory, Aberdeen Proving Ground, MD (2001)
4. Gao, J., Lee, J.D.: Extending decision field theory to model operator's reliance on automation in supervisory control situations. IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans 36(5), 943–959 (2006)
5. Green, C.S., Bavelier, D.: Action video game modifies visual selective attention. Nature 423, 534–537 (2003)
6. Grootjen, M., Neerincx, M.A., van Weert, J.C.M.: Task-based interpretation of operator state information for adaptive support. In: Foundations of Augmented Cognition: Strategic Analysis, 2nd edn. LNCS, pp. 236–242. Springer, Arlington (2006)
7. Hilburn, B., Jorna, P., Byrne, E., Parasuraman, R.: The effect of adaptive air traffic control (atc) decision aiding on controller mental workload. In: Human-Automation Interaction: Research and Practice, pp. 84–91 (1997)
8. Kahneman, D.: Attention and effort. Prentice Hall, Englewoods Cliffs (1973)
9. Lee, J.D., See, K.A.: Trust in automation: Designing for appropriate reliance. Human Factors 46(1), 50–80 (2004)
10. Parasuraman, R., Riley, V.A.: Humans and automation: Use, misuse, disuse, abuse. Human Factors 39, 230–253 (1997)
11. Pavel, M., Wang, G., Li, K., Li, K.: Augmented cognition: Allocation of attention. In: Proceedings of 36th Hawaii International Conference on System Sciences, pp. 286–300. IEEE Computer Society, Los Alamitos (2003)
12. John, M.S., Smallman, H.S., Manes, D.I., Feher, B.A., Morrison, J.G.: Heuristic automation for decluttering tactical displays. Human Factors 47, 509–525 (2005)
13. Wickens, C.D.: Processing resources in attention. In: Parasuraman, R., Davies, D.R. (eds.) Variaties of attention, pp. 63–101. Academic Press, Orland (1984)
14. Wickens, C.D., McCarley, J.S.: Applied attention theory. CRC Press, Boca Raton (2007)

A Dynamic Approach to the Physiological-Based Assessment of Resilience to Stressful Conditions

Mikhail Zotov[1], Chris Forsythe[2], Alexey Voyt[1],
Inga Akhmedova[1], and Vladimir Petrukovich[1]

[1] Saint Petersburg State University, Universitetskaya naberejnaya 7/9,
Saint-Petersburg, 199034, Russia
[2] Sandia National Laboratories, USA
zotov@psy.pu.ru,
jcforsy@sandia.gov,
{voytalexey,pemphix,Petrukov_vm}@mail.ru

Abstract. In the presented research, a new algorithm of detection and analysis of non-stationary phases (NSPh), characterizing sudden changes in heart rate variability (HRV) parameters was used. Physiological reactions of air traffic controllers during the performance of training scenario were estimated. 39 participants - 14 experienced air traffic controllers and 25 students performed a 40-minute scenario, which included 3 stressful incidents: a rapid increase in air traffic density, low fuel level and plane engine failure. Students also performed the scenario after brief training. The results have shown that as expertise grows respondents show a significant decrease in duration and change in patterns of non-stationary phases of heart rate arising in response to the stressful incidents. These changes of parameters of non-stationary phases are connected with increased efficiency of air traffic controllers' cognitive performance in stressful conditions. The research has illustrated that the analysis of non-stationary phase parameters complements classical HRV measures and may be used for assessment of physiological responses of operators in Augmented Cognition applications.

Keywords: heart rate variability, cognitive workload, simulation-based training.

1 Introduction

There is a large amount of data describing the usage of heart rate variability (HRV) measures for monitoring of physiological arousal, attention, stress and general cognitive workload of operators in the process of real or simulated professional work [1, 2, 3]. Still, the majority of time- and frequency domain measures of HRV that have been used are coarse-grained and do not enable identification of moment-to-moment changes of operator functional state in the course of activity. However in many cases it is important not so much to measure the generalized level of physiological arousal of an operator during the performance of professional tasks, as

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 657–666, 2011.

to identify the moments in time when arousal sharply increases, which can be caused both by objective factors, such as increasing task demands, and by psychological factors, such as decision-making difficulties [4].

In the 1970-s Russian scientists developed a “biocybernetic” dynamic approach to the assessment of operator functional state, based on the analysis of patterns of physical arousal changes under the influence of physical or mental stress [5, 6].

It is known, that people respond to the increase of cognitive workload with increased heart rate (HR) and reduction of heart rate variability (HRV) [2, 3]. These periods of sudden changes of HR and HRV parameters are usually called *the non-stationary (transitive, unsteady) phases* (NSPh), in contrast to stationary periods, that are characterized by stability of HR and HRV parameters over time [6].

In a number of studies, it has been shown that time and peak characteristics of non-stationary phases, registered under physical or mental stress, can be considered as informative indicators of stress resilience [6, 7]. Meanwhile, absence of reliable algorithms for detection and analysis of non-stationary periods of heart rate complicates the progress of studies in this area.

2 Study

We have developed a new algorithm for detection and analysis of non-stationary phases (NSPh), characterizing periods of sudden changes in HRV parameters. The purpose of the present research was to assess empirically the ability of this algorithm to identify changes in the level of psychophysiological arousal of novice and experienced operators during simulation-based training.

Participants. A total of 39 participants, 14 experienced air traffic controllers and 25 students, aged from 20 to 46 years, took part in this study. The research was conducted at the training centre of State University of Civil Aviation (St.-Petersburg, Russia).

Procedure. A medium-fidelity air traffic control simulator was used in the experiment. The participants performed a 40-minute training scenario which included 3 critical incidents. These incidents had various difficulty levels - low (rapid increase in air traffic density, low fuel level) and high (plane engine failure). Students performed the scenario two times: before and after a short course of training.

A group of experienced instructors estimated the errors made by students during the performance of the scenario. Depending on an error type - minor (for example, a wrong command during the radio exchange) or blunder (for example, infringement of the safe heights, near miss or plane collision), they received a weight factor. Then the average error score was calculated.

The experiment began with a 2-minute registration of an electrocardiogram for the baseline heart rate assessment. Afterward, the participants performed the 40-minute scenario on the simulator.

Stressful incidents with lower difficulty level appeared at 4 and 16 minutes, with higher difficulty level - at 34 minutes of the scenario. A 2-minute resting HR registration was obtained after the end of the session.

Cardiovascular Measures. Electrocardiogram signals were continuously recorded during the performance of the training scenario by the participants using “Kardi2NP” portable loggers (MCS Inc., Russia) with disposable electrodes. A standard three-electrode configuration was used.

R-peaks from the ECG signal were detected offline, artifacts were corrected, and the heart rate signal was converted into an evenly sampled signal with 4 Hz frequency. Then, the data was used for calculation of instantaneous heart rate and heart rate variability.

Algorithm for detection and analysis of non-stationary phase of heart rate. The integral non-stationarity index (NI) of heart rate is based on two sliding windows of 15 sec. duration (fig. 1). At each window shifting, a 0.25 sec. step, the difference between HR parameters calculated within these two windows was estimated.

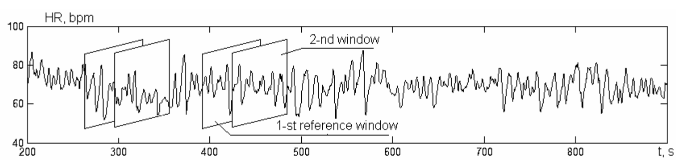


Fig. 1. Windows used in calculation of non-stationarity index

NI is based on the comparison of four HR characteristics: mean value, standard deviation, linear trend slope, prediction error of autoregressive model and is calculated by the formula (1):

$$NI = |RRNN_1 - RRNN_2| \cdot |SDNN_1 - SDNN_2| \cdot SL_1 \cdot ER_{1,2} \tag{1}$$

Where RRNN – mean RR interval for the set window;

SDNN – standard deviation of RR intervals for the set window;

SL – angle of slope of HR-signal linear approximation,

ER – prediction error of the autoregressive model

1, 2 – window number.

Raw values of NI were normalized, then non-stationarity periods of heart rate were identified. An example of NSPh’ detection procedure is presented on fig. 2.

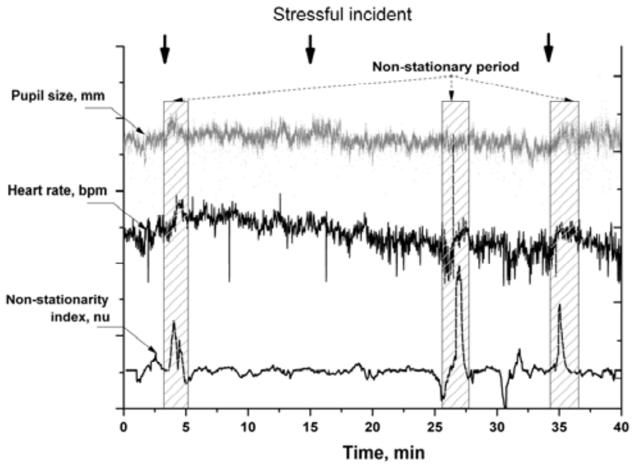


Fig. 2. Non-stationary periods of physiological arousal of an individual during the performance of training scenario

As shown in fig. 2, the algorithm enables identification of moments of sudden changes of the psychophysiological arousal level of an individual, manifested both in changes of heart rate parameters and pupil diameter.

The non-stationary periods of a heart rate were fitted by modeling the mathematical function after recognition. An example of fitting is shown on fig. 3.

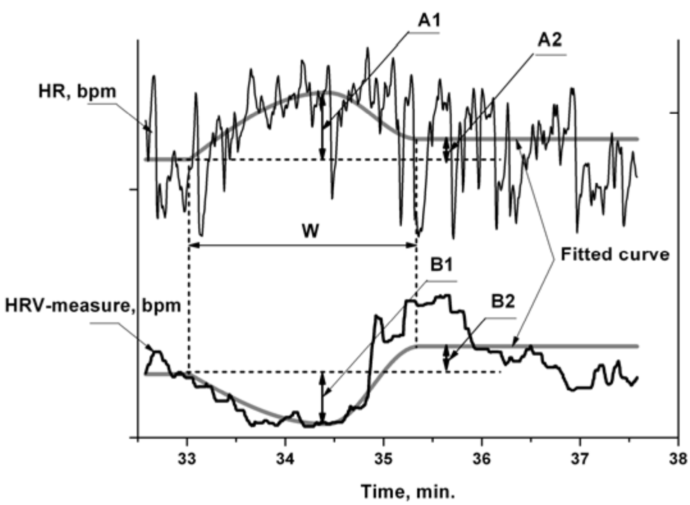


Fig. 3. An illustration of NSPh' fitting procedure

For all NSPh observed during the occurrence of stressful incidents, the following parameters were calculated (fig.3):

W – time duration of NSPh

"A1", "B1" – measures of HR increase and HRV decrease, reflecting intensity of mobilization

"A2", "B2" – measures of differences in levels of HR and HRV before and after NSPh, reflecting «physiological cost» of adaptation to the task demands.

3 Results

3.1 Task Performance Data

Research confirmed, that the experts showed a higher level of successful performance than the novices (F (1,37) = 13.8, p <.01). A significant influence of training (F (1,24) = 14.2, p <.01) was established. After training, students made fewer errors during performance of the scenario on the simulator (fig. 4.)

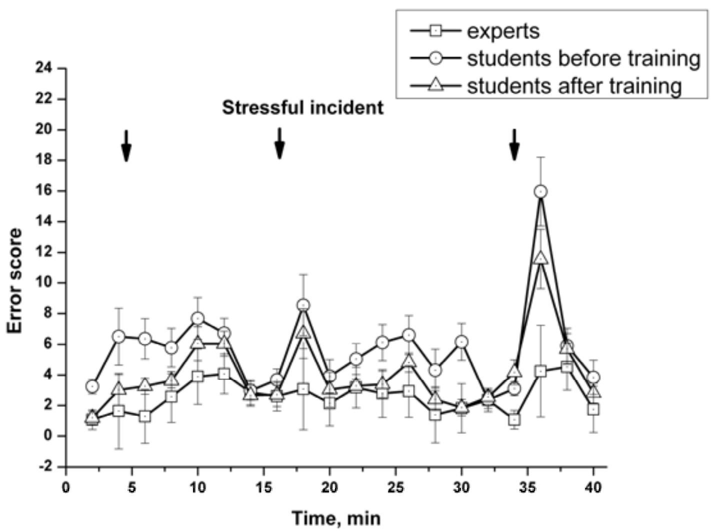


Fig. 4. Average errors score during performance of training scenario by the experts and the students before and after the training

As shown in fig. 4, the students made errors more often when the stressful incidents occurred (4, 16 and 34 minutes). The stressful incident at 34 minutes, connected with one of the plane's engine failure was the most difficult for them. Successfully coping with this incident demanded complex assessment of the situation, recalling appropriate instructions for that case, and fast decision-making to change the mid-air organization to avoid collision of planes. Thus, the given incident was connected with a sharp increase of cognitive workload in students. As shown in fig. 4, students made considerable errors at 34 minutes of the scenario, even after the training.

3.2 Cardiovascular Data

Characteristics of cardiovascular responses of participants during the performance of the training scenario was analyzed. Averaged instantaneous HR values of participants during the performance of the training scenario are presented in fig. 5.

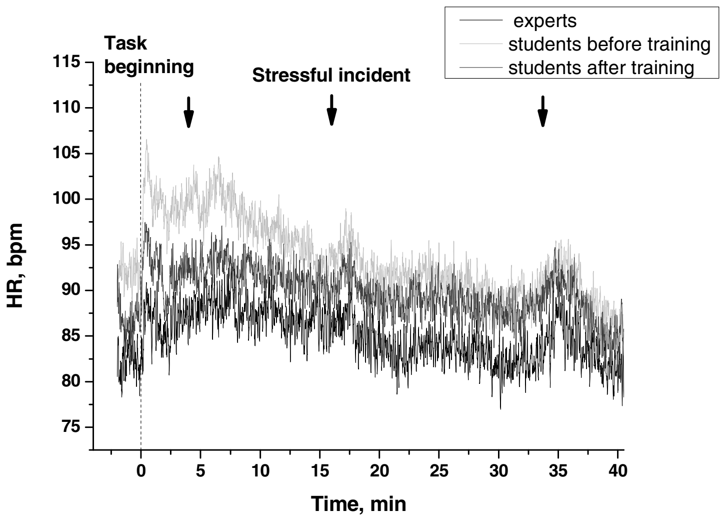


Fig. 5. Averaged instantaneous HR values of participants during the performance of the scenario

As shown in fig. 5, the students in comparison to the experts, showed a higher level of physiological arousal. It can be seen, that all groups of participants responded to the start of the scenario and the stressful incidents with a sharp increase of HR.

Then, average frequency (counted per minute) of NSPh in different moments of the scenario was analyzed (fig. 6).

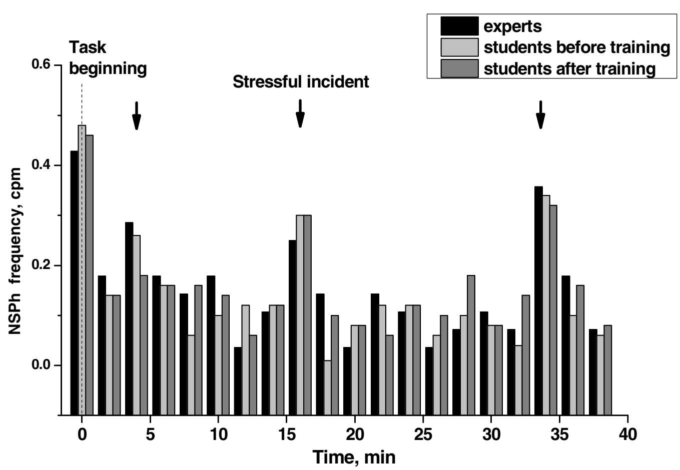


Fig. 6. Average frequency of NSPh during the performance of the scenario by participants

As shown in fig. 6, NSPh were mostly observed in the beginning of the scenario, and also during the stressful incidents occurrence at 4, 16 and 34 minutes in all groups of the participants. Average frequency scores of NSPh for neutral stages of the scenario and stages connected with the occurrence of stressful incidents were calculated. The ANOVA revealed a significant influence of Stage (neutral, stressful) on frequency of NSPh ($F (1,37) = 42.7$, $p <.001$). Between-group differences, and also, training, influences were not significant.

Thus, all the participants, irrespective of expertise level, demonstrated cardiovascular responses to the changing requirements of the task, connected with occurrence of stressful incidents. These results demonstrat the significant differences in characteristics of the cardiovascular responses of experts and students on the critical incidents arising during the performance of the scenario.

NSPh parameters of the participants in the beginning of the session and during stressful incidents at 4, 16 and 34 minutes of the scenario were calculated. An ANOVA with the Stage (stages 1-4) as a within-subject factor and Expertise (novice vs. expert) as a between-subject factor was calculated. The repeated-measures ANOVA with Stage (stages 1-4) and Training (before vs. after training) as within-subject factors was also used.

Significant distinctions between the experts and the students in the duration of NSPh were established ($F (1,37) = 32.7$, $p <.01$). A significant influence of the Stage factor ($F (3,35) = 6.9$, $p <.01$) on duration of NSPh was also revealed. This research has shown, that duration of NSPh significantly decreased ($F (1, 24) =35.2$, $p <.01$) as a result of training (fig. 7).

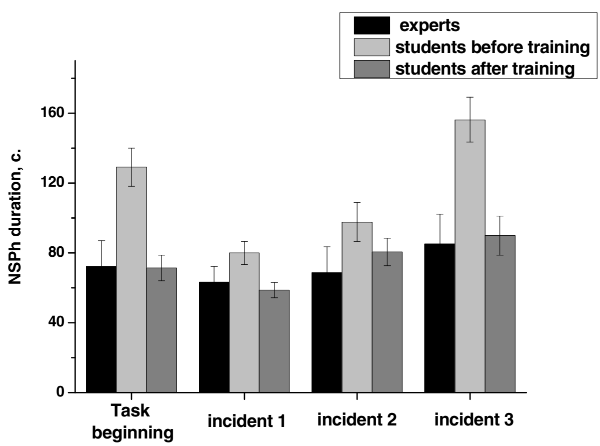


Fig. 7. Average duration of NSPh during the performance of the scenario by participants

As can be seen from fig. 7, experts showed considerably shorter NSPh than students. In all participants, the difficulty of stressful incident influenced the duration of NSPh, which was less in the beginning of the scenario and increased during the difficult stressful situation at 34 minutes.

The significant influence of the interaction of Expertise and Stage factors (F (3,35) = 3.37, p <.05), and also Training and Stage (F (3,22) = 4.6, p <.05) on parameter “A1” reflecting HR increase during NSPh was revealed (fig.8).

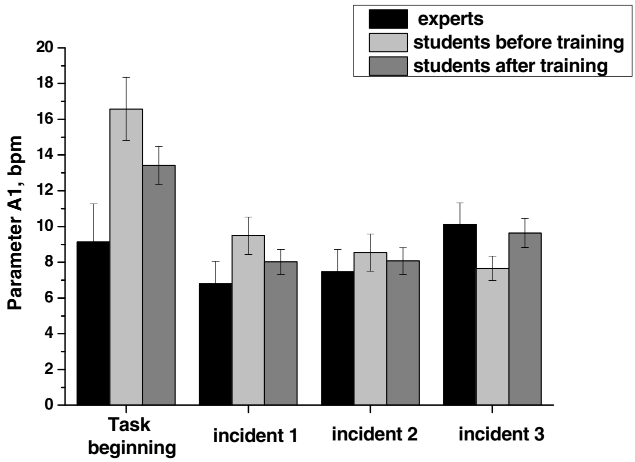


Fig. 8. Average values of HR increasing parameter “A1” in different groups of participants

As shown in fig. 8, the students responded to the start of the scenario with a greater HR increase and with a lower HR increase - to the difficult stressful incidents arising in the course of scenario performance. The experts differentially mobilized efforts in response to the difficulty of critical incidents. They showed maximum HR increase during the most difficult incident at 34 minutes of the scenario.

The research has also demonstrated that students responded to the start of the scenario and stressful incidents with greater decrease of HRV than experts (F (1,37) = 4.46, p <.05).

It was established, that experts, in comparison to the students, showed significantly lower values of “A2” (F (1,37) = 4.9, p <.05) and “B2” (F (1,37) = 5.3, p <.05) parameters, reflecting the physiological “cost” of adaptation to stressful conditions. Training led to a significant decrease in the “A2” parameter (F (1,24) = 10.4, p <.01), while for the “B2” parameter, its influence was not significant (p>.05).

4 Discussion

Most of the heart rate variability (HRV) measures used to estimate arousal and general cognitive workload are calculated for relatively lengthy time intervals and do not allow identification of brief changes of the human functional state. Meanwhile, in many situations, it is important not so much to assess the general arousal level, as to recognize time points when human operators arousal increased due to increased task demands or the influence of stress factors.

It was shown by Russian researchers, that analysis of characteristics of so-called non-stationary (“transitive”, unsteady) phases, characterizing sudden changes of

HRV-parameters, can be successfully used to estimate the resilience to stressors [6, 7]. However the absence of reliable algorithms for NSPh detection has considerably complicated the practical realization of this approach.

In the present research, an algorithm for automatic detection and analysis of NSPh in HR-dynamics was used to estimate psychophysiological responses of experienced air traffic controllers and students during the performance of tasks on a simulator. It was established, that this algorithm enables identification of moments of intense increase of individuals' arousal, connected with stressful conditions. All the participants showed a significantly greater quantity of NSPh in the beginning of stressful incidents, than during other periods of the scenario.

This research demonstrated significant differences in characteristics of cardiovascular responses of the experienced operators and students to the critical incidents. The students in comparison to the experts showed longer duration of NSPh, arising as a response to the stressful incidents. Duration of NSPh decreased in the course of training.

The students in comparison to the experts responded with a greater decrease of HRV at the start of the scenario and stressful incidents. The experts HRV-indicators of arousal stabilized faster and on a lower level, than untrained individuals, so experienced participants demonstrated less "physiological cost" of adaptations to stressful conditions. In the course of training, physiological response characteristics of the students approximated those observed with the experts.

It is obvious that the considered characteristics of cardiovascular responses are closely connected with the efficiency of operators' cognitive activity in the process of coping with stressful situations. Successful performance in stressful conditions demands fast detection and assessment of a situation, effective retrieval of necessary procedures from long-term memory, fast decision-making and effective actions [8]. Measures of participants' cardiovascular responses can be used as indirect measures of mental efforts and time spent for these stages of cognitive activity.

Thus, this research has demonstrated the potential of using the dynamic approach to physiological assessment of the resilience to stressful conditions in developing criteria of assessment of students' physiological responses in simulation-based training.

Acknowledgments. This work was supported by Office of Naval Research Grant N00014-08-1-0731.

References

1. Baevskii, R.M.: Prediction of states on the verge of norm and pathology. Medicine Publ., Moscow (1979) (in Russian), Баевский Р.М. Прогнозирование состояний на грани нормы и патологии. Москва: Медицина (1979)
2. Veltman, J.A., Gaillard, A.W.: Physiological workload reactions to increasing levels of task difficulty. Ergonomics 41, 656–669 (1998)
3. Mulder, L.J.M., de Waard, D., Brookhuis, K.A.: Estimating mental effort using heart rate and heart rate variability. In: Stanton, N., Hedge, A., Brookhuis, K.A., Salas, E., Hendrick, H. (eds.) Handbook of Human Factors and Ergonomics Methods, CRC Press, Boca Raton (2004)

4. Zhernavkov, V.F., Kozlovsky, E.A.: Psychophysiological assessment of pilots' readiness on the flight simulators. Voenizdat. Military Publ. House, Moscow (1981) (in Russian) Жернавков В.Ф., Козловский Э.А. Психофизиологическая оценка подготовленности летчика на пилотажных тренажерах. Москва: Воениздат (1981)
5. Medvedev, V.I.: Human's physiological and psychological functions' resistance under the influence of extreme factors. Science Publ., Leningrad (1982) (in Russian), Медведев В.И. Устойчивость физиологических и психологических функций человека при действии экстремальных факторов. Ленинград, Наука (1982)
6. Sapova, N.I.: Complex evaluation of heart rhythm regulation during measured functional loads// Fiziologicheskiĭ zhurnal SSSR imeni I. M. Sechenova (Sechenov physiological journal of the USSR) 68(8), 1159–1164 (1982) (in Russian); Сапова Н.И. Комплексная оценка регуляции сердечного ритма при дозированных функциональных нагрузках // Физиол. журн. им. Сеченова. 68, 8, С 1159-1164 (1982)
7. Novikov, V.S., Shustov, E.B., Goranchuk, V.V.: Correction of functional states under extreme exposures. Science Publ., St. Petersburg (1998); (in Russian), Новиков В.С., Шустов Е.Б., Горанчук В.В. Коррекция функциональных состояний при экстремальных воздействиях. - СПб.: Наука (1998)
8. Driskell, J.E., Salas, E., Johnston, J.H.: Decision Making and Performance under Stress. In: Britt, T.W., Castro, C.A., Adler, A.B. (eds.) Military life: The psychology of serving in peace and combat. Military performance, vol. 1, pp. 128–154. Praeger Press, New York (2006)

Author Index