1st International Conference on
Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI 2009)
in-cooperation with ACM SIGCHI   Essen, Mon/Tue 21 - 22 September 2009

UNIVERSITÄT DUISBURG ESSEN

SIGCHI

Proceedings of the 1st International Conference on

Automotive User Interfaces and Interactive Vehicular Applications

Albrecht Schmidt, Anind Dey, Thomas Seder
Oskar Juhlin, Dagmar Kern

University of Duisburg-Essen

Sept 2009

AutomotiveUI 2009

First International Conference on Automotive User Interfaces and Interactive Vehicular Applications

September 21-22 2009, Essen, Germany

Conference proceedings

Editors:

Albrecht Schmidt, Anind Dey	*Conference Chairs*
Thomas Seder, Oskar Juhlin	*Program Chairs*
Dagmar Kern	*Publication Chair*

PREFACE

The idea to hold a conference on human-computer interaction issues related to cars emerged some years back. Ubiquitous computing is becoming reality and researchers from computer science and human-computer interaction are moving into new domains. Vehicles and, in particular, cars present an exciting domain that offers many challenging research questions and at the same time new solutions can have a real impact on people's lives. As over the last few years many research projects on vehicle interaction have started and many PhD students work on this topic, we thought it was time to provide a forum for this emerging community. The plans for the conference were well worked out when in late 2008 the economic crisis also impacted the car industry. Despite the risk of running a smaller conference we went ahead and were happy to get many quality submissions and a good number of participants.

A first of hopefully many

With great pleasure we present the proceedings of the First International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AUTO-UI-09, http://auto-ui.org). This new conference addresses human-computer interaction in the context of cars, including new interaction devices and metaphor use, methods and tools appropriate for this domain, and ethnographic work as well as studies that improve our understanding of interaction while operating a vehicle. New applications, as a catalyst for many new forms of interaction in the car, are a further part of the conference proceedings. For its inaugural year, AUTO-UI-09 is being held at the University of Duisburg-Essen in Germany. Major sponsorship is being provided by the University of Duisburg-Essen and the conference is in cooperation with ACM SIGCHI, with its proceedings to be archived in ACM's Digital Library. We have embraced this topic in the hope to foster a new community with aspirations to initiate a recurring annual gathering on Automotive User Interfaces and Interactive Vehicular Applications.

Automotive User Interfaces

Advances in technology have transformed cars into complex interactive systems. Drivers interact with a variety of controls and applications to operate a vehicle. Besides mastering the primary driving task, drivers make use of entertainment, information and communication systems in the car. Technical systems in modern cars support communication, sensing and consuming media. With these novel technologies, many opportunities arise for creating attractive in-car user interfaces. Nevertheless the challenge of creating such interfaces in a compelling and safe to use manner has grown ever greater. Especially in the automotive context, users expect interfaces that are intuitive and straightforward to use, without having to read a manual. The overall experience of driving a car is more and more influenced by the man-machine interface, and hence creating attractive user interfaces is of great importance for a successful product. Traditional means for user interface development taken from desktop computing are often not suitable, as many other conditions have an influence on the design space for automotive user interfaces. In comparison to many other domains, trial and error while the product is already in the market is not acceptable as the cost of failure may be fatal. User interface design in the automotive domain is relevant across many areas ranging from primary driving control, to assisted functions, to navigation, information services, entertainment and games.

Submission and review process

Authors were invited to submit papers that are 2, 4 or 8 pages long, where the length of the paper should fit the content. The call was open for academic papers, design sketches, interaction concepts, and industrial case studies. We received in total 40 papers of various length, the majority 8 pages and 4 pages long. The majority of authors came from the USA, Germany, Austria, United Kingdom, Sweden, Israel, Korea, and Japan. The quality, novelty, and originality of submitted work well exceeded our expectations. All papers received at least 3 independent reviews; the majority of papers had 4 or more reviews. The reviews were completed by experts on the program committee and, if required, additional expert reviews were requested. Based on these reviews, the chairs selected the final program, which consists of 12 long papers and 10 short papers. These contributions are included in the proceedings.

Acknowledgments

We greatly appreciate and warmly thank the many people who have worked to make this conference possible. We recognize and appreciate all the work the Technical Program Committee members and additional expert reviewers put in to promote the conference, review papers, and select the work that composes these proceedings.

We appreciate the fact that many people helped to make the local organization possible. We are grateful to all the student volunteers for their help to make this conference a pleasant experience. In addition, the chairs appreciate the support from their home institutions. Finally, we thank Lauren Thompson and Adrienne Griscti of ACM for their support.

Albrecht Schmidt and Anind Dey, Conference Chairs
Essen, Germany, September 2009

CONFERENCE ORGANIZATION

Conference Chairs

Albrecht Schmidt	University of Duisburg-Essen, Germany
Anind Dey	Carnegie Mellon University, USA

Program Chairs

Thomas Seder	GM, USA
Oskar Juhlin	Interactive Institute & Stockholm University, Sweden

Publication Chairs

Dagmar Kern	University of Duisburg-Essen, Germany

Web Masters

Wolfgang Spießl	BMW Group Research and Technology, Germany
Florian Alt	University of Duisburg-Essen, Germany

Local Organization

Dagmar Kern	University of Duisburg-Essen, Germany
Nicole Recksing	University of Duisburg-Essen, Germany

Program Committee

Peter Abowd	Altia
Elisabeth Andre	University of Augsburg
Klaus Bengler	Technische Unversität München
André Berton	Daimler AG
Susanne Boll	University of Oldenburg
Barry Brown	UC San Diego
Verena Broy	BMW Group Research and Technology
Liselott Brunnberg	DSV Stockholm University
Gary Burnett	Nottingham University
Andreas Butz	Ludwig-Maximilians-Universität München
Vinny Cahill	Trinity College Dublin
Fang Chen	Chalmers, University of Gothenburg
Luca Chittaro	University of Udine
Nigel Davies	Lancaster University
Anind Dey	CMU
Alois Ferscha	Universität Linz
Danny Grant	Immersion Corporation
Martin Hofmann	Carmeq GmbH
Giulio Jacucci	Helsinki Institute for Information Technology
Matt Jones	Swansea University
Dagmar Kern	University of Duisburg-Essen
Matthias Kranz	Technische Unversität München
Antonio Krüger	University of Münster
John Krumm	Microsoft Research
Andrew Kun	University of New Hampshire
Natascha Künstner	Volkswagen AG
Kevin A. Li	UC San Diego
Roberto Montanari	Università di Modena e Reggio
Christian Müller	DFKI
Jonathan P. Munson	IBM
Peter Oel	Volkswagen Coporate Research
Tim Paek	Microsoft Research
Mark Perry	Brunel University
Aaron Quigley	University College Dublin
Bryan Reimer	MIT
Andreas Riener	Universität Linz
Albrecht Schmidt	University of Duisburg-Essen

Dieter Schramm	University of Duisburg-Essen
Mikael B. Skov	Aalborg University
Wolfgang Spiessl	BMW Group Research and Technology
Aaron Steinfeld	CMU
Ivan Tashev	Microsoft Research
Antonella Toffetti	Centro Ricerche Fiat S.C.p.A.
Marcus Tönnis	Technical University of Munich
Ingo Totzke	University of Würzburg
Manfred Tscheligi	University of Salzburg
Michael Weber	University of Ulm
Reto Wettach	FH Potsdam
Nora Zelhofer	Daimler AG

Additional Reviewers

Lisa Cowan	University of California
Manuel Cruz	Immersion
Ronald Ecker	BMW Group Research and Technology
Andrew Gosline	Immersion
Niels Henze	OFFIS Research Institute of Information Technology
Martin Pielot	OFFIS Research Institute of Information Technology

TABLE OF CONTENTS

MULTIMODAL INTERACTION AND PHYSIOLOGICAL ASPECTS

INTERACTING WHILE DRIVING

Keynote

MINI Design: From the Original to the Original. The path from *Center Speedo* to *Center Globe*

Gert Volker Hildebrand
BMW Group
General Manager MINI Design

DESIGN SPACE AND HAPTIC FEEDBACK

Design Space for Driver-based Automotive User Interfaces

Dagmar Kern, Albrecht Schmidt
Pervasive Computing and User Interface Engineering Group
University of Duisburg-Essen,
Essen, Germany
{dagmar.kern, albrecht.schmidt}@uni-due.de

ABSTRACT
Over the last 100 years it has become much easier to operate a car. However in recent years the number of functions a user can control while driving has greatly increased. Infotainment, entertainment and comfort systems as well as driver assistance contribute to this trend. Interaction with these systems plays an important role, as on one hand this can improve the user experience while driving but on the other hand it may distract from the primary task of driving. User interfaces in cars differ regarding the number of input and output devices and their placement in the car to a great extent. In this paper, we introduce a first design space for driver-based automotive user interfaces that allows a comprehensive description of input and output devices in a car with regard to placement and modality. This design space is intended to provide a basis for analyzing and discussing different user interface arrangements in cars, to compare alternative user interface setups, and to identify new opportunities for interaction and placement of controls. We present a graphical representation of the design space and discuss its usage in detail based on several examples. To assess the completeness of the proposed design space we used it to classify and compare user interfaces from more than 100 cars shown at IAA2007, cars from the BMW museum, and from the A2Mac1 image database.

Categories and Subject Descriptors
H.5.2 [Information Interfaces and presentation]: User Interfaces - *Input devices and strategies (e.g., mouse, touchscreen),* B.4.2 [Input/output and data communications]: Input/Output Devices, H.1.2 [Models and principles]: User/Machine Systems – *Human factors*

General Terms
Human Factors

Keywords
Design space, automotive user interfaces, car user interfaces

1. INTRODUCTION
Driving a car today entails a lot more than operating the pedals and steering wheel and has extended well beyond operating the primary controls. With the rapidly increasing complexity of automotive user interfaces in the last decades, drivers are now confronted with many new functions inside the car. This trend is fueled by car manufactures who, in addition to improving the safety and efficiency, e.g. by adding collision-avoidance systems, also aim to improve the comfort and entertainment opportunities within their cars [4].

AutomotiveUI'09, September 21-22, 2009, Essen, Germany
ACM 978-1-60558-571-0/09/009

Thus, a car has become more than just a means of transportation; for many people, especially with longer commutes, it is now a multifunctional living space. With the help of technologies like MP3 players, GPS navigation systems and mobile phones, people use their cars as a space for media consumption, as a personal communication center or as an inter-connected workplace. Many people spend 1 hour or more per day in their car [17] doing boring routine driving tasks on their way to work and back. To make this time more valuable and driving safe it is important to provide good user experiences inside the automobile.

The use of new functionalities inevitably increases the driver’s interaction with the user interface and decreases the driver’s focus on driving, which is still the primary task and should have the highest priority. This makes it important to take driver distraction [18] into considerations while designing new user interfaces for cars. Independent of which kind of functionality is introduced into the car, the associated workload level (physical, visual and mental) has to be considered for safety reasons [7]. Thus, new functionalities in cars should be as minimally distracting as possible. With the design space we provide a visual representation that allows for the designer to see how adding a new control may interfere with existing controls as they occupy the same region in the design space.

In the earlier days of the car, a one-to-one mapping from control to function was common, but with the growing number of functions inside a car, e.g. about 700 functions in a BMW series 7 [5], which are also interdependent, this is no longer possible. There is a trend in automotive systems where different functions are combined in a hierarchical menu structure, which are commonplace in graphical user interfaces for computers. Such structures require the user to search through different menus to find a desired function. This creates either visual or auditory distraction or increases the cognitive load for the task. In some cases, this is not ideal, e.g. searching for the menu function that changes the radio volume might be annoying for the driver. Thus, there is a tradeoff between how many functions are quickly accessible and how overloaded the user interface is. This trade-off can be observed in many current car interface designs.

In this paper, we introduce a design space for driver-based automotive user interfaces that provides an overview of input and output devices in cars with respect to their placement, which part of the body they interact with, which kind of feedback they provide and to which task-class they are assigned.

For generating this design space, we analyzed 706 photographs of 117 models from 35 different manufacturers taken at the international automobile exhibition (IAA[1] 2007) in Frankfurt. The photos are available at https://www.pcuie.uni-due.de/AUI/IAA2007. Additionally we accessed the suitability of the design space by picking a random set of pictures from A2Mac1[2] image database and by modeling selected historic cars.

The central contribution of the paper is a comprehensive design space for driver-based automotive user interfaces that is grounded in an analysis of a large number of existing cars, including historic cars and concept car.

The paper is structured as follows. After discussing the background and related work, we present a graphical representation of the design space. We discuss in detail input modalities, output modalities, and position of the controls. Using two actual cars we show how the design space can be used for comparison. Additionally we show an overview representation that allows to describe a set of cars, and we show how this can be applied.

2. BACKGROUND AND RELATED WORK

2.1 Driving Task

The complex driving task can be divided into three classes primary, secondary and tertiary [16]. Primary tasks describe how to maneuver the car, e.g. controlling the speed or checking the distance to other cars or objects. Secondary tasks are functions that increase the safety for the driver, the car and the environment, e.g. setting turning signals or activating the windshield wipers. Tertiary tasks are all functions regarding entertainment and information systems.

Figure 1: Distribution of primary, secondary and tertiary tasks (based on [20])

2.2 Input Devices

Based on the classification system from Geiser [16], Tönnis et al. [20] assigned input devices to the three classes. They distinguish between primary, secondary and tertiary devices and assign them to specific locations of the car (see Figure 1). Primary devices are used to maneuver the car, e.g. the steering wheel and the pedals. They are usually mapped one-to-one with their functionality and provide haptic feedback. Primary devices are arranged close to the driver so that they are easy to reach. Secondary devices are, for example, stalk controls for the turn signal or windscreen wipers. They are also at an easy-to-reach distance, often mounted on the backside of the steering wheel. Tertiary devices are used for the info- and entertainment systems. Many manufacturers combine a large number of enter- and infotainment functionalities into one system, e.g. the Audi MMI system [2] or the BMW iDrive [6], which consists usually of two parts: a single controller and a display. Tertiary devices are often placed in the center stack. With multifunctional steering wheels, a few tertiary devices intrude into the domain of secondary devices, e.g. radio controls on the steering wheel for faster access to frequently-used functions.

2.3 Output Devices

Output devices are used to provide feedback to the user about the current state of the system e.g. about the current speed, if the direction indicator is turned on, or which radio channel is currently playing. Feedback is important but prioritized differently for the three different driving tasks. Feedback about the primary task must be immediate and clear, whereas the information about which radio channel is playing is less important. Output devices for providing three kinds of feedback are available in cars. They provide visual, auditory and haptic/tactile feedback or even a combination of them. A detailed discussion about issues concerning these displays can be found in [20].

2.4 Design Guidelines and Standards

There is a big difference in designing user interfaces for the computer domain, where the user pays full attention to the interaction, and for cars, where user's main focus has to be on the primary driving task. Interacting with tertiary user interfaces never has the highest priority for the user when the car is moving. There are international standards available [13, 14] that give interaction design recommendations and enforcements, e.g. the user interface must not force the user to take both hands off the steering wheel.

Furthermore, there are a few guidelines that offer support to designers during the design process, e.g. [1, 12, 19]. They describe how to make entertainment und infotainment system safe and easy to use for all drivers. They include concrete design recommendations, e.g. text size or the placement of displays always taking safety and usability issues into account.

2.5 Design spaces

The importance of understanding design spaces for user interfaces is emphasized by HCI researchers. Foley et al [15] provide a classification of input devices using the graphic subtask they were capable of performing. Buxton [8] introduced a taxonomy of input devices. His classification includes the physical properties and the number of spatial dimensions the devices sense. In Card et al's [9] design space, input devices are compositions of one-dimension sensors. Ballagas et al. [3] have taken up these design spaces and provide a design space of ubiquitous mobile input. In the output domain, there are also design spaces regarding structuring information visualization [10].

A design space for automotive user interfaces differs from the aforementioned design spaces in two main areas. First, all devices are fix-mounted in a car, and it is therefore essential to take placement of the devices into consideration. Second, the driver is limited in her mobility but can act with the left or the right hand, as well as with the left or right foot.

[1] http://archiv.iaa.de/07/index.php?id=home2007&L=1

[2] A2mac1 Automotive Benchmarking: http://a2mac1.com/

3. DESIGN SPACE FOR DRIVER-BASED AUTOMOTIVE USER INTERFACES

In this section, we present our design space for automotive user interfaces, which gives a common basis to discuss existing arrangements of user interfaces in cars and aims to find new spaces for them. We focus on user interfaces that are operated by the driver, but the proposed design space can be extended to include passenger-based user interfaces. Following the view of Tönnis et al. [20], that cars are “complex computer systems with very particular input and output devices and mobile functionality”, we decided to create a design space that includes all input and output devices, their connection to each other and their placement.

Our design space is based on an analysis of 706 photographs taken at the IAA 2007. We collected photographs of 117 models from 35 different manufacturers, tagged and categorized them, and looked for similarities and differences. First, we identified the different input and output modalities that can be found in almost all of the observed cars. Then, we analyzed the position and interaction model for input and output devices. The photos can be accessed at https://www.pcuie.uni-due.de/AUI/IAA2007.

The following assumptions and statements refer to left-hand cars, but they can be easily applied to right-hand cars by substituting “left” for all occurrences of “right” and “right” for “left”.

3.1 Input Modalities

We found eight different input possibilities. The most commonly used group are buttons, which are present in different sizes and shapes. Nearly all buttons in modern cars are soft buttons (see Figure 2-a). That means there is no permanent haptic feedback available; instead, a visual feedback is often used. For example, when the high beams are turned on, this is indicated lighting up a button. In the past, mechanical buttons were used, e.g. to turn on the lights. These buttons provided haptic feedback, e.g. when a button was pressed, it felt pushed in (see Figure 2-b). Thus, the driver could determine the state of the button without looking at it. Sliders form the next group. They are often used for adjusting the direction of the fan (see Figure 2-g). We distinguish two different kind of knobs, those that are continuous (see Figure 2-d), e.g. to control radio volume, and those that are discrete (see Figure 2-c), e.g. a knob used to adjust the temperature. Stalk controls are often attached to the steering wheel to indicate or to activate windscreen wipers (see Figure 2-e). On a multifunctional steering wheel, thumbwheels are often used to control volume (see Figure 2-j). Classical pedals are still available in the car: gas, brake and (in cars with stick shift) clutch (see Figure 2-i). In the last few years, more and more manufactures have added a multifunctional controller to their cars. A multifunctional controller can be turned, pressed and sometimes shifted in four or even eight directions, e.g. BMW iDrive or Audi MMI (see Figure 2-f). These controllers are combined with high-resolution displays, and together, they are used as a control unit for entertainment and infotainment systems in the car.

New interaction techniques like speech and gesture recognition, as well as indirect interaction like fatigue detection using an eye tracker or cameras, have also found their way into the car. These new interaction techniques provide means for hands-free interaction so that drivers no longer need to search for and touch specific devices while driving. However, speech recognition often requires the driver to push a push-to-talk button before it can be used.

Touchscreens, the last input opportunity, are at the border to the output modalities, because they combine both input and output modalities in a single device (see Figure 2-h). The application areas for touchscreens are enter- and infotainment systems as well as comfort systems like air-conditioning systems.

3.2 Output Modalities

The output modalities are limited by the human senses, specifically sight, hearing, touch and smell. There are a lot of visual indications available in the car to give feedback about current functional states. These indications vary from simple indicator lamps to high-resolution displays. Looking closer at the simple indicator lamps, e.g. those used to indicate that the high beams are turned on, you can find two different ways to present information. One way is to turn on a light above a description (see Figure 3-d), and the other way is to illuminate a symbol whose shape indicates the meaning (see Figure 3-e).

Visual representations are also used to give information that is directly correlated to the driving task, e.g. actual speed. Both analog and digital representation are used for these purposes (see Figure 3, a-b). Analog representations can also be divided into displays that use a physical dial and pointer and displays that replicate the dial and pointer virtually (see Figure 3-c). Virtual representations allow for more dynamic use of the space in the middle of the dial to show other information. Digital displays have been used since the end of the 1970s to show alphanumerical information, e.g. the current radio channel or traffic information (see Figure 3-g).

Figure 2. Input modalities: a) button b) button with haptic feedback c) discrete knob d) continuous knob e) stalk control f) multifunctional knobs g) slider h) touchscreen i) pedals j) thumbwheel.

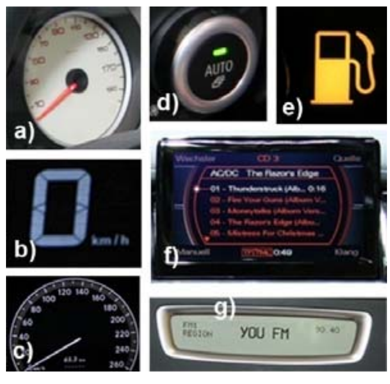


Figure 3. Output modalities: a) analog speedometer b) digital speedometer c) virtual analog speedometer d) indicator lamp e) shaped indicator lamp f) multifunctional display g) digital display

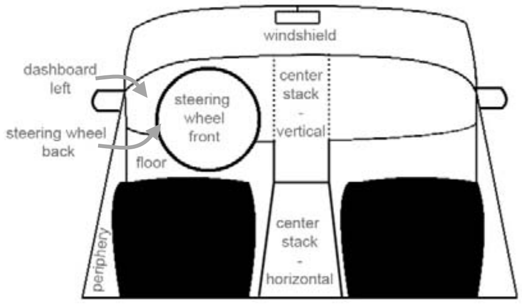


Figure 4. Division of driver's interaction environment.

At the end of 1990s, multifunctional LCT or TFT displays started to appear in cars, and car manufactures started to integrate comfort, entertainment and infotainment functions into single systems. These systems are controlled by buttons on each side of the screen, by a central controller or by touchscreen.

Sense of hearing is addressed by loudspeakers, which are integrated into the car or attached to an external device, e.g. a portable navigation system. This modality has long been used for entertainment purposes and has more recently been used for giving aural feedback, especially with voice-operated systems.

Information can also be delivered to the driver by using the sense of feel or touch. Some car manufactures have recently added vibration feedback to the steering wheel or to the driver's seat to warn the driver, e.g. of lane departures when no turn indication has been made [11]. In the earlier days, cars already relied on sense of touch with mechanical buttons whose physical state gave direct feedback.

Output modalities that use the sense of smell have yet to be established. However, one can imagine that this sense could be used for more ambient information. For example, when the motor temperature is increasing, the odor inside the car could change.

3.3 Positioning Input and Output Devices

The arrangement of input devices in cars is limited by ergonomic factors. All input devices have to be within reach for the driver, so that she can safely manipulate them with the left or right hand or left or right foot while driving. Except for touchscreens, output devices do not necessarily have to be within a safe reaching distance, but they do need to be in the driver's field-of-view.

We identified the following main interaction areas between the driver and the car (see Figure 4):

- **Windshield**: used for example for head-up displays
- **Dashboard**: for driver-based user interfaces we focus on the left part of the dashboard that is directly in front of the driver in left-hand cars.
- **Center Stack:** divided into the vertical part (on the right side of "dashboard left" in front of the driver) and the horizontal part (between the front seats)
- **Steering wheel**: divided into front and back side of the steering wheel
- **Floor**
- **Periphery**: includes the side-/rear-view mirrors

3.4 Graphical Representation

We propose two different graphical representations, one for categorizing a single car and the other for analyzing a set of cars that can be used for comparing cars from different manufacturers or car models from different years.

3.4.1 *Categorizing a single car*

In our two-dimensional graphical representation, we focus on the placement and the task classification of input and output devices based on what body part would interacts with them. We regard the driver as the main user and create the interaction descriptions from the driver's point of view.

The first dimension of the graphical representation indicates the placement of devices: windshield, dashboard (left), center stack, steering wheel, floor, and periphery. The other dimension is given by input and output modalities, where input is divided into left or right (hand or foot) as the main interaction initiators. We added one more column for input devices to represent additional modalities like speech to the design space. Since the voice has no direct spatial representation, it is associated with the periphery area. The output modalities are divided into the three senses: sight, hearing, and haptic (for feel and touch). If new interaction methods cannot be located in the current dimensions, a new column can be introduced, e.g. gesture as input or air-flow/olfactory as output, to represent a new modality.

Each input or output device can be added into the grid shown in Figure 5. The symbolic representation of different device types allows the design space to be extended with new modalities. For example, a sensor to measure skin conductivity that is mounted on the steering wheel would be represented by a new symbol and placed in the section representing the steering wheel. The structure of the design space would remain the same, allowing the new modality to be compared with the others without limiting the design space to the current set of modalities.

We divided the devices into the three task categories defined by [20], primary, secondary and tertiary tasks. These categories are color-coded in the graphical representation. Info- and entertainment systems, as well as comfort functions like air conditioning, could be clearly classified as tertiary tasks, but driver assistance systems like Adaptive Cruise Control are not so easy to classify. Tönnis et al. [20] suggested classifying them as secondary-task devices, but we believe they are rather used for primary tasks, because they influence the driving task directly.

Numbers inside the symbols indicate the occurrences. Dotted lines illustrate connections between input and output devices. Lines ending with arrows represent direct connections (e.g. stalk control for headlight gives visual feedback with indicator lamp) and ending with dots represent indirect connections to an output domain (e.g. volume knob controls audio volume and gives no localized feedback). Numbers on the lines indicate how many controls are connected to the output devices/domain.

In Figure 5, two graphical representations for a 2007 BMW 520d and a 1956 BMW 507 are shown. Each input and output device is classified in the graphical representations. Classifying the head-up display was a unique case, since it fits into all three task categories. Thus, its display-symbol is divided into three parts, one for each task. Section 4 contains discussion comparing the two cars based on these classification results.

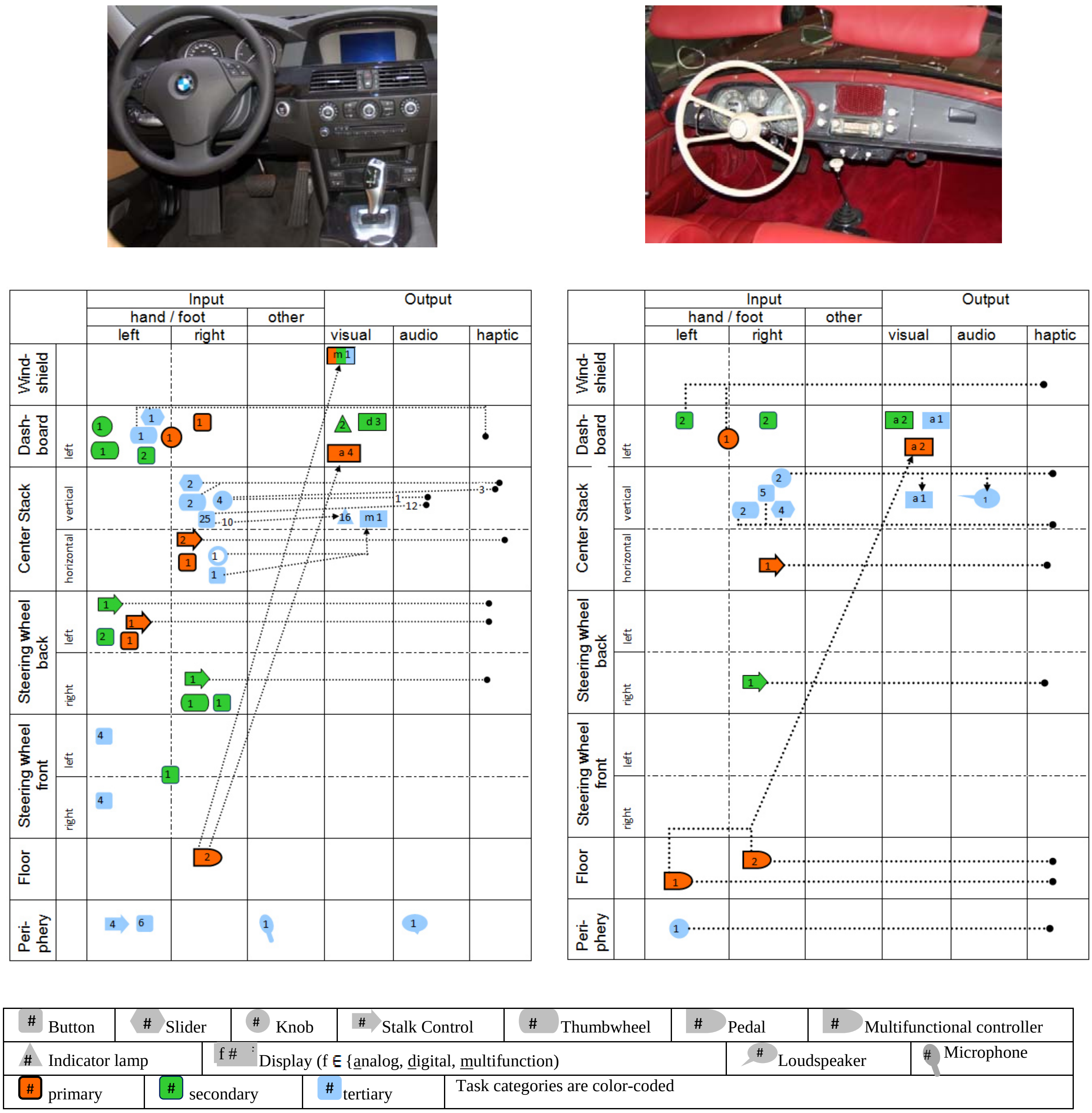


Figure 5. Graphical representation of our design space for driver-based automotive user interfaces. The classifications were created for a 2007 BMW 5 series and a 1956 BMW 5 series. Both cars have the steering wheel on the left side. The design space consists of the different interaction areas in a car, to which the input modalities are assigned. Output is divided into visual, audio and haptic. Numbers inside the controls indicate the occurrences. Primary, secondary and tertiary tasks are color-coded.

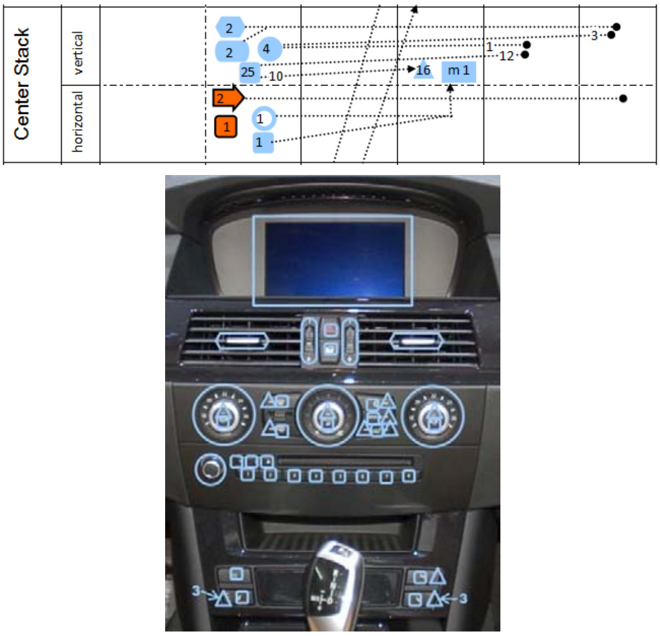


Figure 6. Detailed view of the center stack area. Corresponding markers (vertical) are shown in the photograph.

Figure 6 illustrates the connection between the symbols used in the graphical representation to real devices for the BMW 520d. In the photograph, the devices are marked by the same symbols. In this BMW series 5, 25 buttons are available, from which 10 provide visual feedback with an indicator light, and 12 are associated with the radio and provide audio feedback. The remaining 3 buttons influence the air conditioning system but provide no direct feedback. The two sliders and thumbwheels provide haptic feedback through their current positions. One continuous knob is used to control the volume, and the other three discrete knobs control the air conditioning system and provide visual feedback by indicating at which temperature they are set. The LCD screen shows visual feedback and is controlled by the iDrive controller, which is mounted in the center stack. Each of these input devices can be specified further using Card et al.'s design space for input devices [9].

This center stack example further illustrates that it is possible to analyze a select part of the design space. Still, it must be taken into account that some input-output connections may get lost.

3.5 Analyzing a Set of Cars

For providing a more general view, an abstract representation of the design space is illustrated in Table 1 and 2. The areas are not separated into subareas but instead represented by a triple, which stands for (primary, secondary, tertiary). This abstraction can be used to categories a set of cars, as in Table 1 with different BMWs or Table 2 with different Renault cars.

Analyzing the abstract views of the design space classifications, we found that the BMW models were all very similar in their arrangement of input and output devices, while Renault offered a wider selection of arrangements, especially with devices for tertiary tasks (e.g. the number of controls on the vertical center stack that can be controlled by the driver's right hand in the range 6 to 41). From these abstract views, similarities and differences can be extracted. For example, the floor and horizontal center stack areas are very similar. The number of devices in the floor area only differ for automatic or manual-transmission cars, which was the same for both manufacturers. The variation in the numbers of devices correlated to the number of their functionalities. For example, additional devices for Adaptive Cruise Control (ACC) were needed or a multifunctional controller was used for models have with a multifunctional display.

	Input			Output
	left		right	
Windshield				(0,1,1)
Dashboard	(0,4,2)	(1,0,0)	(1,0,0)	(4,2-5,0)
Center stack			(3,0,35-50)	(0,0,17-20)
Steering wheel (back)	(0-2,3-5,0)		(0,3,0)	
Steering wheel (front)	(0,0,4)	(0,1,0)	(0,0,4)	
Floor	(0-1,0,0)		(2,0,0)	
Periphery		(0,0,11)		(0,0,3-4)

Table 1. Classification of BMW models series 1, 3, 5 and M3

	Input			Output
	left		right	
Windshield				(0,0,2)
Dashboard	(0,2-4,2-11)	(1,0,0)	(0-2,0-2,0)	(0-4,0-3,0)
Center Stack			(2,0,6-41)	(0-4,0-3,1-18)
Steering wheel (back)	(0,3,0-4)		(0,5,4)	
Steering wheel (front)	(0-2,0,0)	(0,1,0)	(0, 0,0-2)	
Floor	(0-1,0-1;0)		(2,0,0)	
Periphery		(0,0,8)		(0,0,3-5)

Table 2. Classification of Renault models Clio, Espace, Kangoo, Koloes, Laguna, Megane, Modus, Twingo.

4. USING THE DESIGN SPACE

4.1 Historical Analysis and Trends

Our proposed design space can be used to analyze trends and explore historical changes. Regarding historical changes, we found that few controls stay where they were, especially for control of primary tasks. Primary-task controls have not changed at all in the last years (e.g. steering wheel or pedals). The trend towards automatic-transmission cars decreased the pedals to two. Another trend, towards facilitating the driver while driving, leads to an increase of devices for primary tasks e.g. for (Adaptive) Cruise Control. In the secondary task domain, there is a trend away from analog speedometers towards digital speedometers in both discrete and continuous types. Some manufactures also changed the position of the visual output of the speed from the driver's side to the middle.

A huge increase in the number of devices for tertiary tasks can also be observed, which is strongly related to the increase in comfort, entertainment and infotainment functionalities in cars e.g. air conditioning, integrated support for mobile phones, navigation systems, and MP3 players.

Figure 5 clearly illustrates the difference in the number of devices. The 2007 BMW series 5 has 113 devices (13 primary, 15 secondary, 85 tertiary), resulting in the input triple 80 (9, 11, 60) and output triple 33 (4, 4, 25). In contrast, the 1954 BMW series 5 has 29 devices (7 primary, 9 secondary, 13 tertiary), with input triple 21(5, 5, 11) and output triple 8 (2, 4, 2). Another big difference can be seen in the feedback opportunities of the buttons. All buttons in the 1954 car has haptic feedback while buttons in the 2007 car has visual feedback. Furthermore, the steering wheel area is becoming more important. Whereas the 1954 BMW only has secondary controls mounted on the back of the steering wheel, the 2007 BMW has controls for all three task classes on the front and back of the steering wheel.

4.2 Analysis of IAA2007

Using our proposed design space, we were also able to analyze the photographs taken at IAA 2007 in more detail.

One trend that we found is that the space on the steering wheel is often used for controls, e.g. for hands-free interaction with mobile phones or controlling the entertainment system. 78% of the cars have controls on the steering wheel. Another trend is the use of displays in cars for navigation systems and other comfort functionalities. 72% of the researched cars already have a built-in display. Display types are evenly balanced between touchscreens and non-touchscreens (46% have a touchscreen). Touchscreens are mostly found in American and Japanese cars, while German cars almost exclusively followed the concept “display controlled with controller”.

An indication of future trends could also be seen in the presented concept cars. Citroen, for instance, has a display and the main controls on the steering wheel in their concept car “Cactus”[3]. In general, we observed that the display space in concept cars is much bigger than in current cars. Displays for front-seat passengers are also prevalent.

4.3 Looking for New Ideas

With the introduction of automatic-transmission cars, the clutch pedal disappeared, freeing up space for other controls. It would be interesting to see if the left foot could be used for interaction with controls in this space, e.g. for zooming in/out in a navigation system. Currently, input modalities on the steering wheel consist of buttons and thumbwheels. The Citroen concept car Cactus, however, already has a display mounted on the steering wheel. It might be interesting to look more into new input and output opportunities on a steering wheel. Handwriting input on a steering wheel, for example, may be easier than in the center stack for left-handed people in cars with the steering wheel on the left side or for right-handed people in a car with the steering wheel on the right side.

With head-up displays, the windshield is also becoming an important new area for output modalities. In addition to providing visual feedback for systems, the windshield area may also hold opportunities for spatial audio.

The front-seat passenger area also provides open space that is not directly represented in the design space, because it can’t be concretely used by the driver, but one can imagine having an additional screen there where the passenger can interact with in-car systems, e.g. enter entries in the navigation system, and send the results to the driver’s screen.

It is also visible from the design space that new modalities (e.g. haptics) can find spaces that are not yet occupied by other controls.

5. CONCLUSION

In this paper, we presented a design space for driver-based automotive user interfaces with respect to the placement of devices in the car and the body parts that interacts with these devices. Our design space is based on an analysis of photographs taken from 117 different cars from 35 manufacturers. We discussed different input and output modalities in cars and presented a graphical representation for categorizing individual cars that should help user interface designers analyze existing layouts, generate new ideas, and find unexplored areas for future designs. Furthermore, we provide a more abstract graphical representation for comparing a set of cars to find concrete similarities and differences between different manufactures or different types of cars, e.g. comparing middle-class and luxury cars.

The design space is based on the analysis of left-hand cars but it can be used for right-hand cars as well. When comparing cars, it is easiest to analyze only one type of car (left or right handed) with this design space. Cross-comparisons are also possible but require changing left and right columns for the dashboard and the center stack for either the right-hand or left-hand cars.

We discussed the usage of the design space by looking at historical changes and trends as well as differences between the 117 cars based on photographs taken at IAA2007.

We showed that this design space can be used as a tool for comparing different user interface options and layouts as well as a means to facilitate a structure discussion of existing and future car user interfaces.

In the future we plan to include additional measure, possibly automatic, that detect potential design flaws that would impact driver performance. We envision a software tool that assists designers in choosing and placing controls into the design space. This software tool might allow the designer to mark specific controls in a picture or in a design sketch, from which a graphical representation could automatically be generated. The tool might also provide immediate estimated feedback on the impact of the control placement on the driver, e.g. with regard to visual load or cognitive load.

Both graphical representations shown in Fig. 5 can be found at https://www.pcuie.uni-due.de/AUI/**.** This wiki can be used to exchange design spaces with others. The design spaces are ordered by label and year of construction. Additionally, a design space template is available at https://www.pcuie.uni-due.de/AUI/**.**

6. ACKNOWLEDGEMENTS

This work was performed in the context of the DFG (Deutsche Forschungsgemeinschaft) funded research group 'Embedded Interaction'.

7. REFERENCES

[1] Alliance of Automobile Manufacturers: Statement of Principles, Criteria and Verification Procedures on Driver

[3] http://wikicars.org/en/Citro%C3%ABn_Cactus-C

Interaction with Advanced In-Vehicle Information and Communication Systems, 2003.

[2] Audi glossary MMI-Multi Media Interface: http://www.audi.co.uk/audi/uk/en2/tools/glossary/operation_comfort/MMI___Multi_Media_Interface.html

[3] Ballagas, R., Rohs, M., Sheridan, J., Borchers, J., The Design Space of Ubiquitous Mobile Input. In Joanna Lumsden, editor, *Handbook of Research on User Interface Design and Evaluation for Mobile Technologies*. IGI Global, Hershey, PA, USA, 2008.

[4] Bishop, R, Intelligent vehicle technology and trends. London: Artech House Publishers, 2005.

[5] BMW 7 Series BuisnessWeek http://www.businessweek.com/lifestyle/content/sep2008/bw2008099_699585_page_2.htm

[6] BMW Technology Guide iDrive: http://www.bmw.com/com/en/insights/technology/technology_guide/articles/idrive.html

[7] Burnett, G.E, Designing and evaluating in-car user-interfaces. In J. Lumsden (Ed.) *Handbook of Research on User-Interface Design and Evaluation for Mobile Technology*, Idea Group Inc. 2008.

[8] Buxton, W., Lexical and pragmatic considerations of input structures. SIGGRAPH Comput. Graph., 17(1), 31-37, 1983.

[9] Card, S. K., Mackinlay, J. D., Robertson, G. G., A morphological analysis of the design space of input devices. ACM Trans. Inf. Syst., 9(2), 99-122, 1991.

[10] Card, S. K., Mackinlay, J., The structure of the information visualization design space, Proceedings of the 1997 IEEE Symposium on Information Visualization (InfoVis '97), p.92, October 18-25, 1997

[11] Citroen UK Technology – Safety – Lane Depature Warning System: http://www.citroen.co.uk/technology/safety/ldws/

[12] Commission of the European Communities: Commission Recommendation of 22 December 2006 on safe and efficient in-vehicle information and communication systems: Update of the European Statement of Principles on human machine interface, 2006.

[13] DIN Deutsches Institut für Normung. Road vehicles - Ergonomic aspects of transport information and control systems - Dialogue management principles and compliance procedures, Mar. 2003. Ref.No.: DIN EN ISO 15005:2003.

[14] DIN Deutsches Institut für Normung. Road vehicles - Ergonomic aspects of transport information and control systems - Specifications and compliance procedures for in-vehicle visual presentation, Oct. 2003. Ref.No.: DIN EN ISO 15008:2003-10.

[15] Foley, J.D., Wallace, V.L., Chan, P., The human factors of computer graphics interaction techniques. IEEE Comput. Graph. Appl. 4, 11 (Nov 1984), 13-48.

[16] Geiser, G. Man Machine Interaction in Vehicles. ATZ 87, 1985. Pages 74 - 77.

[17] National Household Travel Survey: Commuting for Life, November 2006, http://nhts.ornl.gov/briefs/Commuting%20for%20Life.pdf

[18] Pettitt, M.A., Burnett, G.E. and Stevens, A. Defining driver distraction. Proc. World Congress on Intelligent Transport Systems, 2005

[19] Stevens, A., Quimby, A., Board, A., Kersloot, T. and Burns, P. (2002). Design guidelines for safety of in-vehicle information systems. Project report PA3721/01, TRL Limited.

[20] Tönnis, M., Broy, V., and Klinker, G.. A Survey of Challenges Related to the Design of 3D User Interfaces for Car Drivers. In Proceedings of the 1st IEEE Symposium on 3D User Interfaces (3D UI), March 2006.

Assessing Subjective Response to Haptic Feedback in Automotive Touchscreens

Matthew J. Pitts Mark A. Williams Tom Wellings Alex Attridge
WMG, University of Warwick
Coventry
CV4 7AL, United Kingdom
+44 (0) 24 7657 5936
{m.pitts; m.a.williams.1; t.wellings; a.attridge}@warwick.ac.uk

ABSTRACT
The increasing use of touchscreen interfaces in vehicles poses challenges to designers in terms of optimizing safety, usability and affective response. It is thought that the application of haptic feedback to the touchscreen interface will help to improve the user experience in all of these areas. This paper describes the initial outcomes of a study to investigate user responses to haptic touchscreens using a simulated driving scenario based on the Lane Change Test, along with representative use case tasks. Results indicate preference for multi-modal feedback and user acceptance of the haptic feedback technology. Effects relating to multi-modal interaction and attentional demand are also observed.

Categories and Subject Descriptors
H.5 [Information Interfaces and Presentation]: User Interfaces - *Auditory (non-speech) feedback, Haptic I/O, Input devices and strategies (e.g., mouse, touchscreen), User-centered design*

General Terms
Experimentation, Human Factors

Keywords
Touchscreen, Haptic Feedback, HMI, Automotive

1. INTRODUCTION
The use of touch screen interfaces in both mobile devices and automotive technology is rapidly increasing [1] as more vehicle manufacturers adopt touchscreen-based HMI solutions for their latest vehicle line-ups [2][3]. Optimising usability and acceptance poses challenges to designers in both fields [4].

While touchscreens have usability benefits compared to centralised controllers as used by Audi and BMW for example [5], the interface places significant visual attention demand on the driver due to the lack of tactile and kinesthetic feedback [6]. Historical data shows that eye glances away from the road contribute to 60% of crashes, near-crashes and incidents [7]; re-introducing haptic feedback to provide confirmation of inputs may negate the requirement for secondary glances, thereby reducing the overall attention requirements of the touchscreen interface and improving both safety and the user experience.

There are potentially additional benefits in terms affective response to an interactive haptic interface. Research into the use of touch as an enhanced marketing tool found that touch created a enjoyable hedonic experience for the consumer [8]; given that preferences for feel characteristics in pushbutton vehicle interfaces are known to exist [9][10], user-selectable haptic effects would allow the user to personalise their experience to match their own tastes and requirements, thus enhancing their experience [11].

A study into haptic feedback in mobile devices with touchscreen interfaces [12] compared text entry tasks using a software keyboard with and without haptic feedback enabled. Results showed an improvement in error rates and reduced subjective workload with the addition of haptic feedback. In another study, haptic feedback was shown to reduce error rates and task completion time in a scrolling task on a handheld device [13]. Serafin et al [14] showed subjective preference for tri-modal (visual, audible and haptic) feedback from a touchscreen interface in on-bench and static vehicle trials; however, these trials were conducted in the absence of any external tasks requiring attentional resource.

2. EXPERIMENTAL APPROACH
A study was proposed to investigate the response of drivers towards touchscreens fitted with haptic feedback capability in an automotive scenario. The objective of the study was to ascertain the benefits of haptic feedback compared to existing modes of feedback commonly employed on in-vehicle systems, i.e. visual and audible feedback.

The research questions were as follows:

- Does touchscreen task performance improve with audible or haptic feedback?
- Do users show a subjective preference for audible or haptic feedback on touchscreens?
- Does the presence of audible or haptic feedback affect the demand level of touchscreen tasks?

AutomotiveUI'09, September 21-22, 2009, Essen, Germany
ACM 978-1-60558-571-0/09/009

- Is there a relative preference for either audible or haptic feedback?

The hypothesis was formed that the presence of haptic feedback would improve both objective and subjective measures of performance and affect.

As the objective of the study focused on in-vehicle touchscreen use, it was important to consider the context of the evaluation. In order to provide a degree of context and to introduce an element of cognitive workload, the evaluation tasks were conducted in a simulated driving environment based on the Lane Change Task software [15].

The driving task requires the user to change lanes on a straight road in response to signs positioned at the side of the road; as the signs are regularly spaced the driver is subject to a constant workload requirement. This approach also allows for collection of data on the performance of the lane change task (e.g. mean lane deviation) which may indicate differences between experimental conditions.

Figure 1 - Evaluation setup

The touchscreen evaluation tasks themselves were based on real-life use cases for an automotive application, described in section 2.3.2. The interface application was designed to log key presses and timings, allowing calculation of error rate and task completion time metrics to evaluate performance across experimental conditions; this approach has previously been used to illustrate benefits of haptic feedback on mobile devices [13].

In addition to the objective metrics described above, users were required to provide subjective measures of task performance and affect following each evaluation stage. This paper will concentrate on the collection and analysis of this subjective information, with further analysis of the objective data remaining as future work.

2.1 The Haptic Touchscreen Interface

The experiments were conducted using a Touchsense 8.4" LCD touchscreen demo unit from Immersion Corporation [16] – this device is supplied pre-fitted with haptic feedback actuators and control hardware and forms the visual and haptic display elements, as well as the touch input device. The graphical interface used for the trials was based on a production vehicle touchscreen GUI and was programmed in Adobe Flash CS3 and ActionScript 3.0. All interface functions were operated with 'pushbutton'-type controls.

2.2 Pre-Trial Study

It was necessary to select one effect for use in the main trial in order to remove effect type as a variable and minimise negative affective responses. A pre-trial desk-top study to determine preference was conducted using 34 respondents from the automotive industry. Of these, 17 (50%) respondents described themselves as 'experts' in touchscreen interface design.

The Touchsense unit features a palette of pre-programmed haptic effects which can be called from software. These are grouped into five types: "Pulse Click", "Crisp Click", "Smooth Click" "Double Click", and "Complex". Effects within the 'Click' groups vary by magnitude and repeat rate, while the "Complex" effects exhibit much wider variations in both magnitude and character; these were therefore excluded from the evaluation. In order to reduce individual differences in touchscreen usage, all respondents were required to operate the screen with their left hand as per an in-car scenario (right-hand drive). Furthermore, respondents wore ear defenders during the evaluation to reduce cross-modal influence from the audible output of the haptic touchscreen actuators.

Respondents were presented with a series of screens, each having five buttons programmed with different feedback stimuli taken from one of the four 'Click' effect groups. The presentation order was randomised to reduce magnitude order effects. The respondent was asked to choose their most preferred 'feel' from the group of five, before moving onto the next screen where they were presented with effects from a different effect group. Once the respondent had chosen their preferred effect from each group, they were presented with a fifth screen comprising their previous preference selections and asked to make a final choice to determine overall preference.

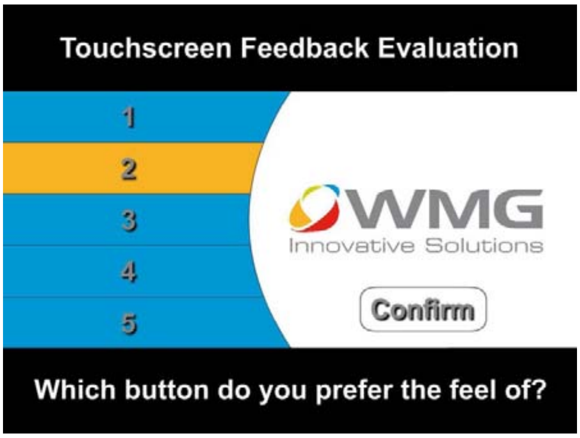


Figure 2 - Pre-trial interface screenshot

The pre-trial study also provided valuable insights into issues surrounding interaction with and implementation of the haptic touchscreen, including effect perceived quality and latency.

The results of the pre-trial indicated a preference for the "Crisp Click" effect type, with 16 of the 34 respondents selecting effects from this group as most preferred (see Figure 3). A binomial test of this result showed significance ($p<0.05$). There was no significant preference for one discrete effect, with three effects receiving similar scores. The effect used for the main study was chosen from these three after discussion within the research group.

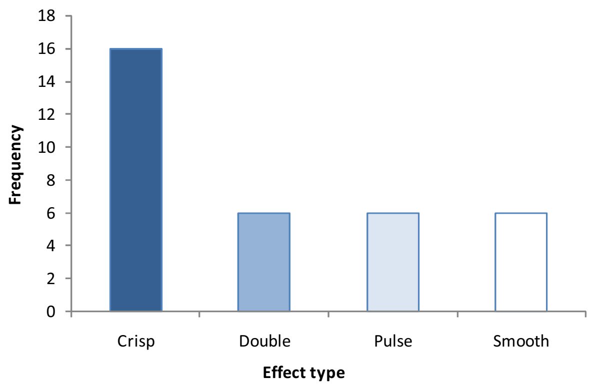


Figure 3 – Histogram of preferred haptic effect type

2.3 Main Study

2.3.1 Respondent selection

A total of 54 respondents participated in the study, with 48 completing the evaluation; six respondents were withdrawn from the study after either showing symptoms of simulator sickness or exhibiting poor driving performance. Selection criteria determined that all respondents were car drivers and had experience of in-car touchscreen use. The demographic breakdown is given in Table 1.

Table 1 - Respondent demographics

	Female		Male		Totals	
Age range	**Count**	**% of Total**	**Count**	**% of Total**	**Count**	**% of Total**
18-25	2	4%	3	6%	5	10%
26-35	4	8%	4	8%	8	17%
36-45	12	25%	5	10%	17	35%
46-55	2	4%	6	13%	8	17%
56+	1	2%	9	19%	10	21%
	21	44%	27	56%	48	100%

There was an exact 50%:50% split between users of portable touchscreen devices (such as handheld navigation units) and factory-installed touchscreen systems.

2.3.2 Experiment design

To test the hypothesis, respondents were presented with a series of use-case trials, based on operations which may be performed on an in-car touchscreen interface. Each set of trials was completed four times, once for each of the following combinations of feedback:

- Visual feedback only
- Visual + Audible feedback
- Visual + Haptic feedback
- Visual + Audible + Haptic feedback

One haptic feedback effect was used throughout the study to remove feedback type as a factor – this was a 'Crisp Click' type effect chosen based on the results of the pre-trail study. The audible stimulus was the acknowledgement 'beep' used on the touchscreen interface of a production premium saloon.

A screenshot of the evaluation interface is shown in Figure 4. As mentioned previously, the use cases were selected to encompass a range of functionality across the system, including climate control, audio system and telephone tasks, requiring different numbers of button presses and levels of menu navigation. These are summarised in Table 2.

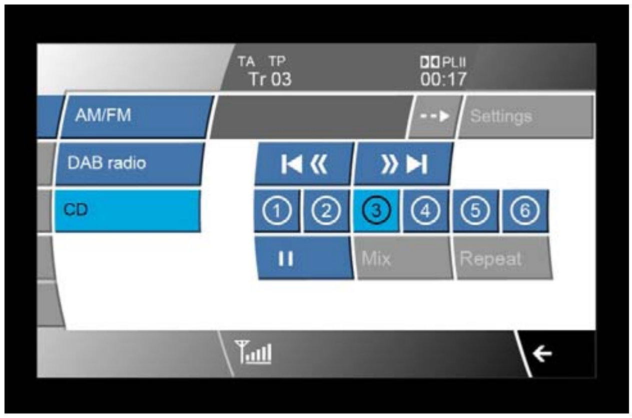


Figure 4 - Evaluation interface screenshot

For each of the feedback states, the use cases were modified where possible to reduce learning effects, for example by requesting a different DAB preset or fan speed setting. The order of presentation of use cases was predetermined, randomised between feedback states. The presentation of feedback states was counterbalanced for presentation order.

Table 2 - Use cases

Task	Button presses required	Menu levels
Set seat heating/cooling	3	0
Tune FM radio to given frequency	3	1
Select DAB station preset	3	2
Play track 4 from given CD	7	2
Set fan speed	4	1
Set climate control to auto/off	2	1
Dial phone number & start call	13	1
Select number from phone book	4	3

2.3.3 Training

A multi-stage training process was applied prior to the start of the evaluation in order to minimize learning effects. Firstly, respondents were shown a simple interface on the haptic touchscreen consisting of four buttons, each programmed to deliver a different combination of feedback stimuli, as shown in Figure 5. Audible signals were delivered over headphones which also provided acoustic isolation from the audible output of the haptic touchscreen actuators. For the purpose of simplicity, haptic feedback was referred to throughout the experiment as "Touch feedback". Respondents were asked to confirm that they could sense each stimulus, and that they understood the terminology used.

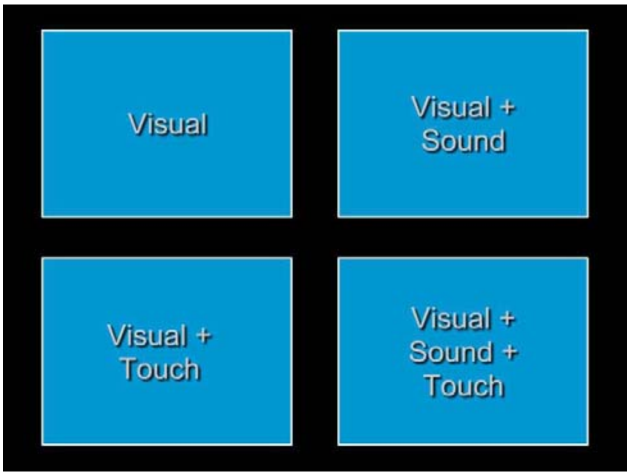


Figure 5 - Feedback introduction screen

Respondents were then introduced to the interface that would be used for the evaluations and instructed as to its functionality. After a period of familiarisation, the respondent was asked to demonstrate the completion of each of the use cases involved in the task.

Once familiar with the evaluation interface, the respondent was introduced to the driving task. The respondent was given instruction on the operation of the driving simulator, then asked to undertake a trial run. Additional instruction was provided for the initial part of the run until driving performance was deemed satisfactory. Due to the basic nature of the driving task the majority of respondents reached this status early in the trial run. The final stage of the learning process was to undertake a mock evaluation, whereby the respondent was required to operate the touchscreen while performing the driving task. Use cases were selected at random and instructions given verbally over the headphones.

2.3.4 Questionnaire design

Following each set of evaluations, respondents' subjective impressions were recorded using a questionnaire. Three types of rating scale were used [17]:

- 9-point hedonic rating scale - used to assess the overall liking for touchscreen use
- 9-point rating scale with verbal anchors at end and mid-points. This rating scale was used to assess usability elements of the task: confidence in choice, difficulty of operation while driving, interference with the driving task. For the trials including haptic feedback, additional questions were included on feedback realism (compared to real switch) and strength of the feedback stimulus
- 5-point Likert scale. This method was used to assess impressions of the technology concept across the different feedback states.

At the end of the evaluation, an additional questionnaire was presented. This consisted of two sections: in the first, respondents were asked to indicate their most and least preferred feedback combinations. The second section contained two questions aimed to assess the level of acceptance of haptic touchscreens, using a five-point Likert scale to measure the level of agreement with the statements: "Touch feedback makes the touchscreen more pleasurable to use"; and "Touch feedback makes the touchscreen easier to use"

3. Results

Of the 48 respondents who completed the study, five indicated in post-completion comments that they were not able to feel the haptic feedback stimulus; these respondents' data were therefore excluded from the analysis on the basis of their responses being unreliable. A further three respondents displayed extreme outliers in their responses and were also removed from the analysis.

Data from the remaining 40 participants was analysed to determine statistical significance across feedback types using the non-parametric Friedman's test. The paired Wilcoxon signed-ranks test with Bonferroni correction is used to determine post-hoc pair-wise significance at the 95% family-wise confidence level ($p_{crit} = 0.0085$).

A selection of findings from the analysis is shown below.

Figure 6 shows the mean hedonic rating for each feedback state. There is a clear trend for improved rating with multi-modal feedback which is shown to be significant ($p < 0.001$). The mean score of 6.00 for 'Visual only' feedback corresponds to the rating 'Like slightly' on the hedonic scale, while the mean value of 7.60 for the 'Visual + Audible + Haptic' state lies between the anchor points 'Like slightly' and 'Like very much'.

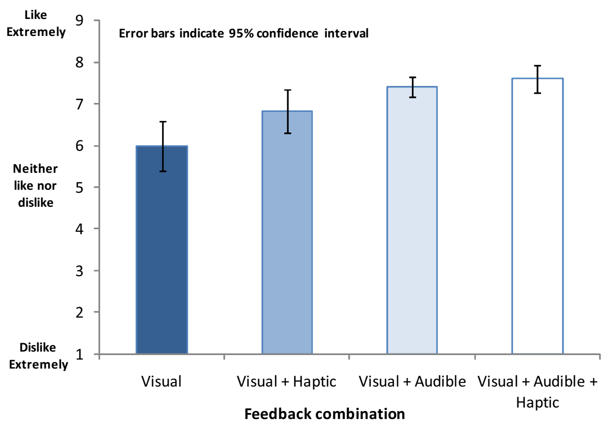


Figure 6 - Mean hedonic rating scores for each feedback combination. Sample size = 40

Post-hoc tests indicate that hedonic rating is improved from the 'Visual only' state with the addition of audible or combined audible and haptic feedback and that 'Visual + Audible' feedback shows an improvement over 'Visual + Haptic' (Table 3).

Table 3 - Mean scores, standard deviation and Wilcoxon Signed Rank pair-wise p-values for Q1: Hedonic Rating. Sample size = 40

Mean and Standard Deviation

	V	VH	VA	VAH
Mean	6.00	6.83	7.40	7.60
SD	1.91	1.66	0.78	1.08

Pair-wise p-values

	V	VH	VA	VAH
V	-	>0.05	**<0.0001**	**<0.0001**
VH		-	**0.0009**	0.0136
VA			-	>0.05
VAH				-

P_{crit} = 0.0085 (Family-wise $\alpha = 0.05$, 2-sided test)

Values in bold are significant

The trend across feedback types is repeated for confidence rating (Figure 7), with the 'Visual only' state attracting the lowest mean score and 'Visual + Audible + Haptic' the highest: a mean of 7.00, which lies between 'Moderately' and 'Extremely confident' on the rating scale. Differences across feedback states were shown to be significant ($p < 0.001$), with post-hoc tests showing improved confidence from the 'Visual only' state with the addition of audible or combined audible and haptic feedback, and improvement from the 'Visual + Haptic' state with the addition of audible feedback (Table 4).

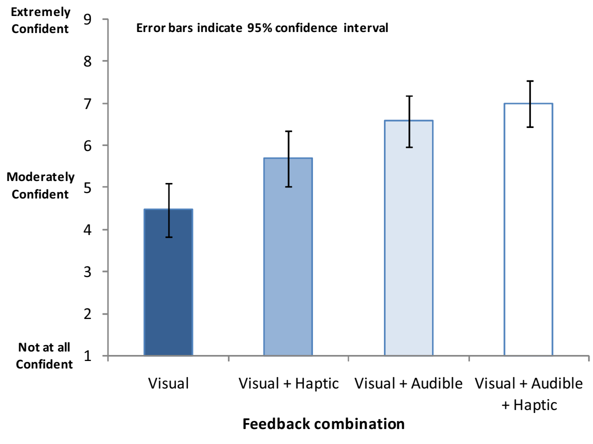


Figure 7 - Mean confidence rating scores for each feedback combination. Sample size = 40

Table 4 - Mean scores, standard deviation and Wilcoxon Signed Rank pair-wise p-values for Q2: Confidence in button press. Sample size = 40

Mean and Standard Deviation

	V	VH	VA	VAH
Mean	4.48	5.70	6.58	7.00
SD	2.06	2.14	1.95	1.80

Pair-wise p-values

	V	VH	VA	VAH
V	-	0.0099	**0.0002**	**<0.0001**
VH		-	>0.05	**0.0030**
VA			-	>0.05
VAH				-

P_{crit} = 0.0085 (Family-wise $\alpha = 0.05$, 2-sided test)

Values in bold are significant

Ratings of touchscreen task difficulty show a small but significant difference across feedback states ($p < 0.05$), with means ranging from 4.00 for 'Visual only' to 4.98 for 'Visual + Audible + Haptic' – an increase of only one scale point with the addition of multimodal feedback. Increases in standard deviation for multimodal feedback states suggest that some respondents found the touchscreen tasks consistently difficult and did not realise benefits from multimodal feedback. Post-hoc analysis showed significant differences for the 'Visual / Visual + Audible' and 'Visual / Visual + Audible + Haptic' pairs only. Rating scores for 'Interference with the driving task' follow the same pattern of mean scores and significant differences, with the highest mean rating of 4.40 for 'Visual + Audible + Haptic' indicating a 'more than moderate' level of interference.

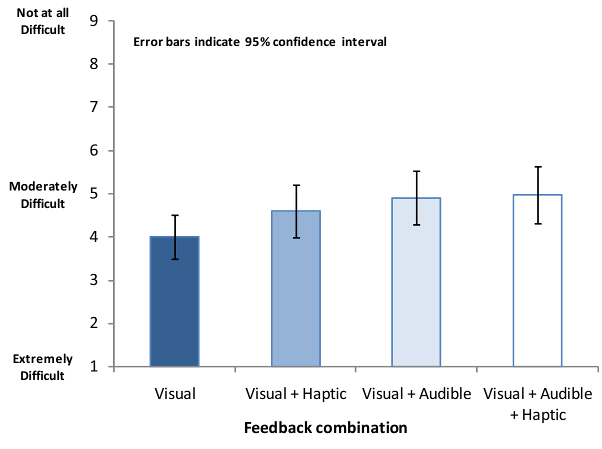


Figure 8 - Mean difficulty rating scores for each feedback combination. Sample size = 40

Table 5 - Mean scores, standard deviation and Wilcoxon Signed Rank pair-wise P-values for Q3: Difficulty in operating touchscreen while driving. Sample size = 40

Mean and Standard Deviation

	V	VH	VA	VAH
Mean	4.00	4.60	4.90	4.98
SD	1.63	1.96	2.00	2.13

Pair-wise p-values

	V	VH	VA	VAH
V	-	>0.05	**0.0044**	**0.0043**
VH		-	>0.05	>0.05
VA			-	>0.05
VAH				-

$P_{crit} = 0.0085$ (Family-wise $\alpha = 0.05$, 2-sided test)

Values in bold are significant

The reported strength of the haptic feedback stimulus also showed significant differences in mean rating with and without audible feedback ($p < 0.001$), indicating that the haptic effect was perceived as 'more strong' in the presence of audible feedback. This suggests a multi-modal effect whereby the presence of the audible feedback reinforces the perception of the haptic stimulus. The mean rating of 3.51 for 'Visual + Haptic' indicates that, on average, the strength of the haptic effect was less than optimal, as a score of 5 indicates 'Just right'. Note the smaller sample size of 35, due to this question being added part way into the study.

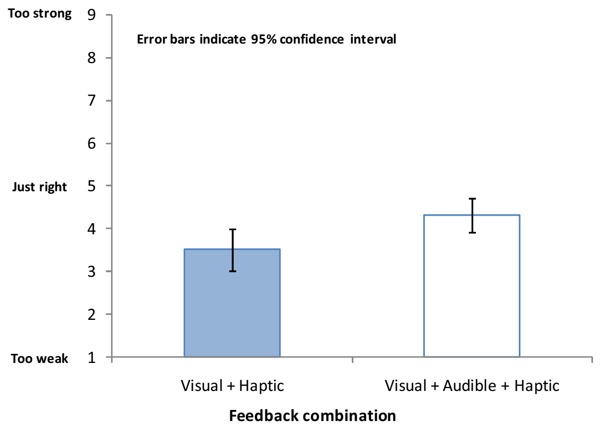


Figure 9 - Mean feedback strength rating scores for each feedback combination. Sample size = 35

While previous results do not indicate significant benefits for the addition of haptic feedback, alternative measures show user acceptance of the technology. Figure 10 shows the number of times each effect state was chosen as most or least preferred, with the least preferred choices shown as negative. A clear preference for combined visual, audible and haptic feedback can be seen, with 24 choices from 40 respondents – double that of the 'Visual + Audible' state; indicating that haptic feedback is seen as desirable by the user.

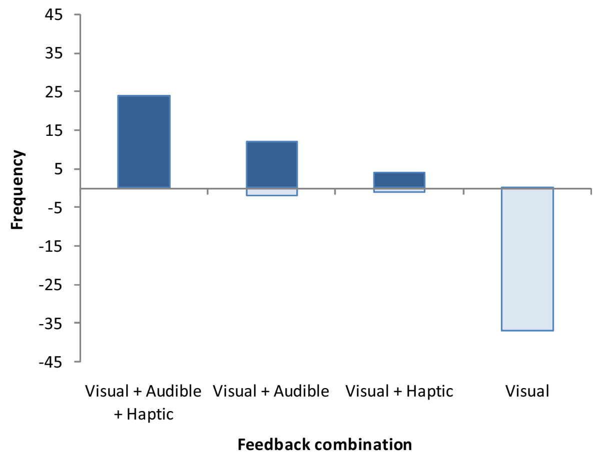


Figure 10 - Histogram of most/least preferred feedback state. Most preferred shown as positive, least preferred shown as negative. Sample size = 40

Further evidence for acceptance of haptic feedback is given by the responses to the questions 'Touch feedback makes the touchscreen more pleasurable to use' and 'Touch feedback makes the touchscreen easier to use'. The mean scores for these questions are 4.14 and 4.34, where a score of 4 corresponds to 'Agree' on the Likert scale.

3.1 Order effects

Results for questions relating to hedonic rating, touchscreen task difficulty and driving task interference were each analysed for differences due to the order of presentation of the feedback states using the Friedman's test ($\alpha = 0.05$). A significant order effect was found for driving task interference ($p < 0.001$), indicating that participants experienced less interference with the driving task when operating the touchscreen as the study progressed. Figure 11 shows the variation in mean interference rating with presentation order.

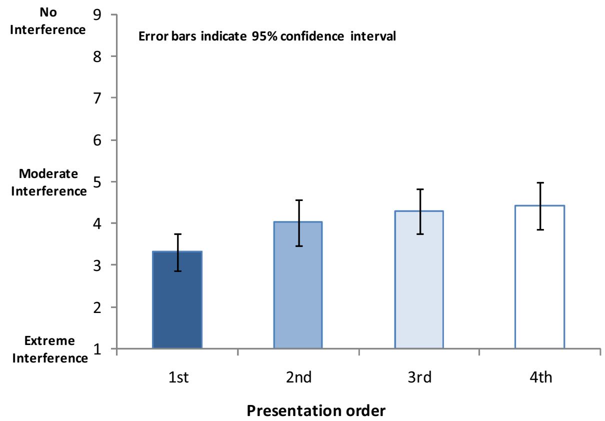


Figure 11 – Variation in mean interference rating with presentation order. Sample size = 40

4. Discussion

Hedonic ratings scores indicate a preference for tri-modal feedback, showing a trend across feedback states which is repeated for ratings of confidence. While combined visual, audible and haptic feedback attracts the highest mean scores, no significant differences are shown for the addition of haptic feedback to other feedback states. This concurs with the findings of Serafin et al [14], which indicated preference for "enhanced" (multi-modal) feedback.

Alternative measures were also used to assess users' affective response to haptic feedback. Combined visual, audible and haptic feedback was chosen as 'most preferred' by 24 of 40 respondents, double that of 'Visual + Audible'. This, along with results indicating that users 'agree' that haptic feedback makes the touchscreen both easier and more pleasurable to use indicate strong user acceptance of the technology. Kern et al [18] found that, while tactile feedback did not show benefits in driving performance, anecdotal evidence from participant comments suggested strong liking for combined audible and tactile feedback, citing advantage gained from reinforcement of perception of the signal when received simultaneously in two modalities.

Looking at the results for the 'Haptic feedback strength' question suggests that, when experienced without its audible counterpart, the haptic feedback stimulus was not sufficiently strong to provide a positive confirmation to the respondent. The haptic effect chosen was selected on the basis of a pre-study trial involving expert and non-expert users; one might assume that this process would reject effects that are 'Too weak'. Indeed, all respondents in the main trial confirmed that they could perceive the haptic feedback during the learning phase. However, a number of respondents also indicated that they had difficulty feeling the feedback during the evaluation tasks. The suggestion is therefore that simultaneous performance of the driving and touchscreen tasks imposes an attentional load which reduces the respondent's ability to perceive haptic stimuli – this agrees with by Leung et al [19], who observed similar differences in haptic sensitivity when participants were distracted.

Previous discussions with applications engineers have highlighted the potential for tuning haptic stimuli to account for background (vibration) noise in the vehicle environment, but the effect of attentional demand had not been discussed. The ability to tune effect magnitude would also compensate for individual differences in sensitivity to haptic stimuli, as well as allowing a user to tune the device to suit their personal preferences, thus maximising the affective benefits of a tactile interface discussed earlier.

The difference in rating for perceived haptic effect strength in the presence of the audible stimulus also suggests a multi-modal interaction effect. An interesting avenue of further study would be variations in perception of haptic effects with age; while vibrotactile sensitivity in the hand is known to decrease in older adults in a similar way to visual and auditory acuity [20], multi-modal stimuli have been shown to restore response times of older participants to those demonstrated by young subjects for single stimuli, suggesting that multi-modal feedback can compensate for age-related sensory degradation [21]. Unfortunately it was not possible to draw significant conclusions on age-related sensitivity effects from the study data.

Care was taken with to minimise effects of presentation order, through a counterbalanced experiment design and a multi-stage pre-trial training process. Analysis of order effects indicated that, while presentation order has no effect of hedonic rating or touchscreen task difficulty, there was a significant effect on interference with the driving task over the duration of the study, with the level of interference becoming lower as the study progressed. As the perceived difficulty of the touchscreen task was constant throughout (no significant order effect), it may be assumed that the perceived demands of concurrent performance of the touchscreen and driving tasks diminished as the study progressed. Additional training or practice time may have reduced this effect, although it may also be the case that the nature of the driving task was also a factor; the fact that the highest mean score achieved for the interference measure indicated a 'more than moderate' level of interference suggests that the demands of the combined tasks was relatively high.

A total of 8 respondents were rejected from the analysis due to unreliable responses. It is valid to question the effect that this may have on the balance of the experiment. Again, order effect analysis on the reduced data set showed no significant effect on hedonic or confidence rating, suggesting that removal of these data was not detrimental.

5. Conclusions

Results indicated a preference for multi-modal feedback over visual feedback only. Measures of hedonic rating and confidence did not show significant improvements with the addition of haptic feedback; however, the combination of visual, audible and haptic feedback consistently attracted the highest ratings – this combination was chosen as 'most preferred' by twice as many respondents as 'Visual + Audible'. Furthermore, respondents 'Agree' that haptic feedback makes the touchscreen interface both easier and more pleasurable to use. Differences in the perceived haptic effect strength with and without the addition of audible feedback indicate multi-modal interaction effects, while reported issues with sensitivity whilst engaged in the driving task suggest an effect on feedback perception caused by attentional load.

5.1 Further work

Analysis of the subjective data from this study has yielded some interesting results with respect to affective response to haptic touchscreen technology. Further insight will be gained from analysis of the objective data also gathered during the experiment, which will illustrate the relationship between task performance and affective response.

A follow-up study using an improved driving simulator environment is scheduled for summer 2009. This will allow the validation of existing experimental results and allow hypotheses bases on anecdotal findings of this research to be tested. Furthermore, this presents to opportunity to employ additional objective measures such as eyes-off-the-road time.

6. Acknowledgements

The authors would like to thank Dr. Carl Pickering and colleagues at Jaguar Land Rover Research for their technical support and input into this study. This research was funded by EPSRC and was conducted under the Warwick Innovative Manufacturing Research Centre (WIMRC) project "Designing the next generation of HMI for Future Intelligent Vehicles".

7. References

[1] "Touch-Screen Sales to Double by 2012". http://www.cellular-news.com/story/24451.php. 20th June 2007, Accessed 29th May 2009.

[2] Jurnecka, R.: "All Volkswagens built after 2008 to get touchscreen system". http://blogs.motortrend.com/6220413/car-news/all-volkswagens-built-after-2008-to-get-touchscreen-system/index.html, November 12th 2007, accessed 19/01/09

[3] Newcomb, D.: "The Touch Screen Is Dead (In Acuras), Long Live the Touch Screen (In Other Makes)". http://blogs.edmunds.com/strategies/2008/12/thetouch-screen-is-dead-in-acuras-long-live-the-touch-screen-in-other-makes.html. December 16th 2008, accessed 19/01/09

[4] "Consumers warm to the touch screen phone" Canalys research release 2007/061. http://www.canalys.com/pr/2007/r2007061.pdf; Accessed 29th May 2009

[5] Rydstrom, A., Bengtsson, P.,Grane, C., Brostrom, R., Agardh, J, Nilsson, J.: "Multifunctional Systems in Vehicles: A Usability Evaluation". Proc. CybErg 2005.

[6] Burnett, G.E., Porter, J.M.: "Ubiquitous computing within cars: designing controls for non-visual use" International Journal of Human-Computer Studies, 55. 2001. pg 521-531

[7] Klauer, S.G., Dingus, T. A., Neale, V. L., Sudweeks, J.D., Ramsey, D.J.: "The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data". National Highway Traffic Safety Administration Report DOT HS 810 594. April 2006. Pg 99-109

[8] Peck, J., Wiggins, J.: "It Just Feels Good: Customers' Affective Response to Touch and Its Influence on Persuasion" Journal of Marketing Volume 70, Issue 4. October 2006, Pg56-69

[9] Wellings, T., Williams, M., Pitts, M: "Customer perception of switch-feel in luxury sports utility vehicles", Food Quality and Preference, Volume 19, Issue 8, December 2008, Pg 737-746.

[10] Wellings, T., Williams, Tennant, C: "Understanding customers' holistic perception of switches in automotive human-machine interfaces", Applied Ergonomics, In Press, Available online 17 April 2009, DOI:10.1016/j.apergo.2009.03.004.

[11] MacLean, K.E.: "Designing with haptic feedback". Proceedings ICRA '00: IEEE International Conference on Robotics and Automation, 2000. pg.783-788.

[12] Brewster, S., Chohan, F., Brown, L.: "Tactile Feedback for Mobile Interactions". Proceedings of the SIGCHI conference on Human factors in computing systems, 2007. Pg 159 – 162.

[13] Poupyrev, I., Maruyama, S., Jun, R.: "Ambient Touch: Designing Tactile Interfaces for Handheld Devices". CHI Letters 4 (2), 2002, pg 51-60

[14] Serafin, C., Heers, R., Tschirhart, M., Ullrich, C., Ramstein, C.: "User Experience in the U.S. and Germany of In-Vehicle Touch Screens with Integrated Haptic and Auditory Feedback". SAE world congress 2007, 2007-01-0451

[15] Mattes, S.: "The Lane-Change-Task as a Tool for Driver Distraction Evaluation". In: Strasser, H., Kluth, K., Rausch, H., Bubb, H. (eds.) Quality of Work and Products in Enterprises of the Future, 2003 pg. 57–60.

[16] Touchscreen Feedback Overview, Immersion Corporation. http://www.immersion.com/products/touchscreen-feedback/index.html. Accessed 1st June 2009

[17] Stone, H., Sidel, J. L. (eds): "Sensory Evaluation Practices" (Third Edition) 2004. Pg 79-139

[18] Kern, D., Marshall, P., Hornecker, E., Rodgers, Y., Schmidt, A.: "Enhancing Navigation Information with Tactile Output Embedded into the Steering Wheel". Proceedings of the Seventh International Conference on Pervasive Computing, 2009

[19] Leung, R., MacLean, K., Bertrlsen, M.B., Saubashik, M.: "Evaluation of Haptically Augmented Touchscreen GUI Elements Under Cognitive Load". Proceedings of the 9th international conference on Multimodal interfaces, 2007, pg 374-381

[20] Jones, L.A., Lederman, S.J.: "Human Hand Function", Oxford University Press US, 2006. Pg 152-155

[21] Laurienti, P.J., Burdette, J.H., Maldjian, J.A., Wallace, M.T.: "Enhanced multisensory integration in older adults" Neurobiology of Aging, Volume 27, Issue 8, August 2006, Pg 1155-1163

Towards an H-Mode for highly automated vehicles: Driving with side sticks

Dipl.-Ing. Martin Kienle*, Dipl.-Ing. Daniel Damböck*, Dipl.-Ing. Johann Kelsch°, Dr.-Ing. Frank Flemisch°, Prof. Dr. Klaus Bengler*

* Department of Ergonomics
Technische Universität München, Boltzmannstr. 15, 85747 Garching
+49-(0)-89-289-15408
kienle@lfe.mw.tum.de

° Institute of Transportation Systems, DLR e.V.
Lilienthalplatz 7, 38108 Braunschweig

ABSTRACT
The increasing traffic volume confronts the road user with a challenging task. The high number of traffic deaths might not be reducible with passive safety alone. However systems that actively influence the guidance of vehicles, like assistance and automation systems, can make a difference towards higher safety, comfort and efficiency. Some of these systems completely take over single subtasks like speed or distance control. This, in turn can lead to effects like “out of the loop”, where the driver withdraws from the actual task and even stops monitoring. In order to realize a safe automation system, the project H-Mode follows an approach where both, driver and assistance system are simultaneously affecting the vehicle, whereby the operator is kept in the loop and active. Moreover a haptic-multimodal communication between driver and automation is established by using active interfaces. Regarding this communication alternative control elements, especially two dimensional ones have to be considered.

The study presented in this paper compares conventional interfaces (steering wheel and pedals) with different configurations of an active side stick. It is shown, that two dimensional elements have the potential to combine the driver-automation communication with acceptable drivability.

Categories and Subject Descriptors
H.5.2 [User Interfaces]: Haptic I/O

General Terms
Performance, Design, Ergonomics, Experimentation

Keywords
Haptic interaction, active control element, cooperative automation, highly automated vehicles, haptic interaction, automation, assistance, interaction

1. H-MODE: AN INTUITIVE CONTROL CONCEPT FOR HIGHLY AUTOMATED VEHICLES
Technological progress enables more and more automation in vehicles. In the sky, highly automated planes are flying for decades with a relatively high safety record. On the ground, assistance systems like Adaptive Cruise Control (ACC) or Lane Keeping Assistant Systems (LKAS), which enable partially automated driving, can be bought in many cars. Fully automated vehicles have been demonstrated in public traffic [3], in desert and urban challenges [12] and as demonstrator vehicles “cybercars” in city environment [9]. While fully automated cars might be technologically feasible and legally acceptable much further in the future, highly automated cars, where the automation is capable of driving almost autonomous, but where the driver is still kept active and in the loop, might be possible in a near term future [7].

One of the challenges for highly automated vehicles is to reduce a relatively high complexity of the automation into a manageable complexity for the human. Here aviation can only be a limited role model: In most aircraft, two well-trained pilots keep the system safe, a luxury that is usually not available in ground vehicles. New concepts for an intuitive approach to automation that everybody can operate without extensive training have to be developed and tested.

One potential technique for increasing intuitiveness is the use of design metaphors. In the computer domain, the desktop metaphor took a natural desktop as an inspiration for the organisation of a PC user interface with folders, trash cans etc. For intelligent vehicles, the H-Metaphor (Figure 1) takes the natural example of the rider-horse relationship to describe a cooperative interaction between a highly automated vehicle and a driver (H-Mode).

Initially developed for air vehicles [5][6], it is now systematically applied to cars and trucks [2][8].

Figure 1. Design Metaphors as technique to create mental models (Example Desktop-metaphor and H-metaphor)

One of the essential features of the H-Mode is a bi-directional haptic-multimodal coupling with continuous and/or discrete communication between driver and automation.

In order to provide the driver with a haptic feedback of automation recommendations, active control elements are used as a basis. This means that the H-Mode can be driven with conventional, but active interfaces like an active steering wheel and active acceleration pedal. However new and unconventional interfaces like active side sticks might offer benefits that cannot be reached with conventional interfaces, especially regarding the haptic communication between driver and automation.

Although such new control elements might have advantages when driving with assistance/automation, a minimum of drivability has to be ensured in case of a breakdown or shutdown of the assistance system, leading to manual driving. Therefore the following article focuses especially on different ways to configure active control elements for manual driving, in this case an active side stick. The goal is to achieve a potentially similar driving performance as with conventional interfaces.

2. ACTIVE CONTROL ELEMENTS FOR HAPTIC FEEDBACK

Active control elements provide a way to benefit from haptic feedback. Forces, which can be generated by the integrated actuators can be used to transmit vital information to the operator. Therefore the mechanical connection between machine and operator can be separated and replaced by an electronic one. On the one hand, the accompanied decoupling of these by-wire systems makes it possible to completely redesign the interface. On the other hand the induced reduction of information flow aggravates the user's ability to operate the system. The loss of information flow is thereby due to the fact that the operator can only feel the dynamic of the control element, but not the dynamic of the controlled system itself. Therefore the user has to estimate the system's behavior [11] in order to keep the system within safety limits. For technical purposes active operating elements must be distinguished between two concepts: force and position reflective elements [1][4][10]. In the following these drafts are exemplified with driving a side stick based vehicle.

For driving the vessel the operator creates forces on the stick. The underlying spring characteristic of the force reflective operating element (see Figure 2) determines its movement with addition of the load injected by the operator. Through the stick position the user adjusts the setpoint settings of the vehicle. Consequently the dynamic of the stick is autonomous and does not predicate conclusions about the vehicle's state. This means that, for example in lateral direction, the driver manipulates the steering angle but has no knowledge about its actual state. He can only estimate the wheel position through the sensed accelerations.

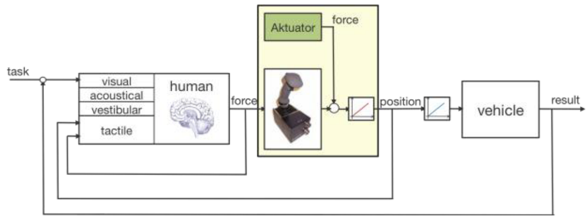


Figure 2. Force reflective element

In contrast position reflective elements (see Figure 3) use the applied forces to generate the setpoint settings. More precisely the forces are measured and translated into control inputs. The feedback information is returned by the position of the element. As opposed to the spring centered stick, where the position results from the balance of forces, the position reflective control element stays fixed for the operator and is only moved by the controlled system.

In doing so, the position of the element represents the actual state and its movement the dynamic of the system itself. Consequently the operator senses the behavior of the system.

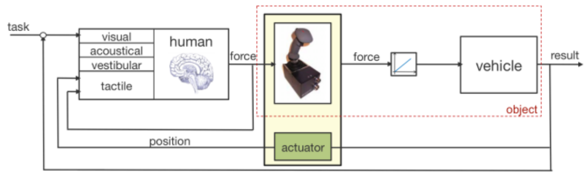


Figure 3. Position reflective element

This configuration works similar to the direct interaction with objects. Raising forces by the user manipulates the item, which responds with movement (see Figure 3).

As a result, position reflective elements seem not only suitable for compensating the decrease of information flow but also enable a specific feedback of essential information that supports the operator fulfilling the task.

In order to keep the vehicle controllable at all times and thus to increase stability, a bottom-up approach is preferred. That is why an experiment is performed without any kind of assistance. The most promising prototype represents the basis for the automation attachment.

3. EXPERIMENT DESIGN

3.1 Experiment Assembly

The experiment takes place at the department of ergonomics, TU München. The static driving simulator includes a mockup with a BWM car and three projection screens, which support 180 degree of sight (see Figure 4).

Figure 4. Static driving simulator

Similar to the steering wheel, substituted by an active wheel, the original accelerator pedal has been replaced by a pedal from Continental. Moreover a side stick from Stirling Dynamics ltd. is integrated into the central console to provide driving capability (see Figure 5).

Figure 5. Central console with side stick from Stirling Dynamics ltd.

The driving simulator software “SILAB”, that is developed by the Institute for Traffic Sciences in Wuerzburg receives all necessary commands from these control elements, simulates the vehicle as well as the whole environment and presents the scenery on the three projection surfaces. Furthermore all essential driving values are logged to provide objective data to evaluate the prototypes.

3.2 Prototypes

In this experiment the following four prototypes are being compared:

- Spring centered force reflective side stick
- Position reflective side stick with yaw rate feedback
- Position reflective side stick with steering angle feedback
- Steering wheel, accelerator and brake pedal

All models are tested in manual driving mode, which implies that no assistance is provided.

The force reflective side stick prototype only uses a spring characteristic, which centers the stick in the middle. Longitudinal movement is interpreted as throttle valve attitude or braking depending on the angle. Because there are no additional forces added this version is comparable with a conventional computer joystick.

Both position reflective side stick models measure the force in longitudinal direction and generate the throttle valve attitude or braking accordingly. The position of the stick is correlatively set to the vehicle’s velocity. In lateral direction, forces are converted into a change of the steering angle. However the lateral feedback of both prototypes differs. The first sets the angle of the element according to the yaw rate, while the second position reflective alternative reflects the steering angle.

The last prototype, which uses a steering wheel, accelerator and brake pedal as control device composes the conventional manner of driving. Objective driving data of the other versions compared to this one shed light on the potential of increasing driving performance by using other control elements.

3.3 Proband Collective

The sample consists of 24 subjects (13 male, 11 female) divided into two groups. Test persons under the age of 18 with minimal driving experience and test persons above, who own a driving license. These two groups with the average age of 15.4 or respectively 29.1 years have to complete a test track with the total length of 18.8 kilometer (5.5 kilometer highway, 13.3 kilometer road) with all four types of control.

4. EXPERIMENT RESULTS

4.1 Subjective Acceptance

After each run the subjects are asked to assess the driven prototype regarding controllability and strain. The study is completed by a final questionnaire in which all kinds of control interfaces have to be judged in direct comparison.

Figure 6 shows the results for the subjective impression of controllability depending on the kind of control. The subjective rating covers a scale from -3 (no control) to +3 (excellent control).

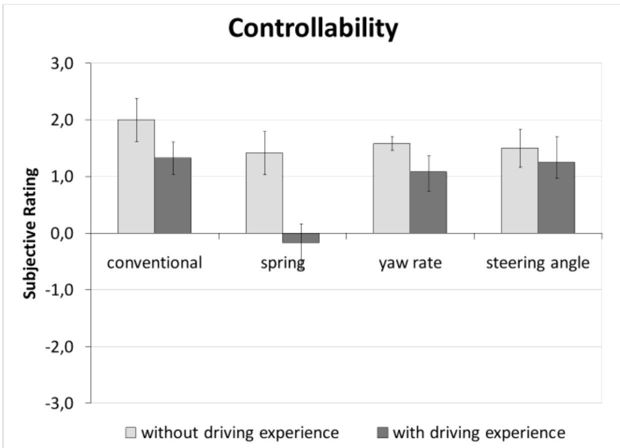


Figure 6. Subjective Rating of Controllability

The results for the group without driving experience show no significant difference between the four types of control interface. That means, that group 1 (without driving experience) has the subjective impression that their performance in car driving is independent from the control element. Much more interesting than that is the fact, that even the group with driving experience states, that the side stick versions with yaw rate or steering angle feedback grant the same controllability as the conventional controls. Only the spring centered side stick version is rated significantly worse. This is due to the fact that this version only gives a feedback about the dynamic properties of the control element itself (spring damper system), but no feedback about the system that has to be controlled.

Regarding the NASA-TLX Overall Workload Index (Figure 7) similar results can be found. Group 1 shows the same strain regardless of the kind of control; whether it is a side stick or steering wheel and pedals. Similar to the controllability results, Group 2 (with driving experience) shows no significant difference between the feedback versions of the side stick and the conventional control elements. The spring centered version of the stick however is rated significantly worse here, too.

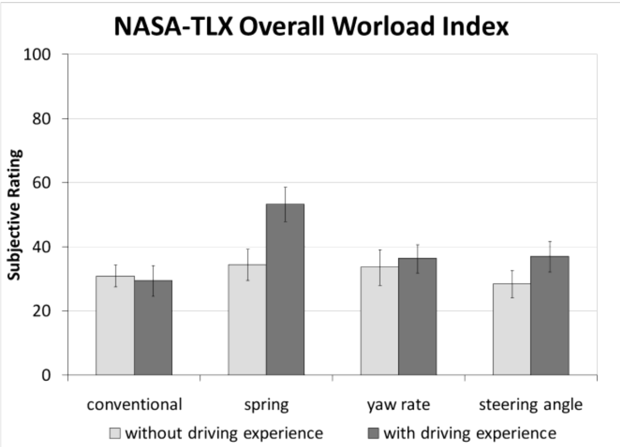


Figure 7. Subjective Rating of the Workload regarding the respective prototypes

4.2 Objective Performance

In addition to the subjective rating the objective driving performance is measured. The assessment of the objective data is divided into longitudinal and lateral driving efficiency.

Figure 8 shows the mean standard deviation of longitudinal speed in a part of the test track where the test persons had to maintain a constant speed of 80 km/h. The mean standard deviation in this case is a characteristic value to assess how good the subjects were able to perform this task.

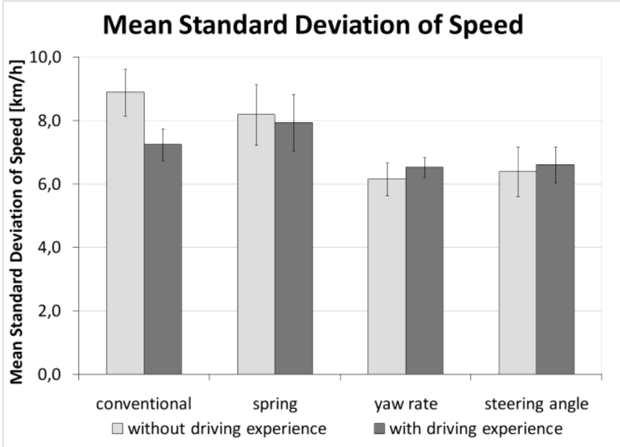


Figure 8. Performance at a longitudinal speed control task

Both, conventional control elements as well as the spring centered side stick give no feedback about the current vehicle speed which leads to a high mean standard deviation of velocity. The feedback versions of the side stick however indicate the driven speed by means of the position of the stick in longitudinal direction. As the figure shows, this feedback leads to a significantly reduced mean standard deviation and therefore to a significant better performance at longitudinal vehicle guidance. This performance enhancement is independent from the level of driving experience.

Representative for the results of the lateral driving performance Figure 9 shows the mean standard deviation of the lateral deviation in right hand bends.

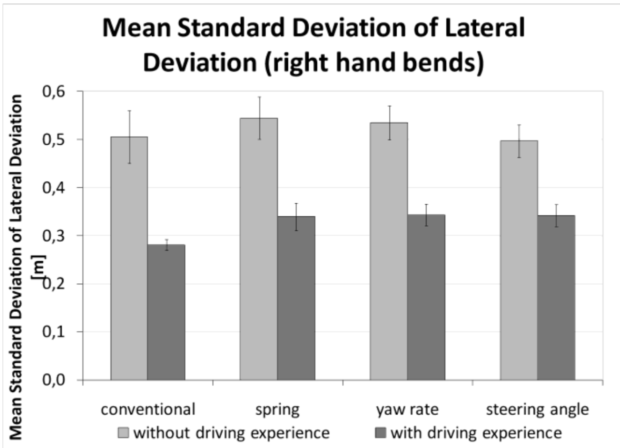


Figure 9. Performance while driving right hand bends

Here statistics show a significantly better performance in lateral control with the conventional control elements compared to the spring centered side stick (α-error = 0.007) and the yaw rate feedback version of the stick (α error = 0.050). The side stick with steering angle feedback however leads to a similar performance as driving with a steering wheel. The differences between the respective interface versions can be found regardless of the level of driving experience.

In left hand bends all of the interface versions show a statistically similar driving performance independently of the level of driving experience. This is most probably due to the bigger and therefore easier controllable radiuses driven in left hand bends.

4.3 Summary

In summary the subjective data of the study shows, that regarding controllability and strain, the position feedback versions of the side stick are rated equal to the conventional control elements.

Moreover the objective data show slight differences between the different interface versions. The side stick with steering angle feedback however shows equal performance in lateral driving tasks as the conventional control elements. In longitudinal driving tasks the position feedback principle even surpasses the performance of the combination steering wheel and pedals.

As a general result it can be said, that the principle of two dimensional control interfaces with position feedback, especially steering angle and speed feedback, is a promising idea to realize the idea of cooperative vehicles.

5. EXTENDING THE PROTOTYPES WITH ASSISTANT INTERACTION

Based on the experimental results, the position reflective side stick with steering angle feedback represents the fundament for additional assistance. As described above, one of the main features of H-Mode is the bi-directional haptic-multimodal coupling with continuous and/or discrete communication between driver and automation. This means that the co-system is able to apply forces to the stick in order to inform the driver about automation recommendations.

The diagram in Figure 10 shows how the system is extended with an arm parallel to the operator, thus allowing the co-system to add signals from the H-Mode automation to the stick.

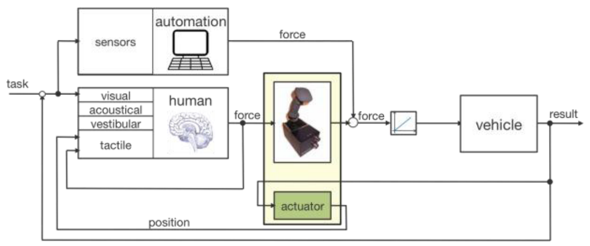


Figure 10. Signal diagram of stick with dynamic feedback and automation

By this means driver and automation system are affecting the vehicle parallel to each other, creating a combined control desire via the convergence point between the active stick and the vehicle. In this way, advantages of redundancy can be used, which leads to a safer overall system. By altering the balance between the human and the automation force the degree of automation can be changed.

6. REFERENCES

[1] Bolte, U. 1991. Das aktive Stellteil – ein ergonomisches Bedienkonzept. Fortschrittsberichte VDI-Reihe 17 Nr. 75. VDI-Verlag, Düsseldorf.

[2] Damböck, D., Kienle, M. 2008. Haptische Interaktion als Grundlage hochautomatisierten Fahrens. Ergonomie aktuell (009)

[3] Dickmanns, E. D. 2002. Vision for ground vehicles: History and prospects. International Journal of Vehicle Autonomous Systems. 1(1). 1-44.

[4] Eckstein, L. 2001. Entwicklung und Überprüfung eines Bedienkonzepts und von Algorithmen zum Fahren eines Kraftfahrzeugs mit aktiven Sidesticks. Fortschrittsberichte VDI-Reihe 12 Nr. 471. VDI-Verlag, Düsseldorf.

[5] Flemisch, F. O., Adams, C. A., Conway, S. R., Goodrich, K. H., Palmer, M. T., Schutte, P. C. 2003. The H-Metaphor as a Guideline for Vehicle Automation and Interaction (No. NASA/TM-2003-212672). Hampton: NASA. Langley Research Center.

[6] Goodrich, K. H., Schutte, P. C. 2006. A Design and Interaction Concept for Aircraft with Variable Autonomy: Application of the H-Mode. 25th Digital Avionics Systems Conference. Portland. USA.

[7] Hoeger, R., Amditis, A., Kunert, M., Hoess, A., Flemisch, F., Krueger, H.-P., Bartels, A., Beutner, A. 2008. Highly Automated Vehicles For Intelligent Transport: HAVEit Approach. ITS World Congress. New York. USA.

[8] Kelsch, J., Flemisch, F., Schieben, A., Schindler, J. Links oder rechts, schneller oder langsamer? Grundlegende Fragestellungen beim Cognitive Systems Engineering von hochautomatisierter Fahrzeugführung. 2006. DGLR Fachausschusssitzung Anthropotechnik "Cognitive Systems Engineering in der Fahrzeug- und Prozessführung". Karlsruhe.

[9] Parent, M. 2007. Advanced Urban Transport: Automation is on the Way. Intelligent Systems 22 (2) pp. 9-11.

[10] Penka, A. 2001. Vergleichende Untersuchung zu Fahrerassistenzsystemen mit unterschiedlichen aktiven Bedienelementen. Dissertation. Technische Universität München. Garching.

[11] Schieben, A., Damböck, D., Kelsch, J., Rausch, H., Flemisch, F. 2008. Haptisches Feedback im Spektrum von Fahrerassistenz und Automation. In Proceedings of 3. Tagung Aktive Sicherheit durch Fahrerassistenz (Garching, Germany, April 7-8, 2008).

[12] Thrun, S., Montemerlo, M., Dahlcamp, H., Stavens, D., Aron, A., Diebel, J., et al. 2006. Stanley: The Robot that Won the DARPA Grand Challenge. Journal of Field Robotics 23 (9) pp. 661-692.

[13] Winner, H., Hakuli, S., et al. 2006. Conduct-by-Wire – ein neues Paradigma für die Weiterentwicklung der Fahrerassistenz. Workshop Fahrerassistenzsysteme 2006. Löwenstein.

REQUIREMENTS, TOOLS AND TECHNIQUES

In-vehicle Technology Functional Requirements for Older Drivers

Andry Rakotonirainy
Centre for Accident Research Road Safety Queensland CARRS-Q
Queensland University of Technology
Australia
r.andry@qut.edu.au

Dale Steinhardt
Centre for Accident Research Road Safety Queensland - CARRS-Q
Queensland University of Technology
Australia
d.steinhardt@qut.edu.au

ABSTRACT
Older drivers represent the fastest growing segment of the road user population. Cognitive and physiological capabilities diminishes with ages. The design of future in-vehicle interfaces have to take into account older drivers' needs and capabilities. Older drivers have different capabilities which impact on their driving patterns and subsequently on road crash patterns. New in-vehicle technology could improve safety, comfort and maintain elderly people's mobility for longer. Existing research has focused on the ergonomic and Human Machine Interface (HMI) aspects of in-vehicle technology to assist the elderly. However there is a lack of comprehensive research on identifying the most relevant technology and associated functionalities that could improve older drivers' road safety. To identify future research priorities for older drivers, this paper presents: (*i*) a review of age related functional impairments, (*ii*) a brief description of some key characteristics of older driver crashes and (*iii*) a conceptualisation of the most relevant technology interventions based on traffic psychology theory and crash data.

Categories and Subject Descriptors
D.m [**Software**]: Miscellaneous; D.0 [**General**]

General Terms
Human factors

Keywords
ADAS,ITS

1. INTRODUCTION
Driving plays an important role in older drivers' mobility as 90% of older drivers rely on a private car as their primary mode of transport [19]. Driving cessation can thus significantly reduce older driver's mobility. Driving cessation is associated with significant negative health consequences such feelings of depression and social isolation.

In North America the proportion of the population over 65 years is expected to double by 2030 [14]. Similarly, the proportion of Australian licensed drivers aged over 65 is predicted to increase from 13 % in 2000 to 22 % in 2030. Of drivers aged over 65 years holding a licence, current research has found 96% report to be active drivers [42]. Older drivers aged over 65 are the most rapidly growing segment of road users in Australia in terms of number of drivers licensed, distance driven, and proportion of the driving population [16]. The ratio of retirees to workers in Europe is estimated to double to 54% by 2050 from four workers to two workers for every retiree [5]. It has been estimated that the working age population in the European Union will decrease by 48 million between 2010 and 2050 (-16%), while the elderly population will increase by 58 million, an increase of 77% [44].

The growing number of older drivers and the significance of the problem that they are facing to maintain their mobility for longer has generated siginificant research interest. Older drivers have low rates of crash per head of population, however their fatal crash rate per mile travelled increases starting at 70 years. This is largely attributable to increased frailty, particularly chest injuries and medical complications rather than over representation in crashes [18].

Ubiquitous/pervasive computing technology such as sensors, actuators, wireless networks and processors are commonly used to assist humans to perform various tasks. Context-awaren systems have become a growing area of study for pervasive and ubiquitous research communities. Unfortunately context-aware systems have not been thoroughly used to assist driving tasks. Technology based interventions such as Advanced Driving Assistance Systems (ADAS) have been hailed as a potential solution to improve road safety including older driver safety. It has been estimated that ITS could reduce fatalities and injuries by 40% across the OECD, saving over USD 270 billion per year [17]. Intelligent Transport Systems (ITS) and Advanced Driver Assistance Systems (ADAS) are growing research fields that use new technology aimed at improving road safety.

Computing assistance can improve situational awareness and reduce older driver errors. Although context-aware systems have great potential to save lives and prevent injuries on the road, they have not yet been integrated to safety critical applications for older drivers. With existing high demands on a driver's visual attention, many ADAS have been designed with a HMI that simplifies driver interactions with the view to reduce cognitive or visual demands. Speech based or tactile interfaces have been designed to reduce the effect of distraction [34].

However, scientific data is still lacking on the design and effectiveness of ADAS interventions, making it difficult to implement relevant policies as to their best use. The design of an ADAS intervention to improve older driver safety necessitates a clear understanding of the context in which crashes occur and the context in which it can assist. To address these concerns, this paper presents (*i*) an overview of age related driving impairments (*ii*) data analysis of road crashes involving older drivers to identify risk factors (*iii*) a review of relevant psychology theories to assess their suitability and effectiveness of in-vehicle technology to remediate identified crash patterns, and (*iv*) a discussion on the adequacy of existing technology to assess older drivers. Finally, recommendations regarding future research to improve older drivers safety are given.

2. OLDER DRIVERS' FUNCTIONAL IMPAIRMENTS

Driving is a complex task which requires cognitive and motor coordination to react and adapt behaviour to changing situations. It is widely recognised that older drivers suffer from age-related impairments to motor, sensory and cognitive abilities. Issues cited in past research include reduced mobility, reduced flexibility, reduced range of motion, slower reaction times [9], reduced visual acuity, prolonged visual accommodation and adaptation times, reduced peripheral vision, increased glare sensitivity [10], reduced ability to deal with high cognitive load driving tasks [29] and greater susceptibility to distraction [7]. Studies on closed roads have suggested that elderly drivers have slower reaction times, less accurate car following pattern and poorer merging behaviour at junctions than young drivers [45].

In studies related to the use of navigation systems by older drivers, [3] and [12] reported that older drivers had difficulty with the dual task of following a route guidance system while driving. Distraction caused by such systems may thus differentially affect older drivers negatively. Older drivers have been shown to spend significantly more time looking at navigation displays than younger drivers [28] [8].

Analysis of elderly drivers has shown that a battery of tests covering attentional, perceptual, cognitive and psychomotor performance are all significantly correlated with unsafe driving incidents as reported by police, family members and licensing agencies[21]. These tests specifically included selective and divided attention, field dependence, short term memory, digit matching, and simple reaction time. Visual tests of acuity were not as strongly correlated with unsafe driving incidents in this instance while psychomotor and cognitive skills were most highly correlated. While research has often defined an older person as those older than a specific chronological age, it is often of more relevance to consider age-related changes in physical, psychological and cognitive ability as a marker of when someone should be classed as an "older driver" [13]. It should also be noted that functional limitations and age related disorders do not necessarily lead to unsafe driving behaviour if a driver can self-regulate by avoiding complex driving situations such as night driving or intersections [23].

This suggests that there is no unique intervention that can uniformly help older drivers as a group. Intervention should be aware of limitations of a given driver with the view to assist him or her.

3. OLDER DRIVERS ROAD CRASH PATTERN

Several studies have identified factors contributing to older drivers crashes in driving simulators and on roads. The analysis of on-road crash involving elderly has shown that they are different from those of the overall driving population. This section presents the crash data analysis results of the Australian state of Queensland to identify the circumstances and contributing factors to crashes which are specific to older drivers.

3.1 Road crash database

The analysis was conducted using data from Queensland Transport's road crash database [41]. The road crash database is an electronic record of police-attended or otherwise reported road crashes that contains considerable information regarding the crash including the date, time, factors contributing to the crash and road characteristics. The level of analysis for this paper was the number of units (vehicles, excluding pedestrians) involved in crashes between 2000 and the end of 2004. Results from serious casualty crashes (those crashes resulting in a fatality of hospitalisation) are presented to exclude the large number of minor incidents. This analysis thus took into account 31,370 vehicles involved in crashes during this time period.

This database has a number of limitations that should be taken into account when interpreting the results. The crashes represent only those that are police reported - though this is likely to be the case for a large majority of serious crashes. The analyses in this paper consider three older age groups of 60 to 74 years, 75-79 years and those aged greater than 80 years, along with a broad younger comparison sample of drivers aged 17-59 years. These age groups were chosen to correspond with the ages at which restrictions begin to be placed on older drivers within Australia.

3.2 Results of road crash analysis

Table 1 shows the number of units (vehicles involved in crashes) broken by age group of the vehicle controller and the traffic features present at the site of the crash. Older drivers aged 60 and above were over-represented at a statistically higher level in crashes involving all forms of traffic control, with the proportion of crashes involving traffic control increasing steadily as age increased. This pattern also applies to those traffic scenarios involving give way or stop signs. There was however no significant difference between the age groups in terms of the proportion of crashes at controlled traffic lights, though a small trend for greater representation in the older age groups was present.

Table 2 presents the contributing factors of serious casualty crashes by age group. As before, statistically significant

Table 1: Queensland Serious Casualty Crash Units, 2000-2004, by Age Group and Traffic Control

	Age Group				
Variable	17-59	60-74	75-79	80+	Sign.a
Any Traffic Control	(n=31,370) 29.2%	(n=2,937) 35.6%	(n=612) 42.5%	(n=538) 44.1%	p< .001
Give Way Sign	(n=31,370) 9.2%	(n=2,937) 12.4%	(n=612) 14.2%	(n=538) 16.0%	p< .001
Operating Traffic Light	(n=31,370) 15.0%	(n=2,937) 15.6%	(n=612) 17.8%	(n=538) 17.7%	ns
Stop Sign	(n=31,370) 4.0%	(n=2,937) 6.2%	(n=612) 8.5%	(n=538) 8.9%	p< .001

crash distributions were found for a number of key crash circumstances, namely alcohol, fatigue, speeding and failure to give way; with no differences found between the age groups in terms of distraction. The involvement of illegal risk taking behaviours such as speeding and alcohol showed a substantial drop from the younger age group to the three older age groups. The involvement of fatigue showed a steady and significant decline as the age group increased in years, though this proportion was small across all age groups. Corresponding to this finding, the proportion of crashes occurring in the nighttime showed a marked decrease for the older age groups as compared to the 17-59 group.

Of particular note however was the overrepresentation of older drivers in Failure to Give Way crashes. This type of driving error is commonly made by senior drivers. Crashes involving age groups over 60 years of age were between 2 and 3.8 times more likely to involve a failure to give way than the 17-59 years age group. The proportion of such crashes showed a notable increase from the 17-59 years age group to the 60-74 years age group, as well as between the 60-74 and 75-79 years age groups.

Our results conform with existing research findings stating that older drivers are more likely to crash at intersections and other complex traffic situations [29] [20].

4. DRIVING BEHAVIOUR THEORIES

Drivers operate in highly dynamic contexts. Driving is a complex, continuous, multitask processing that involves driver's cognition, perception and motor movements. Section 3 showed that complex driving situations increase the likelihood of older driver's errors during decision making. Context-aware systems for cars are one method to provide a greater awareness of relevant information about the driving situation in order to assist the driver in the decision making process.

In-vehicle context aware systems aim to take into account more contextual information related to the driving task in order to produce adapted or customized actions. Driving tasks are classified into two categories, both of which can be assisted by a context-aware system:

- Primary task: Tasks restricted to longitudinal/lateral vehicle control and vigilance.
- Secondary task: Other tasks that do not require continual performance.

Driving a car requires a balanced and dynamic allocation of attention between the primary and secondary driving tasks. Performing the primary and secondary tasks are part of driving behaviour and involve decision makings followed by actions.

Theoretical models abound in literature as a means to explain and predict driver behaviour. Existing driver behaviour models are largely subjective and based on self-report scales [30]. They strongly emphasize the driver's cognitive state and have incorporated important behavioral concepts such as motivation, task capability [11], belief (theory of planned behavior) [1] or risk assessment. However, motivational models such as risk compensation [43], risk threshold [25] or risk avoidance remain highly subjective concepts. Subjective risks have been identified as a core concept influencing decision making [43] [31]. However [25] rejects such concepts and argues that the driving task is about maintaining a safety margin. Fuller [11] models driver's decision making as an interface between task difficulty and driver's capability. A useful model which is able to bring together a number of these concepts is that of the Michon Model.

Michon has defined a model to express the cognitive process of driver decision making [24]. This model allows quantitative measurement and covers some concepts covered in functional models. Each level of the model corresponds to a decision making level requiring a different type of information. Michon's model corresponds roughly to the information processing model defined by [31] whose hierarchical model describes three levels of information characterized by their degree of complexity. These are namely knowledge, rule based and skill based. The three levels defined by Michon are strategic, tactical and operational [24]:

- The strategic level is the highest level where general goals such as route choice, navigation and timing are set. Driving plans are formed and modified, goals established, prioritized, re-prioritized and satisfied or forgotten in real time as the driver assesses different factors from the environment, driving and vehicle. Expectancies and preferences are also part of this level.
- The tactical level involves decision making related to the management of current driving activity such as maneouvering. Tactical actions follow a pattern specific

Table 2: Queensland Serious Casualty Crash Units, 2000-2004, by Age Group and Contributing Factors to Crashes

	Age Group				
Variable	17-59	60-74	75-79	80+	Sign.a
Alcohol	(n=31,370) 10.4	(n=2,937) 3.7	(n=612) 3.9	(n=538) 1.9	p< .001
Fatigue	(n=31,370) 5.2	(n=2,937) 4.3	(n=612) 4.6	(n=538) 3.5	p< .001
Speeding	(n=31,370) 4.9	(n=2,937) 0.5	(n=612) 0.2	(n=538) 0.7	p< .001
Failure to Give Way	(n=31,370) 6.7	(n=2,937) 13.7	(n=612) 21.2	(n=538) 25.1	p< .001
Distraction	(n=31,370) 0.2	(n=2,937) 0.1	(n=612) 0.2	(n=538) 0.3	non-sig
Time of Day Day (6:00am - 5:59pm) Night (6:00pm - 5:59am)	(n=31,370) 70.4 29.6	(n=2,937) 85.9 14.1	(n=612) 91.2 8.8	(n=538) 92.2 7.8	p< .001

to drivers and can be assimilated to a profile. For example, the following distance chosen to remain behind another vehicle is determined by each driver's profile (e.g. aggressivity).

- The operational level involves vehicle handling or executive actions which implement the maneouvers decided at the tactical level. This level is performed almost without conscious thought. The result of such actions are directly measurable as vehicle dynamics.

Augmenting drivers situational awareness can operate at the strategic, tactical or operational level. The effectiveness of technological interventions at each level of Michon's decision making hierarchy is not well documented. However, it is well accepted that technological intervention could have dual opposite effects such as:

- making the driver aware of critical safety information well ahead and providing the driver with enough time to react safely.
- distracting the driver from the main critical driving task by overwhelming the driver with irrelevant, inaccurate or confusing information.

Context-aware systems often assume that users have the cognitive abilities to acquire the produced context-aware information. Such assumptions may be valid in desktop environments but are fundamentally inadequate and potentially unsafe in driving conditions. Conveyed awareness information requires driver's attention in order to register it. Registering information cognitively is not an effortless task.

5. DISCUSSIONS AND POTENTIAL TECHNOLOGY

In the US, approximatively 50% of all traffic crashes and 50% of injury crashes occur at intersections and 27% of intersection fatalities involved people 65 years of age or older (FHWA,08). The current data from the Queensland region is in line with a number of previous findings in that complex road environments are highly represented in crashes, with the older age groups of 75 years of age or older showing a marked increase in proportional crash involvement at crossroad intersections and where "failure to give way" was a contributing factor. Any ITS technology which could reduce the complexity and demands of such driving tasks could thus potentially improve older driver safety.

For ITS and especially in-vehicle technology to be effective, its operational/functional demand must be compatible with the motor (e.g. range of motion, dexterity, coordination, reaction time), physiological (e.g. visual, hearing) and cognitive abilities (e.g. divided/selective/sustained attention, tracking, memory, perception) of road users. This is particularly relevant to the growing driver population of older drivers. Existing technologies can provide such functionalities. Functionalities is about what the device does and what does it perform. The previous sections identified the functional needs in terms of contributing factors to crashes and older drivers functional impairments. The identified functionalities to be provided to the driver could be presented in different HMI forms. The design and the ergonomy of such technology are very important however this discussion focuses on the functional requirements.

An assistive device facilitates drivers task performance by providing real time advice, instruction, warning or even by taking control of the vehicle's dynamics. They operate in advisory, semi-automatic or fully automatic modes. The advisory and semi-automatic modes require human interventions with the associated human computer interface. An ITS intervention demanding a significant level of attention or motor activity (e.g neck torsion) from older drivers would not enhance older drivers safety. Additional advisory cues could also confuse the driver as older adults have difficulty in tasks that involve suppressing or inhibiting the influence of irrelevant information [46].

Older drivers are more likely than younger drivers to be at fault in crashes typically because they failed to yield the right-of-way, disregarded the traffic signal, or committed other traffic violation [20]. They have been shown to underestimate the speed of approaching vehicles at intersections [40]. These type of behaviour does not necessarily mean that they deliberately break the laws or engage in unsafe actions. Rather, the literature suggests that factors such as inattention, perceptual lapses, misjudgment, slow reaction time, illiness, poor vision could be implicated [20]. For example, their failure to give way could be attributed to a failure to notice other vehicles as opposed to a willful disregard to road rules. These behaviours have been attributed to various deficiencies in vision, attention, information processing and field independence. Older drivers have difficulty in processing peripheral stimuli to detect targets with high salience for the driving task. Different cognitive theories of ageing could be used to explain the elevated number of older driver related crashes at intersections. Older drivers experience performance decline in situations requiring selective attention, sustained attention, and dual task completion [4]. They also have greater difficulty in processing peripheral stimuli. These tasks require fast, dynamic and flexible attentional shifts which are essential to perform safe intersection manoeuvres. The above limitations together with a slower reaction time may contribute to a higher exposure to crash risks on intersections.

Existing approaches to assisting older drivers focus on simplifying the ergonomics of in-vehicle technology such as navigation system [27], [15]. Although such approaches could improve driver's interactions with navigation systems, we argue that navigation systems do not address older driver exposure to crashes directly. Our crash data analysis show that the elderly drivers exposure to crash increases when performing a particular maneuver on a particular road geometry such as crossroads or T-junctions. The crash risk associated with such situation cannot be remedied directly with navigation systems. The use of navigation systems influences decision making related to route choice, and are therefore situated at Michon's strategic level. Maneuvering on intersections is a combination of both the tactical and operational levels of decision making. A navigation system is unlikely to have impact on these two levels.

A gradual assistive device appear to be the most suitable to intervene at different phase of an intersection manoeuvre. The system could firstly improve the driver's awareness of threatening vehicles with multi-modal warning mechanisms. Such a mechanism should be able to call attention to approaching difficulties, signal risky events and help the driver to focus on the most critical task. If it is not manually impossible to avoid a crash (time to collision less than 2 seconds) then the assistive device should take control of the vehicle to attempt to avoid the crash. A combination of existing Advanced Driving Assistance Systems (ADAS) such as object detection, collision avoidance systems and lane departure systems could be integrated and extended to provide such services.

The availability of wireless communication protocols between vehicles and infrastructure (V2I) or between vehicles (V2V) offer great potential to assist drivers on intersections [36]. A vehicle could notify its presence and location to surrounding vehicles using V2V. Future research may seek to specifically identify the characteristics of those intersections that are a high risk for older road users and consider a combination of road-infrastructure and in-vehicle device interventions.

Due to the frailty of older drivers and their high exposure to crash on intersections, ITS technology that could protect them during crashes and help them to manoeuvre safely in intersections would provide the most significant benefits as illustrated in Figure 1. However other interventions that would assist vehicle control (passive or active technology) could also bring some benefits to a lesser extent. It has been shown that older drivers have difficulties in maintaining path, speed, changing lanes, performing precise control, backing and smooth stop. Existing ADAS technology address such issues, however such ADAS were not designed for older drivers.

The evaluation of in-vehicle devices should also consider user acceptance. Older drivers are most likely to suffer the effects of poorly designed ITS [39]. Oxley [26] studied the user acceptance of in-vehicle Navigation, Rear Collision Warning, Mayday system, Night Vision Enhancement and showed that older drivers exhibit a high degree of willingness to consider the use and purchase of, ITS applications. Olders drivers' opinions towards ITS have been shown to be generally high [35].

An important point to note is that the introduction of any ITS system is often accompanied with an increase of potential distractors. This may have increased relevance for the current discussion given the aforementioned potential difficulty of older drivers to cope with complex systems and attend to multiple traffic cues. A simulator experiment with an on-board display system showing the relationships between the driver's vehicle, other vehicles and roadside objects was shown to be effective in increasing the driving performance of younger drivers, but not older drivers [37]. The system made little impact on the problems of car manoeuvring faced by older drivers, which was attributed to an implied increase in cognitive performance from using the system as well as driving. A second test utilising a heads up display (HUD) which provided additional information on the degree to which the vehicle should be turned at each stage was successful in improving driving performance for both young and older drivers alike, without the subsequent increase in cognitive load for older drivers. It can be suggested from this research that systems which provide specific feedback on a display that does not distract from the driving task would minimise any cognitive load impact of ITS systems for older drivers.

As a final note, the current crash data analysis also suggests that considering the needs of elderly drivers in ITS systems may also have a positive impact in assisting drivers of all ages, as demonstrated by the high involvement of all age groups in more complex road environments [10]. It is therefore suggested that while specific interventions should be developed for older drivers, those targeting intersections would have benefits across a wide age range of road users.

6. CONCLUSION

Several studies have identified ADAS that might be able to assist older drivers [10] [22], [38]. Pauzie [27] has shown that

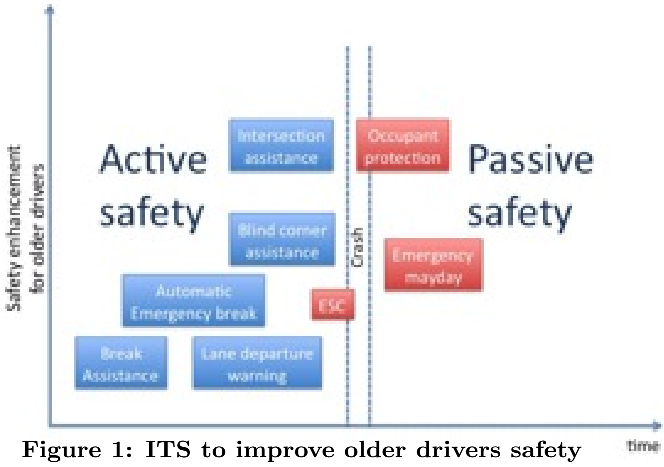


Figure 1: ITS to improve older drivers safety

the ergonomics of in-vehicle technology play an important role in older driver safety safety. A simplified task, simplified dialogue, better legibility and intelligibility of information could improve older driver's performance. This paper focused on the functional requirements of in-vehicle devices to improve the safety of older drivers. We have shown that older drivers are more exposed to crashes in complex driving situations such as intersections. We have argued that Michon's [24] tactical and operational levels are the relevant levels involved in decision making on intersections. Therefore the most promising technology to improve older drivers safety are those affecting tactical and operational levels. The upshot of our findings is that technology based interventions that have impacts on the strategic level, such as navigation systems, are likely to have less safety benefits than those operating at the tactical and operational levels.

There is some encouraging evidence that low-cost safety improvement at intersections such as enhanced traffic signal conspicuity could improve older driver safety [2]. As olders drivers' opinions towards ITS is generally high [35], there are opportunities to enhance their safety with in-vehicle technology. Much research remains to be done to establish the benefits of ADAS for older drivers [6]. The benefits that existing cooperative systems such as V2V or V2I could bring to older drivers have not been fully evaluated. This is despite the fact that V2V and V2I could improve safety on intersections and therefore could be beneficial to the elderly. There is a need to investigate new ways of prompting older drivers to take action, considering their capabilities. For example motor priming and cognitive priming are un-tapped HMI approaches that have not been explicitly experimented in vehicles. ITS is one type of intervention that should be complemented by others including education about self regulation of driving (e.g avoiding intersections, night driving [23]). Continuing research on the extent to which older drivers appropriately use technology and self-regulate their driving is warranted. Much remains unknown about the specific circumstances leading to older driver's crashes and research needs to be conducted in a naturalistic setting as opposed to driving simulators.

7. REFERENCES

[1] I. Ajzen. *From intentions to actions: A theory of planned behaviour*, chapter Action control. From cognition to behaviour, pages 11–40. Springer Verlag, 1985.

[2] J. Bagdate. Low cost intersection improvements reduce crashes for senior driver. Technical report, AAA Michingan, Deaborn, MI, 2004.

[3] P. A. J. Barham, P. R. Oxley, B. Ayala, and J. Alexander. The ergonomic and safety implications of in-car att devices - evidence from field trials with elderly drivers. In *IEE Colloquium on Design of the Driver Interface*, January 1995.

[4] T. S. Braver, D. M. Barch, B. A. Keys, C. S. Carter, J. D. Cohen, J. A. Kaye, J. S. Janowsky, S. F. Taylor, J. A. Yesavage, and M. S. Mumenthaler. Context processing in older adults: Evidence for a theory relating cognitive control to neurobiology in healthy aging. *J. Exp. Psychol. Gen.*, 130(746-763), 2001.

[5] G. Carone and D. Costello. Can europe afford to grow old? International Monetary Fund Finance and Development Magazine, 2006.

[6] R. Davidse. Older drivers and adas - which systems improve road safety. *IATSS Research*, 30(1), 2006.

[7] D. de Waard, F. J. J. M. Steyvers, and K. Brookhuis. How much visual road information is needed to drive safely and comfortably? *Safety Science*, 42:639–655, 2004.

[8] T. Dingus, J. Antin, M. Hulse, and W. Wierwille. Attentional demand requirements of an automobile moving-map navigation system. *Transportation Research*, 23A(4):301–315, 1989.

[9] B. M. Dobbs. Medical conditions and driving - a review of the scientific literature. Technical Report DRNH22-94-G-05297, Washington: U.S. Department of Transportation., 2003.

[10] B. Färber. Can telematics improve mobility of the elderly? *Gerontechnology*, 1:287–290, 2002.

[11] R. Fuller. Towards a general theory of driver behaviour. *Accident Analysis and Prevention*, 37(3):461 – 472, 2005.

[12] P. Green. Variations in task performance between younger and older drivers: Umtri research on telematics. In . *Proceedings of the Association for the Advancement of Automotive Medicine Conference on Aging and Driving*, 2001.

[13] L. Hakamies-Blomqvist. Older drivers' accident risk: conceptual and methodological issues. *Accident Analysis and Prevention*, 30:293–297, 1998.

[14] E. Hildebrand and E. Griffin. The effectiveness of airbags for the elderly. In *Contemporary issues in road user behaviour and traffic safet.*, pages 263–274. D.A. Hennessy and Wiesenthal and D.L. Hauppauge, NY, us: nova science publishers edition, 2005.

[15] S. J. Kim and A. K. Dey. Simulated augmented rality windshield display as a cognitive mapping aid for elder driver navigation. In *CHI 2009*, pages 133–142. ACM, Aprill 4-9 2009.

[16] M. King, A. Nielson, G. Larue, and A. Rakotonirainy. Projecting the future burden of older road user crashes in Queensland. In *Road Safety Research, Policing, Education Conference*, Melbourne, 2007.

[17] M. Peden, R. Scur■eld, D. Sleet, D. Mohan, A. Hyder, E. Jarawan, and C. Mathers. World report on road tra?c injury prevention. Technical report, World Health organization, 2004.

[18] G. Li, E. Braver, and L.H.Chen. Fragility versus excessive crash involvement as determinants of high death rates per vehicle-mile of travel among older drivers. *Accident Analysis and Prevention*, 35, 2003.

[19] R. Martinez. Older drivers and physicians [editorial]. *Journal of the American Medical Association*, 274(1060), 1995.

[20] D. R. Mayhew, H. M. Simpson, and S. A. Ferguson. Collisions involving senior drivers: High-risk conditions and locations. *Traffic Injury Prevention*, 7(2):117–124, 2006.

[21] A. J. McKnight and A. S. McKnight. Multivariate analysis of age-related driver ability and performance deficits. *Accident Analysis and Prevention*, 31:445–454, 1999.

[22] C. Mitchell and S. Suen. Its impact on elderly drivers. In *13th International Road Federation IRF World*, Toronto Canada, 16-20 June 1997.

[23] L. J. Molnar and D. W. Eby. The relationship between self-regulation and driving-related abilities in older drivers: An exploratory study. *Traffic Injury Prevention*, 9(4):314–319, August 2008.

[24] J. A. Michon. A critical view of driver behavior models: What do we know, what should we do? chapter Human Behavior and Traffic Safety, pages 485–520. Plenum Press, 1985.

[25] R. Naatanen and H. Summala. *Road user behaviour and traffic accidents*. Elsevier, New York, 1976.

[26] P. Oxley. Elderly drivers and safety when using it systems. In *IATSS Research*, volume 20, pages 102–110, 1996.

[27] A. Pauzie. In-vehicle communication systems: the safety in-vehicle communication systems: the safety aspect. *Injury Prevention*, 8:26–29, 2002.

[28] A. Pauzie, C. Marin-Lamellet, and R. Trauchessec. Analysis of aging drivers' behaviors navigating with in-vehicle visual display systems. In *Conference Proceedings In Vehicle Navigation and Information Systems*, pages 61–67. Warrendale, PA: Society of Automotive Engineering., 1991.

[29] D. Preusser, A. Williams, S. Ferguson, R. Ulmer, and H. Weinstein. Fatal crash risk for older drivers at intersections. *Accident Analysis and Prevention*, 30:151–159, 1998.

[30] T. A. Ranney. Models of driving behavior: a review of their evolution. *Accid Anal Prev*, 26(6):733–750, 1994 Dec.

[31] J. Rassmussen. Mental models and the control of action in complex environments. In *Mental Models and Human-Computer Interaction 1,*, pages 41–69, New York, 1990. Elsevier Science.

[32] G. Wilde. The theory of risk homeostasis: implications for safety and health. *Risk analysis*, 2(4):2090225, 1982.

[33] J. Wood. Aging, driving and vision. *Clinical and experimental optometry*, 85(4):214–220, July 2002.

[34] M. A. Regan, J. D. Lee, and K. Young. *Driver Distraction: Theory, Effects, and Mitigatio*. CRC Press, 2009.

[35] M. A. Regan, E. Mitsopoulos, N. Haworth, and K. Young. Acceptability of in-vehicle intelligent transport systems to victorian car drivers. RACV report 02/02, Monash University Accident Research Centre, November 2002.

[36] F. D. Salim, S. Loke, S. W, A. Rakotonirainy, B. Srinivasan, and S. Krishnaswamy. Collision pattern modeling and real-time collision detection at road intersections. In *Proceedings of The 10th International IEEE Conference on Intelligent Transportation Systems, , ,*, Seattle, Washington, USA, Sep 30 – Oct 3 2007. IEEE Computer Society Press.

[37] T. Sato, H. Kawashima, and D. Tatsuru. Age differences in behavioral and visual characteristics while driving with narrow road driving assistance system. In *IEEE Intelligent Vehicles Symposium -IEEE Intelligent Transportation Systems Council*, University of Parma, Parma, Italy, 2004.

[38] S. Shaheen and D. Niemeier. Integrating vehicle design and human factors: Minimizing elderly driving constraints. Technical report, ITS UCDAVIS, 2001.

[39] N. Stamatiadis. Its and human factors for the older driver: the u.s experience. *Transportation Quarterly*, 52(3):91–101, 1998.

[40] L. Staplin and R. Lyles. Age differences in motion perception and specific traffic maneuver problems. *Transportation Research Record,*, (1325):23–33, 1991.

[41] Q. Transport. Queensland crash database. https://www.webcrash.transport.qld.gov.au/webcrash2, 2009.

[42] J. Langford, J. Charlton, B. Fildes, J. Oxley, and S. Koppel. Report on vicroadsÕ older driver surve. Technical report, Project 830, 2004.

[43] G. Wilde. The theory of risk homeostasis: implications for safety and health. *Risk analysis*, 2(4):2090225, 1982.

[44] L. Wolfgang. European demographic data sheet 2006. Technical report, International Institute for Applied Systems Analysis and Population Reference Bureau., Vienna and Washington, DC: Vienna Institute of Demography,, 2006.

[45] J. Wood. Aging, driving and vision. *Clinical and experimental optometry*, 85(4):214–220, July 2002.

[46] R. Zacks and L. Hasher. Cognitive gerontology and attentional inhibition: A reply to burke and mcdowd. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences,.*, 52:274–283, 1997.

A cognitive schema approach to diagnose intuitiveness: An application to onboard computers

Sandrine Fischer
Dept. of Risk Engineering
University of Tsukuba
fischer@css.risk.tsukuba.ac.jp

Makoto Itoh
Dept. of Risk Engineering
University of Tsukuba
itoh@risk.tsukuba.ac.jp

Toshiyuki Inagaki
Dept. of Risk Engineering
University of Tsukuba
inagaki@risk.tsukuba.ac.jp

ABSTRACT
Intuitive use is met when prior knowledge is transferred to new task environments. The empirical fact that transfer relies on schemas led us to diagnose intuitiveness based on schema induction. Two cognitive tasks were designed to make novice users perceive versus induce all the states of a prototype onboard computer. Subsequent interaction performances with the system validated the induction effect of the procedure and its interaction with familiarity, known as a primary factor of intuitive use. Implications for the diagnosis and the design of intuitive interfaces are discussed.

Categories and Subject Descriptors
H.5.2 [**User interfaces**]: Theory and methods.

General Terms
Measurement, Performance, Experimentation, Human Factors, Theory.

Keywords
Human Computer Interaction, Intuitive Use, Cognitive Modeling, Schema Induction, Design Evaluation.

1. INTRODUCTION
1.1 Context
With the increase of advanced technologies in everyday life, users await intuitive devices that can be understood and used with no particular effort. This demand is especially difficult to meet when several technologies are gathered in a same device. Automotive industry is uppermost concerned, when developing onboard computer that aggregate multimedia, communication, maintenance, driving assistance and telematic services.

Previous studies on intuitive interaction addressed remote controllers, VCR and digital cameras [1;2]. These devices, although hard to handle at the first attempt, are far simpler than onboard computers which largely exceed 100 states (*cf.* Audi's MMI and BMW's IDRIVE). States of onboard computers[1] typically display 3 to 15 graphical objects (*e.g.* labels, menus and icons) and form complex states-transitions networks.

Novice users, who do not know yet the system, must find by themselves the sequences of transitions and states leading to its functionalities. This activity can be assimilated to means-end analysis [3], where user iteratively judges which available object best reduces the distance to the desired state. Whether these means-end judgments can be performed intuitively is the matter of the present study.

1.2 A schema account of intuitive interaction
1.2.1 Current approaches of intuitive interaction
Intuition is a mechanism by which the solution of a problem is perceived without effortful analysis [4]. It has been empirically attributed, in psychology, to cognitive style [5-7] and to prior knowledge [8;9]. The HCI community recently adopted this concept to evaluate and design interfaces.

Main contributors, namely Blackler and colleagues in Australia and the IUUI (Intuitive Use of User Interfaces) Research Group in Germany, consensually attribute intuitive interaction to the unconscious application of prior knowledge to a new task or to a new environment. Blackler showed, from correlational analyses, that devices were more intuitive when their features (e.g. functionalities, graphical objects, commands) had already been employed in similar or in different devices [2;10]. Intuitive interaction typically requires the "transfer" of relevant prior experience "between products, and probably also between contexts" [10].

The two research groups employed slightly different models of design to anchor this conception. Blackler focused on the location, the appearance (e.g. shape, color, labeling) and the function of interfaced features [2]; the IUUI Research Group, on the conceptual, semantic, syntactic, lexical, and pragmatic (physical) "layers" of design [11]. These models led to reinstate classical and convergent recommendations such as [1;10]:

- employ shared labels and stereotypes when designing familiar functionalities, use affordances and semantics when designing unfamiliar functionalities, and identify

[1] Onboard computers are variably referred as multistate interfaces or as multifunction systems in the present paper.

external consistencies and metaphors originating from other domains, for designing innovative technologies;

- respect ISO standards such as the suitability for task, the conformity with user expectations, the self descriptiveness, affordances and Gestalt laws;
- focus on physical to semantic coupling and image schemas (*e.g*. visual clues of space, containment, process, force, etc.).

Whether these recommendations actually make easier the design of intuitiveness is yet questionable. Indeed, affordances, metaphors, consistencies and stereotypes are not operational enough to be properly managed and their definitions often lead to circular statements. For instance, intuitive use is supported by *self-descriptiveness,* itself presented as the implementation of *obvious* and *immediately clear* contents [10]. Expectedly, these constructs do not reliably impact performances [12].

Also, there is currently no mean to arbitrate which features or layers of a given interface should benefit from affordances, consistencies, stereotypes or metaphors. Blackler studied the intuitiveness of remote controllers, VCRs and digital cameras by inspecting each of their features' familiarity of [10]. More precisely, the interview determined whether the features' location, appearance and function had been used or encountered in similar and in different devices and contexts. This investigation was "very time consuming" and might be heavy to conduct on multifunction systems. Instead of declining prior familiarity, we could directly measure the transfer mechanisms previously reported to support intuitions. Actually, transfer has extensively been studied in cognitive psychology and elucidated, about 30 years ago, by the construct of cognitive schemas.

1.2.2 The schema hypothesis

The domain of analogical reasoning is concerned on how a procedure learned in a given context can be transferred to another and even unfamiliar one.

Transfer is studied in a two-phase protocol. Participants study a "source" problem and its solution, before receiving a "target" problem to solve. For instance, participants read a text explaining that an army should be spread in small units to attack a fortress surrounded by mines (source). Participants subsequently had to explain how to treat a tumor with X-rays without damaging healthy tissues (target problem or task) [13]. Despite very different contexts (military and medical), the source and the target both admit the "divide and disperse" solution.

Read the source for comprehension, summarize it or even read two analog source problems poorly led to solve the target problem [13]. Participants tried to analyze the problem from a medical perspective, instead of simply reusing the divide and disperse principle. Gick and Holyoack resumed this issue by requiring novices to compare two analog source problems [14]. Verbal protocols collected during this comparative study task revealed that participants mentioning the structure shared by the two source problems better solved the target one. Structural representation of the source was referred as a cognitive schema.

Replication experiments showed that schema induction is a by-product of comparative processing that supports transfer between different domains [15]. Schemas well sustain the previously stated idea that intuitive use relies on knowledge transfer.

1.2.3 A schema based model of intuitive interaction

Schema theory has been formalized by Norman and applied to Human Computer Interactions in a framework named ATS (*Activation Trigger Schema*) [16;17].

According to Norman, seven stages of activity determine the interaction. As illustrated Figure 1, the user perceives and interprets the current state of the interface in order to evaluate whether it is different or distant from its goal. If it is the case, the intention is formed to modify the state by handling an available object or command. To do so, the user specifies and executes one or several operations on the system's commands (*e.g*. mouse, stylus, etc.).

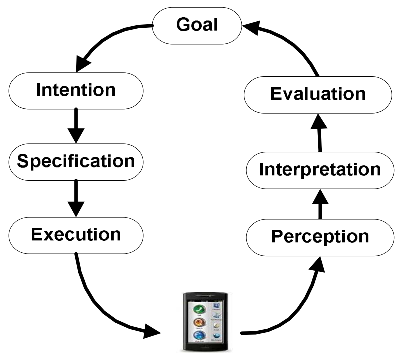


Figure 1. Norman's Action Cycle (1984)

Schema theory postulates that perception, interpretation, specification and execution can be shortcut when prior schemas are triggered. Action is direct, automatic -we might here say intuitive- if each stage benefits from prior schemas. Conversely, when no schema is triggered, the user has to analyze the interface content. This effortful mechanism is necessary until new *ad hoc* schemas are constructed. It is consequently important that the "system image" or the interface that fails to trigger prior schemas, at least supports the construction or induction of new ones.

ATS matches Blackler and the IUUI Research Group's step to model the compatibilities of design layers or dimensions with prior knowledge. Additionally, the framework offers interesting possibilities to diagnose affordances [18], human errors [19] and usability [20;21]. However, these implementations barely deal with the schema theory which actually interests us. A more promising approach was found in the domain of multimedia learning, where schema operations fostered by a source material are measured behaviorally, by an induction procedure.

1.3 Behavioral diagnosis

1.3.1 The induction paradigm

Multimedia are materials that combine text, illustrations, animations or simulation to describe a technical system such as an automotive engine, an air pump, a process plant simulation, an air traffic control simulation, etc.

Whether a multimedia helps to induce schemas is measured by requiring novices to study it before solving target problems or tasks. For example, novices study the illustration of an air pump, before being requested to explain what could be done to make a pump *more effective*, *more reliable*, etc. [22]. Such problems can only be solved if appropriate schemas are induced during the study phase[2] [23].

Read or listen a multimedia enables novices to recall its content (e.g. words, sentences) but not necessarily to solve target problems [24]. Similarly to analogical reasoning, induction is met when novices establish relationships among the source's objects and with familiar knowledge.

Mayer demonstrated the potential of comparative (*i.e.* comparing together different parts of the source) and integrative study tasks (*i.e.* comparing the source with prior knowledge) to induce programming schemas. Participants were requested to study a database instruction language (source) before using the database in a series of counting and sorting tasks (target). Four study groups were constituted [25] :

- The control group received a booklet listing the language's instructions with the instruction to read it.
- An "advance organizer group" also received the booklet, as well as familiar and concrete examples of database tasks.
- The "model elaboration group" also received the booklet, as well a sheet explaining the model of a computer, with the instruction to search its similarities with the database instructions.
- The "comparative elaboration group" was instructed to list the similarities and differences among the booklet's instructions.

The three experimental groups performed better on the transfer tasks than the control one, although they recalled fewer instructions from the booklet. Interestingly, the advance organizer group solved the most difficult target tasks, indicating that adding familiar and concrete information to the source to study particularly fosters induction.

1.3.2 Inductive tasks interact with expertise

Induction effect –*i.e.* the difference in target problem solving between inductive and non inductive study conditions- directly depends on prior experience. Mayer's meta-analysis of multimedia studies reveals that [26] :

- only novices benefit from inductive instructions,
- while experts or familiar persons reach high performances independently of the study and of the material. Their schemas are, as a matter of fact, rich enough to perform the problem test under most conditions.

This interaction pattern was repeatedly obtained in multimedia learning [23;27;28], analogical reasoning [29-31] and text comprehension studies [32]. It thereby seems relevant to differentiate intuitive transfer of prior schema (no difference after inductive and non inductive study conditions) from induction of new schemas (significant difference between an inductive and a non inductive study condition).

1.3.3 Induction procedure for onboard computers

The induction paradigm has been so far applied to documents, booklets, videos, simulations, etc., but never to materials reaching one hundred states. We developed two study tasks susceptible to make novices encode literally versus inductively such material. These tasks required to judge whether a given target matches a given state. Participants read a target, and then a state possibly containing (match) or not (non match) the target.

In the inductive condition, the target was a sentence describing a functionality in familiar (as less technical as possible) and concrete (explicitly detailing the context of the activity) terms. Example of function targets are: "*Calculate the distance covered with the car during the precedent weekend*"; "*Save the car's current GPS location in the address book*", etc. This Function Matching Task was designed to both foster the comparative and the integrative processing known to support schema induction. It indeed incidentally required to interpret and to compare the state's objects together and with a target which familiar and concrete labeling naturally activates prior knowledge.

In the non inductive condition, the target was a word (*e.g.* "*Next*"", "*Map*"). The Word Matching Task could be performed by simply scanning the state's words. This rather perceptive condition fits with definition of intuition as the immediate sensing and perceiving a schema solution. Successful solving of transfer tasks after this study task is in that attributable to intuition.

1.4 Empirical study

We experimentally addressed whether matching all the states of an interface with a function (Inductive Group) in comparison to a word (Perceptive Group) fosters schema induction and interacts with familiarity. A Control Group that only performed the transfer task scenario was also constituted to have baseline interaction performances.

We defined familiarity, based on Blackler's prior research [2], as the prior use of a feature in similar or other contexts. Additionally, we took into account the participants' cognitive style. The numerous studies dedicated to scale how individuals tend to intuit or to analyze problems [9] did not consider, to our knowledge, Human Computer Interactions. Nevertheless, as intuitive scales correlate with cognitive tasks by lowering the

[2] This intra-system transfer is slightly different from the "inter-domain" transfer studied in the domain of analogical reasoning. In analogy studies, participants are taught the procedure we want them to transfer to a new domain. In multimedia stydies, participants have to induce the procedures required by subsequent transfer tasks by themselves.

processing and the assimilation of data [5;33] it is probable that "Experientials" (*i.e.* persons who tend to rely on their intuitions) will less perform the inductive operations appealed by the Function Matching task compared to "Rationals" (*i.e.* persons who tend to solve by analyzing). This was controlled by differentiating experiential and rational participants based on a Japanese version of Epstein's Rational Experiential Inventory (REI) [34;35].

2. METHODS

2.1 Participants

Forty three Japanese students, novice in the use of onboard computers, received 820 yens (approx. 9 $) to participate in a 45 minutes experiment.

2.2 Material

We first designed the material in English before translating it into Japanese. Instructions, targets and interface used well-shared and as less technical as possible wordings. Two Japanese students with no background in informatics and in automotive were independently recruited to improve the material by simplifying its formulations during informal interviews.

2.2.1 The tested interface

The interface to diagnose was a prototype onboard computer developed under C# and named DoIt#.

Developing this prototype enabled us to automatically record the participants' actions and corresponding time code in a log file. States-actions sequences were reconstructed using the task modeling tool AMME [36] as well as specific VBA macros. Prototyping also permitted us to test the procedure on common (*e.g.* temperature setting, dialing a call, display of gas level, defrosting, etc.), advanced (*e.g.* locker anti alcoholism, dust filters, traffic status, etc.) and "prospective" functionalities that do not yet exist in the market and, consequently, go beyond the participants' prior knowledge (*e.g.* wireless download of advertising and information tags, rear view mirror display of driving instructions, etc.).

DoIt# was composed of two windows: a command panel and a state window. The command panel had five menu buttons ("*Onboard Computer*", "*Navigation*", "*Air Conditioning*", "*Audio*" and "*Telephone*") as well as four navigation buttons ("*Up*", "*Down*", "*Enter*" and "*Escape*"). The state window displayed options lists, icons, pictures and virtual input devices.

Interaction with DoIt# mostly required to scroll options lists by clicking, with the mouse, the *Up* and *Down* buttons, and to explore lower or upper-level menus by clicking *Enter* or *Escape*. DoIt#'s functionalities could be achieved in 3 to 5 actions. For example, participants had to reset the odometer by activating the "*Onboard Computer*" and successively selecting the options labelled "*Driving Indices*", "*Mileage recorder*" and "*Reset*".

2.2.2 The matching task

Each state of DoIt# was captured and associated to a word and a function target.

Half pairs of the Matching Task were matched and the other half mismatched. Matching pairs were constituted by randomly selecting an object in a state of DoIt#, from which was taken a word, and derived an explicit and detailed description. Negative trials were constituted by inventing a target word and a target function absent thought realistic and coherent with the state.

Word and function targets are rather easy to generate. The experimenter simply needs, in the case of word targets, to select a word, and in the case of function targets, to describe and detail a functionality without repeating the state's wordings and without using technical terms. This requires a good knowledge of the concerned system and technology, as well as popularization skills. Function targets were improved by two reviewers, non-specialized in the technology domain, who pointed out and rephrased the difficult wordings.

A LabVIEW application was developed to display serially pictures of targets and states above three control buttons. Participants were instructed to (see Figure 2):

1. read a target, click on the button "GO"
2. read the state and clicking on "YES" if it matched or contained the target or on "NO" otherwise.

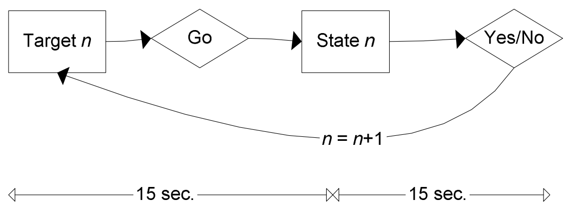


Figure 2. Timeline of the Matching Task

Figures 3 and 4 illustrate pairs of target-states for the two experimental conditions.

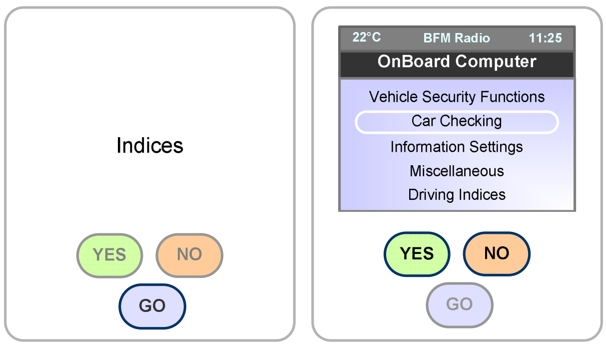


Figure 3. Illustration of a target-state pair for the Word Matching Task (Perceptive Group)

As stated previously, the Word Matching task (see Figure 3) requires neither to understand nor to compare the state's graphical objects (e.g. "*18°C*", "*Onboard Computer*" "*Vehicle*

Security Function", etc.). Participants only need to scan the words individually until finding (or not) the target one.

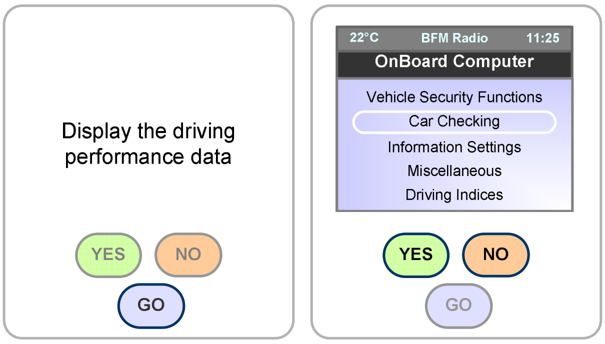


Figure 4. Illustration of a target-state pair for the Function Matching Task (Inductive Group)

Conversely, the Function Matching requires from the participants to understand the words, the object they belong to, their role in the interface and their relationship to the target. The participant also needs to compare the objects between them and with the target when significations are close.

2.2.3 REI and Familiarity questionnaires

The Rational – Experiential Inventory and the Familiarity questionnaires were administered by Excel. REI is a bipolar subjective scale comprising 20 experiential statements (*e.g.* "*I try to avoid situations that my intuitive impressions*") and 20 rational statements (*e.g.* "*I enjoy thinking in abstract terms*", "*I think that it is foolish to make important decisions based on feelings*"). Participants rated each statement on a 5-point scale ranging from *completely false* to *completely true*. Participants scoring beyond 120 points were considered as Rationals, whereas participants scoring over 120 points were considered as Experientials in subsequent analyses [6].

Our prior familiarity questionnaire listed all DoIt#'s functionalities. Participants were instructed to report for each described functionality whether it had been used or seen, in a similar or in a different context. We thus could code *a posteriori* the tasks that had been done or seen by all the participants as Familiar, and the others, as New.

2.3 Procedure

The four experimental phases (REI, Matching Task, Transfer Tasks and Familiarity questionnaire) were embedded in an animated PowerPoint presentation in order to minimize exchanges between the participants and the experimenter. This administration mode enabled us to test participants on three PC simultaneously.

Each participant started by filling the Rational / Experiential Inventory. Then, the participant watched either the Word or the Function Matching Task instructions with some recommendations to properly explore the states. The participant watched next a presentation of DoIt#'s main commands. After being showed two examples of task by the experimenter (*e.g.* call a recent dialed number and check the inbox messages), the participant received a task scenario printout. The ten target tasks were to (1) display the number of covered kilometres, (2) set the guidance to a friend registered in the address book, (3) request to avoid toils, (4) display the guidance instructions in the rear-view mirror, (5) set the temperature to 18°C, (6) launch the anti-drowsiness alert, (7) activate the over-taking assistant, (8) activate the filtering of inside air, (9) set the ventilation on silent mode, and (10) calculate the total break time during the trip. No mention to the speed or to the accuracy was made to let the participant act at his or her pace. At last, the participant filled the familiarity questionnaire by ticking the functionalities he or she had used or seen before the experiment.

2.4 Experimental design

The experimental protocol aimed to verify that:

- the Function Matching Task generates an induction effect, i.e. higher task performances for the Inductive Group compared to the Perceptive Group (Induction Hypothesis),
- both groups exhibit similar performances when performing Familiar Tasks (Interaction Hypothesis).

The level of Induction (Control *vs.* Perceptive *vs.* Inductive Group) was manipulated as a between-subject factor, the REI (Experientials *vs.* Rationals) was controlled as a between-subjects factor and the Task Familiarity (Familiar *vs.* New) was controlled as a Within-Subject and Between-Task factor.

The Control Group comprised 11 participants (5 Experientials and 6 Rationals), the Perceptive Group, 16 participants (7 Experientials and 9 Rationals) and the Inductive Group, 16 participants (8 Experientials and 8 Rationals). Among the 10 tested tasks, 4 were coded as Familiar and 6 as New.

3. RESULTS AND DISCUSSION

We first will examine whether matching tasks foster induction and interact with prior familiarity. This hypothetico-deductive perspective is followed by a qualitative analysis of raw data to question the diagnosis potential of the induction procedure.

We analyzed the mean number of erroneous transitions, calculated by the mean number of transitions minus the number of optimal transitions per participant and per task. Uncompleted tasks and tasks for which the participant asked the experimenter's assistance were excluded.

3.1 Hypothetico deductive validation

Participants made 10.1 (SD = 11.8) errors per task. Participants made 8.0 errors (SD = 8.8) for the Familiar Tasks and 13.2 for the New Tasks (SD = 14.7). The Control Group (Mean = 11.6; SD = 12.3) made barely more errors than the Perceptive Group (Mean = 11.0; SD = 13.0), which made more errors than the Inductive Group (Mean = 7.9; SD = 9.6). Also, Experientials (Mean = 11.4; SD = 13.6) made more errors than Rationals (Mean = 9.1; SD = 10.11).

The 3 (Level of Induction) x 2 (Task Familiarity) x 2 (REI) ANOVA revealed a main effect of the Task Familiarity, $F(1,229) = 10.286$; $p < 0.005$, as well as of the Level of

Induction, $F(2,229) = 3.684$; $p < 0.05$, validating the Induction Hypothesis. The REI factor did not reach significance, $F(1,229) = 2.641$; *n.s.* The only significant interaction was between the 3 factors, *i.e.* Task Familiarity x REI x Level of Induction, $F(2,229) = 4.070$; $p < 0.05$.

The Induction and the Interaction hypothesis were specifically examined in separate post-hoc analyses of the Experientials' and the Rationals' performances for the two experimental conditions (Perceptive and Inductive Groups; see Figure 6). The ANOVA of the Experientials' performances revealed the unique effect of the Task Familiarity, $F(1,70) = 0.323$; $p < 0.05$. The ANOVA of the Rationals' performances revealed a very significant effect of Task Familiarity, $F(1,94) = 17.178$; $p < 0.001$, and of Level of Induction, $F(1,94) = 12.039$; $p < 0.001$, and a significant interaction between the two factors, $F(1,94) = 7.576$; $p < 0.01$.

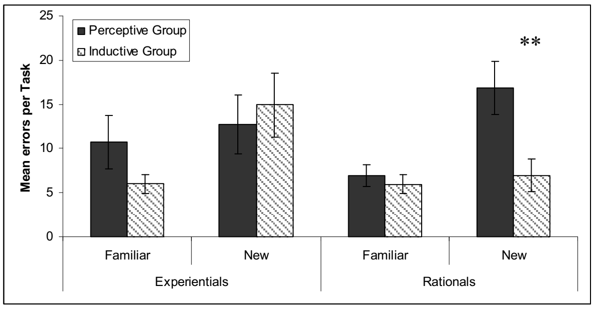


Figure 6. Mean errors per task by REI for the Perceptive and the Inductive Groups

The Rationals hence exhibited the attended Induction and Interaction hypotheses, *i.e.* only benefited from the inductive condition for New Task. The procedure however statistically failed to make Experientials induce schemas for New Tasks.

Yet, Experientials tended to make more errors and to benefit from induction during Familiar Tasks, which contradicts the current conception that familiarity supports intuitive interaction. We foresee here that Task Familiarity might, in fact, be deleterious for users that rely on their intuitions and prior knowledge to solve new tasks and problems.

3.2 Diagnosing perspective

Schema induction can be further addressed by discussing the benefits of schema induction over classical user tests and familiarity evaluation, and by relating prospectively the observed patterns of performance to design recommendations.

3.2.1 Schema induction versus task familiarity

The fact that Task Familiarity statistically improves performances does not imply a one-to-one correspondence between these two variables. Raw data per task and per condition indeed revealed discrepancies between Task Familiarity and performances (see Figure 7).

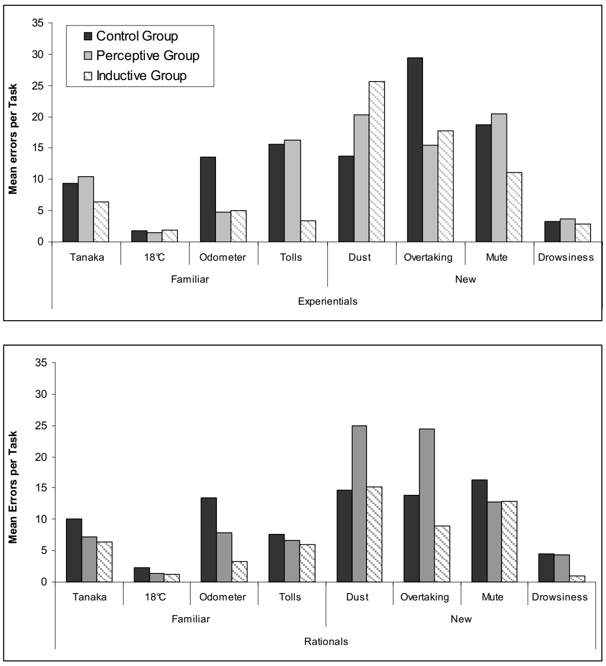


Figure 7. Mean errors per task by REI and by Level of Induction

For example the most intuitive tasks, *i.e.* 18°C and drowsiness were respectively coded as Familiar and New. The three other Familiar Tasks were relatively less intuitive to set in DoIt#. Prior familiarity subjectively reported by the participants poorly accounted for raw data performances. This indicator does not seem reliable enough to replace empirical evaluation.

3.2.2 A step towards design recommendations

The present experiment was designed to test two behavioral patterns known in other domains of psychology to account for induction and transfer.

- Few errors for both the Perceptive and the Inductive Groups indicate a transfer of prior schemas.
- More errors for the Perceptive compared to the Inductive Group reflect a positive induction effect (*cf.* Overtaking by Rationals, Figure 7).

Yet, raw data exhibited two additional patterns of performances.

- Fewer errors from the Perceptive in comparison to the Inductive Group reflect a negative induction effect (cf. Dust by Experientials).
- High errors for both experimental groups indicate that the Function Matching Task was inoperant in inducing new schemas (cf. Mute by Rationals).

Four implications for the design can be prospectively stated.

First, tasks that foster prior schema transfer are intuitive and do not need to be redesigned.

Second, tasks exhibiting a positive induction effect should gain in intuitiveness if familiar, descriptive or contextual information is added in the concerned states. Such intervention should actually lead to consider affordances and stereotypes.

Third, the negative induction might reflect a conceptual inconsistency between the schemas induced by the Inductive Group and of Doit#'s design. For example, the schemas induced by the Inductive Group might lead to consider erroneously that a menu does not fit well with the current goal and to search elsewhere. Studies should be performed to address further this negative induction effect. It would especially be interesting to determine whether designers should here adopt the Inductive Group's logic of interaction, or whether they should keep but improve their design with contextual clues.

At last, tasks for which the inductive condition was inoperant are certainly those to amend in priority. They might require metaphors and abstractions from other domains, instead of local and domain specific information, to gain in intuitiveness.

The induction procedure obviously enables to go beyond classical user tests (e.g. Control Group) in that it indicates among the low performance tasks those which actually can be induced, those which actually suffer from inconsistency and those actually too difficult to support any schema operation.

4. CONCLUSION

We identified from the multimedia learning literature that:

- schema induction can be obtained by using a comparative task applied to familiar and detailed content,
- inductive tasks interact with prior knowledge, expertise and familiarity factors.

Our empirical contribution consisted in adapting these two principles to the particular case of multistate interfaces.

The proposed procedure is relatively easy to perform in a design process. First, the main advantage of behavioral methods is to minimize the intervention of experimenters during the collection, the processing and the interpretation of data. Here, the tests do not require from the experimenter a specific expertise in Human Factors and are fully instrumented. Second, time resources are reasonable as about twenty persons can be tested in a couple of days. Moreover, data processing can be largely automated from the moment that actions on commands are recordable.

The schema induction procedure differentiated prior schema transfer from new schema induction. It also seemed to account for inconsistent and inoperant induction effects. The overall method appears relevant to study whether stereotypes, affordances, metaphors or consistencies contribute to intuitiveness (more transfer effect), assimilation (more induction effect), or whether they are inoperant.

The study also revealed two interesting facts about intuitiveness and about the role of cognitive schema in Human Computer Interactions.

First, familiar interfaces might be deleterious for the users who tend to rely on their intuitions. Though the distinction between Rationals and Experientials did not affect significantly the overall performances, we should keep in mind that experiential users are, in fact, misled by familiar contents. Additional studies should specificallty address this issue and state to what extent experientials fail by intuition. If such hypothesis is confirmed, research on intuitiveness should include strategies to limit familiarity and to prevent experiential users from interacting with interfaces in an instinctive but erroneous fashion, *i.e.* to adopt strong but wrong behaviors.

Second, several tasks lead participants to make about thirty errors. As DoIt#'s main menus counts less than 20 states each, it is probable that most states were seen several times, but that participants failed to understand and to remember them. This remark corroborates that:

- Human Computer Interaction is essentially a reactive activity [17],
- performance remains low as long as the states' schemas are no induced [17],
- means-end analysis interferes with the induction of new schemas [37].

Schemas, which the present study demonstrated the operationality, seem all the more reason to be a key step to intuitive interaction.

5. ACKNOWLEDGMENTS

The present research was credited by the GCOE Cybernics Program, in the University of Tsukuba. We would like to acknowledge Alejandra Quiros for her assistance during the experiment preparation and the tests.

6. REFERENCES

[1] Blackler, A., Popovic, V., and Mahar; D. 2007. Empirical investigations into intuitive interaction: A summary. MMI-Interaktiv. 13, 4-24.

[2] Blackler, A., Popovic, V., and Mahar D. 2003. The nature of intuitive use of products: An experimental approach. Design Studies. 24(6), 491-506.

[3] Newell, A., and Simon, H.A. 1972. Human problem solving. New Jersey, Prentice Hall.

[4] Kahneman, D. 2003. A perspective on judgment and choice: Mapping bounded rationality. American Psychologist. 58(9), 697-720.

[5] Woolhouse, L. S., and Bayne R. 2000. Personality and the use of intuition: Individual differences in strategy and performance on an implicit learning task. European Journal of Personality. 14(2), 157-169.

[6] Epstein, S., Pacini, R., Denes-Raj, V., and Heier, H. 1996. Individual differences in intuitive-experiential and analytical-rational thinking styles. Journal of Personality and Social Psychology. 71(2), 390-405.

[7] Reber, A.S., and Walkenfeld, F.F. 1991. Implicit and explicit learning: Individual differences and IQ. Journal of Experimental Psychology: Learning, Memory and Cognition. 17, 888-896.

[8] Reber, R., Ruch-Monachon, M.A., and Perrig, W.J. 2007. Decomposing intuitive components in a conceptual problem solving task. Consciousness and Cognition. 16(2), 294-309.

[9] Tversky; A., and Kahneman, D. 1973. Availability: A heuristic for judging frequency and probability. Cognitive Psychology. 5, 207-232.

[10] Blackler, A., and Hurtienne; J. 2007. Towards a unified view of intuitive interaction: Definitions, models and tools across the world. MMI-Interaktiv. 13, 36-54.

[11] Naumann, A., Hurtienne, J., Israel, J. H., Mohs, C., Kindsmüller, M. C., Meyer, H. A., Hußlein, S., and IUUI Research group. Intuitive use of user interfaces: Defining a vague concept. In D.Harris (Ed.).Engineering Psychology and Cognitive Ergonomics, HCII 2007, Vol.13, Springer, Heidelberg (2007), 128--136. 2009.

[12] Ozok, A.A., and Salvendy, G. 2003. The effect of language inconsistency on performance and satisfaction in using Web: Results from three experiments. Behaviour and Information Technology. 22(3), 155-163.

[13] Gick, M.L., and Holyoak, K.J. 1980. Analogical problem solving. Cognitive Psychology. 12(3), 306-355.

[14] Gick, M.L., and Holyoak, K.J. 1983. Schema induction and analogical transfer. Cognitive Psychology. 15, 1-38.

[15] Gentner, D., Loewenstein, J., and Thompson, L. 2003. Learning and transfer: A general role of analogical encoding. Journal of Educational Psychology. 95(2), 393-408.

[16] Norman, D.A. 1984. Stages and levels in human-machine interaction. International Journal of Man-Machine Studies. 21(4), 365-375.

[17] Norman, D.A. 1988. The psychology of everyday things. Basic Books, New York.

[18] Hartson, H.R. 2003. Cognitive, physical, sensory, and functionnal affordances in interaction design. Behaviour and Information Technology. 22(5), 315-338.

[19] Fields, B., Wright, P. C., and Harrison, M. 1997. THEA: Human error analysis for requirements definition. Technical Report 294. University of York Computer Science.

[20]Andre, T.S., Hartson, H.R., Belz S.M., and McCreary, F.A. 2001. The user action framework: a reliable foundation for usability engineering support tools. International Journal of Human-Computer Studies. 54, 107-136.

[21] Rizzo, A., Marchigiani, E., and Andreadis, A. The AVANTI Project. 1997. Prototyping and evaluation with a Cognitive Walkthrough based on the Norman's model of action. ACM Conference on Designing Interactive Systems '97 .

[22] Mayer, R.E., and Anderson, R.B. 1991. Animations need narrations: An exprimental test of dual-coding hypothesis. Journal of Educational Psychology. 83(4), 484-490.

[23] Kalyuga, S., Chandler, P., Sweller, J., and Tuovinen, J. 2001. When problem solving is superior to studying worked examples. Journal of Educational Psychology. 93(3), 579-588.

[24] Peper, R.J., and Mayer, R.E. 1986. Generative effects of note-Taking during science lectures. Journal of Educational Psychology. 78(1), 34-38.

[25] Mayer, R.E. 1980. Elaboration techniques that increase the meaningfulness of technical text: An experimental test of the learning strategy hypothesis. Journal of Educational Psychology. 72, 770-784.

[26] Mayer, R.E. 2001. Multimedia learning. Cambridge University Press, New York.

[27] Kalyuga, S., Chandler, P., and Sweller, J. 2000. Incorporating learner experience into the design of multimedia instruction. Journal of Educational Psychology. 92, 126-136.

[28] Kalyuga, S., Chandler, P., and Sweller, J. 1998. Levels of expertise and instructional design. Human Factors. 40(1), 1-17.

[29] Novick, L.R. 1988. Analogical transfer, problem similarity, and expertise. Journal of Experimental Psychology: Learning, Memory and Cognition. 14(3), 510-520.

[30] Brown, A.L. 1989. Analogical learning and transfer: What develops? in Similarity and analogical learning. Cambridge University Press, New York.

[31] Hsu, Y.C. 2005. The long-term effects of integral versus composite metaphors on experts' and novices' search behavior. Interacting with Computers. 17, 367-394.

[32] McNamara, D.S. 2001. Reading both high-coherence and low-coherence texts : Effects of text Sequence and prior knowledge. Canadian Journal of Experimental Psychology. 55(1), 51-62.

[33] Stanovich, K.E., and West, R.F. 1999. Discrepancies between normative and descriptive models of decision making and the understanding/ acceptance principle. Cognitive Psychology. 38, 349-385.

[34] Pacini, R., and Epstein, S. 1999. The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. Journal of Personality and Social Psychology. 76, 972-987.

[35] Naitou, M., Suzuki, K., and Sakamoto, A. 2004. Development of Rational and Intuitive Information-Processing Style Inventory. The Japanese Journal of Personality. 13(1), 67-78.

[36] Rauterberg, M. 1993. AMME: An Automatic Mental Model Evaluation to analyse user behaviour traced in a finite, discrete state space. Ergonomics. 36(11), 1369-1380.

[37] Sweller, J. 1988. Cognitive load during problem solving: effects on learning. Cognitive Science. 12(2), 257-285.

The theater-system technique: Agile designing and testing of system behavior and interaction, applied to highly automated vehicles

Anna Schieben Matthias Heesen Julian Schindler Johann Kelsch Frank Flemisch

Institute of Transportation Systems, DLR e.V.
Lilienthalplatz 7
D-38108 Braunschweig
+49-(0)531-295-3426

anna.schieben@dlr.de

ABSTRACT
In this paper, the theater-system technique, a method for agile designing and testing of system behavior and interaction concepts is described. The technique is based on the Wizard-of-Oz approach, originally used for emulating automated speech recognition, and is extended towards an interactive, user-centered design technique. The paper describes the design process using the theater-system technique, the technical build-up of the theater-system, and an application of the technique: the design of a haptic-multimodal interaction strategy for highly automated vehicles. The use of the theater-system in the design process is manifold: It is used for the concrete design work of the design team, for the assessment of user expectations as well as for early usability assessments, extending the principles of user-centered design towards a dynamically balanced design.

Categories and Subject Descriptors
D.2.2 [Software Engineering]: Design tools and techniques.

General Terms
Design, Human Factors.

Keywords
Theater-system technique, Wizard-of-Oz technique, highly automated vehicles, haptic interaction, design process, user-centered design, balanced design.

1. AGILE DESIGNING AND TESTING UNDER RESOURCE CONSTRAINTS IN GENERAL
Human-machine systems are shaped by technological progress and a natural selection in the market place: Good products earn enough money to be further developed, less adequate products disappear. As the number of design alternatives can be large and the development costs for a certain product high, it can make sense to boost the "natural" selection of the market place with an accelerated selection in an agile development and assessment process. Efficiency to explore larger portions of the design space is crucial for agile design and testing techniques [1]. Many times, the user is the one who will decide, or contribute to the decision for or against a new technical system. It can therefore be beneficial to let the user actively participate early enough in the design process (see the principles of user-centered and participatory design [2].

One method for agile designing and testing of interface and interaction concepts is the theater-system technique that allows the involvement of users from the beginning of the design process in a very tangible way [3]. The following paper addresses how the theater-system technique works in general and takes the technological development in the vehicle domain as an application example to show how the technique is used for agile designing and testing of haptic-multimodal interaction for highly automated vehicles.

2. APPLICATION DOMAIN: HIGHLY AUTOMATED VEHICLES AND HAPTIC-MULTIMODAL INTERACTION
The current trend in the vehicle industry is to bring more and more assistance systems and automation on board of the vehicles like Adaptive Cruise Control (ACC) or Lane Keeping Systems (LKS). This results in so-called highly automated vehicles [4]. From the perspective of a human-machine interface designer, the increasing automation in the vehicles comes along with the need of an adequate interaction design that allows the driver as well as the automation to guide the vehicle in a cooperative way (Figure 1). Both, the driver and the automation build up intentions and act on the vehicle guidance.

The requirements for the interaction design for such highly automated vehicles are mainly to keep the driver in the loop, to ensure he is aware of the current automation mode and to support the driving in different automation levels as well as the transitions between these levels. One approach to meet these requirements is the use of a haptic interaction strategy that is enriched with visual and auditory elements – a haptic-multimodal interaction.

AutomotiveUI'09, September 21-22, 2009, Essen, Germany
ACM 978-1-60558-571-0/09/009

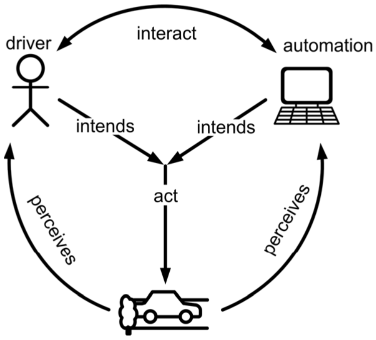


Figure 1: Interplay between the driver and the automation guiding the vehicle by cooperative control

Haptic interaction with a highly automated vehicle mainly happens via force feedback devices. Force feedback devices allow to display a variety of haptic signals, such as continuous forces, vibrations or discrete signals like double ticks for example on the steering wheel or the accelerator pedal [5]. With the help of the haptic feedback the driver can always be provided with information about the current actions and intentions of the vehicle automation, for example via steering wheel movements or forces on the pedals. However, the flow of information is not only directed from the automation towards the driver but also vice versa from the driver to the automation. For example, the driver could have the option to activate and command maneuvers by applying tics or forces on the steering wheel.

For the design of such haptic-multimodal interaction for highly automated vehicles we use the theater-system technique in all different stages of the design process.

3. THE THEATER-SYSTEM TECHNIQUE IN GENERAL

The theater-system technique is based on the idea to do a rapid prototyping of system behavior and haptic-multimodal interaction long before the complex software for such a prototype is build up. The theater-system technique is based on the Wizard-of-Oz technique (WoOz), where a human "wizard" hidden behind a curtain is emulating the functionality of a machine [6]. Originally, the technique was used for automatic speech or gesture recognition and picked up in other domains.

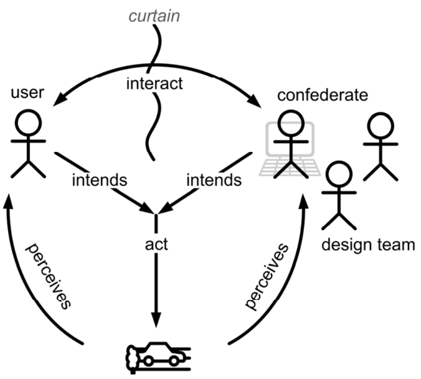


Figure 2: Interplay between the user, the confederate as member of the design team and the task (here a vehicle guidance task)

The theater-system technique extends the WoOz technique in a way that there is no longer a hidden wizard but that the curtain between the user and a member of the design team (confederate) can also be open, and both user and confederate can play through different use cases as if they would play a role in a theater (Figure 2). Whereas the WoOz technique is used for the evaluation of functionality, the theater-system can be used both for evaluation and design.

A typical design process with the theater-system is shown in Figure 3. Based on the initial ideas and an early analysis of the design challenge, an appropriate infrastructure has to be set up or adapted. This includes the adaptation of the theater-system itself for the emulation of the automation behavior and interaction in the chosen scenarios and tasks.

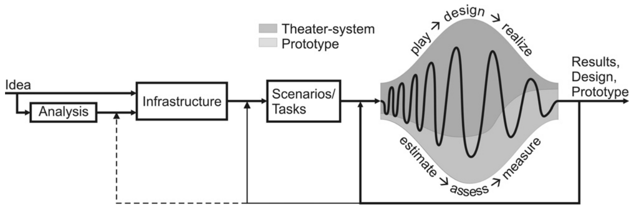


Figure 3: Schematic depiction of the design process using the theater-system in early and software prototypes in later stages of the iterative process

During the iterative design process, prototypes played by the confederate in the theater-system and software prototypes work as complement: Starting with a more open play with design variables and estimation of their effects with the confederate, design options are designed in detail and their effects assessed, until the design can be condensed, realized in software and its effect measured. This loop of infrastructure & scenario adaption, play, detail design, selection and realization can be iterated as often as necessary.

4. APPLICATION EXAMPLE: THE THEATER-SYSTEM TECHNIQUE FOR HIGHLY AUTOMATED VEHICLES

The technique is used so far for designing interaction for vehicles (DLR, TU Munich) and cockpit interaction for aircrafts and helicopters in simulation environments (NASA, DLR). In addition, one aspect of the WoOz/ theater-system technique, here the emulation of the behaviour of the assistance and automation functions, has already been applied to a real car for driving tests on public roads [7].

At the Institute of Transportation Systems at DLR Braunschweig (DLR-TS), the work with the theater-system focuses on haptic-multimodal interaction for highly automated vehicles [8]. The technique is used during the early design work by the design team and for discussion with external partners like vehicle manufactures and users, for the assessment of user expectation as well as for first usability assessments of the interaction design for highly automated vehicles.

4.1 Technical setup of the theater-system at DLR-TS

The theater-system at DLR-TS consists of two static low fidelity simulators located next to each other in a distance of about two meters. Both simulators include force feedback control devices coupled with each other mechanically or electronically as a redundant set of controls. In a current implementation of the theater-system two electronically coupled force feedback side-sticks, and two mechanically coupled force feedback steering wheels as well as two force feedback pedals which are also coupled mechanically, are realized (Figure 4).

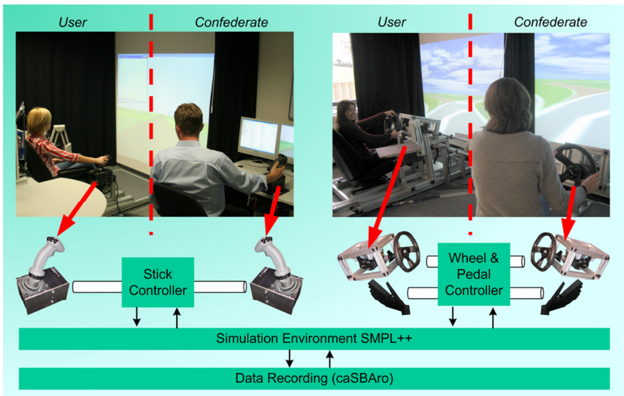


Figure 4: Implementation of the theater-system with coupled inceptors at DLR-TS

For the work with the theater-system one member of the design teams plays the confederate. The confederate is, similar to the wizard, responsible for emulating the vehicle behavior and interaction. The coupled inceptors allow the confederate, who is sitting in the right simulator of the theater-system, at any time to feel the tactile and haptic input of the user, sitting in the left simulator of the theater-system.

Another important feature of the theater-system is indicated by the vertical dashed line in figure 4. The line of sight between the user and the confederate can be obstructed by a curtain (similar to WoOz). In addition, the curtain can be open. Then, the confederate and the user can directly communicate and interact with each other. This enables an open dialogue between the confederate and the user. In this dialogue, expectancies on the behavior of highly automated vehicles or concrete design ideas can be queried by the confederate (see section 4.3).

By closing the curtain the theater-system can be used similar to the WoOz technique for exploration or testing of haptic-multimodal design elements that are not yet implemented in software. Therefore, prototypes of interaction designs are simulated either completely by the confederate or by a combination of confederate inputs and already implemented software parts of the automation (see section 4.4).

4.2 Design work in the theater-system

For the design work, the members of the design team use the theater-system for the generation and test of design ideas. Several design options are played through and documented with video records, data records and in text and pictures. In general, the haptic interaction follows a design scheme that was developed during the work for several projects: Continuous signals like forces are used to display the current behavior of the automation. Vibrations are used for warnings and alerts. Discrete signals like tics are used for communicating intentions of the automation or to trigger maneuvers. These haptic signals can be combined to more complex interaction patterns. For example, the so called "virtual gravel pit" is a combination of forces and vibrations that is displayed in case of an unintentional lane departure. The haptic feeling is similar to a real gravel pit; the vehicle is jounced and slowed down. For urgent warnings and for the communication of future events or intentions of the automation the haptic interaction is enriched with visual and acoustic signals.

4.3 Assessment of user expectations

After collecting first ideas of the design within the design team, the theater-system can be used for assessments of user expectations. For this, the curtain of the theater-system is open. During the expectation assessment the confederate does not simulate any predefined system behaviour but asks the user about his expectations. The assessment is conducted in form of a semi-structured interview during which the confederate leads the user through a sequence of predefined traffic scenarios. The user can express his expectations verbally, but of even more importance for the design process, the user can use the theater-system to directly show the confederate which kind of haptic interaction he or she expects in the given scenario. For example, the following dialog could be heard when discussing the design of a haptic interaction when exceeding the speed limit:

Confederate: "What would you expect if you exceed the current speed limit?"
User: "Maybe a force on the accelerator pedal that pushes me back, followed by a vibration if I do not react."
Confederate: "How would that feel? Like this?" (Conf. demonstrates a soft force threshold on the pedal) "Or more like this?" (Conf. demonstrates a hard force)
User: "I'd prefer it a little stronger force, more like this…" (User demonstrates directly on the coupled pedal what he expects).

The user expectancy assessment allows to get some insights either in the naive expectations that drivers have about the general functionality of vehicle automation and the way this automation interacts with the driver, or in the expectation which users derive from a design metaphor [3].

The advantages of this approach are:

- The confederate directly grasps what kind of haptic interaction the user expects and how it feels like.
- He directly perceives in how far the user expectations differ from the primarily intended design.
- He directly experiences new, possible design variations that the design team did not think of.

4.4 Usability assessment in the theater-system

Besides the user expectancy assessment, the theater-system is used for usability assessments of interaction designs before implementing the design into software prototypes. For the usability assessment approach, the theater-system is used similar to the WoOz technique. The confederate is intensively trained to emulate a specific haptic interaction. During the usability assessment, the curtain of the theater-system is closed and the confederate wears ear plugs to avoid vocal communication. The confederate emulates the automation behavior and interaction while the user drives through different scenarios. Depending on the predefined design this could be for example a slight lane keeping force or in case of lane departure tics or vibrations on the steering wheel. During the runs driving and interaction data, acceptance ratings and thinking aloud protocols are assessed for further analysis. Even though, the confederate can not reproduce the system behavior as standardized and consistent as a software prototype, the approach has one important advantage regarding the understanding of different user behavior: Before analysing any data, the confederate gets a first, intuitive impression of the interaction of the user with the system and potential conflicts by feeling the input of the user on the steering devices.

Based on the outcomes of the expectation assessment and the usability assessment the design is improved and modified. This modified design is then transferred into first software prototypes.

4.5 Confederate = Human-machine interface designer

For the implementation into software, the confederate who has internalized the complex behavior of the prototype in every situation ideally does or leads the implementation of the software prototype. That way, every small part of the prototype, any “feelage” of haptic interaction, can be replicated almost as originally designed. As the confederate may not be a computer scientist, it is therefore necessary that the programming framework is easy to use and easy to understand. DLR-TS uses the Straightforward Modular Prototyping Library in C++ (SMPL++) for this purpose, which is developed by DLR, NASA Langley and several university partners since 2001 [1]. SMPL++ as a rapid prototyping framework already includes several tools for the agile development of prototypes: E.g. SMPLcaSBAro (Computer Aided Situation Behaviour Analysis Replay/Online) serves as a recording tool with capabilities of monitoring recorded data from an overview perspective as well as from a very detailed perspective down to each record entry. This feature, based on the “Pointillistic Analysis” [9] enables the confederate to directly check the behavior of the software prototype. Another valuable tool is the SMPLControlPanel. It can be used for monitoring as well as for changing each variable of the software prototype during runtime, so it is an ideal tool for fine-tuning. Altogether, the complete SMPL++ toolbox combined with the haptic memory of the confederate enables the rapid prototyping of high quality prototypes.

5. DISCUSSION AND OUTLOOK

The theater-system technique is currently used successfully in several projects like H-Mode, IMOST, HAVEit and CityMobil that focus on assistance and automation for vehicles for urban and highway applications. For example, in the project IMOST a haptic-multimodal interaction strategy for a system that assists drivers on highway entries is developed with the help of the technique.

As the DLR-TS theater-system uses only a low fidelity simulation and does not provide, e.g. any vehicle movements, the prototypes are further tested in more realistic environments like the DLR-TS motion-based simulator or the research vehicle FASCar.

Altogether, the technique has a high potential to bridge different domains and perspectives, e.g. a user-centered and a technical perspective. We will continue to use and improve this technique as integral part of an ergonomic tool and technique portfolio that is a basic prerequisite for a better, well-balanced design of human- machine systems.

6. ACKNOWLEDGMENTS

The approach was partially developed during a National Research Council associateship of Frank Flemisch at NASA Langley, and refined with support of the Deutsche Forschungsgemeinschaft DFG in the project "H-Mode for vehicles".

7. REFERENCES

[1] Schindler, J., and Flemisch, F. 2007. Prospektive Gestaltung von Fahrzeugassistenz: Agile Prototypentwicklung und integriertes Testen am Beispiel des DLR-SmpLab; In Proceedings of the 7. Berliner Werkstatt Mensch-Maschine-Systeme (Berlin, Germany, October 10-12, 2007).

[2] Norman, D. A. 1988. The Psychology of Everyday Things. New York, Basic Books.

[3] Flemisch, F.O., Goodrich, K.H., and Conway, S.R. 2005. At the crossroads of manually controlled and automated transport: The H-Metaphor and its first applications (progress update 2005); In Proceedings of the ITS Europe 2005 (Hannover, Germany, June, 1-3, 2005).

[4] Flemisch, F. O., Kelsch, J., Löper, C., Schieben, A., and Schindler, J. 2008. Automation spectrum, inner/outer compatibility and other potentially useful human factors concepts for assistance and automation. In D. de Waard, F. O. Flemisch, B. Lorenz, H. Oberheid and K. A. Brookhuis (Eds.), Human Factors for Assistance and Automation. Maastricht: Shaker Publishing.

[5] Schieben, A., Damböck, D., Kelsch, J., Rausch, H., and Flemisch, F. 2008. Haptisches Feedback im Spektrum von Fahrerassistenz und Automation. In Proceedings of 3. Tagung Aktive Sicherheit durch Fahrerassistenz (Garching, Germany, April 7-8, 2008).

[6] Kelley, J.F. 1984. An iterative design methodology for user-friendly natural language office information applications. ACM Transactions on Office Information Systems, 2:1, pp. 26–41.

[7] Kiss, M., Schmidt, G., and Babbel, E. 2008. Das Wizard of Oz Fahrzeug: ein Werkzeug für Rapid Prototyping and Usability Testing von zukünftigen Fahrerassistenzsystemen. In Proceedings of 3. Tagung Aktive Sicherheit durch Fahrerassistenz (Garching, Germany, April 7-8, 2008).

[8] Flemisch, F., Schindler, J., Kelsch, J., Schieben, A., and Damböck, D. 2008. Some Bridging Methods towards a Balanced Design of Human-Machine Systems, Applied to Highly Automated Vehicles; In Proceedings of Applied Ergonomics International Conference (Las Vegas, USA, July 14-17, 2008).

[9] Flemisch, F., and Onken, R. 2002. Open a window to the cognitive work process! Pointillist analysis of man-machine interaction. Journal of Cognition, Technology and Work. 4:3.

Towards a Flexible UI Model for Automotive Human-Machine Interaction

Guido de Melo, Frank Honold, Michael Weber
University of Ulm
Institute of Media Informatics
89081 Ulm
{guido.de-melo, frank.honold, michael.weber}@uni-ulm.de

Mark Poguntke, André Berton
Daimler AG
Research and Development
Infotainment and Telematics
89081 Ulm
{mark.poguntke, andre.berton}@daimler.com

ABSTRACT

In this paper we present an approach for creating user interfaces from abstract representations for the automotive domain. The approach is based on transformations between different user interface abstraction levels. Existing user interface representation methods are presented and evaluated. The impact of specific requirements for automotive human-machine interaction is discussed. Considering these requirements a process based on transformation rules is outlined to allow for flexible integration of external infotainment applications coming from mobile devices or web sources into the in-car interaction environment.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces

Keywords

human-machine interaction, user interface modeling, user interface generation, UML, Cameleon Reference Framework

1. INTRODUCTION

1.1 Overview

The increasing development and ubiquity of infotainment applications plays an important role for automotive manufacturers. Mobile devices, like smart phones or mp3-players, are widespread and increasingly used. In an in-car environment the use of external devices distracts the driver from the important task of driving. Also, there are legal regulations in many countries that prohibit the use of mobile devices while driving. The convenient integration of different external devices and services, such as smart phones and on-line services, in the in-car environment is desired. Since the development time and life cycle of automotive software is usually much longer than the life cycle of consumer electronics or web-based services, a flexible solution for the adaption of new applications to the existing automotive human-machine interface (HMI) is required. In order to integrate applications of different devices to the automotive HMI, it is necessary to map the device capabilities to the interaction devices within the automotive environment, e. g. central control unit and head-unit display. Additionally, voice control of external applications should be possible via the in-car speech dialog system.

Each automotive manufacturer provides their own HMI specified by colors, images, font styles, interaction concepts and flows. Since there are important automotive-specific requirements like font size settings for minimal driver distraction, the aim is to provide an appropriate HMI concept meeting these requirements, e. g. according to the European Statement of Principles on HMI for in-vehicle information and communication systems (ESoP). Also, for safety reasons, and as a distinctive feature, the control over the automotive HMI has to be completely handeled by the car software. The latter is hard to achieve if external applications are to be integrated which provide their own user interfaces. More abstract representations of user interface concepts build a basis for different concrete user interfaces. Devices should provide abstract descriptions of their functionality and capabilities. And the head-unit would transform these descriptions to the automotive HMI as needed. Thus, the integration of external devices would be possible after deployment of the car software, and the integrated system would still be controlled by the manufacturer.

1.2 Scenario

The simple example of integrating a portable music player into the head-unit HMI illustrates important issues to be solved. The driver or passenger intends to use any player in the car. This may be a very simple device providing basic audio player functionalities or a more complex device providing additional features like album cover presentations, different playlists, a dynamic play order based on similarity of songs and so on. Current car head-units provide their own audio player with mp3-support. Connecting an mp3-player to the car will initiate the car's audio player which is then used to play the songs contained on the device. However, device-specific interaction features are not integrated if they are unknown to the head-unit. Furthermore, completely unknown applications like a calendar cannot be integrated at all.

AutomotiveUI'09, September 21-22, 2009, Essen, Germany
ACM 978-1-60558-571-0/09/009

2. ABSTRACT USER INTERFACE REPRESENTATIONS

In order to achieve a seamless integration, the device capabilities and interaction possibilities have to be transferred to the head-unit which then processes this information to map it to the car-specific interaction and presentation devices. Due to the diversity of interaction concepts, e. g. hard keys, touchscreens, speech interfaces or motion sensors, the representation of device capabilities has to be in an abstract manner. This ensures that the interaction possibilities can be transferred independent of specific user interface concepts or modalities. We use the term UI model for an abstract representation of a user interface which is independent of a certain implementation. Requirements for a flexible UI model are presented in the following. Based on these, existing UI representation techniques are presented and evaluated.

2.1 Requirements

A flexible UI model should fulfill a number of requirements. Van den Bergh and Coninx described some less formal requirements for the working environment [4]. The environment shall be expressive: The model shall be comprehensible and allow for complex relationships without becoming cumbersome. Tool support shall be possible since tools can ascertain that models are consistent. They also enable hiding of parts of the model during design. Other important requirements encompass internationalization. A UI shall be adaptable to different languages and cultures. According to Weld et al. units of measurement like speed, date, and time should be provided in a format matching the user's preferences [12].

A user interface shall also be consistent. User interaction shall run along the same lines each time, as Dix et al. describe [2, p. 584]. At the same time a UI has to comply to a set of ergonomic standards, like the ones set forth in ISO 9241-110. An automotive user interface furthermore has to follow certain automotive guidelines, e. g. the before mentioned ESoP guidelines.

The architecture needs to be extendable to achieve a consistent UI. It has to integrate hardware built into the car as well as additional external devices the user wants to employ within the car. Apart from devices, new functionality can also be provided by services from the web. The model itself needs to be extended at runtime to integrate the functionality of new applications.

The model also needs to be independent of the employed hardware. If the user connects their own mp3-player, the whole system shall be able to respond to commands issued via buttons as well as speech. Since the user's devices will typically not provide an own head-unit HMI or speech UI, the system will have to translate between these modalities and each device's service. Accordingly, a central control module is needed that is able to distribute respective presentation and interaction logic to the involved system components.

2.2 Model Components

Apart from requirements for the notation different aspects of the model need to be described. Our model uses a distinction between the application, tasks and the user interface. We employ the Cameleon Reference Framework which specifies four levels of abstraction [1]. The framework is illustrated in figure 1. The different abstraction levels are shown starting with *Tasks and Concepts* (T&C) at the topmost level. Tasks can be modeled using different notations which are evaluated below. Concepts are all domain objects, in our example the task *Play next mp3-track* invokes a method on a song object. The *abstract user interface* (AUI) is a modality-independent model of the UI. *Concrete user interface* (CUI) is the level at which widgets are employed and the *final user interface* (FUI) is the binary code or the UI in a markup language or hardware mapping. Since the task model concentrates on a high-level description of the user's actions, we use this location in the architecture to extend the functionality of the system.

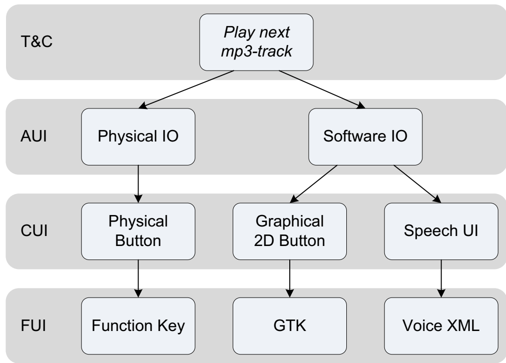


Figure 1: Different abstraction levels in the Cameleon Reference Framework starting with the task *Play next mp3-track*. Arrows denote transformations.

Transformations take place between the different models. The arrows in figure 1 indicate these transformations. The UI is generated from abstract representations by applying respective transformation rules for the HMI design, modality and capabilities of the target interaction devices. These rules have to be implemented on the car head-unit. Thus, the transformations are in control of the manufacturer while still giving them a flexible solution for integrating unknown external services. The approach is described in section 3. Any transformation entails some disadvantages, e. g. there is an additional overhead for maintaining all mapping rules between different models. In addition, to transform from one level of abstraction to another, an accurate mapping has to be found, so that no relevant information is lost in the process. The problem of determining these transformations is called the *mapping problem* and has been widely discussed by Puerta et al. [10].

2.3 Related Work

Recent work on approaches for abstract user interface representations are presented and evaluated in this section.

2.3.1 Existing approaches

Several proposals for abstract description standards exist which can be used to build upon. Concur Task Trees (CTT) [7] is one of the most widely discussed approaches to describe human machine interactions in an abstract manner.

In CTT a user interface model is specified by using several well-defined types of tasks and operators in a hierarchical top down description. Since there are no descriptions of any concrete UI elements, the task description remains platform and modality independent. A concrete UI is generated during the interpretation of a task tree either before executing the application or at runtime.

As demonstrated by Nóbrega et al. [5], the wide-spread modeling language UML [6] provides the same expressiveness as CTT. As established standard in software development, structured contents and problem descriptions can be illustrated in UML by a topology of class diagrams, activity diagrams, and statechart diagrams. Thus, an abstract and formal UI model can be specified in order to be concretized in further process steps. The description of user interactions is platform and modality independent and can be translated to different modalities.

2.3.2 *Evaluation*

Describing an automotive HMI on a task level with the widespread CTT-notation quickly resulted in easily understandable task trees. However, due to increased concurrency, e. g. destination entry while listening to music and accepting an incoming phone call, the tree structure quickly becomes very complex. Also, CTT does not provide a history concept which is often needed in the automotive domain for task switching. Usually, the applications open with the last active state. Thus, modeling with CTT may be useful for some interaction tasks but the approach lacks important aspects needed for more complex task descriptions.

As mentioned before UML provides the same expressiveness as CTT. Furthermore, the UML notation concepts exceed CTT, so that the emerged issues can be addressed by UML. Thus, we considered UML to be the appropriate approach to model the interactions on abstract level.

3. TRANSFORMATION-BASED UI

Since the car's HMI needs to integrate new devices during its lifetime, the preinstalled descriptions of tasks need to be updated from time to time. This can be done automatically by attaching an unknown device which provides its own task descriptions and abstract UI model or by hand through the user. The UI is then generated on demand. For this process transformations are needed between the different abstraction levels. Existing approaches are presented and evaluated in the following. Then, our transformation-based approach is outlined.

3.1 Related Work

Several approaches to generate UIs from abstract representations exist. UIML (User Interface Markup Language) is an XML-based Meta-Interface Model (MIM) [9]. Apart from the UI definition, the runtime behavior of an application can be exemplified. Interfaces for different modalities can be specified in UIML, however each modality-dependent specification is directly bound to the underlying abstract description and restricts the overall flexibility. While UIML is a widespread and advanced approach for modality and platform independent descriptions, it retains one important disadvantage. The described data and their presentations are administered in one document, thus modification at runtime is impossible. The model cannot be extended and modalities cannot be added later on.

UsiXML [11] is an XML-based User Interface Description Language (UIDL). It offers the possibility to describe a UI according to the Cameleon Reference Framework. Employing transformations between the four levels it is possible to transform the basic *Task & Concepts* (T&C) model into several adequate *Final UIs* (FUI) matching different platforms and modalities. Finding all the necessary transformations is a tedious process however, as the mapping problem illustrates.

Another interesting method of how to integrate several external services into the automotive HMI is described by Hildisch et al. [3], who propose to describe all possible abstract UI facets in a semantic ontology hierarchy. The HMI acts as an interpreter mapping the OWL-based interface description to given FUI-elements provided by the HMI system itself. This concept allows generation of UIs which are highly consistent but has the disadvantage of not being able to integrate previously unknown concepts at runtime.

3.2 General Approach

In order to build a system which can be extended at runtime we propose an approach that employs an extendable task model. The task model contains extension markers at which submodels of external devices can be attached.

We propose a transformation-based approach similar to the Cameleon Reference Framework to structure the different levels of abstraction. In contrast to the Cameleon approach, we propose to describe the tasks and concepts as well as the abstract user interface in one step using UML without the need of transformations between these levels. This alleviates the mapping problem since no transformations are necessary for the first level of the framework.

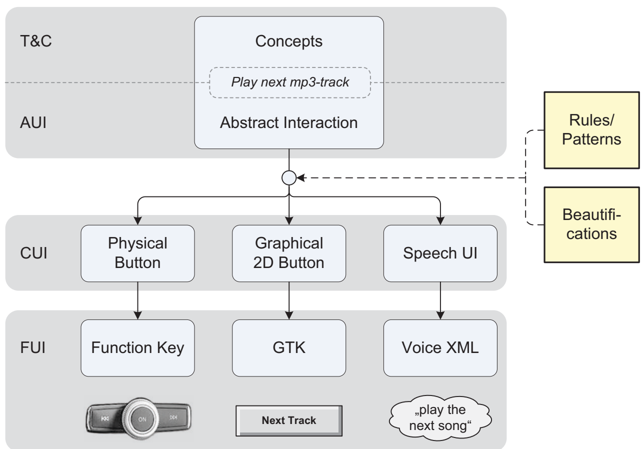


Figure 2: *Play next mp3-track* on different levels of abstraction using the adapted framework.

The adapted framework is shown in figure 2. Again the user wants to carry out the task of playing an mp3-track. The task can be modeled with activity charts and state machines, concepts are modeled as classes. Transformations take place along the arrows, like in the original framework.

3.3 Rule-based Transformations

We aim at flexibility and extendability on the one side and controlled predictable interaction on the other side. The proposed system uses rule-based transformations in order to generate UIs. To this end it employs rules and beautifications to generate consistent UIs of a high standard.

Rules are divided into different categories. They adjust the modality of human-machine interaction to match several situational conditions i. e. provide automated switching between screen-based and voice interfaces. We propose to use patterns for well-known situations and heuristics to decide the modality otherwise. Rules also cover layout and design aspects as well as user customizations. Especially in the layout process patterns can be employed. Additionally, rules can be used to meet special personal requirements (e. g. a larger font size for the elderly).

Another important aspect are beautifications, as described by Pederavia et al. [8]. These are additional rules which are created by a designer who wants to adapt the automatically generated UI of a specific device. The application of these rules is repeated on subsequent UI generations each time the device type is connected to the car. By using beautifications designers can ensure a corporate design or adapt UIs to their preferences.

3.4 Contributions

Our proposed architecture is based on the Cameleon Reference Framework but adapts it for use with UML. This leads to less transformations and thereby alleviates the mapping problem. By employing UML the approach can leverage existing tool support and know-how, thus allowing easier participation in the design process.

The approach allows for extendability of UIs on an abstract level. External devices can be fully integrated into the system at runtime. The added task descriptions are integrated into the UI in a way to make the UI appear to come from a single source. Thus, we achieve a flexible interaction environment to support different capabilities of current and future devices.

4. CONCLUSION AND FUTURE WORK

We presented automotive-specific requirements for a flexible solution to integrate external services into the car. The possibility to completely control the integration of external devices and their user interface into the car was highlighted. Methods for abstract user interface representations were evaluated and our transformation-based approach for building user interfaces from abstract representations was motivated.

Our proposed approach can be employed for all kinds of devices. An already deployed system remains extendable independant of the car's life cycle. This flexible integration of external applications into the automotive interaction environment pushes the development of modern applications for in-car infotainment and their safe use.

The envisioned overall architecture was presented based on well-considered requirements and the evaluation of existing approaches. Further research for a detailed proof of concept is needed and scheduled for the near future.

Acknowledgements

This work is originated in the Transregional Collaborative Research Centre SFB/TRR 62 "Companion-Technology for Cognitive Technical Systems" funded by the German Research Foundation (DFG) in co-operation with Daimler AG.

5. REFERENCES

[1] G. Calvary, J. Coutaz, D. Thevenin, Q. Limbourg, L. Bouillon, and J. Vanderdonckt. A unifying reference framework for multi-target user interfaces. *Interacting with Computers*, 15:289–308, 2003.

[2] A. Dix, J. Finlay, G. D. Abowd, and R. Beale. *Human-computer interaction*. Pearson, 2004.

[3] A. Hildisch, J. Steurer, and R. Stolle. Hmi generation for plug-in services from semantic descriptions. In *ICSE Workshops SEAS '07.*, page 7, May 2007.

[4] Jan Van den Bergh and K. Coninx. Model-based design of context-sensitive interactive applications: a discussion of notations. In *Proc. TAMODIA '04*, pages 43–50, New York, NY, USA, 2004. ACM Press.

[5] L. Nóbrega, N. J. Nunes, and H. Coelho. Mapping concurtasktrees into uml 2.0. In *Proc. DSVIS'05*, pages 237–248, Berlin, 2005. Springer.

[6] Object Management Group. *UML 2.0 Superstructure Specification*, 2005.

[7] F. Paternò, C. Mancini, and S. Meniconi. Concurtasktrees: A diagrammatic notation for specifying task models. In *Proc. INTERACT*, pages 362–369, 1997.

[8] I. Pederiva, J. Vanderdonckt, S. E. na, I. Panach, and O. Pastor. The beautification process in model-driven engineering of user interfaces. In C. Baranauskas, P. Palanque, J. Abascal, and S. D. J. Barbosa, editors, *Proc. INTERACT*, volume 4662/2008 of *LNCS*, pages 411–425, 2008.

[9] C. Phanouriou. *UIML: A Device-Independent User Interface Markup Language*. PhD thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, September 2000.

[10] A. Puerta and J. Eisenstein. Towards a general computational framework for model-based interface development systems. In *Proc. IUI*, pages 171–178, 1999.

[11] J. Vanderdonckt, Q. Limbourg, B. Michotte, L. Bouillon, D. Trevisan, and M. Florins. Usixml: a user interface description language for specifying multimodal user interfaces. In *Proc. Workshop on Multimodal Interaction WMI'2004*, pages 1–7, 2004.

[12] D. Weld, C. Anderson, P. Domingos, O. Etzioni, T. Lau, K. Gajos, and S. Wolfman. Automatically personalizing user interfaces. In *Proc. IJCAI-03*, page 7, 2003.

Developing a Low-Cost Driving Simulator for the Evaluation of In-Vehicle Technologies

Garrett Weinberg and Bret Harsham
Mitsubishi Electric Research Labs
201 Broadway,
Cambridge, MA 02139, USA
+1 617 621 {7547, 7526}
{weinberg, harsham}@merl.com

ABSTRACT
We present a case study concerning the development of a driving simulator at Mitsubishi Electric Research Labs. By relying largely on off-the-shelf components, we have kept the total system cost under USD 60,000, yet attained a level of fidelity comparable with more expensive, custom-built research simulators.

Categories and Subject Descriptors
H.5.2 [INFORMATION INTERFACES AND PRESENTATION] User Interfaces – Benchmarking, Evaluation/methodology, Prototyping

General Terms
Design, Experimentation, Human Factors, Economics

Keywords
Driving simulation, automotive user interfaces, human-machine interfaces.

1. INTRODUCTION
In theory, it is preferable to conduct automotive human-computer interaction research in moving vehicles on real roads or test tracks, as is often done in transportation engineering studies. Practically speaking, however, HCI studies in real vehicles are rare. This may be due to the safety and liability issues inherent in testing unproven technology not specifically related to core vehicle operation. But beyond safety advantages, driving simulators offer HCI researchers distinct advantages over real vehicles in terms of repeatability. By keeping the simulation scenario exactly the same from trial to trial or subject to subject, one can highlight the differences between in-car devices or interfaces with fewer complications and confounds.

We believe it is for this latter reason that driving simulators have emerged in the past several years as vital tools for the evaluation of new in-vehicle technologies. Whereas in the past automotive OEMs and aftermarket device manufacturers might have considered their interfaces' visual and psychomotor demand at design time and then brought products to the market with "fingers crossed," today there is more emphasis on empirically verifying this demand in simulated driving situations [2],[9],[17].

Exactly what a "simulated driving situation" entails, however, varies widely from institution to institution and study to study. At the low-fidelity, low-cost end of the spectrum are studies that involve counting the number of vehicle crashes in a video game session [13] or having subjects carry out abstract steering-like tasks such as tracking a shape's horizontal movement using a wheel [6]. At the high-fidelity, high-cost end of the spectrum are the multi-million-dollar, full-motion platforms that occupy entire hangar-sized buildings [11]. Somewhere in the middle are hundred-thousand-dollar research simulators (e.g. [14]) that offer unparalleled flexibility in terms of scenario creation and playback. However they require an enormous investment of time for object modeling and scripting, and their cost generally does not include equipment (computers, displays, and driving chairs/vehicle cabs).

This paper discusses the construction of a simulator with a degree of realism and flexibility similar to that of mid-level research simulators, but at a far lower cost. It is not the aim of the present work to compare our simulator with other setups on a point-by-point basis. Rather we offer a practical case study in hopes that our techniques and experiences can be valuable as other institutions weigh their options.

In the following sections, the simulator's hardware and software components will be discussed, some supporting tools will be mentioned, and then we will briefly discuss the current limitations of the setup and our plans for addressing these limitations in the future.

2. SIMULATOR HARDWARE
2.1 Computer
A single high-end desktop PC is the basis for our simulator. The CPU is a 3.0 GHz Intel Core 2 Extreme, with 4.0 GB of 2000MHz DDR3 RAM. Two NVidia GeForce 8800 Ultra graphics cards are used for video output, either in standard or parallel-processing (SLI) mode depending on display configuration (see below). We chose Windows XP as the operating system because of driver support and its compatibility with a wide array of gaming and simulation software. The total cost of all computer components was under $2500.

AutomotiveUI'09, September 21-22, 2009, Essen, Germany
ACM 978-1-60558-571-0/09/009

2.2 Input/Output devices

The most important input/output device is a D-Box GP Pro-200 RC gaming chair [5]. This cockpit-style chair (see Figure 1) rests on three hydraulic actuators that move in response to events in the driving simulation. These movements consist of vibration and tilt with two degrees of freedom. The vibration is synchronized with simulated engine RPM and greatly improves the perception of the virtual vehicle's speed. The tilt corresponds to in-game acceleration, braking, and steering/cornering. We find that the vestibular stimulation offered by this tilt feature helps to counteract the "simulator sickness" effect that is the bane of fixed-base, motionless simulators.

A Logitech G25 force-feedback wheel bolted to the D-Box chair affords primary steering input. This is one of the largest and most solidly built game controllers on the market, and comes with a weighted throttle, brake, and clutch pedal assembly as well as a shifter knob. Engine noise and sounds/music generated by in-vehicle interfaces are played through a Creative Inspire 5.1 speaker system. The D-Box chair includes the Logitech G25 and the speaker system, and retailed for $15,000 in 2008.

2.3 Displays

We have experimented with two different display configurations. The first was a Samsung SyncMaster 305T LCD measuring 76 cm diagonally and offering 2560 x 1600 native resolution (SLI-mode video was necessary for smooth rendering at this resolution). This display was placed on a shelf approximately 147 cm off the floor (as shown in Figure 1). This configuration offered a horizontal viewing angle of 42.7° and a vertical viewing angle of 27.7° in the worst case (the adjustable seat slid as far back as it will go, resulting in a viewing distance of 81.8 cm). At this screen distance and position, the most natural in-game camera perspective superimposes some of the vehicle interior (dashboard and forward left pillar) over top of the roads and terrain. We purchased the Samsung display for $1245.

Figure 1: First configuration

We were quite satisfied with the level of textural detail and realism afforded by this high-resolution display configuration (about 60 pixels per horizontal degree). However, we wanted to experiment with larger, potentially more immersive displays. To this end we re-purposed three DLP-based Mitsubishi MegaView displays [10] that had been used for a previous project in the lab and were sitting idle. Each display measures 127 cm diagonally and supports 1024 x 768 resolution. We arranged them in a coplanar 3x1 layout and combined their inputs using a Matrox TripleHead2Go device. This allows them to appear to the Windows display driver as one large, combined 3072 x 768 display rather than three individual displays. In order to bring the subject's eye level in line with the vertical center of the displays (approx. 127 cm off the floor), we placed the D-Box chair on a sturdy wooden platform rather than building expensive custom mounts for the displays. At a viewing distance of 186 cm, again in the worst case, the horizontal viewing angle is 78.6° and the vertical angle is 23.1°. Despite the lower resolution in this case (about 39 pixels per horizontal degree), the driving experience is qualitatively more immersive and realistic in this configuration because of the larger screen size. As shown in Figure 2, the most natural in-game camera perspective for this physical layout is the "hood view."

Figure 2: Second configuration

While it could be argued that using $20,000 commercial-grade displays such as the Mitsubishi MegaViews invalidates the positioning of our simulator as a low-cost alternative, it should be pointed out that a very similar setup could be achieved using consumer-grade equipment. DLP or LCD projectors at 1024 x 768 resolution can be had for under $1000 apiece.

3. SIMULATOR SOFTWARE

After evaluating several open-source and commercial alternatives, the commercial driving game rFactor [8] was chosen as the software platform for our driving simulator. It offers a convincing, realistic driving experience thanks to richly detailed graphics, accurate vehicle physics, and full support of force-feedback steering wheels. And while it does not offer the complete flexibility of an open-source product, the game does allow for a deep degree of modification and customization. There is a large community of enthusiasts who produce everything from custom tracks to custom vehicles and camera angles. The game's "out of the box" support for the D-Box chair is also a distinct advantage. In addition, rFactor provides a plug-in API whereby vehicle telemetry (including position, velocity, and acceleration), and user input (steering angle and throttle/brake positions) can be captured at rates up to 90 Hz.

Our own rFactor plug-in simply dumps comma-delimited raw data to a file for later processing. This processing allows us to report higher-level results using standard metrics from the driving simulation and human factors literature [12][18]. These include, for example, lane position variance, speed variance, and following distance variance.

For one study we used a mixed city/highway course that ships with the paid version of rFactor itself ($40), for another we used a third-party highway-based course that we found on the fan site “rFactor Central” [15], and for a third study we built an entirely custom course from scratch using a basic 3D modeling tool called Bob’s Track Builder [3].

4. SUPPORTING TOOLS

4.1 Eye tracker

There is wide consensus that the measurement of eye glances and fixations is crucial to determining how distracting any given in-vehicle interface is [4], [7]. Distracted drivers tend to reduce their tactical and strategic scanning behavior, narrowing their focus to the area immediately in front of their vehicle and missing peripheral stimuli [1], [12], [18].

For this reason we consider it essential to measure glances and fixations, and to report excessive (e.g. greater than two second) glances away from the forward roadway in our study results. An extremely powerful tool for making these sorts of measurements is Seeing Machines’ FaceLAB system [16]. This system incorporates a dedicated laptop and two Firewire cameras that are placed at either end of a stationary mount, allowing them to triangulate the position of the subject’s head. Infrared light is emitted from a pod at the center of this mount, and the cameras track the glint produced as this light bounces off the corneal surface of each eye. This allows the FaceLAB system to generate both head position and eye gaze vectors.

For each study setup, one creates a model of the primary screen, noting any coordinates of interest (e.g. of the virtual roadway surface or a lead vehicle), as well as of any objects of interest in the real world outside the screen, such as a navigation system display or steering wheel-mounted buttons. The bundled software can thereby create a report showing exactly which screen coordinates or real-world objects a user fixated upon, and for how long.

Not counting the re-purposed MegaView DLP displays, the FaceLAB system was the single most expensive component in our simulator. It cost approximately $40,000, with options, when we purchased it in 2008. Based on our experience so far, it was money well spent.

4.2 Experimental tools

We use a suite of in-house software tools to automatically generate and time the in-vehicle interface tasks that subjects must carry out. These tasks may include, for example, destination entry or music retrieval. A simple USB-based device (Figure 3) displays information to the experimenter so that he or she may prompt the subject to carry out one of these tasks. The experimenter then presses the device’s buttons to mark the beginning and end of the task, and to annotate it in various ways within the task log.

Figure 3: Experimenter’s tool

Another tool merges and synthesizes the various logs – rFactor, FaceLAB, and the task log – creating time series that can be queried during the analysis phase by means of simple SQL statements.

5. ADVANTAGES and LIMITATIONS

The major advantages of our approach versus traditional research simulators are cost and time. Typical simulation software, which starts in the $100,000 range, does not usually include input/output hardware or eye trackers. We built a comparable system with arguably superior motion feedback and rendering quality for under $60,000, *including* the eye tracker.

Table 1. Approximate cost breakdown, as of 2008

Component	Cost (USD)
Computer	2500
Primary display	1245
Driving chair, steering wheel, speakers	15,000
Eye tracker, with options	40,000
Simulation software and modeling tools	100
Total:	**$58,845**

Our choice of rFactor as the simulation engine also meant significant time savings. Rather than painstakingly modeling vehicles and roadways and painstakingly scripting scenarios, we let the worldwide community of rFactor enthusiasts do most of the work for us. If we cannot find a custom course design that suits our needs, we can build one within several hours using Bob’s Track Builder rather than taking the many days necessary to learn and use a full-scale modeling suite such as 3D Studio Max.

The reliance on off-the-shelf components is not without significant disadvantages, however. rFactor is primarily a racing simulation game. Thus, it is difficult to model the complex street layouts and intersections found in urban areas. The game engine furthermore requires that there be a single, designated “best path” around the course. It is unclear, based on our initial investigations, whether this path may branch or double back on itself, as would be required, for example, to enable the simulation of opposing traffic flow.

Our degree of control over other vehicles on the roadway is currently very poor as well. The game’s developers offer very little programmatic control over the computer-controlled “AI” drivers; one can merely tweak relatively opaque “strength” and “aggressiveness” settings in the configuration files. Combined with vehicle handicapping, this has allowed us to slow the AI driver enough so that it may act as a pace car for studies that require such a design. However, we currently have no way of causing AI drivers to perform specific maneuvers at specific times.

6. FUTURE PLANS

In situations where a study’s protocol calls for the subject to react to specific situations at specific times during a scenario, we may populate the simulation with one or more human “Wizard of Oz” drivers who are aware of the study protocol and receive specific instructions or signals as to when and where to carry out specific maneuvers – for example, sudden swerving or braking. As it is by design a multiplayer game, rFactor would support this approach well.

We plan to further enhance the immersion and realism of the driving experience by angling the two side displays toward the subject, such that the subject’s gaze vector remains orthogonal to the surface of the display no matter which display she fixates upon. This will reduce the distortion evident at the periphery of the rendered driving scene, as well as increasing the effective field of view.

Finally, we plan to evaluate our driving simulator against typical research simulators in order to determine the validity of HCI studies performed in it.

7. REFERENCES

[1] Atchley, Paul and Jeff Dressel, “Conversation Limits the Functional Field of View,” *Journal of the Human Factors and Ergonomics Society*, vol. 46, p. 664-673, 2004

[2] Baron, A. and Green, P. “Safety and Usability of Speech Interfaces for In-Vehicle Tasks while Driving: A Brief Literature Review,” (Technical Report UMTRI 2006-5). University of Michigan Transportation Research Institute, Ann Arbor, MI, 2006.

[3] Bob’s Track Builder. http://www.bobstrackbuilder.net/. Retrieved June 1, 2009.

[4] Chisholm, Susan L., Jeff K. Caird, Julie Lockhart, Lisa Fern, and Elise Teteris, “Driving Performance while Engaged in MP-3 Player Interaction: Effects of Practice and Task Difficulty on PRT and Eye Movements,” *Driving Assessment*, 2007.

[5] D-BOX Technologies, Inc. http://www.d-box.com/gaming/en/gaming/pro-gaming-series/gp-pro-200-rc/. Retrieved June 1, 2009.

[6] Forlines, C., Schmidt-Nielsen, B., Raj, B., Wittenburg, K., and Wolf, P. A Comparison between Spoken Queries and Menu-based Interfaces for In-Car Digital Music Selection. In *Proceedings of International Conference on Human-Computer Interaction (INTERACT ‘05)* (Rome, Italy, September 12-16, 2005). IFIP, Amsterdam, 2005, 536-549.

[7] Horrey, W.J, Wickens, C.D., Consalus, K.P. “Modeling Drivers’ Visual Attention Allocation While Interacting With In-Vehicle Technologies”. *Journal of Experimental Psychology: Applied.* 12, 2 (2006), p. 67-78.

[8] Image Space Incorporated. rFactor. http://www.rfactor.net. Retrieved June 1, 2009.

[9] Lai, J., Cheng, K., Green, P., and Tsimhoni, O. 2001. On the road and on the Web?: comprehension of synthetic and human speech while driving. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Seattle, Washington, United States). CHI '01. ACM, New York, NY, 206-212. DOI= http://doi.acm.org/10.1145/365024.365100

[10] Mitsubishi Electric Corp. MegaView Wall. http://www.mitsubishi-megaview.com/products/modelVS-50XL50U.html. Retrieved June 1, 2009.

[11] *Motor Trend*. “Toyota’s Driving Simulator: Not All Fun and Games.” http://www.motortrend.com/auto_news/112_news260711_toyotas_driving_simulator/index.html. Retrieved 1 June 1, 2009.

[12] Muttart, J.W. et al., 2007. Driving Without a Clue: Evaluation of Driver Simulator Performance During Hands-Free Cell Phone Operation in a Work Zone. *Journal of the Transportation Research Board,* Vol. 2018, pp. 9-24

[13] Nuance, Inc. “Amazing Race: Distracted Driving” http://www.youtube.com/watch?v=3v7hpRBCW5c. Retrieved June 1, 2009.

[14] Realtime Technologies, Inc. On-Road Simulator. http://www.simcreator.com/simulators/onroadsim.htm. Retrieved June 1, 2009.

[15] RFactor Central. http://www.rfactorcentral.com/. Retrieved June 1, 2009.

[16] Seeing Machines, Inc. FaceLAB. http://www.seeingmachines.com/product/facelab/. Retrieved June 1, 2009.

[17] Society of Automotive Engineers, 2004. SAE Recommended Practice Navigation and Route Guidance Function Accessibility While Driving (SAE 2364). Society of Automotive Engineers, Warrendale, Pennsylvania, U.S.A.

[18] Strayer, D.L. et al., 2001. Cell Phone-induced perceptual impairments during simulated driving. In *Proceedings of the International Driving Symposium on the Human Factors in Driver Assessment, Training, and Vehicle Design*. Aspen, Colorado, U.S.A.

NAVIGATION AND FURTHER APPLICATIONS

Open Vehicular Data Interfaces for In-Car Context Inference

Matthias Kranz and Eduard Weber
Technische Universität München
Institute for Media Technology
Arcisstrasse 21, 80290 Munich, Germany
matthias.kranz@tum.de

Korbinian Frank and Daniel Hermosilla Galceran
German Aerospace Center (DLR)
Institute of Communications and Navigation
Oberpfaffenhofen, 82234 Wessling, Germany
korbinian.frank@dlr.de

ABSTRACT

In this paper, we present a concept for an open vehicular data interface and describe it's components and architecture. We discuss the enabled applications in the context of advanced driver assistance systems with a focus on human-machine interfaces, vehicle-to-x (V2X) communication and context inference systems. We conclude by a presentation of the initial implementation and deployed system.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; H.3 [**Information Storage and Retrieval**]: Systems and Software

General Terms

Algorithms, Measurement

Keywords

vehicular interfaces, context awareness, context inference

1. INTRODUCTION

Modern vehicle comprise hundreds of sensors, various communication busses and significant processing power, comparable to a modern personal computer. These systems have immense untapped potential for advanced driver assistance systems (ADAS) and context-aware systems and many more innovative automotive applications.

Unfortunately, access to the in-vehicle sensors requires the knowledge of restricted information, such as the CAN matrix. This information is only available to vehicle manufactures and not the the general research community. We therefore are investigating how interested researchers could interface their research vehicles without the need of a full CAN access. This, we think, significantly leverages research in the area of automotive user interfaces and V2X communication.

AutomotiveUI'09, September 21-22, 2009, Essen, Germany
ACM 978-1-60558-571-0/09/009

We present three ways to access in-vehicle sensor data: first, using the standardized vehicle diagnostic bus, providing only limited data which can be used e.g. as ground truth (e.g. for the vehicle speed), second, using our developed general purpose data interface and third, using vision-based OCR of the vehicle's debug system. The latter is part of nearly all vehicles and is intended to provide an easier-to-use access than the vehicle manufacturer's full diagnostic tool set.

We integrate data from all three interfaces to develop context-sensitive driver user interfaces. So far, nearly all user interface components in a vehicle are static: the tachograph (even if displayed on a pixel-based screen instead of being an analog meter), the gear information, and many more. The set of information is never adapted, e.g. in case of potentially hazardous situations, such as driving at high speeds and using the high beam lights at the same time which could indicate another driver changing lanes and obstructing the way for the ego vehicle. Neither is the sound volume of the stereo adapted, nor the telephone muted and incoming calls silently blocked. Kern et al. [8] showed the benefits of integrating context information in the driver assistance systems for providing better user interfaces. We extend this concept to different types of sensors and a broader application range.

2. ACCESS TO RAW DATA

In this section, we introduce three ways of open data access in vehicles that can be used to elicit sensor information.

2.1 OBDII Interface

OBDII and EOBD [13] are standardized interfaces for vehicle diagnostic. Tab. 1 summarizes selected sensors and their meaning w.r.t. to vehicle diagnostics. As we will see later, more meaning and context information can be elicited from this data. The inferable meaning is additionally enlarged by incorporating additional information, as described in the following two sections.

2.2 General Purpose Data Interface

Even though many functions of modern vehicles are managed by microcontrollers, in the end, there is either a a sensor or an actuator controlled, such as the light or the front wiper. This means at some point, there is current flowing and voltage present. We use this fact to create a general purpose data interface to elict information from the vehicle, without the need of the CAN matrix which is usually confidential not accessbile for researchers in general. Fig. 1 shows an overview on the the developed system.

Information	Initial Meaning
Engine Load	Computed from Air Flow Rate into the engine and Intake Manifold air pressure
Engine Speed	Reported by the Crankshaft Position
Sensor Coolant Temperature	Reported by the engine coolant sensor, a thermistor that varies its resistance according to the engine coolant temperature
Throttle Position	Throttle position sensor creates a voltage signal that varies in proportion to the throttle valve opening angle
Intake Air Temperature	Measured by another thermistor located in the Mass Air Flow Sensor unit
Battery Voltage	affects the speed at which the fuel injectors open and must be taken intoaccount in computing the fuel injectorpulse length, or injector open time
Oxygen Sensor	The oxygen density in the exhaust emissions is detected and generates a control signal back to the ECU indicating the burned air to fuel ratio.

Table 1: The On-Board Diagnostics interface, mainly targeted at maintenance, provides information about several sensors that can be used to assess the vehicle's context. The table give a list of the most useful information for context inference `http://www.4x4wire.com/toyota/4Runner/tech/OBDII_ECU/`.

The data is currently used by three systems in the vehicle: first, a simple HMI component, shown in Fig. 6 that visualizes the state of one interface to the driver. This will in future be an fully graphic interface, replacing all other user interface information components, such as the tachograph.

While local sensor information alone is very important, esp. for context inference, we also explore options of integrating this data directly in the ongoing research efforts on vehicle-to-vehicle and vehicle-to-infrastructure communication, subsumed as vehicle-to-x (V2X) communication. The goal of V2X communication is, by cooperative communication, to increase safety, traffic efficiency, and provide novel services [1].

Again, researchers are in need for the in-vehicle sensor data and are relying on the cooperation with a vehicle manufacturer. This might limit options for exploring ideas. By providing a general approach to vehicular sensor data, we think that here, too, research is fostered by our approach.

The output connector for V2X communication enables the collaborative sharing of local data, allowing the near-by vehicles to get a notion of the context of the other surrounding vehicles. Our system is the local correspondent of the distributed collaborative CODAR (Cooperative Object Detection and Ranging) architecture [9, 10].

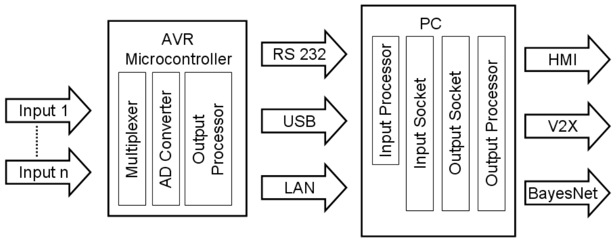


Figure 1: General Purpose Vehicle Data Acquisition Architecture. A set of self-describing components acquire data from analog signal wires and convert the input data and provide pre-processed data over three optional outputs. In case of serial communication, the data is read and re-sent over a socket. The self-description of the node is used for correctly producing the outputs for the data consumers: a driver HMI, a 802.11p-based V2X communication unit and a Bayesian network for context inference.

2.3 Visual Diagnostic Screen Recognition

Our PriCARVe research vehicle enables us, due to the availability of the CAN matrix [14] to access the in-vehicular sensor systems directly. With an additional video-in-motion modification using a VAIS tech CAN module faking a present DVD player on the AVC-LAN bus, we can use the built-in factory touch screen for our adaptive user interfaces. This enables, for example, to reproduce the results of Coroama et al. [3] without the need of any additional external sensing equipment.

Figure 2: Debug Information System: Factory debug system of the 2006 Toyota Prius in our PriCARVE, the Prius Context-Aware Research Vehicle. In addition to the general purpose interface, sophisticated CAN BUS access complements the input data. The debug screen shows speed, the status of the car (ignition, powered), the driving direction (rev is currently off), the status of the vehicle's light and the parking brake.

In the context of this research, we use the availability of

Figure 3: Debug Information System: The debug information system of the Mercedes G400 delivers battery and audio status, the coordinates as measured by the in-vehicle navigation unit, the driving direction as set by the automatic gear and the current speed as measured by the odometry.

the real data as ground truth and can thus directly compare the information to the one computed by the combination of OBDII, general purpose data interface and the visual diagnostic screen.

Visual diagnostic screens, ranging from LED segment displays to fully graphic displays are standard in modern vehicles. They enable the garage to quickly check the main functions of the vehicle, without the need of fully wiring the manufacturer's diagnostic computers to the vehicle. Fig 2 and Fig 3 show two examples of secretly built-in diagnostic screens. The first shows an example of the Toyota Prius' screen, the second an example of a Mercedes G400.

The contents of the screen are only dynamic w.r.t. the data values, the position of the information is static. Using a fixed digital video camera, such as an webcam, we can grab images of the screen. As the a-priori knowledge of the information locations are known, optical character recognition software can extract the information from the images several times per second. We are planning to explore the options of video analysis, though feel more than 5 updates per second are not necessary. This is also the envisoned update rate of EU V2X systems.

3. FROM RAW DATA TO SAFETY RELEVANT CONTEXT

Active Safety Application are any application that try to prevent accidents and therefore have to intervene at the first indication of a potential accident situation (in the remainder also called hazardous situation). To detect these hazardous situations applications have to collect the available *context* information. To be precise, when we use the term "context", we follow the definition given by Dey in [4]:

> Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves.

In particular *high-level context* information like the danger of a situation is of importance for context aware applications in cars. The process of generating this information from the available data described in the last section is called *context inference*. The high-level information we are interested in given the available data is the profile of the track where the car is currently driving (e.g. tunnels, hills, intersections, high-ways), the danger of a situation or the car's status or actions like parking, lane changing or normal driving. Among others we can seen the following dependencies:

- If the blinking lights are on, we receive a GPS signal and are driving at relatively low speed, this might indicate that we are approaching an intersection.
- If on the other side lights are on and we do not receive any GPS signal, we are most probably in a tunnel.
- If one is driving backwards at relatively low speed, we can assume that the car is parking.
- If the engine load is different than usual at the same speed or your current speed differs from your normal speed profile you are probably driving up- or downhill.
- Blinking lights at constant speed indicate a lane change maneuver on a motorway.
- With a high wiper level and lights switched on, one can assume an environmental condition with low sight.

From those dependencies we can model inference rules that are evaluated in real-time while the car is driving. Active Safety Applications can access these data from a predefined interface and enhance the overall traffic safety.

4. CONTEXT INFERENCE WITH BAYESIAN NETWORKS

There are different ways to perform inference, among others logical reasoning with all different types of logics (like Propositional Logics, Description Logics, First and Second Order Logics and many more), context history based approaches and probabilistic algorithms. Probabilistic inference e.g. with *Bayesian Networks (BN)* resulted to provide the best trade off between expressivity, ease of modeling and inference performance [5].

Over the last 15 years, BNs [12] have evolved as a major tool in a wide area of scientific disciplines requiring sound statistical analysis, automated reasoning or exploitation of knowledge hidden in noisy data. These range from fields in medical research, genetics, insurance analysis, and fault handling to automation and intelligent user interaction systems. BNs combine techniques from graphical models with those from Bayesian analysis to provide a formal framework where complex systems can be represented and analyzed.

A BN encompasses a set of *random variables (RV)* that represent the domain of interest and the BN encodes many of the important relationships between these variables, such as causality and conditional dependence and conditional independence. Specifically, their structure bears information about the qualitative nature of these relationships whereas

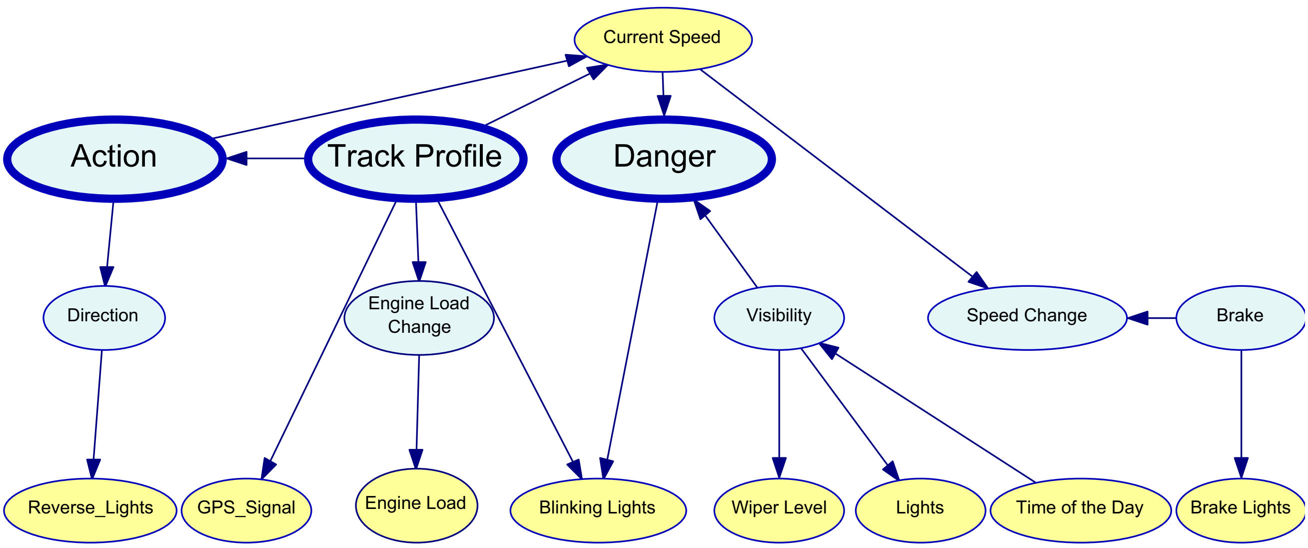


Figure 4: A complete Bayesian Network linking the information that is available through the sensors. Nodes represent Random Variables that model the raw data sources (top and bottom rows) and inferred concepts (in the middle). The target random variables, that represent the value added information for drivers are represented in bigger font size and thicker frame. The directed edges represent causal influence in the direction of the arrow. The "Direction" you are driving to influences for instance the status of your "Reverse Lights".

their network parameters encode the quantitative probabilistic relationships among the variables of interest. Figure 4 shows a BN, that models the status of a car based on the available information described in section 2. It may seem simplistic, but represents a fair trade-off between complexity and quality of the results. This trade-off can even be learnt by automatic processes like in [11] that create structure and transition probabilities of BNs from a given data set. As any inference rule or ontology, it represents a relevant part of the reality, abstracting from the general complexity.

For context inference now, we can represent any inference rule by a BN and evaluate them by calculating the conditional probabilities. RVs represent context attributes of a specified user. Context inference takes into account sensed values for context attributes in the BN as evidence and computes the conditional probability of the target context attribute. The most probable value of this context attribute will be returned together with its probability as confidence level. This computation of the conditional probabilities can for example be done in a message passing algorithm [7], that first transforms the graph into a tree structure of cliques (=combinations of random variables) and then, if evidence is added, passes the new probabilities as messages through the whole tree, so the evidence takes effect in every related node. Among the proposed evaluation algorithms for Bayesian networks, this one offers exact inference, and a well-described implementation that is more efficient than a straight-forward evaluation of conditional probabilities.

It can be shown however that the general problem of inference in a BN is NP-hard in the number of nodes [2], that's why we developed the concept of *Bayeslets* [6]. In these, the concepts of "divide and conquer" as well as object orientation are applied to BNs. Inference is only applied to sub-networks that are thematically closely linked. Only on demand for a higher inference goal, several Bayeslets can be joint. This offers faster inference as only necessary nodes are evaluated taking into account only available sensors. With the predefined interfaces contents of Bayeslets can be shadowed as long as the outcome is applicable to Bayesian inference. Bayeslets furthermore ease personalization and dynamic incorporation of other users' context, which is particularly desirable in large scale highly dynamic environments like in road traffic.

In Figure 5 we show the evaluation of a Bayeslet, i.e. of a part of the BN from Figure 4. The result shows that based on the causal influences defined in section 3 we can infer the current track profile taking into account possible ambiguities, fault rates of sensors and general uncertainty.

5. TEST ENVIRONMENT

Our test environment comprises two equipped vehicles, a Mercedes model G400 and a Toyota Prius model 2006. They are equipped with an automotive computer from DSM that automatically powers up when the car is turned to ignition. It has a connection to the three above named interfaces, OBDII, general purpose data interface and debug screen.

We are using the information data sources and information summarized in Tab. 2 in the current state of our system.

The automotive computer acquires the data from the different interfaces and processes, when necessary, the corresponding output, e.g. for the debug screen.

Every input that can be read then digitally available and provided over sockets, allowing flexible information exchange between the producers and consumers of the data, using a client/server model.

Fig. 6 shows a first visualization of the vehicle's state,

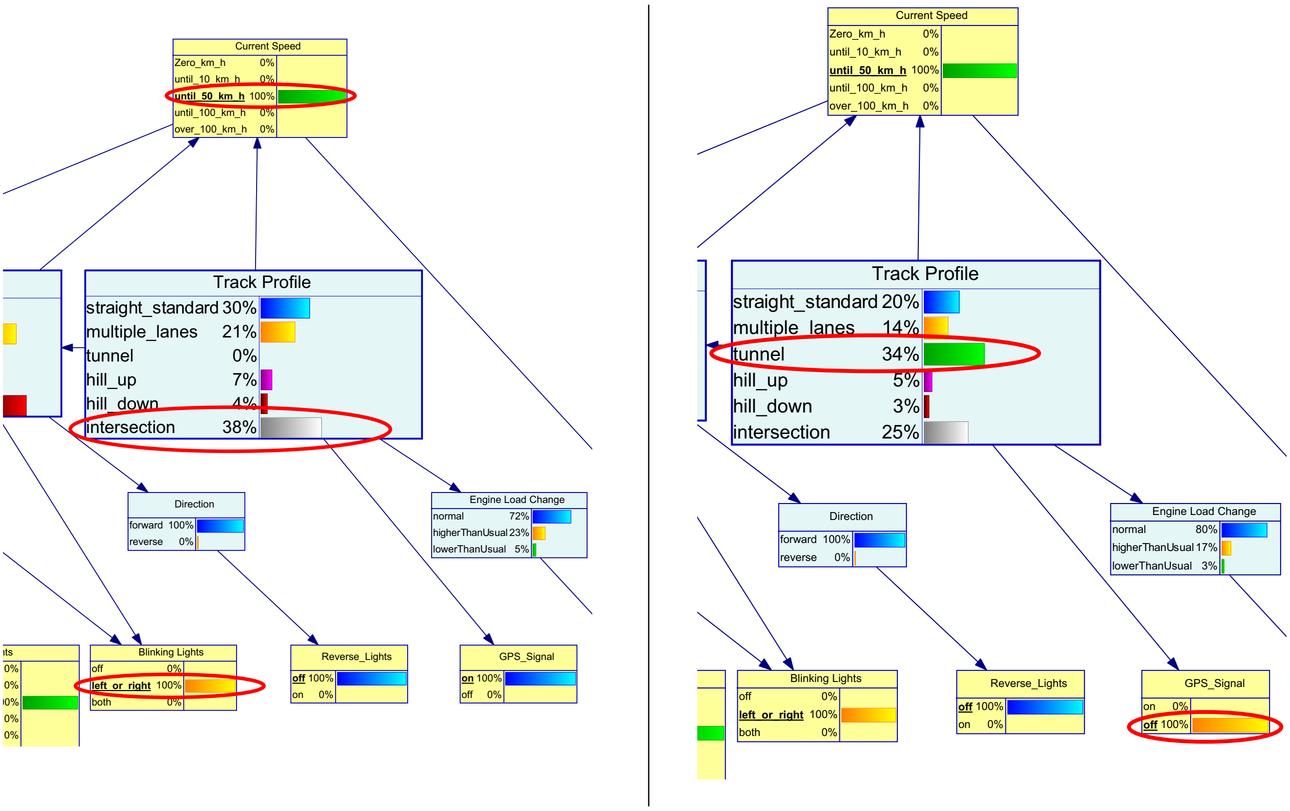


Figure 5: Based on the inference network of Figure 4, the measured information is taken into account and the current "Track Profile" is evaluated. This representation of BNs shows the possible values in the ranges of a RV associated with their probabilities. The measured data have been introduced into the sensor nodes as evidence and propagated throughout the network to calculate the probabilities of "Track Profile". In the left sub-figure you can see that with the blinking lights set to either left or right and relatively low speed, the most probable current track profile is a street "intersection". If the system knows in addition that there is no GPS signal receivable at the moment like in the right sub-figure, the highest probability switches to "tunnel" for the current Track Profile.

Source	Data
OBDII	speed, engine load, fuel tank level
General Purpose Data Interface	front and back wiper speeds, fog, front and rear lights, brake status
Visual Diagnostic Screen Recognition	GPS coordinates, driving direction, phone status

Table 2: Data source and information acquired in our system. Data from all three sources is combined and evaluated by a Bayesian network and for adaptive human-machine interfaces.

based on the developed general purpose data acquisition system. The goal of the data acquisition is to infer the vehicle's and the driver's context and adapt the user interface to the situation. As example, in case of driving at high speeds and when using the high beam lights, the audio could be muted as a potentially hazardous situation could arise and the driver's attention should not be distracted by any radio or CD audio signals. Using context information from *the driver and the vehicle*, a later user interface could adapt to the current situation. For instance, when driving at high speeds, the amount of UI elements could be reduced, and important elements such as the tachograph could be enlarged, the amount and type of feedback could be changed, e.g. from visual to audio or tactile output. This closes the loop to improving traffic safety.

As with the proposed system no legal or insurance aspects are concerned, e.g. with interference with safety-critical systems, such as vehicle stability control (VSC), real world test can be conducted in real traffic contexts on public roads.

6. CONCLUSION AND OUTLOOK

We proposed and implemented an open-access vehicular data interface for in-car context inference and adaptive automotive human-machine interfaces. We used one existing standardized interface and proposed and implemented two additional, simple and transferable interfaces. Thereby, researchers are enabled to modify vehicles into research objects, without the burden of acquiring a CAN matrix and thus leveraging research in this field.

Based on the proposed architecture, we presented an initial user interface as basis for context-aware user interfaces

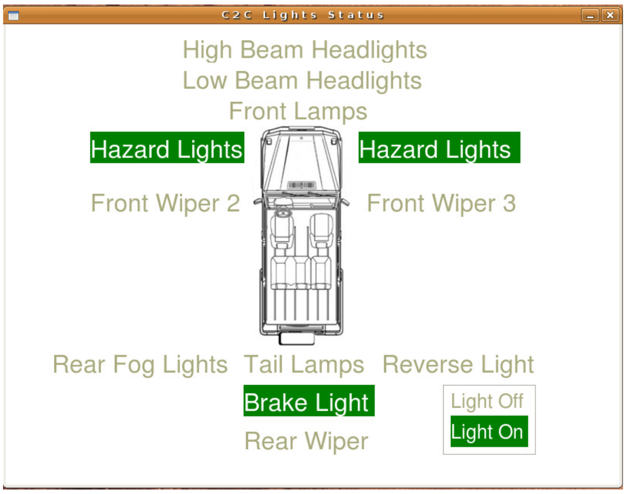


Figure 6: HMI Visualization of selected vehicle signals. This initial bird's eye visualization of the Mercedes G400 model currently visualizes the state of selected signal wires. The information from this and further general purpose data units can be combined and form the basis of later context-aware adaptive user (driver) interfaces.

and developed a Bayesian network for determining the vehicle's and thus driver's context. We thereby verified the validity and potential of our approach.

The connections underlying the Bayesian network have been developed based on our experiences as vehicle drivers. We in the next step will collect real world data sets and analyze the performance of the network with respect to the accuracy of the connections and the reliability of the context predictions.

For our future research, we will explicitly focus on non-GNSS (global navigation satellite system) based driving. We want to elaborate on different questions, e.g. w.r.t. map building: does driving slowly, blinking and speed allow us to correctly infer an intersection? Using the history information of the data sets, can we infer deviations from normal driving behavior, such as making a short stop at the mall when driving home?

Using GNSS information, we will explore if we can improve digital map information with e.g. road profile information, update missing intersections and tunnels and in general improve the quality of community-based free geo-information systems by applying our research.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Community's Seventh Framework Programme [FP7/2007-2013] under grant agreement no. 215098 of the *Persist* (PERsonal Self-Improving SmarT spaces) Collaborative Project.

REFERENCES

[1] Car-to-Car Communication Consortium. Manifesto. Technical report `http://www.car-to-car.org/fileadmin/downloads/C2C-CC_manifesto_v1.1.pdf`, Car-to-Car Communication Consortium, 2007.

[2] G. F. Cooper. Probabilistic inference using belief networks is NP-hard. Technical Report KSL-87-27, Medical Computer Science Group, Knowledge Systems Laboratory, Stanford University, Stanford, CA, May 1990.

[3] V. Coroama. The smart tachograph - individual accounting of traffic costs and its implications. In K. P. Fishkin, B. Schiele, P. Nixon, and A. J. Quigley, editors, *Pervasive*, volume 3968 of *Lecture Notes in Computer Science*, pages 135–152. Springer, 2006.

[4] A. K. Dey. Understanding and using context. *Personal and Ubiquitous Computing*, 5:4–7, 2001.

[5] K. Frank, M. Röckl, P. Gallego Hermann, and M. T. Morillas Vera. Knowledge representation and inference in context-aware computing environments. In J. Lloret Mauri et al., editor, *The Second International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM08)*, pages 89 – 95. IEEE Computer Society Conference Publishing Services (CPS), 09 2008.

[6] K. Frank, M. Röckl, and P. Robertson. The bayeslet concept for modular context inference. In J. Lloret Mauri et al., editor, *The Second International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM08)*, pages 96 – 101. IEEE Computer Society Conference Publishing Services (CPS), 05 2008.

[7] C. Huang and A. Darwiche. Inference in belief networks: A procedural guide. *International Journal of Approximate Reasoning*, 15(3):225–263, 1996.

[8] D. Kern, A. Schmidt, M. Pitz, and K. Bengler. Status- und kontextinformationen für die telekommunikation im auto. In T. Gross, editor, *Mensch & Computer*, pages 119–128. Oldenbourg Verlag, 2007.

[9] M. Kranz, A. Franz, M. Röckl, A. Lehner, and T. Strang. CODAR viewer - a situation-aware driver assistance system. In *Pervasive 2008 Demonstration*, volume 236 of *Advances in Pervasive Computing*, pages 126 – 129. Austrian Computer Society, May 2008.

[10] M. Kranz, M. Röckl, A. Franz, and T. Strang. CODAR viewer - a V2V communication awareness display. In *Pervasive 2008 Late Breaking Results*, volume 236 of *Advances in Pervasive Computing*, pages 79 – 82. Austrian Computer Society, May 2008.

[11] K. P. Murphy. *Dynamic bayesian networks : representation, inference and learning.* PhD thesis, 2002.

[12] J. Pearl. *Causality: Models, Reasoning, and Inference.* Cambridge University Press, 2000.

[13] Society of Automotive Engineers (SAE International). On-board diagnostics system. `en.wikipedia.org/wiki/OBD-II_PIDs`,`http://en.wikipedia.org/wiki/On_Board_Diagnostics`, 2003.

[14] A. Vass. Toyota Prius CAN Project. `http://www.vassfamily.net/ToyotaPrius/CAN/cindex.html`, 2004.

Efficiency of visual time-sharing behavior – The effects of menu structure on POI search tasks while driving

Tuomo Kujala
Agora Human Technology Center
P.O.Box 35
University of Jyväskylä
+358 14 260 4660
tuomo.kujala@jyu.fi

ABSTRACT

In this paper, the effects of two user interface menu structures on a mobile device display, list and grid, are compared in a driving simulation with the measures of visual time-sharing efficiency, visual load, driving performance and secondary task performance. Eighteen participants conducted a set of eight Point-of-Interest (POI) search tasks with the grid- or list-style menus on navigation software during simulated driving. Between-subject analysis revealed that the list-style menu structure supports more efficient and systematic, and thus, safer interaction while driving than the grid-style menu, in terms of time-sharing and total glance time. However, significant effects of the menu structures were not found in secondary task performance, driving performance measured as lane excursions, or in the measures of average duration of, or total number of glances at the display. The results also suggest that the fewer items in a view, the more efficient and safer the interaction in terms of time-sharing. The sensitivity of the time-sharing metrics for revealing tactical level driver distraction in driving simulation can be argued as being at a higher level than the sensitivity of metrics related to lane maintenance, visual load or secondary task performance.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems – *human factors, human information processing.*

General Terms

Measurement, Performance, Design, Reliability, Experimentation, Human Factors, Theory.

Keywords

Driver distraction, time-sharing, visual interaction, displays, menu structures, workload, visual load, driving performance, levels of control, tactics, strategies.

1. INTRODUCTION

The safety effects of in-vehicle information system (IVIS) use while driving is a topic that is gaining more and more attention these days because of the fast development of mobile technology and services [12]. In this line of research, driver distraction is the key concept defined by Lee, Young, and Regan [10] as *a diversion of attention away from activities critical for safe driving toward a competing activity.*

The experimental approach on studying driver distraction has been an area of interest in human factors research since the 1980s. Driving simulation studies have been frequently used in order to avoid real crash risk (see e.g. [3]). A popular paradigm in this line of research has been based on the measurement of driver workload and driving performance at the level of operational control of the vehicle [9]. The basic problems with interpreting the results of these experiments often reside in the not-self paced and time-pressured tasks, and subsequently in the absence of participants' possibilities to prioritize the driving over secondary tasks. The external validity of the conclusions can often be questioned (e.g., [7][17], see also [6]). These studies are valuable for revealing capacity limitations of the drivers in a dual-task situation. However, they do not necessarily tell us if the drivers are able to overcome their capacity limits with tactical behaviors in real traffic to maintain a sufficient level of driving performance.

Recently, new perspectives and models for studying driver distraction on multiple levels have been proposed [9][14]. Lee, Regan, and Young [9] introduced the model of driver distraction comprising of breakdowns at the operational, tactical and strategic levels of control in dual-tasking while driving based on Michon's [11] three-level model of driving behavior. This model induces new types of challenges for experimental research; how can breakdown in control be measured on the levels of tactical and strategic control? These are not necessarily in direct relation to task workload or to the lapses of vehicle control at the level of operational control [9].

In this paper, while focusing on interaction with visual IVIS displays, we can ask; what kinds of display design solutions could support drivers' tactical and strategic skills in overcoming their visual capacity limitations? Task predictability, interruptability, resumability, and ignorability have all been acknowledged as important secondary task qualities for promoting traffic safety [1] [5][9][16], but guidelines, such as the European Statement of Principles [1], are not specific in defining how particular display properties relate to these aspects and how to measure them. Awareness of task demands and one's own capabilities, i.e. situation awareness, are related concepts that are highly relevant for drivers' tactical and strategic abilities in a dual-task situation. How can these be measured in an objective way?

AutomotiveUI'09, September 21-22, 2009, Essen, Germany
ACM 978-1-60558-571-0/09/009

The typical measures in dual-tasking experiments focus on measuring driver workload (e.g. visual load) or performance at the operational level of control (e.g. driving performance). Traditional measures of visual load focus on the average glance durations or on the total glance durations (i.e. total glance time, tgt) and total number of glances at the display. However, it has been observed that in general, drivers tend to keep the average glance durations below 1.6 seconds in all circumstances and increase the frequency of glances instead of increasing the lengths of individual glances while the visual secondary task demands increase [20]. This is natural behavior, if we acknowledge that drivers, *in general*, try to behave as rational and intentional human beings in traffic.

In this paper, visual time-sharing, or time-sharing in short, is defined as allocation of visual attention in time between tasks. Time-sharing-metrics have been suggested and also utilized to a minor degree to provide information on the glance duration distributions towards a visual secondary task, and thus, on the total efficiency of the allocation of visual attention [2][4][18][21][22]. Very short glances at an in-vehicle display can indicate inefficient search behaviors, as well as rare but significantly long glances that can also increase the level of crash risk [4]. Thus, time-sharing metrics could presumably provide us information on driver distraction at the tactical level of dual-task control. For example, a significant difference in the variance of glance durations on two display designs could tell us that the design with lower variance gives better support for controlled visual search behavior, given the same variability of the driving task's visual demands. In addition, the significance of even one "overlong" glance at an in-vehicle display in the wrong situation cannot be emphasized enough [4]. The traditional measures of average or total glance durations cannot provide us with information on the frequencies of these often rare occasions.

For industrial purposes, fast but sensitive and reliable methods for revealing differences in the distraction potentials of visual IVIS displays are obviously required. Sensitivity means that the metrics can discriminate between designs reliably with statistical significance already with small sample sizes, and thus enables cost-efficient studies. The methods should also provide us with information about driver behavior on multiple levels of driver distraction [9], not merely on the level of operational control, for enabling higher external validity of the conclusions.

The experiment presented in this paper relates to a real design problem in the design of navigation software for a mobile device. The problem goes; which menu structure should be used in the driving mode of the software: list or grid (see Figure 1)? Does this decision have some potential effects on traffic safety? Intuitively, one could argue that the grid-style menu supports faster interaction by shorter paths to more items than the list-style menu. In addition, larger icons can be used and a single view can show more items at once than the list-style menu, thus enabling lower menu structures. All these aspects could support faster, and thus, perhaps safer interaction while driving. On the other hand; the list-style menu could support more predictable interactions because of the more straight-forward two-way movements in the menu. However, in bench-tests without driving, the interaction with either menu does not seem to be significantly more complex than with the other.

Figure 1. Two alternative menu structures, list and grid, for the driving mode of the navigation software.

In this paper, the following questions are addressed:

-Which menu structure, grid or list, supports safer interaction with a mobile device while driving? Is there a significant difference between the two designs with any of the measures?

-Do the amount of items in a view, or the levels of menu, have moderating effects on the previous issues?

In addition, the sensitivity of lane maintenance, secondary task performance, visual load, and time-sharing metrics are compared. What types of measures could indicate significant effects already with small sample sizes? Are the metrics of time-sharing efficiency suitable and sensitive enough for assessing distraction effects of in-vehicle display designs at the level of tactical and strategic control?

2. METHOD

Two hypotheses were made prior the experiment based on our earlier research. Firstly, interaction with the list-style menu is assumed to be safer while driving than with the grid-style menu, because it could support more systematic visual interaction. The visual demands of the interaction are thus more easily learnable, and thus more predictable, interruptible and resumable than when interacting with the grid-style menu. This should be visible with the measures of time-sharing efficiency, but not necessarily with the measures of lane maintenance, visual load or secondary task performance. We expected larger variances in glance duration distributions, larger maximum glance durations, and greater amounts of very long, as well as very short glances towards the display with the grid-style menu. Secondly, the fewer items in a view and the lower the menu structure, the more efficient the interaction is supposed to be in terms of time-sharing.

2.1 Participants

Volunteers were invited to participate through public university e-mail lists. The 6 female and 12 male right-handed participants were from the ages of 20 to 35 years old, and had normal or corrected vision. All had a valid driving license and possessed lifetime driving experience of at least 10 thousand kilometers, ranging from 10,000 to 500,000 km. Drivers with a very low level of experience and aged drivers were not selected to the sample for mitigating the known effects of low level of driving experience [21] and aging [22] on time-sharing efficiency. The experiment was conducted in Finnish with fluent Finnish-speakers.

Participants were randomly selected from the volunteers, but they were divided in two pair-matched groups according to gender, levels of driving experience and age (see Table 1). The group with the grid-style menu had an average lifetime driving experience of 103,000 km (SD=159), and an average age of 25.1 years (SD=2.8). For the *List*-group the corresponding averages were 95,000 km (SD=124) and 25.7 years (SD=5.0).

Table 1. Classes of pair-matched participants

Number of participants	Driving experience (thousand km)	Age
6	<20	20-25
2	<20	26-35
2	20-50	20-25
2	50-100	20-25
2	>100	20-25
4	>100	26-35

2.2 Tools and environment

The experiment was conducted in the three-display driving simulation environment of the Agora User Psychology Laboratory (see Figure 2).

Figure 2. The driving scene from a participant's point of view.

The central equipment included consent forms, Nokia N95 8GB mobile device with 2.8" display in a dashboard holder, SMI iView X HED helmet-mounted eye-tracking system with 50Hz sampling rate, two video cameras for recording the driving scene with sound and for back-upping the eye-tracking, as well as two laptops for capturing the video material. The distance between participants' eyes and the windscreen projected driving scene was fixed at ca 100 cm, but the distance of the pedals and the steering wheel with the device holder from the participant were adjustable. Thus, the mobile device's distance from the participant's eyes varied between 55 to 70 centimeters depending on arm lengths.

The driving simulation software is an open-source based car simulation of which motion formulae is based on actual engineering documents from the Society of Automobile Engineers (www.racer.nl). The trials were driven with a simulated Ford Focus with automatic shifting on a road-like environment simulating the Polish countryside. A simulated racetrack was used for practice. The driving scene was projected onto the wind screen of the fixed-base vehicle cockpit and included a speedometer and a tachometer.

2.3 Design and procedure

The experimental design was a mixed-factorial design (see Table 2). The menu structure was a between-subject variable and the levels of menu, and the number of items in the view, were the within-subject variables.

Table 2. The experimental design

Menu (between-subject)	Levels (within-subject)	Items (within-subject)
List	3	2
Grid	4	4
	>4	6
		9

The experiment started with the signing of a consent form, and by receiving general instructions. Practice in driving on a looped track of around 5 minutes was provided for the participant. After the rehearsal the participant completed a baseline driving task of around 10 minutes for getting more practice and for baseline-dual-task driving performance comparisons. The participant got to complete one search task without driving with the search tasks on the grid- or list-style menus before the dual-task trial. The dual-task trial lasted for 6 to 10 minutes depending on the participant's task completion times. After driving, the participant was interviewed in order to explore the participants' strategy space and to classify the drivers' ways of interacting. Both menu structures were shown to the participants during the interviews. The main questions of interest in the interviews were: "Did you feel time-pressure or need to hurry in the search tasks?"; "How did you perform the search tasks; did you have or did you develop certain ways of interacting during the trial?"; "Which menu structure would you prefer to use while driving?"; and finally, "Could you imagine yourself conducting this type of search activity while driving?".

The driving task instructions were to keep the velocity of the vehicle between 40-60 km/h, and to keep the two Head-Up-Display meters between the white lane markings. The participant was also instructed to stop the vehicle immediately if he/she saw a deer. Driving practice included a deer, but the actual trials did not. However, the participants were not made aware of this beforehand. There was oncoming traffic in the form of four cars at preset points on the road.

The search tasks were *self-paced* and the participant was instructed to keep *priority on driving*. Driving task priority was emphasized by promising 10 movie tickets in total to the most accurate drivers. Driving task accuracy was defined as the total duration spent out of the lane or above/below the instructed speed zone. Tasks were given verbally by the experimenter while driving, allowing for a very short pause of a few seconds between tasks after a successful task. The participant could ask the task to be repeated with saying “repeat”, if he/she forgot or did not hear the task.

Participants were given the scenario that they are travelling in the Polish countryside by car and searching for Points-of-Interest (POIs) nearby. The search tasks are listed in Table 3. The number of items in a view varied within the tasks depending on the level of the menu. Task orders were randomized.

Table 3. The search tasks

Task #	Task (path (# of items))	Levels
1	Find the way to the nearest hotel (Options-Search(9)-Hotels)	3
2	Find the way to the nearest shop (Options-Search(9)-Shops)	3
3	Find the way to the nearest rest area (Options-Search(9)-Automotive(6)-Rest areas)	4
4	Find the way to the nearest library (Options-Search(9)-Services(9)-Libraries)	4
5	Find the way to the nearest railway station (Options-Search(9)-Transport(4)-Railway stations)	4
6	Find the way to the nearest McDonald’s (Options-Search(9)-Restaurants(18)-McDonald’s [required scrolling])	>4
7	Find the way to the theatre named Kto (Options-Search(9)-Entertainment(2)-Theatres(18)-Kto [required scrolling])	>4
8	Find the way to the museum named Dom Jana (Options-Search(9)-Sights(6)-Museums(27)-Dom Jana [required scrolling])	>4

2.4 Variables and analysis

Independent variables included the menu structure, the levels of menu, and the number of items. Efficiency of time-sharing, visual load, driving performance and search task performance were selected as dependent measures.

Time-sharing efficiency was measured by the maximum and standard deviations of glance durations (at the display), by the frequency of over-1.6-second and over-2-second glances in total and in curves, and by the frequency of under-0.4-second glances. 1.6 seconds has been observed to be the limit under which drivers generally prefer to keep their glances at in-vehicle displays in all circumstances [20]. Over-2-second long glances have been observed to increase crash risk and frequency of near crash situations in real traffic [8]. Additionally, we wanted to include a measure of situation awareness. The metrics of overlong glances while driving in curves served this purpose. The movement of gaze from the driving scene to the display and back was scored into the glance duration, and as such, under 0.4 second glances leave very short time for gathering any useful information from the display, especially if assuming some task set switch costs (see e.g. [13]). A typical shift of gaze between the display and the driving scene took 160 ms. The effects of the levels of menu and the number of items were analyzed for maximum and standard deviations of glance durations when applicable (enough glances). Interaction effects of menu structure and the within-subject-variables on these measures were also analyzed.

Visual load was measured by the total number and mean duration as well as total duration of glances at the display. Driving performance was measured as total frequency and duration of lane excursions, and additionally involved baseline-dual-task comparisons. The within-subject effects of the levels of menu on driving performance were excluded in the analysis. Total frequency and duration of speed area violations were scored automatically from the simulation log file for the accuracy comparisons between participants. Secondary task performance was measured as frequency of errors and task completion times with driving excluded, that is, total glance times at the display by task. Error in a search task was defined as a wrong selection.

Effort was invested to control some undesired variables. There was an effort to accommodate for learning effects and individual differences in skills via driving practice and practice for the search task. As mentioned, the menu-groups were balanced by gender, driving experience and age. Order effects were eliminated by randomizing orders of the search tasks (5 different orders, same orders for the pairs). Driving task difficulty while dual-tasking was controlled by random task starting points, which depended on the participants’ performance. In addition, every other participant in the group drove the same road in the dual-task trial as the others, but in the opposite direction. This kept the driving task demands (road curvature) at the same level for everyone but gave more randomness to the task starting points. The driving speed was kept fixed between 40-60 km/h by instructions. Movie tickets were promised to the most accurate drivers in order to make the participants prioritize the driving and to encourage greater effort, giving the absence of real danger in a driving simulation. The deer observation task was instructed to make the participants observe the environment in a more natural way than merely observing the lane markings and the speedometer.

Noldus Observer XT software was used for scoring behaviors frame-by-frame (25 frames per second), for search task performance, lane excursions, and eye-movements. A glance at the secondary task display was scored following the SAE J2396 definition [15]. The analysis of overlong glances in curves was done via an automatic script that compared the steering wheel movements recorded in the log file of the driving simulation to the synchronized eye-tracking data file. The limit for driving on a curve was defined to be the absolute value of 1.00 or more of the steering wheel position in terms of the simulation’s log file data, in which 0.00 was the calibrated center point. The frequency and durations of lane excursions were analyzed for equal journey lengths between the two trials. Due to there being an end in the road, there was a time limit of about 10 minutes for the completion of the search tasks. Three participants did not have enough time to start the last tasks in their trials. This was taken into account in the analysis by excluding the corresponding task data from their pairs in the other group. The time limit was not instructed to the participants. For statistical analysis, two-tailed *t*-tests were used for between-subject comparisons and mixed-model ANOVA (menu x level, 2 x 3; menu x items, 2 x 4) to find within-subject as well as interaction effects. An alpha level of .05 was used in the statistical testing. The interviews were analyzed from the videos and the participants’ answers were classified.

3. RESULTS

3.1 Driving performance

The means for the total number of lane excursions were 7.78 (SEM=2.12) for the list and 20.11 (SEM=6.71) for the grid. Correspondingly, the means for the total duration of lane excursions were 7.26 s (SEM=2.19) for the list and 30.26 s (SEM=13.41) for the grid. However, these differences were not statistically significant (total number: $t(16)=1.75$, $p=.110$; total duration: $t(16)=1.69$, $p=.130$).

Despite of the higher level of practice in driving after the baseline driving trial, there was a significant effect of the dual-task condition on the total number of lane excursions ($F(1,16)=4.92$, $p<.050$, see Figure 3). Analysis of speed variations did not reveal high numbers of significant speed zone excursions and speed was not included in the analyses.

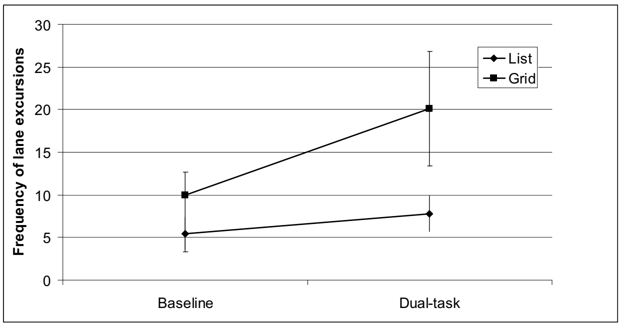


Figure 3. Total number of lane excursions by trial (N=18). Means and Standard Errors of Means.

3.2 Visual load

Total glance times was the only measure of visual load that showed significantly larger values for the grid-style menu ($t(16)=2.91$, $p<.050$, see Table 4). The mean glance lengths at the displays were similar and indicate safe *average* visual behavior [20].

Table 4. Visual load (N=18), means (SEMs)

Menu	Total glance time, s	Total number of glances	Average glance duration, s
List	98.20 (5.22)	93.56 (5.46)	1.07 (.05)
Grid	144.10 (14.90)	121.33 (12.43)	1.06 (.04)

3.3 Time-sharing efficiency

The effects of the menu structures on the participants' time-sharing efficiency are illustrated in the Figure 4.

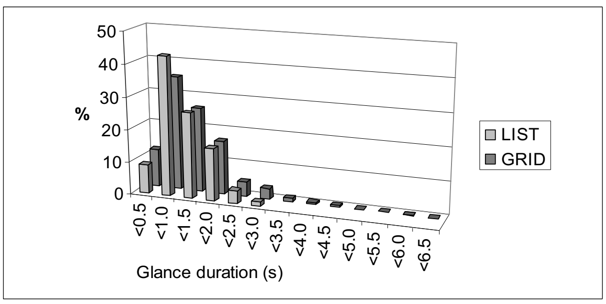


Figure 4. Glance duration distributions (N=18).

The maximum glance durations at the grid-style menu were significantly larger than at the list-style menu ($t(16)=2.93$, $p<.050$, see Figure 5). The means for the individual standard deviations of glance durations were also larger for the grid-style menu, but the difference was not statistically significant ($t(16)=1.93$, $p=.072$) in these sample sizes.

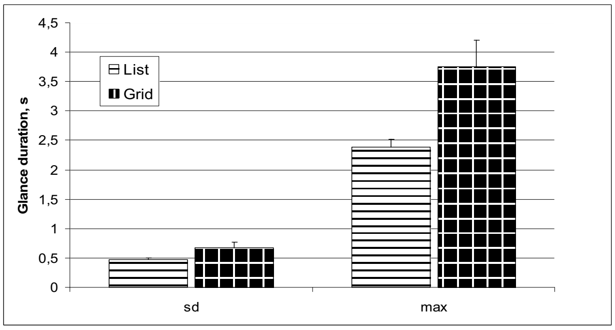


Figure 5. Standard deviation of and maximum glance durations (at the display, N=18). Means and SEMs.

There was a significant effect of the menu structure on the frequencies of over-1.6s-glances ($t(16)=2.82$, $p<.050$) and over-2.0s-glances ($t(16)=2.12$, $p<.050$, see Figure 6). The larger frequencies of over-1.6s-glances ($t(16)=2.22$, $p<.050$) and over-2.0s-glances ($t(16)=2.41$, $p<.050$) in curves indicates, that the participants using the grid-style menu also did significantly more of these glances while they were not driving on a straight road. In addition, there was a significantly larger number of under 0.4-second-glances on the grid-style menu ($t(16)=2.12$; $p<.050$). A summary of the time-sharing metrics is presented in the Table 5.

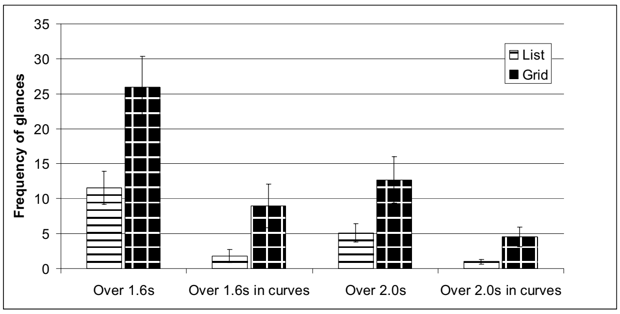


Figure 6. Frequencies of overlong glances in total and while driving in a curve (N=18). Means and SEMs.

Table 5. Time-sharing metrics (N=18), means (SEMs)

Measure	List	Grid
Standard deviation of glance durations, s	.47 (.03)	.67 (.10)
Maximum glance duration, s	2.38 (.13)	3.75 (.45)
Over-1.6s-glances	11.56 (7.02)	25.89 (4.52)
Over-1.6s-glances in curves	1.78 (.81)	9.00 (3.11)
Over-2.0s-glances	5.11 (1.36)	12.67 (3.30)
Over-2.0s-glances in	1.00 (.33)	4.56 (1.44)

curves		
Under-0.4s-glances	3.22 (.85)	7.00 (1.57)

3.3.1 Number of items

The number of items on a view had a significant effect on the maximum glance durations at the display (F(3,14)=26.02, p=.000, see Figure 7). Between-subject effects of the menu structure were not significant, but there was significant interaction effects between the menu structure and the number of items (F(3,14)=3.70, p<.050). There were significantly larger maximum glance durations on 9 item views with the grid-style menu compared to the list-style menu, *t*(16)=2.93, p<.050.

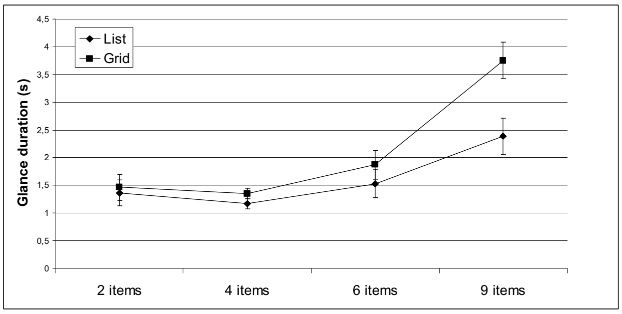


Figure 7. Maximum glance durations by the number of items (N=18). Means and SEMs.

On the task level, the standard deviations of glance durations in task 8 (27 items in the last menu) were significantly larger with the grid-style menu (M=.74, SEM=.12) than with the list-style menu (M=.48, SEM=.04), *t*(14)=2.14, p<.050.

3.3.2 Number of menu levels

The number of menu levels had a significant effect on the maximum glance durations at the display (F(2,15)=8.67, p<.010, see Figure 8).

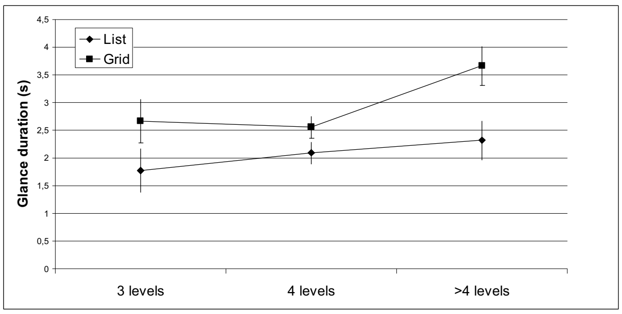


Figure 8. Maximum glance durations by the number of menu levels (N=18). Means and SEMs.

There was also a significant between-subject effect of the menu structure on the maximum glance durations (F(1,16)=5.32, p<.050). However, significant interaction effects of the variables were not found. Standard deviations of glance durations did not show significant effects of the menu levels, but indicated significantly larger values for the grid-style menu (M=.72, SEM=.10) than for the list-style menu (M=.47, SEM=.03) in the >4 level tasks, *t*(16)=2.38, p<.050.

3.4 Search task performance

The menu structure did not significantly affect task performance on any other task than #2 (tgt: *t*(14)=2.28, p<.050). The most difficult tasks in total seemed to be the tasks 7 and 8 (see Table 6).

Table 6. Task times and errors in the search tasks

Task #	N	Menu	Task time (as tgt, mean (SEM), s)	Errors (sum)
1	9	List	7.86 (1.40)	0
1	9	Grid	9.65 (.78)	2
2	8	List	6.81 (1.28)	1
2	8	Grid	16.94 (4.26)	6
3	9	List	13.01 (2.56)	2
3	9	Grid	18.31 (3.56)	10
4	9	List	12.80 (1.92)	3
4	9	Grid	13.65 (1.46)	4
5	9	List	7.48 (.73)	0
5	9	Grid	8.66 (1.11)	1
6	8	List	13.71 (.91)	0
6	8	Grid	26.91 (5.95)	4
7	9	List	19.62 (2.35)	9
7	9	Grid	18.77 (4.42)	4
8	8	List	21.58 (2.89)	8
8	8	Grid	40.59 (10.66)	23

3.5 Interviews

The participants reported that the development of "sense of touch" and memory helped them in doing some of the movements without looking at the device, e.g. in the event of the repeated 'Search'-selection. This observation could be confirmed from the video material. Some participants jumped directly between the first and the last item in the menu; in these cases the participants reported that the list better supported the awareness of location in the menu (notice: there was no scroll bar in the scrollable submenus). All of them reported they mainly read the texts. Icons were reported to have been helpful only in the identification of certain targets, e.g. restaurants. Some felt that the texts were easier to distinguish in the grid. There were notes of difficult menu hierarchies in some tasks, especially in deciding to which submenu rest areas, theaters, and museums belonged to.

Twelve out of 18 participants reported that they did not feel time-pressure while completing the search tasks. Four participants reported some time-pressure, but they felt it was self-induced. Two participants reported that they felt time-pressure while completing the tasks, but that they also tried to prioritize driving. When asked for preference, 10 participants preferred the list, 6 preferred the grid, one preferred a combination of grid and list depending on the view, and one felt it did not make a difference. Thus, all of them were not fully aware of the risks of the grid-style menu. Finally, 15 out of the 18 participants could imagine themselves conducting these types of search tasks while driving.

4. DISCUSSION

In this paper, we studied the sensitivity of the metrics of lane maintenance, secondary task performance, visual load, and time-sharing efficiency, to reveal lapses in participants' tactical visual behaviors while driving and dual-tasking with visual secondary tasks. This was achieved through a case study in which we compared the effects of mobile software's menu structures on these measures.

As hypothesized, the collected time-sharing data suggests that the unsystematic nature of visual search and movement that the grid-style menu offers may lead to less efficient, less predictable, less resumable and less interruptible interactions while driving than the interaction with the list-style menu. This was observed especially with displays featuring over 6 items. Although the lane maintenance metrics provided some hints of this, they do not seem to be very sensitive to revealing risky visual behavior or differences in the distraction potential of visual display properties at this level. A possible reason for the non-significant effects of menu structures on lane maintenance could be that the drivers' performance at the level of operational control of the vehicle does not need to be in direct relation with the lapses of tactical level behavior [9]. Driving performance is, besides visual behavior, highly dependent on other simultaneous factors, such as the curvature of the road at any one time (see [4]).

Nor does the metrics of visual load seem to be highly sensitive in revealing occasional risky visual behavior. It was only the total glance times that indicated significantly faster performance for the list-style menu. However, neither does this measure of visual load have to be in a direct relationship to risky visual behavior in all cases. It is highly possible that the total glance times for two secondary tasks are the same, but the glance duration distributions are very different. The crash risk potential in this case is not about the visual load, i.e. the total or mean amount of visual attention required for a secondary task, but about how efficiently you can allocate your visual attention in time between driving and the visual secondary task (see [4]). This time-sharing behavior is a critical component of successful driving, and not merely for lane keeping and collision avoidance, but also for detecting and preparing for potential, unexpected threats in time.

Analysis of search task performance revealed that the difficulty (i.e. complexity) of the secondary task is not necessarily the main factor for unsafe time-sharing behavior. In other words, task complexity can be at the same level between two visual secondary tasks, but still they can have different effects on time-sharing efficiency, and consequently, on potential crash risk [4].

Some design recommendations can be inferred from the results. The list-style menu seems to support safer interaction while driving than the grid-style menu. However, it should be noted, that this recommendation at this point only applies to mobile devices with no touch screen. The selection of an item on a touch screen display can be much more straight-forward. The grid-style menu's higher potential for very short and inefficient, as well as overlong glances in relation to the driving situation, can be explained by the more unsystematic steps of interaction. Thus, IVIS display designs should support systematic ways of interaction. Furthermore, the results suggest that the maximum of between 6 to 9 items on a single view can make the information search less risky. Thus, prioritization of what kind of POIs are the most often needed by the drivers, and implementation of only these, could be beneficial. On the other hand; the data seems to suggest, that the lower the menu structure, the better, and these two requirements are often in contradiction with each other in versatile software. However, the effects of the number of menu levels should be examined more closely in different experimental design, because in this experiment the observed effects of menu levels can be explained by the large number of 9 item views in the over 4 level tasks. An interesting finding was that in a short time the participants learned to find and select the often repeated functions of the software (Options and Search) without any visual attention. One possible reason for the better time-sharing efficiency with the list-style menu could relate to this finding in that the participants were able, after locating the target item, to quickly calculate the required steps to the item, and perform the movement without looking at the device. This was much more difficult for the grid-style menu. Moreover, these findings suggest that the sense of touch is worth supporting whenever possible. The finding relates to Wickens' [19] theory of multiple resources, suggesting that tasks which occupy different sense modalities can often be efficiently performed simultaneously. The participants seemed to be aware of this and utilized this information tactically.

The current findings have some methodological significance. The time-sharing-metrics utilized here seem to be suitable for measuring task predictability, resumability and interruptability. They also seem to be sensitive enough for discriminating between safer and not-as-safe, as well as between efficient and less efficient IVIS visual display designs, reliably already with small sample sizes. The efficiency of time-sharing seems to be more closely related to the tactical abilities of drivers than to the complexity or visual load (i.e. workload) of the secondary task. In this sense, we can discuss of a kind of "paradigm shift" in driver distraction research with experimental techniques. In this line of research, instead of asking; does the secondary task overload the driver, we ask; what qualitative features of IVIS display designs can support drivers' tactical and strategic abilities? The target of analysis in experimentation at this level resides in the efficiency of drivers' visual time-sharing behavior, not in the visual load or driver's performance at the operational level of vehicle control [9]. For example, glance duration distributions, or in particular, the frequencies of very short glance durations are rarely looked at in contemporary distraction research [12].

The experiment invoked new questions of which some are already under on-going research. By utilizing the time-sharing metrics, the comparison of touch screen vs. non-touch screen devices with the grid-style menu could reveal whether the main source of the observed distraction is related to moving in the menu, or merely to searching for information in the grid-style views. Another interesting issue under current experimentation is the means of scrolling on a touch screen with menus consisting of more items than the display can hold at a time. Overall, a large amount of similar issues exists that should be tested with the time-sharing metrics. The cumulating database could also serve for comparisons between visual IVIS designs. Moreover, the metrics themselves may be further developed, and validated e.g. with field studies. Future research efforts should especially include further development of the metrics for situation awareness and systematicity of drivers' visual search behavior on IVIS displays.

Finally, please notice that we do not take any account with these results as to whether this type of activity while driving is risk-free or not, even after a large amount of practice. The focus of inquiry was on first-touch experience with the software while driving. We wanted to see how the different menu structures affect visual search behavior with unfamiliar contents, because search for POIs involves typically unseen, dynamic contents depending on the driver's location. However, the driving speed was relatively low in the experiments, *40-60 km/h* and all of the participants were fairly *experienced young drivers*. At higher speeds, or with less experienced [21] or aged drivers [22], the safety risks of the secondary tasks will presumably be greater than observed. With these experiments, we wanted only to reveal which menu interaction style is *safer* while driving – the list or the grid-style – and get support for our hypotheses as to why it is so. We recommend that drivers are made aware of the potential risks of system use while driving. We further encourage them to be fully stationary while searching for locations, whenever possible.

5. ACKNOWLEDGMENTS

This research was sponsored by Theseus, a research project on human-technology interaction design funded by TEKES (Finnish Funding Agency for Technology and Innovation). The driving simulation environment was developed with a grant from the Henry Ford Foundation. The author would like to thank Professor Pertti Saariluoma and research assistant Ari-Matti Saari for their help in planning and conducting the experiments. Credits for the editing of spelling go to Rebekah Rousi. The author wants to express gratitude especially to Mikko Nirhamo and Timo Tokkonen at Nokia for the valuable collaboration in making the experiments possible.

6. REFERENCES

[1] Commission of the European Communities 2006. Recommendation on safe and efficient in-vehicle information and communication systems: Update of the European statement of principles on human machine interface (Report C(2006) 7125 final). Brussels, BEL: Commission of the European Communities.

[2] Hoffman, J. D., Lee, J. D., McGehee, D. V., Macias, M., and Gellatly, A. W. 2005. Visual sampling of in-vehicle text messages: Effects of number of lines, page presentation, and message control. Transport. Res. Rec. 1937, 22-30.

[3] Horrey, W. J. and Wickens, C. D. 2006. Examining the impact of cell phone conversations on driving using meta-analytic techniques. Hum.Factors 48, 196-205.

[4] Horrey, W. J. and Wickens, C. D. 2007. In-vehicle glance duration: Distributions, tails, and model of crash risk. Transport. Res. Rec. 2018, 22-28.

[5] International Organization for Standardization 2007. ISO 16673 Road vehicles-ergonomic aspects of transport information and control systems-Occlusion method to assess visual demand due to the use of in-vehicle systems. Geneva, CH: International Standards Organization.

[6] Jamson, A. H., Westerman, S. J., Hockey, G. R. J., and Carsten, O. M. J. 2004. Speech-based e-mail and driver behavior: Effects of an in-vehicle message system interface. Hum. Factors 46, 625-639.

[7] Just, M. A., Keller, T. A., and Cynkar, J. 2008. A decrease in brain activation associated with driving when listening to someone speak. Brain Res. 1205, 70-80.

[8] Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., and Ramsey, D. J. 2006. The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data (DOT HS Rep. 810 594). Washington DC: U.S. National Highway Traffic Safety Administration.

[9] Lee, J. D., Regan, M. A., and Young, K. L. 2008. What drives distraction? Distraction as a breakdown of multilevel control. In Driver Distraction: Theory, Effects, and Mitigation. M.A. Regan, J.D. Lee, and K.L. Young, Eds. CRC Press, Boca Raton, FL, 41-56.

[10] Lee, J. D., Young, K. L., and Regan, M. A. 2008. Defining driver distraction. In Driver Distraction: Theory, Effects, and Mitigation. M.A. Regan, J.D. Lee, and K.L. Young, Eds. CRC Press, Boca Raton, FL, 31-40.

[11] Michon, J. A. 1985. A critical view of driver behavior models: What do we know, what should we do? In Human Behavior and Traffic Safety. L. Evans and R. C. Schwing, Eds. Plenum Press, New York, NY, 485-520.

[12] Regan, M.A., Lee, J.D., and Young, K.L. 2008. Driver Distraction: Theory, Effects, and Mitigation. CRC Press, Boca Raton, FL.

[13] Sakai, K. 2008. Task set and prefrontal cortex. Annu. Rev. Neurosci. 31, 219-245.

[14] Sheridan, T. B. 2004. Driver distraction from a control theoretic perspective. Hum. Factors 46, 587-599.

[15] Society of Automotive Engineers 2000. SAE J2396 Surface vehicle recommended practice, definitions and experimental measures related to the specification of driver visual behavior using video based techniques. Warrendale, PA: Society of Automotive Engineers.

[16] Society of Automotive Engineers 2004. Navigation and route guidance function accessibility while driving, SAE recommended practice J2364. Warrendale, PA: Society of Automotive Engineers.

[17] Sodnik, J., Dicke, C., Tomazic, S., and Billinghurst, M. 2008. A user study of auditory versus visual interfaces for use while driving. Int. J. Hum.-Comput. St. 66, 318-332.

[18] Victor, T. W., Harbluk, J. L., and Engström, J. A. 2005. Sensitivity of eye-movement measures to in-vehicle task difficulty. Transport. Res. F-Traf. 8, 167-190.

[19] Wickens, C.D. 2002. Multiple resources and performance prediction. Theoretical Issues in Ergonomics Science 3 (2), 159-177.

[20] Wierwille, W. W. 1993. An initial model of visual sampling of in-car displays and controls. In Vision in Vehicles IV, A.G. Gale, I. D. Brown, C. M. Haslegrave, H. W. Kruysse, and S. P. Taylor, Eds. Elsevier Science Publishers B.V., Amsterdam, NL, 271-279.

[21] Wikman, A.S., Nieminen, T. and Summala, H. 1998. Driving experience and time-sharing during in-car tasks on roads of different width. Ergonomics 41 (3), 358-372.

[22] Wikman, A. S. and Summala, H. 2005. Aging and time-sharing in highway driving. Optometry Vision Sci. 82, 716-723.

Toggle Strategies for the POI selection via the iDrive controller

Ronald Ecker[1] **Verena Broy**[1] **Nina Joshi**[2]

[1]BMW Group Research and Technology
Hanauerstraße 46
Munich, Germany
+49 89 382 55362
Ronald.Ecker@bmw.de

[2]Stanford University
Stanford, CA 94309
ninasj@stanford.edu

ABSTRACT

The importance of spatial and geo-based information has increased over the last few years. The most prevalent example of this kind of information is points of interest (POI) like hotels, restaurants, gas stations, etc. As cars are made for individual transportation, interacting with geo-based information via the In-vehicle Information System (IVIS) should be possible. At present, state-of-the-art IVIS only permit a list based or center based selection on the map, which makes it difficult to handle a high closeness of geo-based data. In this paper, we present alternative approaches for selecting geo-based data with a multifunctional controller. In our work, visual cues help users predict the selection order. An explorative user study showed potential advantages of our concepts.

Keywords

Geo-based information, POI selection, In-vehicle Information System, automotive HMI, iDrive Controller.

INTRODUCTION

Geo-based data representation is gaining more and more importance. Several applications offer possibilities to mark and select points of interest (POI) e.g. restaurants or hotels depending on their position on a map. Especially in portable or in-built navigation systems, users should be able to find places to eat or to refuel in their surroundings or along their planned route in reference to their actual position. Displaying and interacting with POIs implies challenges for In-vehicle Information Systems (IVIS). Some interesting research topics are: how to deal with a large amount of geo-referenced data, how to define filtering methods, and how to select a POI on a map with a multifunctional controller.

Regarding desktop or mobile applications, POIs on a map are selected via direct manipulation by the mouse pointer, the stylus or the user's finger. As common IVIS systems, like BMWs iDrive [3] and AUDIs MMI [1], are manipulated by multifunctional controllers, common direct manipulation concepts are not suitable. Multifunctional controllers normally can at least be pressed, rotated clockwise and counter clockwise [1]. Some can also be pushed in four directions [2]. In actual realizations such commands are used for manipulating the map itself, e.g. zooming by rotating or panning by pushing.

AutomotiveUI'09, September 21–22, 2009, Essen, Germany
ACM 978-1-60558-571-0/09/009

In this paper we present three different concepts that enable users to select POIs on a map. These concepts vary in their visualization as well as in their selection order strategies, which can be toggled through the POIs via the controller. An explorative user study showed that users prefer the concept containing an appropriate visualization of the implemented toggle strategy.

RELATED WORK

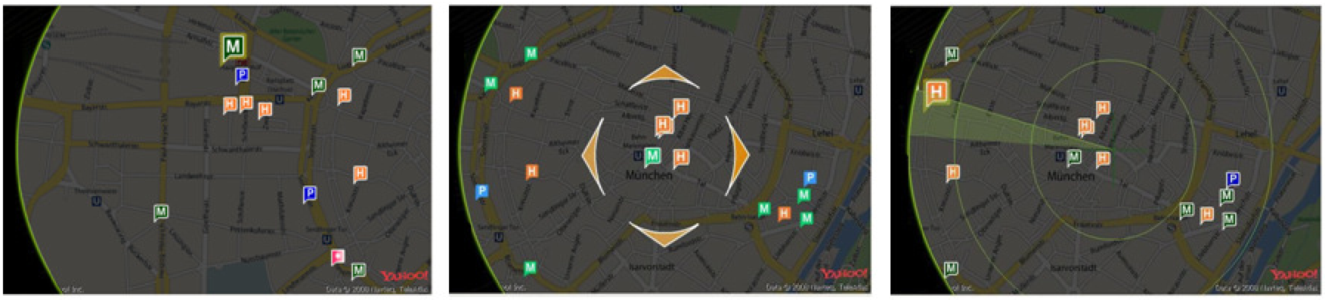
State of the art in-car IVIS provide POI selections in a list [3] (Figure 1), sorted e.g. by the distance to the actual position or center based selection directly on the map itself [1] (Figure 2). Especially when selecting POIs within a high-density area, problems with overlapping icons can arise. Choosing POIs near the screen edge, leads to a long interaction time.

Figure 1: iDrive Navigation screenshot. Splitscreen with list and map representation of POIs.

a) b)

Figure 2: Audi MMI Screenshot of the navigation system. a) Selecting a POI on the map. b) Entering the selection for targeting.

We approach this problem by toggling from one POI to another to reduce completion time and the visual demand for hand-eye coordination. The identified research questions are: which are the most suitable toggling strategies and can the usability be improved by a visualization of these strategies?



a) Toggle b) Spotlight c) Radar

Figure 3: Screenshots of the three alternative prototypes.

DESIGN

All three developed concepts satisfy the basic requirements for selecting POIs with the iDrive controller. In the selection mode we shade the map via a transparent layer to reduce visual annoyance between the map and the displayed POIs. We arranged three categories of POIs that differ in their visualization on the map: hotels (H, orange), motels (M, green) and parking spaces (P, blue).

Our multifunctional controller can be pushed in the cardinal directions, rotated and pressed. Pushing the controller in one of the four directions pans the map like in serial implementations, rotating realizes toggling through the displayed POIs on the map, and pressing in vertical direction selects the focused POI. To label the focused POI it is enlarged to the double size of the other POIs and surrounded by a yellow glow. We thought also about marking the two POIs, which can be reached by one step, by a smaller glow and by a bigger representation but this was more annoying than helpful as informal expert interviews unfolded.

Toggle

In the most basic concept, no visual hints of the selection order are displayed when rotating the controller (Figure 3 a). The selection concept is very simple and toggles through the geo-based data as though one is reading. This means that the POIs are accessed from left to right and from up to down.

Spotlight

The third concept was called spotlight as it displayed a circle in the middle of the map and only POIs surrounded by the circle were selectable via rotating (Figure 3 b). We implemented the same reading strategy, like in the toggle concept, for choosing between POIs inside the spotlight.

Radar

Our radar concept provides a radial selection sequence based on the distance to the center of the map. As additional hints we display circles around the center and a line, which connects the focused POI with the center of the map, see Figure 3 c.

USER STUDY

For the evaluation of the three selection strategies, prototypes were realized in Adobe Flash and ActionScript 2. The interactive maps were implemented with the Yahoo Flash framework. As an input device, an iDrive Controller was connected via a CAN card to our applications. A 19'' LCT screen with a 1280*1024 pixel resolution was used. All three prototypes have a size of 800*480 pixels. For the transition between different tasks, users had to pan. Therefore a map overlay which indicated the next map center of task area was added. Every task is comprised by the selection of several POIs. Therefore a dedicated POI icon was chosen (pink with a star). When the selection was executed successfully, the selected pink POI disappeared and the next appeared (Figure 4).

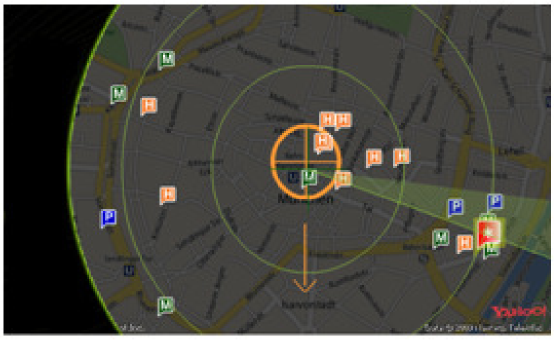


Figure 4: Overlay (orange cross and arrow) which indicates the direction to the next task and selectable POI to participants.

Design of the User Study

To compare all three systems, a within-subject explorative user study was conducted. A Latin square was applied to permute the order in which each participant had to interact with the prototypes. The independent variable was POI selection order strategy in terms of the prototypes (levels: radar, spotlight and toggle). The dependent variables were selection time of the POIs and user preferences.

In total, four different tasks had to be executed. First of all, participants were asked to select three POIs within a group of other POIs (*groupSelect*) followed by a selection of three POIs outlying from others (*lonelySelect*). Afterwards users had to pan to the next task center and select three POIs on a map sector with a few POIs (*fewSelect*). At the end, a selection task with high POI closeness was performed (*manySelect*).

Procedure
At the beginning of the evaluation each user had to solve three training tasks. Afterwards they explored the system followed by a rating: 1 (very good) to 6 (very bad). Then, all system functions were explained. In the next step, participants had to complete the tasks (*groupSelect, lonelySelect, fewSelect* and *manySelect*) as described above while completion time was captured. After these tasks, two questionnaires, SUS [1] and AttrakDiff [5], were answered, and once again, a rating of the system's quality was made and supplemented by a rating of the comprehensibility of the system on a Likert scale from 1 (do not agree) to 6 (highly agree). After participants finished the described procedure with all three prototypes in the above mentioned order, each user ranked the three systems based on his or her preferences (1st, 2nd and 3rd).

Participants
For the user study, twelve volunteers were recruited with an average age of 32 years, two of them female. Everyone had a driving license and experience with navigation systems. One left-handed person attended.

Assumptions
Informal expert interviews showed that POI selection order from the radar prototype seems to be faster and clearer. Based on these results we assumed a higher performance and faster completion times with the radar system. Due to the visual representation of the radar system we concluded that participants prefer the radar prototype.

Results
Total Task Time (TTT)

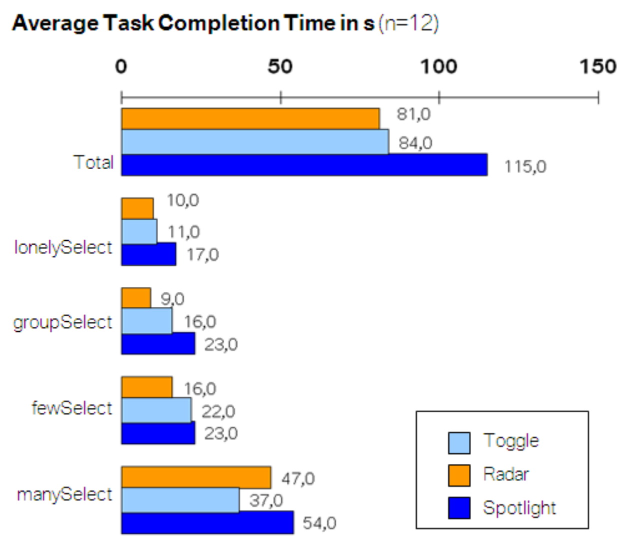


Figure 5: Average task completion time in seconds (n=12).

Each TTT comprises three POI selections. Time was measured from the end of the task instruction to the last interaction step of the third POI selection. TTT for *fewSelect* and *manySelect* contains the additional panning time.

On average, the best overall performance could be achieved under the radar condition (81.0 sec) followed by the toggle condition (84.0 sec). With an average TTT of 115.0 sec, the spotlight system was the slowest for editing the three tasks.

Comparing the different tasks shows a faster interaction time with the radar system except the *manySelect* task. *groupSelect* was finished on average 60% faster under the spotlight and 44% faster than the toggle condition. *fewSelect* was performed on average 30% faster than the spotlight and 28% faster than the toggle condition. Also the selection tasks *lonelySelect* was 10% (toggle condition) and 40% (spotlight condition) faster under the radar condition. The POI selection within the highest POI closeness (*manySelect*) was executed in the shortest time with the toggle prototype (31% faster than spotlight and 13% faster than radar) (Figure 5).

Subjective User Opinion

To retrieve participants' personal preferences concerning the prototypes, two questionnaires were conducted as described above.

The first questionnaire was the SUS (System Usability Scale). It comprises ten questions regarding three dimensions of usability (efficiency, effectiveness and learnability). The result is represented as number between 0 (worst) and 100 (best). On average the twelve participants evaluated the radar prototype with 84 points, the toggle with 74 and the spotlight system with 68 points. Figure 6 shows the evaluation of each dimension. Over all three usability dimensions, users preferred the radar system.

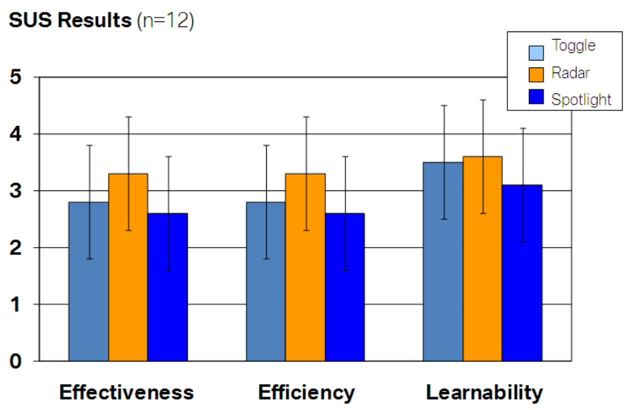


Figure 6: Result of the SUS questionnaire (n=12).

After the SUS, the AttrakDiff questionnaire was filled out. This questionnaire is a semantic differential for evaluating the users' opinion. Normally pragmatic quality, hedonic quality-stimulation, hedonic quality-identity and attractiveness are covered by this questionnaire. We only asked the questions concerning the attractiveness. As exhibited in the illustration in Figure 7 participants favor the radar system.

Regarding the comprehensibility of the three systems, participants preferred the radar prototype. On a 1 to 6 Likert scale, where one stands for very bad and six for very good, users evaluated the radar system as a 5.2 (n=12). The toggle prototype was judged as a 4.9 followed by the spotlight (4.0) (Figure 8).

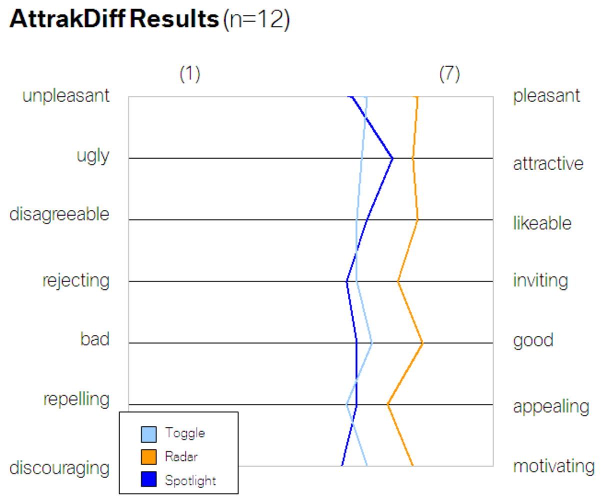


Figure 7: Results of the AttrakDiff questionnaire (n=12, dimension: attractiveness).

The grading of the three variants concerning users' preferences yielded to the following result: radar 2.1, toggle 2.8 and spotlight 2.8. Also the ranking showed that all attendees preferred the POI selection via the radar prototype (Figure 8).

Result Summary (n=12)

	1 Radar	2 Toggle	3 Spotlight
SUS	83,95	73,75	68,3
Comprehensibility (1-6, Likert scale)	5,2	4,9	4,0
School Marks (1-6)	2,1	2,8	2,8
Ranking (1st, 2nd, 3rd)	1,3	2	2,7
Task completion time (sec.)	81	84	115
	+6	0	-4

Figure 8: Summary of subjective and objective user study results.

CONCLUSION AND FUTURE WORK

Our explorative user study showed the tendency that different strategies for the POI selection can influence the user performance as well as the attractiveness of a system. We compared three systems in terms of prototypes. The toggle prototype, which implements the reading order without any visualization, the spotlight prototype with a center based selection order and the radar prototype, which realizes a radial POI selection order combined with an appropriate visualization.

Except the selection in a very high POI density, the shortest task completion time was achieved with the radar prototype. Based on user input, the main advantage of the radar prototype seems to be the predictable selection order and a less disturbing change between the POIs. This could be ascribed to the visualization.

According to the opinion of the participants one reason for the higher completion time within many POIs could be the tail, which is supposed to indicate the turning direction of the iDrive controller. Five people mentioned that this tail irritated them.

Concerning the spotlight prototype eight participants mentioned the pre-selection in the center of spotlight as main advantage of this system.

For the design of future systems a combination of the mentioned advantages could make sense. For example, a pre-selection of POIs via a spotlight combined with a radar-like toggling is a possibility. Another issue for future work will be the validation of these systems in a more realistic driving environment. Therefore a dual task evaluation method will be applied. Either under simulated conditions or a realistic driving study could be conducted.

ACKNOWLEDGMENTS

We thank Nicolas Stadtelmayer for executing the user study.

REFERENCES

1. AUDI Lexikon: MMI. URL: <http://www.audi.de/de/brand/de/tools/advice/glossary/mmi.browser.filter_i_m.html>
2. BMW Techniklexicon: iDrive Controller. URL: <http://www.bmw.com/com/de/insights/technology/technology_guide/articles/controller.html>
3. BMW Techniklexikon: iDrive. URL: <http://www.bmw.com/com/de/insights/technology/technology_guide/articles/idrive.html>
4. Brooke, J. 1996 SUS: a quick and dirty usability scale. In: P. W. Jordan, B. Thomas, B. A. Weerdmeester, A. L. McClelland (eds.). Usability Evaluation in Industry. London, 189–194.
5. User Interface Design GmbH. AttrakDiff™. Ein Service der User Interface Design GmbH. 22.2.2008. URL: <http://www.attrakdiff.de/ >

On Timing and Modality Choice with Local Danger Warnings for Drivers

Yujia Cao
Human Media Interaction
University of Twente
Enschede, The Netherlands
Y.Cao@ewi.utwente.nl

Sandro Castronovo
German Research Center for Artificial Intelligence (DFKI)
Saarbrücken, Germany
Sandro.Castronovo@dfki.de

Angela Mahr
German Research Center for Artificial Intelligence (DFKI)
Saarbrücken, Germany
Angela.Mahr@dfki.de

Christian Müller
German Research Center for Artificial Intelligence (DFKI)
Saarbrücken, Germany
Christian.Mueller@dfki.de

ABSTRACT
We present an experimental study on the effectiveness of five modality variants (speech, text-only, icon-only, two combinations of text and icons) for presenting local danger warnings for drivers. Hereby, we focus on sudden appearing road obstacles within a maximum up-to-date scenario as it is envisaged in Car2Car communication research. The effectiveness is measured by the minimum time necessary for fully interpreting the content. Results show that text-only requires the most time while icon only is perceived the fastest. The two combined versions lie in between. The minimum length for speech is determined by the duration of the utterance, which is longer than perception time of text-only in this case. However, speech could be decoded reliably by nearly all subjects. Results indicate further that a blinking visual cue provided through the periphery visual channel is able to enhance the saliency of visual modalities. Subjective judgements by the subjects furthermore suggest a combined use of visual and auditory modalities.

Categories and Subject Descriptors
H.5.2 [**Information interfaces and presentation**]: User Interfaces, User-centered design

Keywords
automotive, modality choice, timing

1. INTRODUCTION
In-vehicle messages and in particular local danger warnings need to be effective, since the driver has to decode them while being engaged in something else (driving) and because there is typically not excessively much time left to react. The way the information is presented as well as the right timing are therefore crucial factors. One major aspect of the former is choice of the right modality – visual or auditory, which both come with advantages and disadvantages as a number of studies revealed [1, 2, 4, 7]. First of all, there is environmental factors: Visual modalities are superior in delivering information in noisy environments while the performance of auditory modalities is more robust towards variances in lighting conditions. Then, there is consumption of perceptional recourses: Since driving is mainly a visual task, messages that are delivered through the auditory channel can be perceived in parallel with the driving task. In contrast, the perception of visual messages requires to take the eyes off the road. The ability of attracting attention (the level of saliency) has to be taken into consideration as well: Without any additional cue, visual messages are less able to attract attention, especially in a high-load driving condition. When focusing on a busy traffic, drivers might not notice the onset of a new visual message, or they choose to delay attending to it until a moment when they can safely remove their eyes from the road. On the other hand, auditory messages have a preemption effect as to require an immediate perception. When it comes to memory requirements, visual messages allow iterated perception. Auditory information, however, is transient thus might require a repeat function in order to allow recall if it is forgotten.

One might conclude that visual and auditory modalities should complement each other presenting one message. Studies (e.g. [8]) showed that careful combinations outperform the respective inferior modality. However, it has also been found that the result of combining is hardly ever "best of both worlds" [8]. Other studies revealed that people, when occupied by the driving task, tend to only listen to the audio messages and not bother looking at the display on which the same information was presented visually [1, 6]. In addition, a redundant use of both modalities might bear the risk of overloading both perceptual channels or annoying the driver. Therefore, the choice of modality should certainly be done case-by-case. The relevance of the message to the driving task (level of priority) was suggested to be taken into account [2, 8]. When presenting driving-irrelevant messages (such as weather forecast for the coming days), visual modality is more suitable since it is self-paced. The driving task can be better sustained since drivers can take their eyes off the road when the situation allows them to do so. However, if the message to be presented is driving-relevant

AutomotiveUI'09, September 21-22, 2009, Essen, Germany
ACM 978-1-60558-571-0/09/009

(such as warnings of road obstacles), speech should be used in order to allow drivers to obtain a timely awareness of the potential danger ahead.

However, at least one of the drawbacks of visual modality – lack of saliency – might be overcome by means of an additional cue, such as a blinking object located in the peripheral visual field. Peripheral vision is well suited for providing pre-attentive cues because it is sensitive to motion and luminance changes, and it can be picked up in parallel with an on-going foveal vision task [9]. Moreover, there is still the issue of timing. If we are heading towards truly up-to-date warnings, we may not be able to alert the driver well in advance, which would again speak for vision. Imagine a broken vehicle automatically triggering an alert, which is transmitted over ad-hoc car-to-car network and received by a vehicle approaching the place – a scenario that is being investigated in a number of ongoing research projects (e.g. [3]). In a situation like that, we might not have an awful lot of time to generate wordy speech warning messages.

We present an experimental study, which evaluates the choice of modality from a timing perspective. We focus on one aspect of local danger warnings: sudden appearing road obstacles within a maximum up-to-date scenario as described above. The modality choice for presenting this type of messages should assist a fast perception and comprehension of the content of the message. Thereby, the goal of the experiment is to compare the effectiveness of auditory modality (speech in particular) and several enhanced visual modalities. The effectiveness of a modality is measured by the minimum time necessary for fully interpreting the content of the message.

2. EXPERIMENTAL STUDY

2.1 Design/Apparatus

While performing a visual task that required constant attention (simulating driving), subjects were presented with warning messages using different modalities and presentation durations. For each message, the task was to identify whether a repeated version displayed after the offset of the original one was the same or certain details changed. Based on the correctness of the identification performance, we inferred whether or not the presentation duration was sufficient for the subjects to fully perceive the message. The study was performed in a lab room using a PC with a single 20" screen. For the primary task, we chose a "Find the differences" picture puzzle using two versions of a single, very complex photography of a domestic scene. We instructed subjects to perform this task throughout the entire experiment and find a required number of differences. Although the task was interruptible at any time, subjects became very engaged in it because it was very hard to find all differences. The pictures were displayed in a single row on the top left corner of the screen yielding a line of vision that corresponded to looking at the road. Subjects were further instructed to click a button on the screen right below the pictures using the mouse whenever they found a difference. The performance in this task was not analyzed.

Warning messages were displayed on the bottom right corner of the screen respectively played via loudspeakers. A message consisted of three components: 1. type (which kind of

Figure 1: Four types of obstacles were used in the study: break down vehicle, fallen tree, rock, and lost cargo.

of obstacle); 2. location (on which lane respectively shoulder); 3. distance (how far the place is ahead). Each visual warning message was preceded by a visual cue (a blinking color bar of the same width as the presentation area and on top of it) and remained on the screen for a certain number of seconds. Two seconds after the message disappeared, it was repeated on the top right corner of the screen together with a choice of three buttons: "same", "different", "not sure". SAME and DIFFERENT cases occurred with a 1:1 ratio and a random order. In the latter case, either type, location, or distance was changed. Since the time interval between the offset of the original message and the repeated message was only 2 seconds, the identification task did not require a long-term memorization of the message. However, it did require subjects to realize what is on the road, where it is and how far it is.

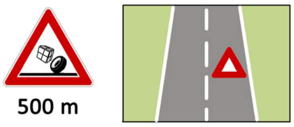
Using a within-subject design, all subjects performed all five presentation styles (see Table 1). The order of the five styles was counterbalanced by a size-5 Latin square. For each visual style, the presentation duration was decreased with a step of 1 second after every three warnings until the subjects started to make errors in identifying the repeated message. We took the minimum presentation duration as a measurement of the effectiveness of this visual presentation style. This includes the time needed to switch the foveal visual attention to the message, perceive the message and understand the meaning. Since in this experiment, subjects always switched their attention immediately when they noticed the blinking motion, our measurement did not include the delay of attentive switch, nor the time needed to prepare an action upon the presented situation. For speech, the length of the utterance was taken as the minimum perception time. Additionally, we surveyed the subjective preferences towards variants of presentation styles.

2.2 Stimuli

Within visual modalities, we further distinguished between textual modality (text, numbers) and graphical modality (icon image), due to the differences in their presentational power. Text is suitable for conveying abstract information, such as the relationships between events. Numbers are suitable for providing precise quantitative understandings of numerical data while images are superior in describing concrete concepts and information of a high specificity nature, such as concrete objects. In general, graphical modalities are more vivid than textual information, thus are likely to receive greater weight during decision making processes. In particular, shapes and colors have great salience to human information processors due to the sharp contrast they are able to create [5].

Four types of obstacles were used: broken vehicle, fallen tree, rock, and lost cargo (see Figure 1). Five modality variants

Table 1: Modality variants used in the experiment.

Variants	example
text only	Lost cargo 500 m right lane
icon only	
mixed 1	
mixed 2	
speech	"Lost cargo in 500 m on the right lane"

were used: speech and four variants of visual presentation (see Table 1). With the visual ones, the distance information was always presented by numbers (e.g. 500 m, 1 km). The obstacle type and the location could be conveyed by either text or icons, resulting in a text only condition, an icon only condition and two mixed conditions (Figure 1). The icons, the wording of text and speech were selected based on a pre-user study with various designs in order to ensure the intuitiveness of the presentation. The textual information (text and speech) were presented in German. Speech was generated using a text-to-speech software.

2.3 Subjects

Ten subjects (2 women and 8 men) voluntarily participated. All of them are German native speakers, between 25 and 45 years old, and working in a technical field (some in speech /dialog-related topics some not). This has to be taken into account when interpreting the results. Interesting correlations could be found especially with subjective preferences.

2.4 Results

2.4.1 Time measurements

The results of the time measurement are illustrated in Figure 2. For visual variants, the center points of the error bars indicate the average of the minimum perception time, and the length of the bars shows the standard error. Text-only required the most time: the average presentation duration which enabled subjects to recall the messages correctly was 3.6 seconds. Icon-only was perceived the fastest. Here, on average only 1.8 seconds were needed to reliably interpret the message and compare it with the subsequent prompt. Not surprisingly, the two mixed version lie in between. However, considering that in both cases only one out of three informational components was replaced by an icon (either type or location), the improvement from 3.6 (text only) to 2.6 for MIXED1 and respectively 2.4 seconds for MIXED2 is remarkable. For the speech condition, the minimum presentation length is determined by the time duration of the utterance, which was 5 seconds on average (Figure 2). The speech messages could be decoded reliably by all subjects, except one who had difficulty to follow the numerical information (distance) in the utterance.

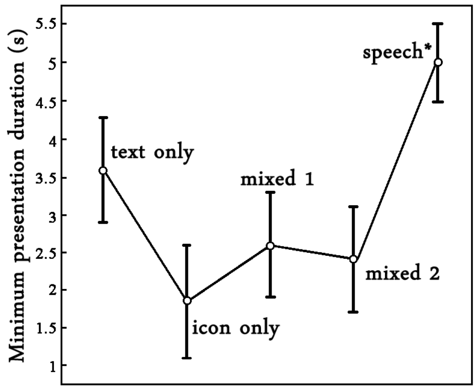


Figure 2: Timing results in seconds for visual modality variants. (*) For speech, the minimum presentation length is determined by the utterance duration.

Table 2: Helmert contrasts between visual variants.

Contrast	Sig.
text vs. icon/mixed1/mixed2	$F(1,9)=30.00$, $p<0.01$
icon vs. mixed1/mixed2	$F(1,9)=8.65$, $p<0.05$
mixed1 vs. mixed2	$F(1,9)<1$, $p>0.05$

Repeated-measure ANOVA further showed a significant modality effect on the minimum presentation duration measurement ($F(3, 27)=11.46$, $p<0.001$), indicating that the usage of modality could significantly influence the amount of time needed to perceive and comprehend the same information content. Helmert contrasts (Table 2) further revealed that text-only required a significantly longer perception time compared to the other three; icon-only allowed a significantly faster perception compared to the rest two, and no significant difference was found between the two mixed conditions. Regarding the time needed to decode the message, these results confirmed that text was the least and icon was the most efficient for presenting both obstacle type and location. Although people read the shape of familiar words rather than every single letter, well-designed icons still allow easier perception than text. Being consistent with previous findings, our result showed the representative power of graphical modalities to present concrete concepts such as the obstacle type and location. It also stands in line with the suggestion from [2] that information of higher priority should become more symbolic.

The periphery visual cue was shown to be able to effectively attracts attention. Under the condition that visual presentation gets attention immediately after onset, the time needed to perceive and comprehend the message was much shorter

Table 3: Subjective voting for the best visual variant and visual vs. auditory comparison.

vision variants	text	icon	mixed 1	mixed 2
number of votes	0	8	2	0
vision vs. speech	speech	vision	combination	
number of votes	4	2	4	

than the duration of speech, especially when visual messages were well-designed. Although visual presentation still requires eyes off the road, they might be considered as a good option when a warning message is presented with a short notice (such as shortly before the obstacle). Certainly, speech is suitable when the message is presented long enough ahead and speech should be kept short and precise. Note that our measurement of perception time does not take into account the time needed to prepare an action upon the presented message. However, this time duration is not modality dependent, which means that the difference in the perception time induced by the usage of modality will still be valid even if the reaction time is taken into account.

2.4.2 Subjective judgements

Table 3 summarizes the subjective judgements of the modality variants. When asked to choose the visual variant that they found the easiest to perceive and understand, 8 out of 10 subjects chose ICON ONLY, which is consistent with the minimum duration measurement. They commented that it was time-consuming to read a lot of text. Besides, when interpreting the message, they usually illustrate the text in their mind which requires additional cognitive effort. The reason of disliking the two mixed designs was mainly that the spatial separation of the three information components required longer perception time. The other 2 subjects preferred MIXED1. In contrast to the majority, they explained that they tend to use sub-vocal speech to encode information components into the short-term memory, especially for the location of an obstacle. Therefore it was much more convenient when the location was presented with text. Interestingly, they are the only ones who daily work with language related topics, such as text retrieval and dialog management. These subjective reports indicate that graphical modalities are generally more effective for presenting concrete concepts. However, this conclusion might be moderated by the professional training background of a subject, which might influence the modality used to encode information in the short-term memory.

When asked to compare visual presentations with speech, 4 subjects preferred speech. They stated that speech is more compatible with the on-going visual searching task. Two subjects preferred visual presentations. They said that the visual prime immediately shifted their attention onto the message. However, when they were engaged in the visual searching task, they had a tendency to ignore the speech even though they heard it. The remaining four subjects preferred to be provided with both visual and auditory messages. They stated that, although they listen to the speech, they prefer to have visual presentation as well in case they need to recall details. Moreover, they could choose to look at the visual presentation while the speech output is still ongoing, which is faster for long utterances.

3. CONCLUSIONS

In this user study, we investigated the effectiveness of five modality variants in presenting local danger warning messages for drivers. Hereby, we focussed on sudden appearing road obstacles within a maximum up-to-date scenario as it is envisaged in Car2Car communication research. The effectiveness was measured by the minimum time necessary for fully interpreting the content. Results show that text-only requires the most time while icon only is perceived the fastest. The two combined versions lie in between. The minimum length for speech is determined by the duration of the utterance, which is longer than the perception time of text-only in this case. However, speech could be decoded reliably by nearly all subjects. Result further indicate that the blinking visual cue provided through the periphery visual channel was able to enhance the saliency of visual modalities, thus made them more suitable to present messages of a high priority. When visual messages were attended immediately, the perception time could be much shorter than the duration of speech (5 seconds) for the same information content. This suggests that visual modalities with prime might have advantages over speech when a warning message needs to be presented on a short notice. Speech, however, is certainly suitable when time is sufficient to present the warning. Based on subjective preferences, it might as well be wise to use both visual and auditory modalities. Moreover, our results suggest that spatial integration of information components can reduce the perception time. Generally speaking, our results confirm earlier studies in that it is effective to present concrete information with graphical modalities.

Acknowledgements. This work was funded by the Dutch Ministry of Economic Affairs (project Interactive Collaborative Information Systems, grant nr: BSIK03024) and the German Ministry of Education and Research (project Car-Oriented Multimodal Interface Architecture, grant number 01IW08004).

4. REFERENCES

[1] N. Bernsen and L. Dybkjaer. Exploring natural interaction in the car. In *CLASS Workshop on Natural Interactivity and Intelligent Interactive Information Representation*, pages 75–79, 2001.

[2] A. G. C. Kaufmann, R. Risser and R. Sefelin. Effects of simultaneous multi-modal warnings and traffic information on driver behaviour. In *European conference on human centered design for intelligent transport systems*, pages 33–42, 2008.

[3] S. Eichler, C. Schroth, T. Kosch, and M. Strassberger. Strategies for context-adaptive message dissemination in vehicular ad hoc networks. In *International Workshop on Vehicle-to-Vehicle Communications (V2VCOM)*. IEEE Computer Society, 2006.

[4] W. Horry and C. Wickens. Driving and side task performance: The effects of display clutter, separation, and modality. *Human factors*, 46(4):611–624, 2004.

[5] N. Lurie and C. Mason. Visual representation: Implications for decision making. *Journal of Marketing*, 71:160–177, 2007.

[6] M. Moldenhauer and D. McCrickard. Effect of information modality on geographic cognition in car navigation systems. In *IFIP TC.13 Conference on Human-Computer Interaction (INTERACT)*, 2003.

[7] N. Sarter. Multimodal information presentation: Design guidance and research challenges. *International journal of industrial ergonomics*, 36(5):439–445, 2006.

[8] B. Seppelt and C. Wickens. In-vehicle tasks: Effects of modality, driving relevance, and redundancy. Technical Report AHFD-03-16/GM-03-2, GM Corp, 2003.

[9] C. Wickens and J. McCarley. *Applied attention theory.* CRC Press, 2008.

Proceedings of the First International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI 2009), Sep 21-22 2009, Essen, Germany

Towards System-Mediated Car Passenger Communication

Angela Mahr, Margarita Pentcheva, Christian Müller
German Research Center for Artificial Intelligence (DFKI)
Building D3$_2$, Campus, Stuhlsatzenhausweg 3
66123 Saarbrücken, Germany
{Firstname.lastname@dfki.de}

ABSTRACT

In this paper, we outline a system that supports the communication between passengers by transmitting speech (and maybe also video) of the communication partners back and forth. A study is presented that addresses the questions: 1. Is listening to noisy speech coming from the backseat really distracting the driver? Subjects are rating the truth of common-sense statements played from the back of the car (clear, noisy) while driving with a drive simulator. NOISY is rated significantly more distracting than CLEAR while objective driving performance only degrades for men but not for women.

Categories and Subject Descriptors

H.5.2 [**Information interfaces and presentation**]: User Interfaces, User-centered design

General Terms

DESIGN,HUMAN FACTORS

Keywords

infotainment, automotive, multi-party, lane change

1. INTRODUCTION

Research on automotive assistance and infotainment systems has traditionally focussed on one person: the driver. However, even with commuting to work, an average of 10.2 % chose to carpool in the US in 2004, as a study by American Community Survey revealed [1]. With leisure time traveling, the number of cars with more than one passenger can be expected to be much higher. Hence, taking passengers into account when designing in-car systems is a reasonable thing to do. In our research, we lay the foundations for a new generation of context-aware multimodal interfaces for car passengers that support the interaction of the passengers with the car, the interaction of the passengers and the road environment as well as the interaction between the passengers, mediated by the system. This study contributes to the latter aspect.

AutomotiveUI'09, September 21-22, 2009, Essen, Germany
ACM 978-1-60558-571-0/09/009

2. DESIGN SKETCH

When taking multiple users into account, the problem of appropriate alignment of multimodal output presentation has to be addressed. To this end, appropriate contents need to be routed to output devices bound to specific seat areas but also adapted to the role of the speaker. Drivers might, for example, decide that their children should just have access to some specific areas of the infotainment system like for example for watching DVD or playing games. Thus, a multimodal dialog platform must support role-specific access to the underlying services and appropriate configuration possibilities. Due to the requirement of minimal distraction, the way of presenting information to the driver is especially crucial. Hence, we propose mainly auditive output: The driver interacts mainly using speech and auditive output; the co-driver has an additional small screen that can – under certain conditions – be shared with the driver. The passengers on the backseat have access to full blown entertainment touch screens. For technical details regarding the microphone technique one might have a look at the DVE system recently available in the VW Multivan [5].

Thinking on how to support interaction (communication) between passengers, the most immediate problem appears to be that conversations between people in the front and people in the back are difficult: due to driving noise, the acoustic characteristics of the car as well as the fact that everyone faces one direction, it is hard to understand what is being said. One straightforward way to support the communication with such a setup would be to transmit speech (and maybe also video) of the communication partners back and forth. With the right equipment, this could function in the same way as noise canceling headphones, which cancel the ambient noise and thereby enhance speech. The study presented here addresses the question: Is listening to noisy speech coming from the backseat really distracting the driver? And would this conversational task during driving affect men and women to the same degree?

3. EXPERIMENT

We recorded 63 yes/no common-sense statements that were either true or false like for example "In Norway a scarf and gloves are useful winter clothings", "At the beach you should always take care of snowslides". Out of these sentences two lists with equal proportions of true and false sentences were made. Half of the recordings were overlaid with car noise (NOISY), the other half was not preprocessed (CLEAR). Between subjects the choice of list that was to be presented clearly was balanced. The sentences were played from the

Table 1: Subjective rating by the subjects how demanding the task was.

	CLEAR	NOISY	sign. diff
hard to listen	1.88	2.75	p < 0.01
distracted me from driving	2.75	3.83	p < 0.01
compromised my driving	2.88	3.63	p < 0.05

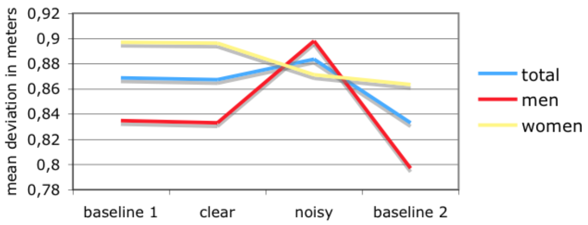


Figure 1: Objective measure of distraction using lane change task.

backseat while the subject was driving (with a drive simulation software) in order to assess the level of distraction. The subjects were instructed to say whether the statement is true or false. In addition to the objective measure, we asked the subject to rate the subjective distraction after each condition. Prior to the experiment, a motivating example was provided to the subjects: "Imagine that you, as a driver, have a conversation with the passengers in the back seat. It is loud in your car (vehicle/motor or ambient noises, radio) and you have to make great efforts to hear your partner." After the experiment, we asked the subjects whether they would appreciate a system like that. We measured the driver distraction using the standardized "lane change task" (LCT) [3], a simple laboratory dual-task method that is intended to estimate driver distraction (ISO Draft International Standard 26022). We used the following statement for the subjective rating of distraction: 1. The communication task distracted me from driving; 2. The communication task compromised my driving performance; 3. I found it hard to listen to the presented questions. The possible scale of answers was from 1 ("I do not agree") to 5 ("I fully agree").

24 subjects (11 men and 13 women), were paid to participate in a user study. The age range was between 21 and 60 with an average age of 35.9 years for men and 33.2 years for women. The entire experiment took about one hour to complete. However, a significant part of that time (appr. 30 min) was designated to a different study (not presented here). The recordings were played from the back by another experimenter. After warmup and first baseline, the main part of the experiment started. The order of the (CLEAR and NOISY) conditions was balanced between subjects and gender. A second baseline was measured afterwards, followed by the other part of the experiment (different study).

Table 1 shows the results of the subjective rating. NOISY was rated demanding, distracting from driving, and compromising driving performance. The differences between CLEAR and NOISY were statistically significant for each of the questions ($p < 0.01$ resp. $p < 0.05$). Figure 1 shows the mean deviation in meters between a normative model and the actual driving in LCT. A repeated measures ANOVA was carried out with the relevant covariates age (significant correlation with driving performance in the baselines, $r = .4$, $p < .05$) and order of the experimental conditions (due to expected learning effects): The main effect for condition was not significant, $F(3,60) = 1.16$, *ns.* as well as the main effect for gender, $F(1,20) = 1.61$, *ns.* As expected the main effect of age was significant $F(1,20) = 15.32$, $p < .01$. The interaction between gender and condition was not significant over all conditions either $F(3,60) = 1.83$, *ns.* But as indicated earlier the central questions are, whether drivers were distracted by the CLEAR or the NOISY condition. In order to test this, orthogonal contrasts were conducted. The first comparison of the two baselines was not significant $F(1,20) < 1$, *ns.* Then the NOISY condition was contrasted with the CLEAR condition and we found no significant difference $F(1,20) = 2.14$, *ns.* As a last contrast we compared both baselines with both speech conditions, but didn't find a significant difference either $F(1,20) = 1.98$, *ns.*

A probable explanation for not having found the expected and subjectively rated stronger interference of the NOISY condition with the driving task than for the CLEAR condition might be the lack of sensitivity of the lane change task. One more suggestion we have had was that there might be an interaction of condition and gender concerning the NOISY condition, so we conducted a contrast for the interaction comparing the NOISY condition with the CLEAR condition. We found a significant interaction for this contrast, $F(1,20) = 5.59$, $p < .05$. This made us have a closer look at each gender separately and the same contrasts were repeated for each of the groups. For men the comparison of the NOISY condition with all other conditions was significant $F(1,10) = 5.21$, $p < .05$, but for women it was not $F(1,12) < 1$, ns. This means, that performance degrades in the NOISY condition only for men but not for women. The analysis of the subjective rating regarding the gender aspect revealed no significant interaction, $F(1,22) = 2.35$, $p = .14$, but a similar pattern.

4. CONCLUSIONS

We outlined a system that supports the communication between passengers by transmitting speech (and maybe also video) of the communication partners back and forth. We reported results of a study indicating that listening to noisy speech from the back is distracting drivers (subjective rating), especially men (subjective rating and drive simulator).

5. REFERENCES

[1] American community survey 2004: Percent of workers who traveled to work by carpool. http://www.statemaster.com/. 09/05/27.

[2] M. Burmester, R. Graf, J. Hellbrück, and M. Ansgar. Usability — Der Mensch im Fahrzeug. In *Infotainmentsysteme im Kraftfahrzeug*. Springer, 2008.

[3] S. Mattes. The lane change task as a tool for driver distraction evaluation. In H. Strasser, H. Rausch, and H. Bubb, editors, *Quality of Work and Products in Enterprises of the Future*. Ergonomia, 2003.

[4] C. Müller-Bagehl and P. Endt, editors. *Infotainment/Telematik im Fahrzeug – Trends für die Serienentwicklung*. Haus der Technik Fachbuch 38. Expert, Renningen, 2004.

[5] S. Steenhagen. The first automotive implementation of a digital voice enhancement system, 2005. (09/07/31) http://www.appliedsignalprocessing.com/dve2.htm.

Acceptance of Future Persuasive In-Car Interfaces Towards a More Economic Driving Behaviour

Alexander Meschtscherjakov, David Wilfinger, Thomas Scherndl, Manfred Tscheligi
ICT&S Center, University of Salzburg
Sigmund Haffner Gasse 18, 5020 Salzburg, Austria
{alexander.meschtscherjakov, david.wilfinger, thomas.scherndl, manfred.tscheligi}@sbg.ac.at

ABSTRACT
Measuring user acceptance to avoid system rejection by the users in pre-prototype stage of product development is of high interest for both researchers and practitioners. This is especially true when technology uses strategies of persuasion in an emotional laden environment like the car. This paper presents the results of an online survey aiming at evaluating the acceptance of future persuasive in-car interaction approaches for a more economic driving behaviour. Five different persuasive interface concepts are presented and studied towards their acceptance. The results show an overall acceptance of the system concepts and the usefulness of the presented method. We show that individual expectations of the systems' disturbance and risk have an effect on the acceptance of technology and the behavioural intention to use.

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *evaluation/methodology*

General Terms
Design, Human Factors.

Keywords
In-car interfaces, economy, technology acceptance, persuasion.

1. INTRODUCTION
Similar to every aspect of our lives, the pervasion of the automotive context with advanced technologies is increasing. Having in mind that every year a diversity of technologies is developed and reaches customers and potential users, it is highly interesting for both researchers and practitioners, whether these technologies will be accepted by the target group or not. A lack of acceptance will lead to a rejection of the system by new users and to a strong dislike by existing users of comparable technology, who have a high intention on using newer developments. The risk of rejecting systems, which aim at changing attitudes or behaviour, is even higher in emotion-laden contexts like the car, because driving assistance is easily mistaken as a critique on driving behaviour. Therefore it is of high importance to find issues in system design that decrease the acceptance of novel systems in a very early stage of development.

Energy efficiency in the automotive context is currently in the focus of research and industry activities. Besides technological innovations, the drivers' behaviour is a potential area of improvement. This can be fostered by novel interface concepts, which require the users to accept the new technology in order to make it successful. This paper describes how a scenario-based online questionnaire can be used to evaluate user acceptance of persuasive in-car interfaces in a pre-prototype phase of development.

ACM 978-1-60558-571-0/09/009

2. RELATED WORK
In the next sections an overview on academic as well as industrial research for novel in-car interfaces, persuasive technology and technology acceptance is given.

2.1 In-car interfaces
Research on in-car interfaces has recently gained higher attention in the area of human-computer interaction (HCI) [17]. Novel technologies creating attractive in-car user interfaces have become a great challenge [21]. Ablassmeier et al. [1] state that the growing amount of information in cars makes the development of new strategies to cope with this amount of information for drivers necessary. This is when new interaction technologies (e.g. speech interfaces, olfactory interfaces) can provide new possibilities to handle the complexity of the system. For that purpose, several researchers focus on the integration of multimodal interaction in the car. Siewiorek et al. [22], for example, introduce a companion contextual car driver interface that assists the driver.

Research focusing on in-car interface solutions includes different ways of input and output systems that enable new concepts of driver assistance (for example, recent presented concepts were search based interfaces [9], handwritten input [15] and augmented windshield displays [16]). Since the automotive context is already penetrated by a multitude of interaction systems, one way of assisting the driver is through the augmentation of existing in-car interfaces as suggested by Varhelyi et al. [26]. Their active acceleration pedal increases its resistance when the driver exceeds the speed limit. Another interesting approach is the dynamic speedometer, which integrates current speed limits seamlessly into the dashboard [18].

Research activities on supporting the early stages of in-car ICT development is for instance the CARS (Configurable Automotive Research Simulator) project [14]. It enables the application of in-car system prototypes in a driving simulator. The presented work uses a different approach and addresses the study of in-car interfaces in a pre-prototype and conceptual phase.

2.2 Persuasion

The topic of persuasion has been applied to several areas in the realm of HCI. Fogg [7] coined the term Captology ("computers as persuasive technologies") to describe the area where computing technology (web sites, mobile phones, smart environment, virtual reality, etc.) and persuasion (behaviour change, attitude change, motivation, change in worldview, compliance) overlap. Especially Persuasive Technologies, which are defined by Fogg as *any interactive computing system designed to change people's attitudes or behaviours*, seems to be a promising approach to be researched in automotive environments, when it comes to fostering a more fuel-efficient driving behaviour.

Lately, there have been several approaches from car manufacturers to include persuasive technologies into cars to facilitate a voluntary change of behaviour, attitude, or both to reduce greenhouse gas emissions. A good overview on recent work on approaches to help people become more fuel-efficient drivers can be found at [10]. Persuasive in-car technologies in the academic realm have been researched by Tester et al. [23] and Pace et al. [20], among others.

2.3 Technology Acceptance

User acceptance (UA) can be defined as *the demonstrable willingness within a user group to employ information technology for the tasks it is designed to support* [5]. UA is vital for the success of any information technology system. This leads to the need of informing ICT development as early as possible with data about the UA of the system under development.

There are a high number of research efforts concerning UA [5]. The widest used model for describing UA is the Technology Acceptance Model (TAM) introduced by Davis [4]. It is based on the prediction of user acceptance prior to a real system usage and was further developed in order to enable the study of technology acceptance in a pre-prototype phase of system development [3]. Venkatesh et al. [27] present another approach of modelling UA by introducing the UTAUT model using eight scales for assessing UA such as performance expectancy and social influence.

An early assessment of UA in the development of ICTs can give advantages on both cost and effort [3]. Davis and Venkatesh propose to use the TAM questionnaire as early in the product development as possible. Based on the original TAM they use three scales to assess UA of systems in a pre-prototype state: Perceived Usefulness (U), Perceived Ease of Use (EOU) and Behavioural Intention of Use (BI). Perceived usefulness (U) describes *the extent to which the individual believes that using a system will enhance his/her job performance.* Perceived ease of use (EOU) is *the extent to which an individual believes using a system will be free of effort*. In the original TAM Intention of Use (BI) is a function of U and EOU. In the pre-prototype TAM model BI is covered by additional items. Davis and Venkatesh argue this by the change of the direct influence of EOU on BI, which becomes non-significant after users gained hands on experience which would have lead to U being the only factor to compute BI with. They showed that the TAM is a reliable and valid tool to predict actual usage behaviour.

Technology acceptance in the vehicle was already addressed by Comte et al. [2] studying driver acceptance of automatic speed limiters. Main goal of their work was to evaluate acceptability and if drivers perceive speed limiters to be effective in reducing accidents. Acceptance of advanced traveller information systems (ATIS) was researched by Wochinger and Boehm-Davis [28] letting users rate qualities of the system based on their own needs. Kantowitz et al. [13] conducted acceptance research in order to inform developers how ATIS had to be designed to fulfil their purpose without causing bad experiences that might keep people from using them. All those approaches show the need for an easy and fast assessment of UA towards technology in the car.

Acknowledging the fact that the spread of ICT with persuasive elements in vehicles will increase in the future, the development of those systems will have to take UA into account as the systems might interfere with the drivers wish not to be controlled [2]. While this applies to all kinds of new technologies, it is of special importance in the car environment, which is safety critical and traditionally laden with emotions. A rejection of a system might have a wide range of effects from an image loss of the brand to a severe loss of security when frustrated drivers drive less safe. Applying strategies of persuasion in the car is highly critical as the car is often a very emotional object for the owner. This can increase the effect of persuasion strategies on user experience factors (e.g. fun/enjoyment, comfort, trust) in both positive and negative ways and influence drivers performance [12]. Therefore systems that aim at in-car persuasion have to be designed with an evaluation of design decisions as early as possible to reduce the risk of negative effects.

3. RESEARCH GOALS

Given the above described state-of-the-art, appropriate methods to research future persuasive in-car interfaces in a pre-prototype stage are missing. We intend to fold this gap by our approach. We therefore present the following research goals:

Research Goal 1 (RG1)
The first research goal was to evaluate the user acceptance of persuasive in-car interfaces that are designed to support a fuel-efficient driving style. Based on iterative design in the user centred design cycle, it is valuable for the design of interactive systems to study users' reaction and interaction as early as possible in the development phase. That ideally happens also in the moment of concept creation when a minimal effort was invested in implementing the concepts into prototypes. The presented interface concepts are all in this state.

Research Goal 2 (RG2)
Based on the measured acceptance of the presented interface concepts the aim of the second research goal was to analyse if other factors have influence on the user acceptance. Specifically, it was researched if there are identifiable influences of driver properties and expectations toward the technology concerning safety, disturbance and assistance. Furthermore, we investigated the effect of driver's general attitude toward technology in cars and sociodemographic variables (age, gender, frequency of driving a car) on the acceptance rating. This kind of information could give additional insights for the development process and the design of system properties for special target groups.

Research Goal 3 (RG3)
Finally, the third research goal of the presented work was to find out if the usage of an online TAM questionnaire for persuasive in-car interfaces in a pre-prototype level would lead to results that support future design decisions.

4. METHODOLOGY

Based on the results Davis and Venkatesh [3] it was decided to use this questionnaire for the assessment of persuasive user interfaces for the vehicle in a pre-prototype state. Five interfaces that were designed to support an ecologic driving style were derived from literature. For each system a maximum 100-word description and an image that illustrated the system (see chapter 5 for details) were made. The descriptions were formulated neutrally and the illustrations used the same graphical style. To improve comprehensibility both text and graphics were presented to 3 fellow researchers and based on their suggestions reformulated and redesigned.

For the assessment of the persuasive interface user acceptance a questionnaire was developed that included the following items: Based on the TAM by Davis and Venkatesh two questions asked for Behavioural Intention of Use (BI). The scales for Perceived Usefulness (U) and Perceived Ease of Use (EOU) consisted of four questions each (e.g. *Assuming I had access to ..., I intend to use it.)*. The TAM questionnaire proved to be very suitable for assessing the user acceptance, but does not provide detailed information on the reasons why a system is rated high or low. Therefore three additional questions were added that were computed separately from the TAM analysis and were expected to give further insights on the perception of the systems. One question asked for an expected disturbance (Disturbance) by the system. The second additional question addressed perceived security risks (Risk) caused by a system usage while driving. Question three was pointed towards the suitability of the system to serve its purpose, namely to support drivers in a more ecological driving behaviour (Suitability).

The five presented systems were randomized in their order of presentation to avoid biases. Additionally, the questions in the questionnaire were counterbalanced for every system to avoid artefacts caused by the question order. While the same questions were asked for each of the five systems, participants also filled out questions before they were confronted with the systems. These questions asked for gender, age, car usage frequency (driving) and the duration of the driving license ownership. Nine questions were asked for the general attitude towards new technology (Attitude). Two final questions were asked for a ranking of the systems from 1-5 and how easy it was to imagine the use of the described systems. The questionnaire was distributed in the form of an online questionnaire and communicated over various mailing lists in order to reach an audience as broad as possible.

5. PERSUASIVE INTERFACES

To evaluate the acceptance of future in-car interaction approaches for the purpose of persuading car drivers to drive in a more economic way, we decided to research already existing - but not yet deployed - approaches. This gave us the possibility to focus on studying UA rather than the design of new solutions. We therefore have accomplished an extended literature research. We aimed at identifying different approaches from academic as well as commercial sources, which are designed to support more fuel-efficient driving. The systems were chosen based on two preconditions. First, the systems had to be in a pre-prototype stage. Second, the systems had to be understandable and imaginable by users who took part in the study.

After identifying several approaches from various sources, we decided to extract five different designs all fostering fuel-efficient driving. Each system presented in the following sections is based on already existing ideas but was redesigned by the authors, resulting in the fact that each system combines different properties of earlier identified approaches. The five systems were described in an online survey including the visualizations presented in this paper. Although most of the identified systems from the literature are illustrated with high-fidelity graphics, we decided to create our own graphical representations of the system to be more consistent in the representation of the different approaches. The following subsections characterise these five approaches. For each system, we first provide a description including a figure. Both the text and the figure resembles the information, which was given in the online questionnaire. Secondly, we describe systems from industry, which motivated our approaches. Thirdly, we discuss persuasive aspects of each system.

5.1 Automatic Eco System (EcoMatic)

Description

The Automatic Eco System (EcoMatic) is a fully automated in-car appliance supporting fuel-efficient driving. The system can be manually activated and deactivated by pressing an EcoButton (see Figure 1). When the system is operating, it automatically reduces fuel consumption by adjusting various parameters within the car. As an example, the automatic climate control is switched into an economy mode and the engine is automatically changed into an idle mode at red traffic lights. At the end of each trip the amount of saved fuel in comparison to the standard fuel consumption is shown (see Figure 1).

Figure 1: EcoMatic. Eco Button (right) and feedback display (left). English translation: "Fuel saved: 6.2 litres"

Motivation

This system is motivated by Honda's ECON Mode button, which is part of Honda's Ecological Drive Assist System [11]. The ECON Mode automatically achieves energy-saving control of the air-conditioning unit and extends the idle stop time. Contrary to our system, it does not provide any information about the amount of saved fuel during the last trip. Toyota plans to equip their new hybrid Prius with three driving modes: ECO (low acceleration power), EV (medium acceleration power, electric-only), and POWER (high acceleration power) [25]. These modes have influence on the acceleration pedal sensitivity.

Discussion

Since the automatic Eco System is operating autonomous, without any user input (except switching it on an off), it is not designed to change the operators driving behaviour. The persuasive element in this system is the displayed information at the end of each trip. This information aims at persuading the driver to use the EcoMatic by showing the concrete benefits of the system. The feedback relevant for economical driving is given once after a trip. One property with this system worth considering is that it might affect the user's comfort level, e.g. by switching the air-conditioning in an economy mode.

5.2 Eco Accelerator Pedal (EcoPedal)

Description

The Eco Accelerator Pedal (EcoPedal) is similar to a traditional acceleration pedal with one distinction: It aims at reducing fuel consumption by simply pushing back against the driver's foot when it detects wasteful acceleration (see Figure 2). This means that the driver feels an increased pressure against his foot, when - for instance - he wants to push down the EcoPedal to its limit. Nevertheless, it is always possible for the driver to push down the pedal as far as he wants, even if the fuel consumption is wasteful.

Figure 2: EcoPedal. Dotted red line: border of fuel economy deterioration. Left indicates an economical; right a wasteful acceleration with increased pedal pressure.

Motivation

A similar system has been introduced by Nissan [19]. Their eco-pedal system activates a counter pushback control mechanism if the system detects excess pressure, each time the driver steps on the accelerator. The optimum acceleration rate is calculated using data on the rate of fuel consumption and transmission efficiency during acceleration and cruising.

Discussion

The EcoPedal aims at changing driver's behaviour by providing feedback at the right moment (when the driver is going to accelerate the car at cost of high fuel consumption) at the right place (within the car at the acceleration pedal itself). Since both the right moment and the right place are crucial elements of persuasive technologies [7], the EcoPedal seems to be a promising approach to assist drivers to become more fuel-efficient. On the other side, one has to keep in mind that the acceleration pedal is one of the most important interfaces for the driver's primary task.

5.3 Eco Speedometer (EcoSpeedometer)

Description

The EcoSpeedometer provides real-time fuel-efficient driving guidance. It is a display seamlessly integrated into the traditional speedometer providing visual feedback whether the driver is driving fuel-efficient at the moment or if the current driving style is wasteful. When driving fuel-efficient, the EcoSpeedometer glows green, when driving inefficiently, it glows orange (see Figure 3).

Figure 3: EcoSpeedometer. Left (green) indicates an economical; right (orange) a wasteful driving behaviour.

Motivation

Real-time visual feedback systems on the momentary driving behaviour have been proposed by several car manufacturers. Honda's Ambient Meter [11] is the background on the speedometer, which notifies the driver of the current driving conditions using colour (green for high fuel-efficient driving, blue-green for moderate fuel-efficiency and blue for wasteful driving). The Nissan eco-driving indicator [19] supports the above-described eco-pedal system. Incorporated on the instrument panel it glows green when the driver is driving within the optimal fuel consumption range. Toyota's Hybrid System Indicator [25] display also indicates whether the driving style is within an economical range.

Discussion

The EcoSpeedometer resembles the EcoPedal in the way that it aims at changing the driver's momentary driving behaviour by providing instant feedback. Contrary to the EcoPedal, the feedback of the EcoSpeedometer is persistent. It therefore provides positive feedback (green light). Additionally, we assume it to be not as distracting for the driver's main task as the EcoPedal. On the other hand, it might not be intuitively clear for the driver how to perform a more fuel-efficient driving style since the feedback is rather abstract.

5.4 Eco Display (EcoDisplay)

Description

The EcoDisplay visualizes the fuel-efficiency accumulative for the current trip by displaying a set of green leaves. The more fuel-efficient the driving is, the more green leaves are shown (see Figure 4). When the driving habits become wasteful again, leaves begin to vanish. For each trip, an EcoScore is calculated and displayed at the end of each trip together with information on mileage and average fuel consumption. Additionally, a ranking of comparable trips is shown (see Figure 4). This EcoScore gives the driver the possibility to compare different trips or to compete against other drivers.

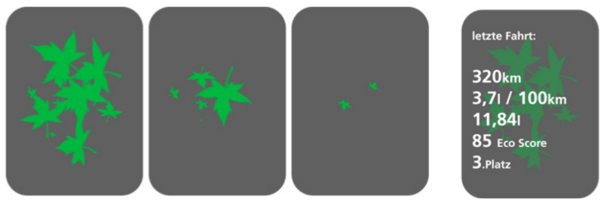


Figure 4: EcoDisplay. From left to right: high, medium, and low economic driving style, feedback display. English translation "Last trip: 320 km, 3.7 l/100 km, 11.84 litres, 85 Eco Score, 3rd rank."

Motivation

A similar system was introduced by Ford with their SmartGauge with EcoGuide dashboard [8]. Long-term fuel efficiency is displayed by "growing leaves" and vines. The growing amount of leaves creates a visual reward for the driver's efforts to drive more fuel-efficient. When the car is turned off, summary information from the just-completed trip, as well as long-term comparative data, is displayed. Honda's Ecological Drive Assist System [11] includes the Eco Guide, which shows growing leaves in three stages as driving practices become more fuel-efficient over time. At the end of each trip, an Eco Score shows the drive cycle results, as well as "lifetime results", represented as leaves on the Eco Guide. The Multi-Information Display allows drivers to view fuel economy figures for the past three trips, as well as instantaneous and average fuel economy statistics.

Discussion

The EcoDisplay represents the most playful approach of the five presented systems. The visualization using leaves as a reward for fuel-efficient driving serves as a symbol for a greener earth. The possibility to reach high-scores and to compare the individual driving performance on an eco-scale with other drivers wants to turn driving into a green game. Contrary to the EcoSpeedometer, which also provides a visual feedback on driving behaviour, the EcoDisplay does not immediately react to wasteful driving behaviour but shows fuel efficiency over a certain period over time – in this case the duration of the last trip. It therefore aims at a long-term behaviour change by providing information over a longer driving cycle.

5.5 Eco Advisor (EcoAdvisor)

Description

The EcoAdvisor analysis driving behaviour along with car specific status information and presents hints to foster a more fuel-efficient driving. The hints are presented verbally before a trip or during driving in appropriate moments utilizing the in-car entertainment system (see Figure 5). Hints regarding the status information of the car are, for example, to increase the tire pressure to a certain amount or to remove unnecessary weight from the car. Hints to improve fuel-efficient driving are e.g. to switch into a higher gear or not to drive at full throttle.

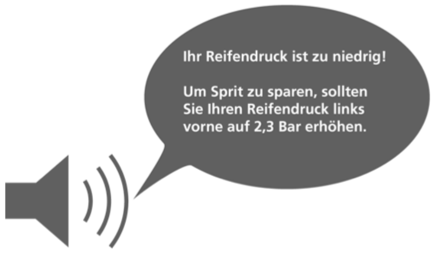


Figure 5: EcoAdvisor. English translation: "Your tyre pressure is too low! To save fuel you should increase the air pressure of your front left tyre to 2.3 bar."

Motivation

The idea of the EcoAdvisor was encouraged by Fiat's eco:Drive system [6], which is a computer application with the aim of improving fuel-efficient driving. While the car is being driven, data pertaining to the vehicle's fuel consumption, exhaust emissions, and the driver's acceleration, braking and shifting patterns are recorded on a flash drive. Later, the data is downloaded to a PC equipped with the eco:Drive software, which can suggest specific changes to driving behaviour [10].

Discussion

The EcoAdvisor is the only auditory system presented here. It gives feedback in form of concrete improvement suggestions at appropriate moments. From a persuasive perspective, concrete improvements and the right timing of giving hints are desirable. Using the auditory channel has the advantage of avoiding visual cluttering, but does not allow the user to actively choose the moment of information gathering.

5.6 Persuasive System Attributes

As stated above all five systems have in common that they are aiming at persuading driver's to change their driving behaviour to drive more fuel-efficient. Besides this, we chose interfaces with different persuasive attributes.

The most obvious difference lies in the sensory channel addressed by the systems. The EcoPedal gives a tactile feedback, whereas the EcoAdvisor an auditory one. The EcoDisplay, the EcoSpeedometer and the EcoMatic, by means of the feedback display at the end of each trip, address the visual channel. Another difference can be found when looking at how the systems are integrated into the dashboard. Three systems make use of already existing in-car interfaces (EcoPedal: acceleration pedal, EcoSpeedometer: speedometer, EcoAdvisor: sound system), whereas the EcoMatic and the EcoDisplay create new interfaces.

One of the most important characteristics of persuasive strategies is the intervention at the opportune moment, which is also referred to as kairos [7]. A persuasive intervention at this moment increases the likelihood of a successful outcome, resulting in the desired behavioural change. Regarding kairos, three different kinds of systems can be found. Two systems (EcoPedal, EcoAdvisor) give feedback on the driving behaviour in appropriate moments (at the exact moment when the driver is going to drive wasteful), two systems (EcoSpeedometer, EcoAdvisor) give constant feedback, whereas the EcoMatic gives feedback only once after each trip.

Another distinction can be seen in the style of feedback and whether it aims at changing the momentary or long time driving behaviour. The EcoAdvisor and the EcoPedal provide concrete suggestion on how to drive more fuel-efficient, whereas the other three approaches (EcoMatic, EcoSpeedometer, EcoDisplay) offer only generic information. The EcoPedal, the EcoSpeedometer, and the EcoAdvisor give feedback on momentary driving behaviour, whereas the EcoMatic and the EcoDisplay provide feedback over a longer period of time.

6. RESULTS

57 participants (31 female, 26 male) took part in the study. They were recruited through email invitation and filled out the online survey. The questionnaire was created using LimeSurvey (www.limesurvey.org), it was online for 2 weeks and it took approximately 10 minutes to answer the questions. All of our participants owned a driver license (duration range: 1 to 40 years), the average age of the participants was 30.04 years (SD = 9.51) with a range from 19 to 58 years. Half of our participants stated that they use a car at least several times a week (57.2%), 21.4% use a car at least several times a month and 21.4% use a car less frequent.

At first, we computed the TAM scales Behavioural Intention of Use (BI), Perceived Usefulness (U) and Perceived Ease of Use (EOU) for each system and checked their reliability. To measure the internal consistency of the scales, we computed the Cronbach Alphas for each scale. They ranged for BI from .883 to .940, for U from .880 to .939 and for EOU from .841 to .941, therefore indicating a generally high internal consistency of the individual scales.

In terms of research goal 1 (RG1), we conducted a repeated-measure ANOVA with system as within-subject factor and the Behavioural Intention of Use as dependent variable in order to assess the significance of possible differences. The ANOVA showed a significant main effect for the within-subject factor ($F(3.30,152.203) = 20.061$, $p<.001$, $\eta^2 = .304$). To show between which systems differences emerge, a post-hoc-test (Bonferonni) was conducted. It indicated that the EcoPedal was rated significantly ($p<.01$) lower than the EcoMatic, EcoSpeedometer and the EcoDisplay, but not lower than the EcoAdvisor. The EcoSpeedometer was significantly ($p<.01$) higher rated than the EcoDisplay, EcoAdvisor and EcoPedal. The EcoMatic did not differ from the EcoSpeedometer but was rated significantly (p<.01) higher than the other systems.

Similar findings emerged for the scale Perceived Usefulness i.e. there was a significant difference between the systems ($F(4,180) = 12.206$, $p<.01$, $\eta^2 = .213$). Firstly, the EcoPedal was rated lowest again in comparison to all the other systems and the difference was significant (p<.01). Secondly, the EcoMatic and the EcoSpeedometer were perceived significantly more useful than the other systems (p<.01).

At last, we compared the Perceived Ease of Use score across the different systems and found again a significant main effect ($F(4,176) = 15.427$, $p<.01$, $\eta^2 = .260$). In this case, the EcoSpeedometer was perceived as the easiest system to use (p<.01), whereas no difference between the other systems emerged. For an illustration of the findings see Figure 6.

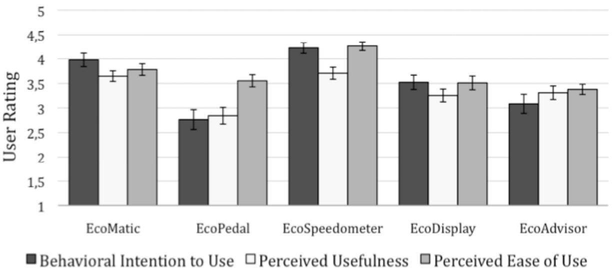


Figure 6: Mean Scores of Behavioural Intention of Use (BI), Perceived Usefulness (U) and Perceived Ease of Use (EOU) for the different systems (including error bars)

Analyzing which individual factors influenced the participants' evaluations, we concentrated on the best (EcoSpeedometer) and worst (EcoPedal) rated systems. Not surprisingly the rating of all three factors (Risk, Disturbance and Suitability) were highly similar to the TAM rating, which seems to be a cue for face validity. The EcoSpeedometer was considered both as least fraught of risk, least disturbing, and most suitable for assisting economic driving. The EcoPedal was rated as relative high in risk and disturbance. The Suitability factor was rated above average but nevertheless significantly lower than the EcoSpeedometer (see also Figure 7).

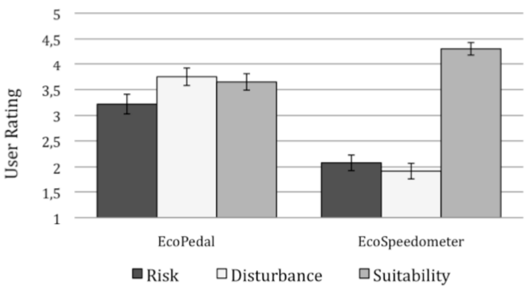


Figure 7: Mean Scores of Risk, Disturbance and Suitability while system usage for EcoPedal and EcoSpeedometer

For both systems we conducted linear regressions to predict the individual TAM scores. As possible predictors, we included age, gender, frequency of driving a car, the scale concerning general attitude toward car technology and questions toward the possible disturbance (Disturbance), risk for safety (Risk) and perceived suitability of the system (Suitability). All of these predictors were entered in the regression equation using the stepwise method.

For the prediction of the three TAM scales of the EcoPedal, Disturbance and Suitability emerged as best and only predictors explaining 68.4% of the variance. While Disturbance had a strong negative influence ($\beta=-.576$), the Suitability had a moderate positive effect ($\beta=.319$) meaning that the indicated Intention to Use (BI) was higher when the user did not perceive the system as disturbing and the user expected the system to be suitable supporting economic driving. A very similar pattern emerged for the Perceived Usefulness scale (Suitability: $\beta=.490$; Disturbance: $\beta=-.444$) explaining 73.8% of the variance. However, there was no significant effect of Disturbance on Perceived Ease of Use, but again an effect of Suitability emerged ($\beta = .581$).

For the EcoSpeedometer, the predictors Suitability ($\beta=.762$) and general attitude towards new technology (Attitude) ($\beta=.256$) emerged as the best predictors for BI. For U, we found Suitability ($\beta=.644$) and Attitude ($\beta=.278$) to be predictive. EOU was predicted solely through Attitude ($\beta=.330$). The effects of the other postulated predictors did not reach significance and were generally low ($|\beta| < .17$). The results of the linear regression are also summed up in more detail in Table 1.

Table 1. Predictors for the factors Behavioural Intention to use (BI), Perceived Usefulness (U) and Perceived Ease of Use (EOU) of the EcoPedal (* p<.05; ** p<.01)

EcoPedal			
Factor	**Predictor**	**β**	**t**
BI (R^2 = .684)	Disturbance	-.576	-4.888**
	Suitability	.319	2.712**
U (R^2 = .738)	Suitability	.490	4.568**
	Disturbance	-.444	-4.143**
EOU (R^2 = .338)	Suitability	.581	4.845**
EcoSpeedometer			
BI (R^2 = .672)	Suitability	.762	9.082**
	Attitude	.256	3.055**
U (R^2 = .507)	Suitability	.644	6.255**
	Attitude	.278	2.700**
EOU (R^2 = .09)	Attitude	.330	2.373*

Concerning our third research goal (RG3), whether an online survey is an apt methodology for evaluating persuasive in-car technologies, we would argue that our approach seems to be promising (see **Fehler! Verweisquelle konnte nicht gefunden werden.**).

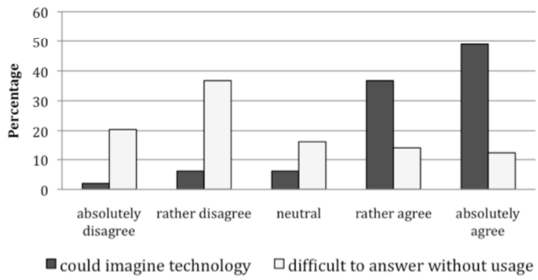


Figure 8: Frequency distribution of answers to the questions "I could imagine the presented technologies because of the description and pictures" and "It was difficult to answer the questions without actually using the technology"

The major part of the participants (85.7%) indicated that they could imagine the technology after reading the description and seeing the pictures. The opinion toward the question whether it was difficult to answer the questions without actual usage of the technology was more heterogeneous: although 57.1% disagreed with this statement, 22.8% answered affirmative to this question.

As mentioned in chapter 4 we asked participants at the end of the survey to rank the systems from 1 to 5 (see Figure 9). The figure shows how often each system was assigned to rank 1-5. Comparing the results of the TAM questionnaire and the participants' ranking of the different systems, it can be concluded that parallels between these two measures emerge: the EcoSpeedometer was evaluated best (median rank of 1 and mean rank of 1.75) and the EcoPedal worst (median rank of 4 and mean rank of 3.977) in the direct evaluation, as well as in the TAM. It has to be noted that the EcoAdvisor's mean and median rating did not differ much from the EcoPedal, which is consistent with the findings from the TAM score that the differences of Behavioural Intention to Use and Perceived Ease of Use between these two systems were not significant.

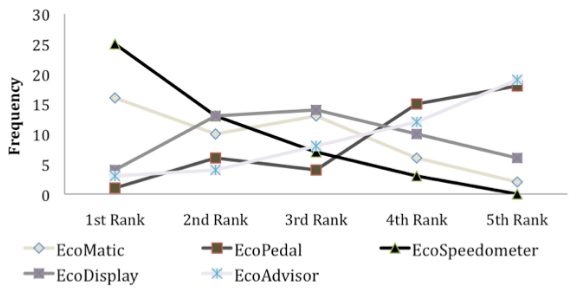


Figure 9: Ranking frequency of the systems.

7. DISCUSSION

In this section we will discuss the results of our study for each research goal:

Research Goal 1 (RG1)
RG1 was to evaluate the user acceptance of persuasive in-car interfaces that are designed to support a fuel-efficient driving style.

For all five systems differences regarding User Acceptance (UA) as well as the three UA factors Behavioural Intention to Use (BI), Perceived Usefulness (U), and Perceived Ease of Use (EOU) were found. All factors were rated positive for the EcoMatic, the EcoSpeedometer, the EcoDisplay as well as the EcoAdvisor. Only for the EcoPedal Behavioural Intention to Use (BI) and Perceived Usefulness (U) were rated negatively.

The user acceptance for the EcoSpeedometer was rated highest for all three factors, followed by the EcoMatic. Especially the intention to use (BI) and the perceived usefulness for these two systems were rated significantly higher than for the other systems. The EcoDisplay as well as the EcoAdvisor were rated almost the same on the scales U and EOU, but the EcoDisplay was rated higher than the Ecoadvisor on the intention to use scale. The EcoPedal was rated lowest on the factors intention to use (BI) and perceived usefulness (U). Compared to these scales, the perceived ease of use (EOU) factor is rated positive and rather high. This leads to the assumption that the users would find it comprehensible and easy to use, but neglect its usefulness and would rather not use it. This finding goes along with Davis and Venkatesh's [3] assumption, that the factor U strongly affects BI but has only a low effect on EOU.

Due to the fact that most systems did not require sophisticated interactions, we expected that the factor perceived ease of use (EOU) would be rated higher for all systems than the other factors. This was not the case in our study. We assume that the users did not rate only the handling of the system but also the transparency of system behaviour. This assumption is strengthened by the fact that the EOU for the EcoSpeedometer, a bicolour interface with a plain behaviour, was rated relatively high.

Research Goal 2 (RG2)
RG2 was to analyse how driver's general attitude toward technology in cars, sociodemographic variables and factors like driver properties and expectations toward the technology concerning safety, disturbance and assistance have influence on the user acceptance.

Sociodemographic variables generally played no role for the systems' evaluations. This could be partly due to the fact that the influence of the questions concerning disturbance, safety and expectancy of assistance were explaining a major part of the variance. Indeed, we found a high influence of expected disturbance and expectancy of assistance. Participants felt especially disturbed by systems with tactile and/or auditory feedback, raising the question whether this finding can be generalized and should be considered in future research and design. Safety issues were mostly no problem for our participants, but nevertheless it has to be noted that a rating of 3 (see Figure 7) or more seems problematic for a system. It remains unclear whether the systems were indeed rated as a possible risk while driving a car, or this finding can be explained as a result of a more general halo effect [24]. This would mean that the appearance and evaluation of an object in other dimensions has an effect on the rating of other system's properties.

Research Goal 3 (RG3)
RG3 was to research if the usage of an online TAM questionnaire for in-car persuasive interfaces in a pre-prototype level would lead to results that support future design decisions.

The fact that we gained different, interpretable results for the different systems encourage our assumption that user acceptance in a pre-prototype phase of persuasive in-car interfaces using the TAM with different scenarios in an online survey is reasonable. 85.7% of the users indicated that they could imagine the technology after reading the description and seeing the pictures. 26.5% found it challenging to answer questions about the systems without actually using them. It seems to be an economic approach with ecological validity to gain early feedback, which seems especially important for prototypes in development. However, this approach does not change the importance of other methods (e.g. focus-groups) to gain valuable insights on users' thoughts about technology of tomorrow.

8. FUTURE WORK

To support the reliability of our results firstly focus groups and secondly the evaluation of hands-on prototypes will be conducted. A comparison of our results and the real driving experience using the systems would be interesting. Furthermore it seems promising to investigate the relatively more negative rating of the EcoPedal and EcoAdvisor. The negative ratings could be partly due to the fact that feedback was given tactile and auditory. Whereas the first feedback may be interpreted by users as working against one's own intention, the latter may be negative because it can be interpreted as similar to the nagging and complaining of a co-driver. Our next steps will be to classify the systems based on system properties like feedback style. Another approach will be to systematically covariate the systems persuasive properties in order to identify which persuasive strategy is most accepted by the users. Additionally to that, a closer look will be taken on the drivers' general attitude towards technology and potential influences on user acceptance.

9. REFERENCES

[1] Ablassmeier, M., Poitschke, T., & Rigoll, G. 2006. A new approach of a context-adaptive search agent for automotive environments. In Proc. CHI '06 Extended Abstracts, New York, NY, USA, 1613-1618. ACM.

[2] Comte, S. Wardman, M., Whelan, G. 2000, Drivers' acceptance of automatic speed limiters: implications for policy and implementation, Transport Policy 7(4), 259-267

[3] Davis, F. D. & Venkatesh, V. 2004. Toward preprototype user acceptance testing of new information systems: implications for software project management. Engineering Management, IEEE Transactions on, 51(1), 31-46.

[4] Davis, F. D. 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly, 13(3), 318+.

[5] Dillon, A. & Morris, M. G. 1996. User acceptance of new information technology: theories and models. http://dlist.sir.arizona.edu/1261/

[6] Fiat, eco:Drive, http://www.fiatecodrive.com (last accessed on 25th May 2009)

[7] Fogg, B.J. 2003. Persuasive Technology: Using Computers to Change What We Think and Do. Morgan Kaufman, San Francisco, CA, USA.

[8] Ford, SmartGauge with EcoGuide, http://media.ford.com/images/10031/SmartGauge.pdf (last accessed on 25th May 2009)

[9] Graf, S., Spiessl, W., Schmidt, A., Winter, A., and Rigoll, G. 2008. In-car interaction using search-based user interfaces. In Proc. CHI '08. ACM, New York, NY, 1685-1688.

[10] Green Car Advisor, http://blogs.edmunds.com/greencaradvisor/2009/04/eco-driving-systems-now-your-car-can-gently-nag-you-into-being-more-fuel-wise.html (last accessed on 25th May 2009)

[11] Honda, Ecological Drive Assist System, http://world.honda.com/news/2008/4081120Ecological-Drive-Assist-System/ (last accessed on 25th May 2009)

[12] Hu, J., Winterboer, A., Nass, C. I., Moore, J. D., & Illowsky, R. 2007. Context & usability testing: user-modelled information presentation in easy and difficult driving conditions. In Proc. CHI '07, New York, NY, USA, 2007 1343-1346. ACM.

[13] Kantowitz, B. H., Hankowski, R. J., and Kantowitz, S.C. 1997. Driver acceptance of unreliable traffic information in familiar and unfamiliar settings. Human Factors 39, 164-176.

[14] Kern, D., Müller, M., Schneegaß, S., Wolejko-Wolejszo, L., & Schmidt, A. 2008. CARS - Configurable Automotive Research Simulator. In Automotive User Interfaces and Interactive Applications - AUIIA 08, Workshop at Mensch und Computer 2008.

[15] Kern, D., Schmidt, A., Arnsmann, J., Appelmann, T., Pararasasegaran, N., and Piepiera, B. 2009. Writing to your car: handwritten text input while driving. In Proc. CHI '09 Extended Abstracts. ACM, New York, NY, 4705-4710.

[16] Kim, S., Dey, A. K. 2009. Simulated augmented reality windshield display as a cognitive mapping aid for elder driver navigation. In Proc. CHI '09. ACM, New York, NY.

[17] Krum, D. M., Faenger, J., Lathrop, B., Sison, J., and Lien, A. 2008. All roads lead to CHI: interaction in the automobile. In Proc. CHI '08 Extended Abstracts. New York, NY. ACM.

[18] Kumar, M. and Kim, T. 2005. Dynamic speedometer: dashboard redesign to discourage drivers from speeding. In Proc. CHI '05 Extended Abstracts. New York, NY. ACM.

[19] Nissan, ECO Pedal, http://www.nissan-global.com/EN/NEWS/2008/_STORY/080804-02-e.html (last accessed on 25th May 2009)

[20] Pace, T., Ramalingam, S., and Roedl, D. 2007. Celerometer and idling reminder: persuasive technology for school bus eco-driving. In Proc. CHI '07 Extended Abstracts. New York, NY, 2085-2090. ACM.

[21] Schmidt, A., Boll, S., Spießl, W., & Kranz, M. 2008. Automotive User Interfaces and Interactive Applications. In Workshop at the Conference Mensch und Computer 2008.

[22] Siewiorek, D., Smailagic, A., & Hornyak, M. 2002. Multimodal Contextual Car-Driver Interface. Multimodal Interfaces, In Proc. Multimodal Interfaces, 367-373. IEEE.

[23] Tester, J., Fogg, B., and Maile, M. 2000. CommuterNews: a prototype of persuasive in-car entertainment. In Proc. CHI'00 Extended Abstracts. New York, NY, 24-25. ACM.

[24] Thorndike, E. L. 1920. A constant error in psychological rating. Journal of Applied Psychology, 4, 25–29.

[25] Toyota, 3rd Generation Prius, http://www.toyota.com/vehicles/minisite/newprius/ (last accessed on 25th May 2009)

[26] Várhelyi, A., Hjälmdahl, M., Risser, R., Draskóczy, M., Hydén, C., & Almqvist, S. 2002. The effects of Large Scale use of active accelerator pedal in urban areas. In International Cooperation of Theories and Concepts in Traffic safety (ICTCT) Workshop, Nagoya, Japan , 2002.

[27] Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D. 2003 "User acceptance of information technology: toward a unified view," MIS Quarterly (27:3). 425-478.

[28] Wochinger, K., Boehm-Davis, D. 1997. Navigational preference and driver acceptance of advanced traveler information systems. In Ergonomics and Safety of intelligent Driver interfaces. NJ, 345-362.

MULTIMODAL INTERACTION AND PHYSIOLOGICAL ASPECTS

Enhanced Auditory Menu Cues Improve Dual Task Performance and are Preferred with In-vehicle Technologies

Myounghoon Jeon, Benjamin K. Davison, Michael A. Nees, Jeff Wilson, & Bruce N. Walker
Sonification Lab, School of Psychology
Georgia Institute of Technology
654 Cherry Street
Atlanta, GA 30332 USA
+1 404 894 8265
{mh.jeon, davison, mnees}@gatech.edu, jeff.wilson@bitc.gatech.edu, bruce.walker@psych.gatech.edu

ABSTRACT
Auditory display research for driving has mainly focused on collision warning signals, and recent studies on auditory in-vehicle information presentation have examined only a limited range of tasks (e.g., cell phone operation tasks or verbal tasks such as reading digit strings). The present study used a dual task paradigm to evaluate a plausible scenario in which users navigated a song list. We applied enhanced auditory menu navigation cues, including *spearcons* (i.e., compressed speech) and a *spindex* (i.e., a speech index that used brief audio cues to communicate the user's position in a long menu list). Twenty-four undergraduates navigated through an alphabetized song list of 150 song titles—rendered as an auditory menu—while they concurrently played a simple, perceptual-motor, ball-catching game. The menu was presented with text-to-speech (TTS) alone, TTS plus one of three types of enhanced auditory cues, or no sound at all. Both performance of the primary task (success rate of the game) and the secondary task (menu search time) were better with the auditory menus than with no sound. Subjective workload scores (NASA TLX) and user preferences favored the enhanced auditory cue types. Results are discussed in terms of multiple resources theory and practical IVT design applications.

Categories and Subject Descriptors
H.5.2 [**Information Interfaces And Presentation (e.g., HCI)**]: User Interfaces – Auditory (non-speech) feedback, graphical user interfaces (GUI), interaction styles (e.g., commands, menus, forms, direct manipulation), user-centered design, voice I/O

H.5.1 [**Information Interfaces And Presentation (e.g., HCI)**]: Multimedia Information Systems – audio input/output

D.2.2 [**Software Engineering**]: Design Tools and Techniques – user interfaces

General Terms
Design, Experimentation, Human Factors, Performance

Keywords
auditory display, dual task, infotainment, IVTs (In-Vehicle Technologies), spearcon, spindex, TTS (Text-to-Speech), auditory menus, multiple resources

1. INTRODUCTION
Emerging wireless and digital technologies have allowed for an abundance of information to be delivered in mobile devices. This information portability has extended to the automobile cockpit in the form of so-called in-vehicle technologies [IVTs, see 1, 2]. IVTs can include such diverse digital media as pictures, video, and audio, and IVTs have been developed to deliver driving-relevant information (e.g., navigation instructions; weather and traffic updates), in-vehicle entertainment (e.g., digital music and video or television), and productivity applications (e.g., cellular phone and wireless internet) for the driver and passengers (see, for example, www.centrafuse.com).

Complex in-vehicle technologies may increasingly distract drivers, and research has suggested that problems of driver inattention have become worse [3-5]. A critical concern that has been validated in research involves the extent to which visually-demanding tasks like driving are prone to interference from secondary tasks such as those encouraged by IVTs. Secondary tasks have been shown to negatively affect driving performance, and subjective workload increased while driving and performing a secondary task [6, 7].

Despite the potential pitfalls of IVTs with respect to driver distraction, it has been argued that such technologies can be safely integrated into automobiles, and good practice guidelines have even been proposed [8]. Research has found that younger adults accomplished a task that required reading text messages aloud from an IVT system with surprisingly little impact on simulated driving performance, although this promising finding did not hold for older adults [9]. Given that IVTs and other secondary distractions appear to be a common component of the modern automobile, the appropriate design of safe IVTs remains a challenge that must be addressed by further research, and auditory information presentation represents an obvious alternative to visual information presentation for IVTs.

1.1 Auditory and Multimodal Presentation for IVTs
Information presentation via audio has been shown to facilitate performance with interfaces where visual overload may otherwise occur [10, 11]. Research has further suggested that auditory and multimodal IVTs may overcome some of the problems associated with visually-taxing IVTs. For example, Liu found that both driving and secondary task performance

were better using auditory, and particularly multi-modal, in-vehicle information displays [12].

Multiple resources theory [see, e.g., 13] often has been invoked to explain the apparent benefit of dividing information display across modalities during multitasking. Multiple resources theory would predict that concurrent auditory and visual tasks draw upon separate pools of modality resources and thus should be time-shared efficiently (i.e., without interrupting each other) to the extent that they also do not require the same processing code (cognitive representational) resources, stages of processing, or response modalities (manual versus oral). Other studies [1, 14], however, have suggested that a discrete auditory task *preempts* or causes a brief lapse in the performance of a continuous visual task while the auditory stimulus is attended to, perhaps owing to the auditory modality's superior ability to attract attention [15, 16]. As such, an auditory cost has been found in a number of studies that examined the modality of in-vehicle information displays [1, 17-19]. The results from these studies suggested that the potential modality benefits of auditory (rather than visual) presentation of secondary task information might be mitigated by processing mechanisms (such as *preemption,* described above) and display characteristics. Related research has shown that even hands-free, auditory cell phone conversations impair driving [20]. In other studies, both an auditory cost and an auditory benefit for in-vehicle information displays has been shown [21-23], while much research has shown the intuitively predicted auditory benefit for both tasks [12, 24-28].

Taken together, these findings suggest qualified successes for the implementation of auditory displays in IVTs, but the precise circumstances in which auditory cues help or harm performance of a visual primary task and the exact locus of interference remain to be determined. The current study examined the impact of a number of recently developed enhanced auditory cues that are currently being considered for implementation in an existing IVT on performance of a perceptual-motor visual primary task.

1.2 Enhanced Auditory Cues in Menu Navigation

The use of sound to communicate information about the driving task itself [e.g., warnings relating to the vehicle status or the presence of an approaching vehicle, see 29] must be distinguished from the use of sound as a means of interacting with the IVT systems (i.e., "infotainment" systems). The content in infotainment IVTs is often organized into a menu structure through which the driver (or passenger) must navigate in order to select the desired option (e.g., to play a particular song or to retrieve directions to a particular restaurant). Relatively little research has examined the use of sound in this particular context, even though audio might improve overall performance and safety (as well as user workload, stress, and satisfaction ratings) as compared to visual menu structures.

Typically, sound is used in such menus simply by speaking aloud the menu items via text-to-speech (TTS) synthesis, but more can be done to enhance auditory menus. Non-speech cues, for example, can supplement spoken menu items. The present report focuses on the use of non-speech cues to enhance a spoken auditory menu. Our recent research in this area has specifically examined spearcons and spindex cues, described below.

Figure 1. View of conducting dual tasks. Participants navigated a song list while playing a ball-catching game.

1.2.1 Spearcon: Compressed Speech Sounds

Spearcons (speech earcons) are brief sounds that are produced by speeding up spoken phrases, even to the point where the resulting sound is no longer comprehensible as a particular word [30]. These unique sounds are analogous to fingerprints because of their acoustic relationship with the original speech phrases. Spearcons are easily created by converting the text of a menu item to speech via TTS and speeding it up using a pitch-constant compression algorithm, a process that allows the system to cope with dynamic menus. Typically, spearcons are prepended to (or may even entirely replace) the spoken menu item, which allows faster learning and navigation of the auditory menu.

Spearcons have shown better performance and learning rates than other well known nonspeech auditory cues such as auditory icons [31], earcons [32], and TTS alone. For example, Walker et al. [30] showed that spearcons resulted in faster and more accurate performance than other auditory cues for a search task. Spearcons also improved navigation efficiency over auditory menus using only TTS or no sound when combined with visual cues [33-35]. Other studies [36, 37] have demonstrated that spearcons are as learnable as speech, but auditory icons and earcons were more difficult to learn.

1.2.2 Spindex: Speech Index

A spindex [speech index, see 38] is created by associating an auditory cue with each menu item, and the cue is based on the pronunciation of the first letter of each menu item. For instance, the spindex cue for "All the above" would be a sound based on the spoken sound "A". The set of spindex cues in an alphabetical auditory menu is analogous to the visual index tabs that are often used to facilitate flipping to the right section of a thick reference book such as a dictionary or a telephone book, and analogous visual indices have been used, for example, in newer Apple iPods. The benefit of an auditory index (spindex) can be explained by the fact that users employ a combination of rough and fine navigation strategies in the search processes [39]. In the rough navigation stage, users invoke top-down knowledge about the serial order of the alphabet to exclude non-targets until they approach the alphabetical area proximal to the target. After users perceive that they reach the target zone, they need more precise and detailed information to select the target. The spindex-enhanced auditory menu can contribute per-item speedups during the rough search stage while still supporting detailed item information via the TTS phrase in the final search stage.

Spindex cues are natural sounds (based on speech) and part of the original word, thus they do not require training to learn the

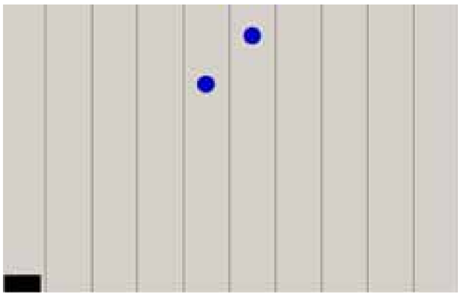


Figure 2. Screen capture of the primary task (game). Balls fall from the top of the screen, and the task is to "catch" them by moving the "bucket" (bottom left).

Figure 3. Screen capture of the secondary task (song list navigation) from the IVT. The task is to navigate to, then select, the target menu item ("Use somebody" in this case).

mapping from the sound to its intended meaning. In previous research, participants showed better performance in a TTS + spindex condition than in a TTS-only condition. Moreover, spindex-enhanced menus were learned more quickly, with peak performance reached in about half the number of trials, as compared to TTS-only menus. A subsequent study showed that alternative designs (*decreased* and *attenuated* types, discussed below) further improved user acceptance and performance [for more details of the spindex cue types, see 40].

1.3 The Current Study and Hypotheses

With respect to the menu-oriented tasks often required to select content in IVTs, relatively little research has examined the potential for audio cues to reduce conflicts with a visual primary task. Conflicting results have suggested that auditory secondary tasks may sometimes preempt performance of a visual primary task, while other results have shown an advantage for auditory presentation of a secondary task in the presence of a visual primary task. Furthermore, the extent to which enhanced auditory cues (spearcons and spindex) may improve IVTs has yet to be established. To investigate these issues, the current study devised a plausible secondary task in which participants navigated a song list on an in-vehicle head unit. For this scenario, a divided attention paradigm [41] was used to examine the effectiveness of five types of auditory cues on performance for both a primary visual attention task (a simple ball-catching game that required perceptual vigilance and nearly constant manual control) and a concurrent secondary menu search task.

We predicted that the displays with auditory cues would shorten the navigation time in the secondary task, and also that the primary task (a visual task with perceptual and manual control components) should be less affected by the secondary task when auditory cues are used. The combined workload of the task configuration was predicted to be attenuated by the use of auditory cues. With respect to the relative effectiveness of auditory cues, we predicted that enhanced auditory cues (i.e., those using spearcons and spindex cues) would outperform traditional TTS cues.

2. METHOD

2.1 Participants

Twenty-four undergraduate students (10 female; mean age = 20.2, *SD* = 1.2) participated in this study for credit in psychology courses. Participants reported normal or corrected-to-normal vision and hearing and gave informed consent.

2.2 Apparatus

Figure 1 shows the experimental apparatus. The primary task stimuli were presented using a Dell Dimension XPS T600 computer, running Windows XP on a Pentium 3, 598 MHz processor and 512 MB of RAM. A 17" monitor was placed on a table 50 cm in front of the seated participant. For the secondary task, stimuli were presented using an in-vehicle head unit, running Windows VISTA on a Pentium 4, CF as the HMI / car PC software, 1.83 GHz processor and 1 GB of RAM. A Sigma Tel High Definition audio output device was used for sound rendering. Participants listened to auditory stimuli using harman/kardon HK195 speakers located 30 cm behind the primary task monitor. The head unit included a 6.5" resistive touch screen panel. The head unit was located on an in-vehicle head unit position (approximately 34 cm below and 37 cm to the right from the center of the primary task monitor) [1] (See Figure 3).

2.3 Stimuli

2.3.1 Primary Task

The primary task was a visual perceptual-motor vigilance task and was piloted to be of sufficient difficulty to observe dual task decrements when the secondary menu task was introduced [for a discussion of the importance of task difficulty in dual task scenarios, see 42]. The simple computer game (see Figure 2) was programmed in Visual Basic 6.0 and consisted of balls that dropped along 10 vertical columns from the top of the screen at a rate of approximately 1 ball per second, with a black box that participants moved along the bottom of the screen. The purpose of the game was to catch all of the balls with the box before they reached the bottom of the screen. When a ball was successfully captured, the box flashed from black to green. To control the box, participants placed the index and middle finger of their left hand on the right and left arrow keys on the keyboard, respectively. Five pilot subjects allowed us to establish the baseline performance of the primary task at 92.11% (*SD* = 5.31) accuracy for catching the balls over a 1 minute trial.

2.3.2 Secondary Task

The secondary IVT menu navigation task was designed as a song selection task. A song list was created with 150 song titles gathered from the Billboard Hot 100 & Pop 100 (2009, 2008) (http://www.billboard.com/bbcom/index.jsp) and iTunes Top 100 (http://www.apple.com/itunes/top-100/songs/). A visual menu (see Figure 3) was created in C# using the CentraFuse SDK programming tools for use as a plugin for the Centrafuse

2.1 head unit user interface (www.centrafuse.com). The menu items were in alphabetical order, and the participant was able to scroll downward and upward in the menu by pressing arrow buttons on the touch screen. One arrow press moved the selected item down by one menu position, and the display advanced upon any arrow press where the next item was on a different page. The participants' objective was to reach the given target name in the list menu as fast as possible. Participants logged their selection as the current active item by pressing a "select" button (top right of figure 3). If the participant reached the top or bottom of the menu, the list did not wrap around.

In addition to the visual display, each menu item could also have auditory cues (depending upon the experimental condition) that played when the menu item was highlighted. When the arrow button was pushed, the button-pressing handler triggered the auditory sound playback action. The sounds were prerecorded as a single file for each menu item (with negligible loading delay). In order to maintain a code-based performance similarity between the no sound and sound conditions, a non-audible sound file of similar playback length was played for each menu item the no sound condition. The auditory cues included speech (TTS) and non-speech enhanced auditory cues as described below (also see Table 1).

2.3.2.1 Text-To-Speech Cues

TTS files (.wav) were generated for all of the song titles using the AT&T Labs TTS Demo program with the male voice *Mike-US-English* (http://www.research.att.com/~ttsweb/tts/demo.php). Menu items in this condition simply consisted of an auditory TTS phrase that played for each menu item as the participant navigated the song list.

2.3.2.2 Spearcon Cues

Spearcons were created from the TTS files of each name by running them through the GT Sonification Lab's spearcon generation algorithm, in the form of a MATLAB script that compresses each TTS cue logarithmically while maintaining original sound frequency. Logarithmic compression is currently considered the preferred compression technique for creating spearcons, because it compresses longer phrases more than shorter phrases. Shorter words tend to sound more like "clicks" if they are compressed too much and lose their original acoustic properties. In this condition, spearcons were prepended to each TTS menu item, with a 250 ms silent interval between the spearcon and the beginning of the TTS phrase

2.3.2.3 Spindex Cues

Since the *attenuated* spindex design has been shown to be the most preferred and simplest to implement [40], we used that version in this experiment. The attenuated version contained cues that were attenuated by -20 dB from the first menu item in a letter category. Spindex cues were created by generating TTS files for each letter (e.g., "A"). Each spindex cue consisted of only one syllable, pronouncing each of 26 letters which represented the initial letter of the names. Spindex cues used in the list were presented before the TTS cues, with a 250 ms interval between the spindex cue and the TTS phrase [33, 38]. If a participant touched an arrow button very fast, the spindex cues were generated preemptively without a lag between items.

2.3.2.4 Mixed Cues

We also created mixed cues with combined TTS, spearcons, and spindex cues. For this, we employed the *minimal* spindex type because event this showed the same level of performance on auditory menu searches as the other spindex types [40]. The minimal spindex cues were used only when the user crossed category boundaries in the search list (e.g., for the first menu item starting with A, then the first item starting with B, and so on). Therefore, the spindex cues were added to only the category boundaries of the spearcon version of the auditory menu.

Condition	Auditory Cue Order (250 ms delay between)
No sound	(empty sound played)
TTS-only	TTS
Spearcon + TTS	Spearcon, TTS
Spindex + TTS	Spindex, TTS
Spindex + Spearcon + TTS	(Spindex,) Spearcon, TTS

Table 1. Auditory cue orders for each experimental condition of the secondary task.

2.4 Design and Procedure

Before the start of the dual tasks, participants performed the primary task alone for one minute to obtain a baseline for the single task condition. Participants then began the dual task portion of the study. In order to more accurately analyze the timing of both tasks, we synchronized the system clocks of the computers using a network time server. The primary task was initiated, and the target name for the secondary task was presented through the speakers after a delay randomly selected from 5, 10, or 15 seconds from the start of the primary task. The target name was also displayed visually on the first line of the list on the secondary task IVT head unit (e.g., "Use somebody" in Figure 3). After hearing the target menu item, participants navigated the list of songs on the touch screen while simultaneously playing maintaining performance of the visual primary task. They were instructed to always allocate 80% of their effort/attention to the primary task (game) and 20% to the secondary task (navigation) [see, e.g., 43]. After the selection of the target, there was another randomly selected delay of 5, 10, or 15 seconds before the next target item was presented. Menu navigation time was operationalized as the time between the first menu navigation button press, and the pressing of the select button. There were five within-subjects conditions, based on auditory cue type: No sound, TTS-only, spearcon + TTS, (attenuated) spindex + TTS, and (minimal) spindex + spearcon + TTS. One block included five trials of different targets. To evenly spread out the target menu positions across conditions, one target in each block was randomly selected from menu items 1-15, one from 16-30, and so on. Each condition was composed of two successive blocks, and the order of presentation of the cue conditions was counterbalanced across participants. At the end of the block (i.e., after all menu targets had been presented), participants saw a pop-up window and pressed the 'Q' key on their keyboard to quit the primary task. After each condition, participants completed the electronic version of NASA TLX [e.g., 44] to obtain measurements of perceived workload for the overall task combinations. Finally, after completing all conditions, participants filled out a short questionnaire for demographic information, indicated their preferred auditory cue condition, and provided comments on the study.

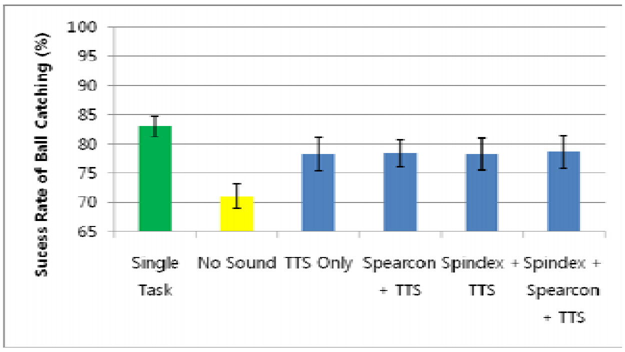


Figure 4. Primary task performance across auditory cue types.

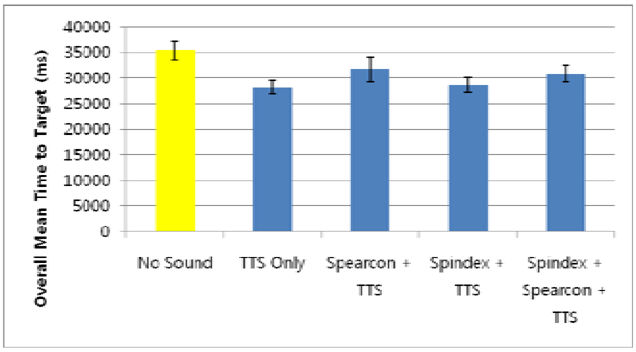


Figure 5. Secondary task performance across auditory cue types.

3. RESULTS

3.1 Primary Task Performance

Figure 4 shows overall mean percentages of success in the primary task for the single task and each auditory cue type. Results were analyzed with a 5 (Auditory cue type) x 2 (Block) repeated measures analysis of variance (ANOVA), which revealed a statistically significant difference between auditory cue types in mean success rate, $F(4, 92) = 8.372$, $p < .001$, $\eta_p^2 = .267$. Also, Block 2 ($M = 78.025$, $MSe = 2.248$) led to significantly higher score than Block 1 ($M = 75.711$, $MSe = 2.454$), $F(1, 23) = 15.737$, $p = .001$, $\eta_p^2 = .406$. The interaction of cue type with block was not significant, $F(4, 92) = 0.263$, $p = .901$. For the multiple comparisons among single task and the auditory cue types, we conducted paired-samples *t*-tests. Participants caught significantly more balls in the single task and all of the auditory-enhanced conditions than in the no sound condition. Success rate in the single task condition ($M = 82.96$, $SD = 8.86$) was higher than that in the no sound condition ($M = 71.01$, $SD = 10.12$), $t(23) = 7.325$, $p < .001$[1]. Also, TTS-only ($M = 78.16$, $SD = 13.54$), $t(23) = -3.753$, $p = .001$, the spearcon + TTS cue ($M = 78.37$, $SD = 11.39$), $t(23) = -5.365$, $p < .001$, the spindex + TTS cue ($M = 78.21$, $SD = 13.10$), $t(23) = -5.509$, $p < .001$, and the spindex + spearcon + TTS cue ($M = 78.59$, $SD = 13.80$), $t(23) = -4.054$, $p < .001$ were also higher than the no sound cue. Primary task performance decreased in the no sound condition relative to baseline, but statistically performance recovered to the single task level in all sound conditions.

3.2 Secondary Task Performance

Errors (selection of non-target) in the secondary task were minimal, so the primary focus of the analyses for the secondary task was on the reaction time. For the sake of completeness, however, a one-way (Auditory cue type) repeated measures analysis of variance (ANOVA) was conducted and revealed a statistically significant differences between auditory cue types in navigation errors, $F(2.939, 92) = 3.613$, $p < .05$, $\eta_p^2 = .136$. For the multiple comparisons among the auditory cue types, we conducted paired-samples t-tests. The TTS-only cues ($M = .29$, $SD = .86$), $t(23) = 3.149$, $p = .004$ and the spindex + spearcon + TTS cues ($M = .33$, $SD = .56$), $t(23) = 3.204$, $p = .004$ showed significantly lower errors than the no sound condition ($M = 1.17$, $SD = 1.20$). The spearcon + TTS cues ($M = .54$, $SD = .98$), $t(23) = 1.871$, $p = .074$ and the spindex + TTS cues ($M = .54$, $SD = .88$), $t(23) = 1.969$, $p = .061$ showed only marginally significant improvements in errors over the no sound condition for the secondary task.

We included only correct responses in reaction time analyses. Figure 5 shows overall mean time to target (i.e., "search time", in ms) in the secondary task for each of the auditory cue types. These results were also analyzed with a 5 (Auditory cue type) x 2 (Block) repeated measures analysis of variance (ANOVA), which revealed a statistically significant difference between auditory cue types in mean search time, $F(4, 92) = 3.530$, $p < .05$, $\eta_p^2 = .133$. Also, Block 2 ($M = 29881.537$, $MSe = 1224.721$) led to significantly shorter search times than Block 1 ($M = 32036.963$, $MSe = 1213.727$), $F(1, 23) = 7.912$, $p < .05$, $\eta_p^2 = .256$. For the multiple comparisons among the auditory cue types, we conducted paired-samples t-tests. Participants searched significantly faster in TTS-only ($M = 28195.23$, $SD = 6791.48$), $t(23) = 3.888$, $p = .001$ and the spindex + TTS ($M = 28607.71$, $SD = 7324.10$), $t(23) = 3.330$, $p = .003$ conditions than in the no sound condition ($M = 35412.12$, $SD = 8996.31$). The spindex + spearcon + TTS ($M = 30871.16$, $SD = 7942.12$) also showed a tendency to be faster than the no sound, although this result was only marginally significant, $t(23) = 1.923$, $p = .067$. The spearcons condition ($M = 31710.03$, $SD = 2459.80$) was not significantly different from the no sound condition, $t(23) = 1.499$, $p = .147$. The interaction of block with cue type was not significant, $F(2.76, 92) = 1.167$, $p = .328$ with a Greenhouse-Geiser correction for sphericity violations employed.

3.3 Overall Workload and Preference

Figure 6 shows the overall workload scores for each of the auditory cue types. All of the auditory cue types decreased the perceived workload of both tasks. These results were supported by a one-way (Auditory cue type) repeated measures analysis of variance (ANOVA), which revealed a statistically significant difference between auditory cue types in workload score, $F(4, 92) = 14.348$, $p < .001$, $\eta_p^2 = .384$. For the multiple comparisons among the auditory cue types, we conducted paired-samples t-

[1] All pairwise comparisons in this study applied a Bonferroni adjustment to control for Type-I error, which meant that we used more conservative alpha levels (for the primary task, critical *alpha* level = .003; for the secondary task and workload scores, critical *alpha* level = .005).

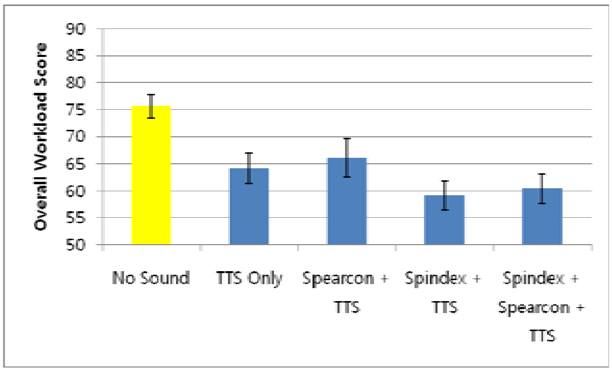


Figure 6. Overall workload score across auditory cue types.

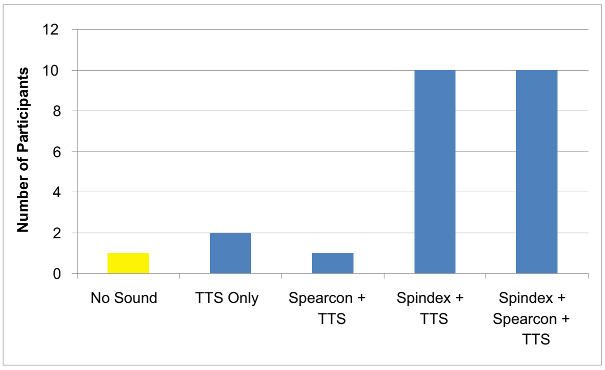


Figure 7. Overall preference across auditory cue types.

tests. The TTS-only cues (M = 64.12, SD = 14.09) showed lower workload than no sound cues (M = 75.64, SD = 10.90), $t(23) = 4.332$, $p < .001$. Also, the spearcon + TTS type (M = 66.35, SD = 17.40), $t(23) = 3.661$, $p = .001$, the spindex + TTS type (M = 59.65, SD = 13.18), $t(23) = 6.650$, $p < .001$, and the spindex + spearcon + TTS type (M = 60.68, SD = 13.57), $t(23)$ = 5.485, $p < .001$ showed lower perceived workload than the no sound type. Further, the spindex + TTS, $t(23) = 2.294$, $p = .031$ and the spindex + spearcon + TTS, $t(23) = 1.933$, $p = .066$ showed marginally lower perceived workload than TTS-only. For the best choice of the auditory cue types, participants clearly preferred the spindex + TTS (N = 10) and the spindex + spearcon + TTS (N = 10) to others (the no sound, N = 1; TTS-only, N = 2; the spearcon + TTS, N = 1) (See Figure 7).

4. DISCUSSION

We evaluated performance, workload, and preference measures for five types of auditory presentation cues for an IVT menu navigation task in the presence of a visual perceptual-motor vigilance primary task. The results showed that the application of the auditory cues for in-vehicle head units could improve both primary and secondary task performance and ameliorate the overall workload. The significant performance improvements over time (i.e., from Block 1 to Block 2) for both primary and secondary task measures suggest that participants may continue to acquire skill with the system and further improve performance on both tasks with more practice using the IVT interface during a visual primary task, although more longitudinal research will be required to examine these practice effects.

In terms of the primary task, all of the auditory conditions outperformed the no audio condition. This suggested that redundant multimodal presentation was less disruptive to performance of the primary task than visual-only presentation. Given the visually intensive nature of the primary task employed here, we expect that these results may generalize to driving scenarios. Specifically, auditory cues for IVTs might allow drivers to devote more attention to the roadway than visual-only menus in IVTs, as all of the auditory cue conditions recovered the primary task performance to the baseline single task level.

With respect to secondary task performance, all of the conditions with auditory cues reduced the mean number of secondary task errors (at statistically significant or at least marginally significant levels) as compared to the condition with no sound cues. Additionally, some auditory cues (TTS-only and the spindex + TTS) showed significantly faster performance than the condition with no sound cue. While the spearcon + TTS and mixed cue conditions only showed marginally faster performance than the no sound condition, the mean difference of nearly 5 seconds may represent a practially relevant finding that would reach statistical significance with a larger sample size.

In addition to our findings with respect to performance, we found promising results that showed an overall reduction in perceived workload and also a subjective preference for enhanced auditory presentations. Participants perceived workload to be lower with auditory cues as compared to no sound, and enhanced auditory cues (particularly the spindex and the spindex + spearcon conditions) resulted in lower workload than TTS-only. It can be inferred that the lower workload in complex multitasking situations might increase the capacity for driving or other visually-demanding tasks to be performed while interacting with IVT menus.

Participants also favored the spindex + TTS and the spindex + spearcon + TTS cues, despite the fact that these conditions showed equivalent levels of performance with TTS-only. The intersection of performance and aesthetics preferences remains a challenge for auditory display design [45, 46], and the user may reject non-preferred or undesirable auditory displays even when performance measures are improved by the use of such displays. We believe that the appropriate implementation of audio in IVTs will require the consideration not only of performance consequences, but also of user preferences and perceived desirability. Any audio design could be more successfully deployed to the extent that it meets user preferences and improves performance [40].

Our results do not offer evidence for a cognitive mechanism of preemption [1, 14] with respect to the effects of the discrete auditory secondary task on the continuous visual primary task. In other words, our data suggested that participants in this study did not preempt or interrupt performance of the visual primary task in order to accomplish the secondary task. Primary task performance was better in the redundant presentation condition than that in the visual only condition, so no auditory cost was observed. Moreover, the use of auditory cues seemed to contribute more to improve the primary task performance than the secondary task performance.

The findings of the present study are perhaps most readily explained by the time sharing predictions of multiple resources theory. For the no audio condition, the primary task and the secondary task conflicted with each other in terms of both processing stage (both required motor response processes) and modality (both required focal vision) resources. We explicitly piloted and calibrated our primary visual task to be particularly demanding of the visual resources, and the addition of the secondary task (which was also demanding, with the overall average time-to-target at around 31 seconds) seemed to have exceeded participants' capacity to effectively time-share the tasks equally across all secondary task conditions. Our primary task performance findings, in particular, suggested that supplementing the visual display of the secondary task with audio may have alleviated some of the demands on focal vision, thereby allowing for better primary task performance (as a function of lowered demands on visual resources), even when motor demands remained constant across conditions. Indeed, dual task performance is worse in many circumstances when two visual tasks must be time shared as compared to a task configuration in which information is divided across modalities [e.g., 41].

5. CONCLUSION & FUTURE WORK

Our results that the auditory modality, and enhanced auditory cues in particular, may allow a user to more safely operate the menus of IVTs in presence of a visually-demanding primary task. IVTs may be more gracefully embedded into a driving task through the application of enhanced auditory cues that can improve the performance and reduce perceived workload. For a more representative primary task, enhanced auditory cues should be evaluated in a high fidelity driving simulation using a wheel remote controller for the navigation task. Other critical issues remain to be examined, including the effects of cabin noise on IVT auditory displays in a real driving situation. The present research, however, has suggested that auditory displays, and particularly enhanced auditory cues such as spearcons and spindex, may improve dual-task performance and also be preferred for interacting with IVTs.

6. ACKNOWLEDGMENTS

We thank CentraFuse for lending the head unit and software and Yarden Moskovitch for gathering data.

7. REFERENCES

[1] W. J. Horrey and C. D. Wickens, "Driving and side task performance: The effects of display clutter, separation, and modality," *Human Factors,* vol. 46, p. 611-624, 2004.

[2] W. J. Horrey, C. D. Wickens, and K. P. Consalus, "Modeling drivers' visual attention allocation while interacting with in-vehicle technologies," *Journal of Experimental Psychology: Applied,* vol. 12, pp. 67-78, 2006.

[3] S. Ashley, "Driving the info highway," *Scientific American,* vol. 285, pp. 44-50, 2001.

[4] T. Dukic, L. Hanson, and T. Falkmer, "Effect of drivers' age and push button locations on visual time off road, steering wheel deviation and safety perception," *Ergonomics,* vol. 49, pp. 78-92, 2006.

[5] C. Patten, A. Kircher, J. Ostlund, and L. Nilsson, "Using mobile telephones: Cognitive workload and attention resource allocation," *Accident Analysis and Prevention,* vol. 36, pp. 341-350, 2004.

[6] T. C. Lansdown, N. Brook-Carter, and T. Kersloot, "Distraction from multiple in-vehicle secondary tasks: Vehicle performance and mental workload implications," *Ergonomics,* vol. 47, pp. 91-104, 2004.

[7] O. Tsimhoni and P. A. Green, "Visual demand of driving and the execution of display-intensive in-vehicle tasks," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting (HFES 2001)*. vol. 2, 2001, pp. 1586-1590.

[8] P. C. Burns and T. C. Lansdown, "E-Distraction: The challenges for safe and usable internet services in vehicles.," in *Internet Forum on the Safety Impact of Driver Distraction When Using In-Vehicle Technologies* www.driverdistraction.org, 2000, July-August.

[9] F. Schieber, A. Holtz, B. Schlorholtz, and R. McCall, "Analysis of visual demands of in-vehicle text displays reveals an age-related increase in time needed to reallocate attention to the road," in *Human Factors and Ergonomics Society (HFES2008)*. vol. 52, 2008, pp. 149-153.

[10] S. Brewster, "Using non-speech sound to overcome information overload," *Displays,* vol. 17, pp. 179-189, 1997.

[11] M. L. Brown, S. L. Newsome, and E. P. Glinert, "An experiment into the use of auditory cues to reduce visual workload," in *ACM CHI 89 Human Factors in Computing Systems Conference (CHI 89)*, 1989, pp. 339-346.

[12] Y. C. Liu, "Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveller information systems," *Ergonomics,* vol. 44, pp. 425-442, 2001.

[13] C. D. Wickens, "Multiple resources and performance prediction," *Theoretical Issues in Ergonomics Science,* vol. 3, pp. 159-177, 2002.

[14] C. D. Wickens and Y. Liu, "Codes and modalities in multiple resources: A success and a qualification," *Human Factors,* vol. 30, pp. 599-616, 1988.

[15] C. Spence and J. Driver, "Audiovisual links in attention: Implications for interface design," in *Engineering Psychology and Cognitive Ergonomics Vol. 2: Job Design and Product Design*, D. Harris, Ed. Hampshire: Ashgate Publishing, 1997, pp. 185-192.

[16] N. A. Stanton, R. T. Booth, and R. B. Stammers, "Alarms in human supervisory control: A human factors perspective," *International Journal of Computer Integrated Manufacturing,* vol. 5, pp. 81-93, 1992.

[17] G. Matthews, T. J. Sparkes, and H. M. Bygrave, "Attentional overload, stress, and simulate driving performance," *Human Performance,* vol. 9, pp. 77-101, 1996.

[18] J. D. Lee, B. F. Gore, and J. L. Campbell, "Display alternatives for in-vehicle warning and sign information: Message style, location, and modality," *Transportation Human Factors,* vol. 1, pp. 347-375, 1999.

[19] T. A. Dingus, D. V. McGehee, N. Manakkal, S. K. Jahns, C. Carney, and J. M. Hankey, "Human factors field evaluation of automotive headway maintenance/collision warning devices," *Human Factors,* vol. 39, pp. 216-229, 1997.

[20] D. L. Strayer and F. A. Drews, "Cell-phone-induced driver distraction," *Current Directions in Psychological Science,* vol. 16, pp. 128-131, 2007.

[21] M. A. Mollenhauer, J. D. Lee, K. Cho, M. C. Hulse, and T. A. Dingus, "The effects of sensory modality and information priority on in-vehicle signing and information systems," in *Human Factors and Ergonomics Society (HFES 1994)*. vol. 38: Human Factors and Ergonomics Society, 1994, pp. 1072-1076.

[22] A. Parkes and G. Burnett, "An evaluation of medium range “advance information” in route-guidance displays for use in vehicles," in *IEEE Vehicle Navigation & Information Systems Conference* Ottawa, Canada, 1993, pp. 238-241.

[23] T. A. Ranney, J. L. Harbluk, and Y. I. Noy, "Effects of voice technology on test track driving performance: Implications for driver distraction," *Human Factors,* vol. 47, p. 439-454, 2005.

[24] G. E. Burnett and S. M. Joyner, "An assessment of moving map and symbol-based route guidance systems," *Ergonomics and Safety of Intelligent Driver Interfaces,* pp. 115-136, 1997.

[25] T. A. Dingus, M. C. Hulse, D. V. McGehee, and R. Manakkal, "Driver performance results from the TravTek IVHS camera car evaluation study," in *Human Factors and Ergonomics Society (HFES 1994)*. vol. 38: Human Factors and Ergonomics Society, 1994, pp. 1118-1122.

[26] K. W. Gish, L. Staplin, J. Stewart, and M. Perel, "Sensory and cognitive factors affecting automotive head-up display effectiveness," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 1694, pp. 10-19, 1999.

[27] L. Streeter, D. Vitello, and S. Wonsiewicz, "How to tell people where to go: Comparing navigational aids," *International Journal of Man-Machine Studies,* vol. 22, pp. 549-562, 1985.

[28] J. Walker, E. Alicandri, C. Sedney, and K. Roberts, "In-vehicle navigation devices: Effects on the safety of driver performance," McLean, VA: Performer: Federal Highway Administration, Office of Safety and Traffic Operations Research and Development, 1991, p. 107.

[29] C. Ho and C. Spence, "Assessing the effectiveness of various auditory cues in capturing a driver's visual attnetion," *Journal of Experimental Psychology: Applied,* vol. 11, pp. 157-174, 2005.

[30] B. N. Walker, A. Nance, and J. Lindsay, "Spearcons: Speech-based earcons improve navigation performance in auditory menus," in *International Conference on Auditory Display (ICAD2006)* London, UK, 2006, pp. 95-98.

[31] W. W. Gaver, "Using and creating auditory icons," in *Auditory display: Sonificaiton, audification, and auditory interfaces*, G. Kramer, Ed. MA: Addison-Wesley, 1994, pp. 417-446.

[32] M. M. Blattner, D. A. Sumikawa, and R. M. Greenberg, "Earcons and icons: Their structure and common design principles," *Human-Computer Interaction,* vol. 4, pp. 11-44, 1989.

[33] D. Palladino and B. N. Walker, "Navigation efficiency of two dimensional auditory menus using spearcon enhancements," in *Annual Meeting of the Human Factors and Ergonomics Society (HFES 2008)* New York, NY (22-26 September), 2008, pp. 1262-1266.

[34] D. Palladino and B. N. Walker, "Efficiency of spearcon-enhanced navigation of one dimensional electronic menus," in *International Conference on Auditory Display (ICAD08)* Paris, France, 2008.

[35] B. N. Walker and A. Kogan, "Spearcons enhance performance and preference for auditory menus on a mobile phone," in *5th International Conference on Universal Access in Human-Computer Interaction (UAHCI) at HCI International 2009* San Diego, CA, USA, 2009.

[36] T. Dingler, J. Lindsay, and B. N. Walker, "Learnability of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech," in *15th International Conference on Auditory Display (ICAD 08)* Paris, France, 2008.

[37] D. Palladino and B. N. Walker, "Learning rates for auditory menus enhanced with spearcons versus earcons," in *International Conference on Auditory Display (ICAD2007)* Montreal, Canada, 2007, pp. 274-279.

[38] M. Jeon and B. N. Walker, ""Spindex": Accelerated initial speech sounds improve navigation performance in auditory menus," in *Human Factors and Ergonomics Society (HFES 2009)* San Antonio, TX, 2009, p. TBD.

[39] P. Klante, "Auditory interaction objects for mobile applications," in *24th annual ACM International Conference on Design of Communication* Myrtle Beach, SC, USA, 2004, pp. 157-163.

[40] M. Jeon and B. N. Walker, "Optimized spindex improves both acceptance and performance with auditory menus," Georgia Institute of Technology Sonification Lab Technical Report, 2009.

[41] A. Treisman and A. Davies, "Divided attention to ear and eye," *Attention and Performance IV,* pp. 101-117, 1973.

[42] D. Gopher, M. Brickner, and D. Navon, "Different difficulty manipulations interact differently with task emphasis: Evidence for multiple resources," *Journal of Experimental Psychology: Human Perception and Performance,* vol. 8, pp. 146-157, 1982.

[43] A.-M. Bonnel and E. R. Hafter, "Divided attention between simltaneous auditory and visual signals," *Percpetion & Psychophysics,* vol. 60, pp. 179-190, 1998.

[44] S. G. Hart, "NASA-Task Load Index (NASA-TLX); 20 years later," in *Human Factors and Ergonomics Society 50th Annual Meeting* San Francisco, CA, 2006, pp. 904-908.

[45] J. Edworthy, "Does sound help us to work better with machines? A commentary on Rautenberg's paper 'About the importance of auditory alarms during the operation of a plant simulator'," *Interacting with Computers,* vol. 10, pp. 401-409, 1998.

[46] G. Kramer, "An introduction to auditory display," in *Auditory display: Sonification, audification, and auditory interfaces*, G. Kramer, Ed. MA: Addison Wesley, 1994, pp. 1-78.

Heart on the road: HRV analysis for monitoring a driver's affective state

Andreas Riener and Alois Ferscha
Johannes Kepler University Linz
Institute for Pervasive Computing
Altenberger Strasse 69
4040 Linz, Austria
{riener, ferscha}@pervasive.jku.at

Mohamed Aly
International School of Informatics Hagenberg
Softwarepark 21, Top 01
4232 Hagenberg, Austria
mohamed-abdulazim.aly@isi-hagenberg.at

ABSTRACT

Driving a vehicle is a task affected by an increasing number and a rising complexity of Driver Assistance Systems (DAS) resulting in a raised cognitive load of the driver, and in consequence to the distraction from the main activity of driving. A number of potential solutions have been proposed so far, however, although these techniques broaden the perception horizon (e. g. the introduction of the sense of touch as additional information modality or the utilization of multimodal instead of unimodal interfaces), they demand the attention of the driver too. In order to cope with the issues of workload and/or distraction, it would be essential to find a non-distracting and noninvasive solution for the emergence of information.

In this work we have investigated the application of heart rate variability (HRV) analysis to electrocardiography (ECG) data for identifying driving situations of possible threat by monitoring and recording the autonomic arousal states of the driver. For verification we have collected ECG and global positioning system (GPS) data in more than 20 test journeys on two regularly driven routes during a period of two weeks.

The first results have shown that an indicated difference of the arousal state of the driver for a dedicated point on a route, compared to its usual state, can be interpreted as a warning sign and used to notify the driver about this, perhaps safety critical, change. To provide evidence for this hypothesis it would be essential in the next step to conduct a large number of journeys on different times of the day, using different drivers and various roadways.

Categories and Subject Descriptors

H [**Information Systems**]: H.5 Information Interfaces and Presentation—*H.5.2 User Interfaces*

General Terms

User-centered design, Affective state recognition

Keywords

On-the-road studies, Driver-Vehicle interface, Electrocardiography (ECG), Emotional state recognition, HRV analysis

1. STATE-OF-THE-ART INTERFACES

The provision of a safe and a comfortable driving experience is a major concern of motor vehicle manufacturers. As the motor vehicle industry develops, more entertainment and information systems are integrated in new vehicles. These systems are aimed to make the driving experience more enjoyable and as safe as possible. However, a driver is expected to focus all his attention on road events at all times. Any activity that a driver engages in other than that is considered to be a distraction. A study conducted by Ranney et. al [25] shows that any form of distraction can cause a crash. 25 percent of the police reported crashes were due to distractions. The study classifies sources of distractions into four different categories; visual (e. g. looking away from the roadway), auditory (e. g. responding to a mobile phone), bio-mechanical (e. g. typing in a destination on a navigation device), and cognitive (e. g. daydreaming or being lost in thought).

Current car systems interfaces have a lot of disadvantages. For instance, the driver must have previous knowledge about the operation of these interfaces. Rydström *et al.* [31] reported that the operation of vehicles using different systems such as the BMW iDrive, Audi MMI or Jaguar touch screen interface took up to four times longer to use for persons unfamiliar with the interfaces than for the drivers knowing them. Additionally, the driver must pay some attention during driving to control these interfaces, which in term is a source of distraction. Another drawback of common driver assistance systems (DAS) is that they get very little or no input about the driver's emotional (or affective) state. Very little attention has been given for studying emotions in the context of driving. Nevertheless, one can envision that affective interfaces might be essential in automotive safety critical and driver assistance applications.

In an attempt to research alternative automotive interfaces, we thought about investigating the relationship between the driver's affective state and routes that are being regularly driven by him. The idea was inspired by the fact that different people feel and react differently to different roads at various times of the day. For example, we assume that most people will feel more stressed on a road with more traffic jams than a road that has a moderate traffic flow. In this paper we investigate our hypothesized claim. We present our first experiments using physiological data acquired from electrocardiography (ECG) and location data obtained using a global positioning system (GPS) device.

1.1 Attention-free Driver-Vehicle Interaction

Current vehicular interfaces are operating on a combination of either haptic, visual, or vocal modalities (Mauter and Katzki [19, p. 78], Bernsen [4, p. 2], Riener [26, p. 61f.]). These interfaces require a lot of knowledge and attention from the driver in order to interact with different car systems. Furthermore, there has been no or only little information considered about the affective state of the driver which can be gathered almost for free. Feeding the driver's affective state into different vehicular control systems might help to provide many possibilities for new vehicular applications dealing with safety, information, navigation, and entertainment. In a study conducted by Green [10], the following recommendations are made to help overcome crashes induced by in-car information systems:

(i) application and extension of driver interface regulations and design guidelines,

(ii) utilization of human factor experts, data, and methods to develop new driver-vehicle interfaces,

(iii) making greater use of usability testing,

(iv) conduction of research on and development of a workload manager which measures driving demands of a road required from the driver.

These are very important remarks that indicate that a lot of research work is yet to be done for improving the safety and usability of driver-vehicle interfaces.

As pointed out before, cognitive distractions can make the driver prone to accidents. Moreover, one can state that the emotional state of the driver falls under the category of cognitive distractions. Studying the driver's emotional state in relation to driving performance had an increasing interest from several researchers. Grimm *et al.* [11] researched on the importance and feasibility of detecting the driver's emotional state. The assumption of their study, based on cited evidence, is that different emotional states affect driving performance. Some of the emotional states were described as being positively improving the driving performance, and others as adversely affecting. A similar study was done by Cai *et al.* [17] on the feasibility of detecting driver emotions using driving simulators. Other studies by Nass *et al.* [21], Jones *et al.* [13], and Jonsson *et al.* [14] showed evidence that automotive safety can be improved by pairing an in-car voice interface output with the emotional state of the driver. Wang and Gong [34] showed the feasibility of emotion recognition in vehicular environments. In the study a driving simulator was used to elicit various emotions using driving courses and guidance voices. Most studies in this area claim that little attention has been given towards studying emotions encountered by drivers throughout the driving process.

Outline

The paper is organized as follows. In section 2 an overview of related work in emotion research and a short background on ECG is presented, in section 3 we present the experimental setting, conducted studies and a discussion of initial results. Finally, section 4 concludes the work and gives some directions for our future research.

2. AFFECTIVE STATE RECOGNITION

2.1 Emotion Research

To the best of our knowledge, there exists no scientifically agreed on definition for the notion of emotions. Finding an accepted working definition of emotions is an important issue and still under research. Understanding emotional components and their generation are made difficult by a lot of factors; describing emotions (or tagging emotions with adjectives) and interference problems (due to social pressures and expectations) are some of these factors.

The widely used definition for emotion recognition in computer disciplines was introduced by Picard [23] in the 1990s. Emotion recognition is defined as *"measuring observations of motor system behavior that correspond with high probability to an underlying emotion or combination of emotions"*. This definition is based on the fact that measuring cognitive influences is currently impossible. Nevertheless, we are able to measure physiological responses that can reflect an emotional state. This definition of emotion recognition simplifies the problem of understanding what an emotional state is. Furthermore, it is suggested to use the terms "emotional state", "affective state", and "sentic state" interchangeably in the context of emotion reasoning and computation.

Body Expressions

Affective computing is not aimed at measuring cognitive influences but to detect emotions from what is referred to as "sentic modulation". The body expresses (or modulates) an emotional state through many channels. What to be considered as a reliable source for understanding sentic expressions seems to be also debatable. A variety of motor system outputs and physiological responses have been studied with respect to emotional influence. Categorization of the main classes or the reliable sources for the purpose of emotion recognition is still debatable. Mauss and Robinson [18] classify the widely investigated channels as follows:

(i) facial expressions and whole body behavior,

(ii) vocal characteristics using features like quality, utterance timing, and utterance pitch contour,

(iii) physiological responses and other motor outputs (arising from biosignals like heart rate, blood volume pressure (BVP), pulse, pupillary dilation, respiration, skin conductance, and temperature),

(iv) subjective experience (based on self report).

Various interpretations and definitions from disciplines like psychology and philosophy are given about the notion of emotions. The recommended definition by Picard gives us a stricter domain for understanding emotions in the field of computer science. Given this definition we need to understand *(i)* what are the causes for emotions or emotion elicitation, *(ii)* what are the channels for expressing emotions, and *(iii)* how to measure emotional responses. We also need a computational model for interpreting the measured responses. Two of the widely used models in emotion research are the discrete emotion model (a basic set of emotions are assumed), see Zinck *et al.* [35, p. 2], and the dimensional model (describes different categories of emotions in three or fewer dimensions; such dimensions include arousal, valence, and control/attention), see Sebe *et al.* [32] or Barrett [2]. Arousal indicates the strength of the emotion (calmness/excitement), valence shows the pleasantness of an emotion (positive or negative), and control/attention addresses the internal or external source of emotion. Nevertheless, the names of the dimensions vary across the literature. For the convenience of mapping dimensional models to discrete emotions, Russel [30] proposed a model which is widely used by researchers in this area.

Emotion Recognition from Biosignals

Facial expressions and the voice are bodily signals that we can control. Emotions that are being conveyed through these channels can be deceiving as they can be faked by the person. For example, think about how good actors can show certain emotions in films or in the theater. Although emotions appear to be realistic, their truthfulness is debatable. The other problem with relying on such signals is the setup needed for data acquisition. Such setups rely on sensors like cameras or microphones which are, particularly in the car, constrained by factors like placement and environment conditions (like lighting, background noise, etc.), see Riener [26, p. 93f.]. For this reason, researchers currently tend to investigate other signals that can also convey an affective state such that a person can have less influence on. Such signals are commonly known as biosignals (or physiological signals) and, according to Benovoy *et al.* [3], are believed to provide more reliable means for determining emotions.

Biosignals are widely related to the autonomic nervous system (ANS), the limbic system, and other parts of the central nervous system (CNS). These systems are responsible for controlling a lot of vital activities and involuntary muscles, and are furthermore known to respond to emotional stimuli. Despite the fact that a lot of sensors exist for the acquisition of biosignals, the usage of data from such signals for emotion recognition is neither an easy nor a direct task. In relation to other approaches there are no "golden rules" yet established for the usage of biosignals for emotion recognition.

2.2 Electrocardiography (ECG)

The ANS controls smooth muscles, cardiac muscles, and secretions from various glands. Two branches of the ANS are the sympathetic and the parasympathetic system. The sympathetic system is needed for "fear, flight, fright" response (high arousal state). It is responsible to prepare the body for a stressful condition. The parasympathetic system works in the opposite way. It is responsible to put the body in a "calmer state" (low arousal state). For the normal activity, a balance is maintained between the sympathetic and the parasympathetic activities. Such variations of ANS activity can be measured using several channels. The following list by Mendes [20] represents a summary of the most widely used noninvasive methods for measuring ANS activity:

(i) electrodermal activity using skin conductance and skin potential,

(ii) cardiovascular activity using electrocardiogram, impedance cardiography, blood pressure, respiration,

(iii) pupillary responses (measurement of pupil diameter),

(iv) skin temperature,

(v) skin blood flow (volume of blood flowing in skin).

Cardiovascular activity has been used by a lot of researchers in emotion research and related fields [16, 29, 5, 28, 8, 24, 33, 12]. Electrocardiography (ECG) is one of the most common ways of measurement. The ECG records these cardiac electrical currents (voltages, potentials) by means of metal electrodes placed on the body (the recording is visualized by means of an electrocardiogram). Normally, the cardiac stimulus is produced in the sinoatrial (SA) node, that is present in the right atrium (RA). The stimulus then is passed through the RA and left atrium (LA). After that the stimulus is passed through the atrioventricular (AV) node and the bundle of His. The stimulus then passes into the left and right ventricles (LV and RV) by way of the left and right bundle branches. Finally, and according to Goldberger *et al.* [9], the stimulus is transferred to the ventricular muscle cells.

For normal cases the process of cardiac stimulus generates patterns as shown in Figure 1. The time interval between two heart beats can be calculated by observing the time between two consecutive R peaks using a QRS detector. This R-R interval is known as the inter-beat time and is used for the measurement of the heart rate.

Heart Rate Variability (HRV)

On the shortest time scale, the time between each heartbeat is irregular (unless the heart is paced by an artificial electrical source such as a pacemaker or due to medical conditions). An important tool to measure this irregularity is heart rate variability (HRV). HRV is a promising tool for applications involving medical diagnoses and stress detection. Kim *et al.* [15, 5] have reported the use of HRV statistics as to estimate mental stress. This can be applied to vehicular applications where the estimation of emotional state is required.

The tool relies on the analysis of the series of R-R interval differences. Time and frequency domain measures provide means for HRV analysis. Measures of time domain include mean, standard deviation, and root mean square of differences of consecutive R-R intervals. Frequency domain analysis represents deviations with respect to frequency. For that, several interesting frequency bands can be analyzed like the very low frequency (VLF) ($< 0.04Hz$), low frequency (LF) ($0.04 - 0.15Hz$), and high frequency (HF) ($0.15 - 0.4Hz$). VLF was indicated as being unreliable for short time intervals. The LF/HF ratio is an indicator for autonomic balance. High values are thought to indicate the dominance of sympathetic activity with vagal modulation and low values indicate dominance of parasympathetic activity. Typically, HRV analysis is done for time windows of 5 minutes or for longer periods like 24 hours. However, there is no standard mentioned for an ideal time window frame (Clifford *et al.* [6, p. 71–83]). For a dimensional model of emotions, this parameter could be a good indicator for arousal but not valence.

Mobile ECG Measurement

One might think that the measurement of ECG can be very tedious as compared to a setup available in hospitals which is mostly based on a standard 12-lead ECG. Today most mobile ECG devices used for measuring heart (or pulse) rate, heart rate variability and other biorhythm related parameters operate with three conductively coupled electrodes ("Einthoven ECG"), attached to the skin of the person and providing direct resistive contact (see Figure 2). But also their application in vehicles is almost unfeasible due to the inconvenience and lack of user friendliness (even the "ultimate" DASs necessitating the driving person to attach three electrodes every time prior boarding would not be accepted).

However, using a system operating on capacitively coupled electrodes, as for instance presented by Aleksandrowicz *et al.* [1], could avoid these restrictions. The introduced system is able to measure ECGs through the clothes, without a direct skin contact. Although the measurement system is, compared to a conventional conductive ECG measurement device, more sensitive to moving artifacts and is furthermore strongly dependent on the subject's clothing, it seems useful for at least high convenient heart rate detection in mobile fields of application. The measurement device additionally avoids skin irritation often evoked by the contact gel between skin and the electrodes. The proposed capacitive measurement system could be for example integrated into a vehicle seat with two electrodes embedded into the back, and the reference electrode integrated into the seat. This system would then operate fully autonomously and attention-free, and thus would be the missing building block for the class of implicit operating sensing systems.

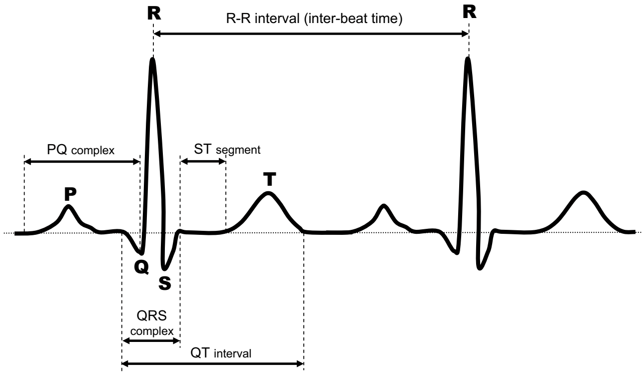


Figure 1: A normal electrocardiogram.

3. FIELD TESTS

In order to study the relationship between the driver's emotional (or affective, arousal) state and a driven route and to test the proposed framework, we conducted experiments measuring pairs of ECG/GPS for a specific route and a fixed daytime (a small variation in driver time is indispensable according to environmental parameters such as weather or traffic jams). Based on this data a "personal affective profile" for a route and a specific daytime can be compiled, indicating the "normal, balanced" state of that person for each position on the (regularly driven) track (=training set). The assumption of the research is that differing affective state values identified during a trip (testing set) represents some kind of abnormality and should be immediately forwarded to the driver as a kind of proactive notification to avoid danger situations.

A second field of application for the emotional profiles would be the utilization for any service provider. For instance, streets or road segments can be classified according to the arousal state of the collection of all drivers using this road on a certain day or at a certain time of the day regularly in order to identify the "danger-level" (or "stress-impact") of a route. For a car insurance company the aggregated state values could be used to calculate the insurance rate for this trip.

3.1 Geographic Regions of the Experiments

The on-the-road driving tests have been conducted in the greater Linz area. In order to avoid the general areas of traffic congestions, two different driving routes (inbound via the city of Altenberg, outbound via Glasau) – according to the personal preference of the test person – have been used for data acquisition. All of the test runs have been processed on these predefined courses with a distance of $20.47km$ (inbound) and $19.53km$ (outbound). Figure 3 illustrates maps of the routes driven in the experiments.

3.2 Data Acquisition

GPS traces and ECG data have been acquired in on-the-road experiments on two predetermined routes (morning and evening route) driven by a single identical person for a period of two weeks (the subject was commuting from his home to work; we only consider the workdays in our experiments). A total of 22 trips with more than 500 kilometers driven were logged and employed in this research study.

For recording electrocardiograms we used a common 3-lead ECG device "HeartMan 301" from HeartBalance AG[1]. This appliance can be easily attached to a human's body, is small-sized, light-weight and records up to 24 hours with one battery pack. Figure 2 illustrates the setup of the device on the subject. The device operates reliably and delivers high precise data in real-time at a sampling rate of $50Hz$. Data sets are either transmitted via a Bluetooth communication interface or stored in the European data format (EDF)[2] on an integrated SmartMedia memory card.

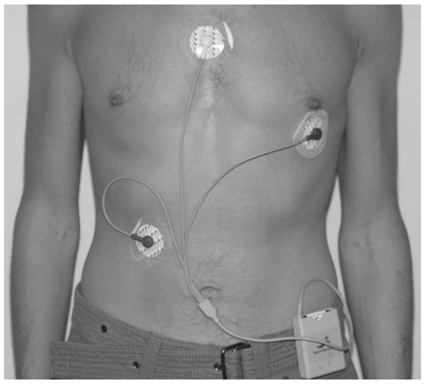


Figure 2: A 3-lead mobile ECG device "Heartman 301" attached to the test driver.

A GPS receiver ATR062x[3] with ANTARIS 4 GPS chipset, mounted nearby the front window, was used to get the vehi-

[1] `http://www.heartbalance.com/hb2/index.php?content=home`, last retrieved July 30, 2009.
[2] `http://www.edfplus.info/`, last retrieved July 30, 2009.
[3] For details on ANTARIS 4 GPS Chipsets and Sin-

Figure 3: GPS traces of the two pre-defined driving routes with subjacent maps. Tthe left image shows the morning journey (20.47km), the right one indicates the evening trip (19.53km).

cle geo-locations. The ATR062x is optimized for automotive and mobile terminal applications. GPS data is logged in the National Marine Electronics Association (NMEA) 1083 format at a rate of $1Hz$. Furthermore, the GPS time field was consulted as external synchronization basis.

3.3 Signal Processing and Feature Extraction

The ECG signal was preprocessed with a high-pass filter of $1Hz$ followed by a low-pass filter of $1,000Hz$. For the next processing steps we used BioSig[4](an open source toolkit for biomedical signal processing) in Matlab. In the beginning we analyzed the dataset mapping between raw ECG and GPS logs but no significant correlation was noticed. Therefore, we decided on using HRV analysis. In order to calculate the R-R interval series, we first must detect the R peaks throughout the entire ECG signal. For that we used a QRS complex detector provided by the toolkit and as described by Nygards *et al.* [22]. The detector returns the fiducial points of R peaks. We then used the integrated heart rate variability toolkit to calculate the LF/HF ratios as an index for autonomic balance.

GPS data was converted from the NMEA format to a simplified comma separated values (CSV) file format. This was done using GPSBabel[5] (an open source toolkit for the conversion between multiple GPS device formats). Transformed data consisted of the car latitude, longitude, speed, course, and a time stamp. The time needed to travel a route varied every day. This is due to factors like driving speed, road conditions, and traffic congestions. Therefore synchronizing data based on exact time was not possible.

In order to overcome the synchronization problem, reference routes for the morning and the evening trips were defined. These reference routes were manually plotted using Google Earth[6]. Moreover, we had to choose a good time window for segmenting and analyzing the data. We experimented with several time window sizes ranging from 1 to 5 minutes. The least time window we can use, that provided us with the best resolution, was 60 seconds (since a journey lasted between 20 to 30 minutes, a large time frame was not able to provide us with variations of LF/HF ratios over distance). With a time window of $60sec.$, the lowest frequency that can be resolved is $1/60 = 0.016Hz$ which is below the lower limit of the LF region. The highest frequency that can be resolved is calculated by applying the Nyquist constraint of $N/2T >= 0.4$, where N is the number of beats and T is the time in seconds [6, p. 79]. Applying this formula leads to a lower limit of $N = 48beats$. Our subject is a healthy adult with an average of 75 beats per minute (bpm), and since we are interested in analyzing the LF and HF bands this time window choice was appropriate.

The distance ranges (with respect to the final destination) traveled within every division were stored along with the corresponding LF/HF ratio. By the end of the experiment we had different distance ranges of $60secs.$ overlapping with each other. Finally, to calculate the corresponding LF/HF ratios of any point of the route the following was done. The distance ranges in which a route point falls were first detected (the distance of a point to the final destination was calculated and the corresponding ranges which it falls in were known by a simple comparison). Figure 4 illustrates the various 60 second ranges for a driver on two different days. Given the known ranges and the LF/HF pairs, the corresponding LF/HF ratio of a point was the mean of the LF/HF ratios across the ranges.

3.4 Discussion

After collecting and processing the datasets, we visualized the aggregated LF/HF ratios along the routes. This was done by means of a quantitative visualization on Google Earth (an illustration of the morning route visualized can be seen in Figure 7) over the driven tracks, and as simple graphs generated by Matlab. Figure 5 and Figure 6 show the corresponding ratios in relation to the distance to destination of the morning and the evening journeys respectively. As described before, we use LF/HF ratios as indicators for autonomic balance. Higher values are thought to exhibit

gle Chip GPS Receivers see `http://www.u-blox.com/products/a4chipsets.html`, last retrieved May 13, 2009.

[4]The BioSig Project, URL: `http://biosig.sourceforge.net/`, last retrieved July 30, 2009.

[5]GPSBabel, URL: `http://www.gpsbabel.org/`, last retrieved July 30, 2009.

[6]Google Earth, URL: `http://earth.google.com/`, last retrieved July 30, 2009.

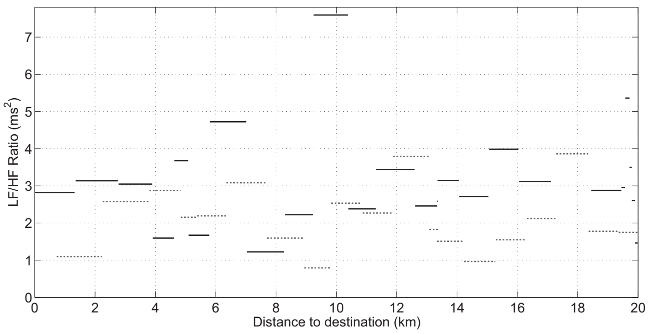


Figure 4: Morning route distance ranges and corresponding LF/HF ratios for two days.

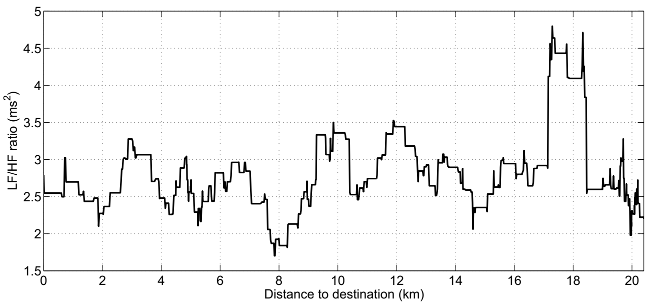


Figure 5: Morning route distance ranges and aggregated LF/HF ratios for two weeks.

higher levels of arousal (implied by increased sympathetic activity) and lower values are opt to demonstrate lower levels of arousal (as a result of the dominance of parasympathetic activity).When calculating the mean of LF/HF ratios for a span of two weeks, we get a characteristic gradient of the curves as depicted in the figures. After analyzing the routes driven and the ratios we came to some interesting observations. In fact we have no means to proof the reasons behind the phenomenon in the data. However, we try to give reasons that might be likely to exhibit the observed measurements. The analysis is done based on road characteristics noted throughout the experiment.

HRV is known to vary according to age, gender, activity, medications, and health [6, p. 71]. It is rather unclear how to differentiate between this causes, e. g. when driving at high speed. Therefore, it is not clear whether the high LF/HF ratios are caused by an increased mental load (attention on the road) or the raised activity of steering the vehicle (braking and accelerating, changing gears, steering).

The Morning Journey

At the beginning of the journey (morning, starting from home) the level of arousal is with a value of $2.6ms^2$ relatively low. The value increases for a short time, probably caused by several dangerous road crossings, and decreases again while driving at low speed in the municipality. The following section (from kilometers 2 to 4), driven on an interurban road with a speed limit of $100km/h$, directs to an LF/HF ratio between 2.5 and $3.3ms^2$. Similar curve shapes can be indicated for the other interurban road sections on the route (regions from kilometers 9.0 to 10.5 and 11.5 to 13.0). The region 4.5 to 6.0 corresponds to the city of "Hellmonsödt", driven at a speed limit of $50km/h$. The $8km$ mark between the city "Hellmonsödt" and the small town "Pelmberg" (long straight street section through a forest) corresponds to the lowest value of arousal for the entire trip ($1.7ms^2$). The road there has very light traffic at that time of the day. The most significant road segment is the distance from kilometers 17.0 to 18.6, the state of arousal here, varying between 4.2 and $4.8ms^2$, is much higher than in any other region of the curve. The reason for this is probably the incipient traffic congestion (dense traffic, but vehicles are still moving) on the borders of Linz (inbound). Driving on workdays and at the same time each day (at around 7:30AM) a traffic jam (standstill) will appear every day between the kilometers 18.6 and 20.1. This behavior is also noticeable in the Google Earth representation in Figure 7 (please note that the labels of the bar graph stands for the LF/HF value scaled by a factor of 1000 – due to a restriction of the utilized software tool). The final segment (low to very-low LF/HF ratio) is driven at walking-speed on the university parking lot, which is almost empty at this time (neither cars nor pedestrians/students).

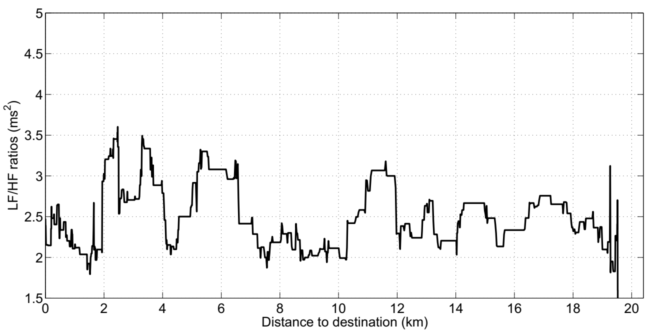


Figure 6: Evening route distance ranges and aggregated LF/HF ratios for two weeks.

The Evening Journey

The LF/HF ratios for the evening route fundamentally follow that of the morning route. The first $1.5km$ of the route, indicated by a very low state of arousal around $2ms^2$, are driven on the parking lot and a following $30km/h$ zone. It is connected to a common "city-traffic" region (route kilometers 2.0 to 6.5), showing a high LF/HF ratio of up to $3.6ms^2$. The reason for this is probably due to city traffic (outbound, around 6:30PM, high traffic density but in general no traffic jam). The region of 7.5 to $12.0km$ indicating the lowest LF/HF ratio is represented by permanent road works ($50km/h$ zone, narrow roads), but at the time of driving regular work has already been stopped for the day. The remaining route (kilometers 12.0 to 19.53) shows no more distinctive features. It can be mentioned that the apex at the end of the route (at kilometer 19.3), where the value of arousal increases from 1.8 to $3.2ms^2$, might be due to a number of hazardous curves that require maneuvering just before reaching the end point.

4. CONCLUSION AND FUTURE WORK

It is undoubted that the cognitive workload of a car driver is increasingly demanded by modern vehicular interfaces and driver assistance systems. The consequence is a possible

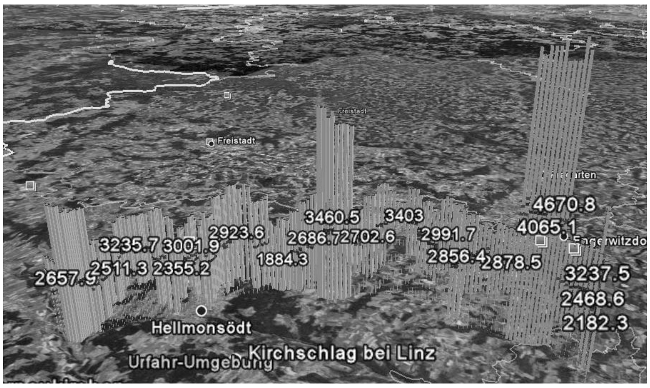


Figure 7: Visualization of the morning route using Google Earth. The values are representations for the arousal state of the driver.

threat, mainly caused by distraction (from the task of driving) due to information overload. In this study we have investigated the proof-of-possibility for the application of heart rate variability (HRV) analysis for representing the driver's affective state in terms of autonomic arousal levels in a noninvasive and a non-distractive way. The experiments we conducted lasted for 2 weeks using one driver commuting every day from his home to work place. We calculated LF/HF values from ECG data, partitioned them into $60sec.$ segments, and mapped them to the corresponding GPS coordinates. This curve, denominated as "personal affective profile", can be used to identify differences for further trips of that driver on the same route in order to notify him (or the driver assistance system) of that change.

In short, the results of the initial tests can be summarized as follows.

(i) The here presented and used metric is only a good measure for arousal. For emotion recognition a metric for representing valence is still required.

(ii) A disadvantage of using ECG (or in particular HRV) is that we had to take larger time intervals (we used $60sec.$ segments). For realtime applications this would be unfeasible (a measure with a quicker response will be needed).

(iii) We presented the potential for using one type of biosignals (ECG) as an indicator for arousal. We might consider comparing it to other ANS measures in future studies.

(iv) Using an ECG device with a sampling rate of $50Hz$ was not feasible for usage with advanced ECG analysis techniques in short time intervals.

(v) We cannot back our observed phenomenons in relation to the road characteristics with a proof. Nevertheless, the stated observations are only remarks on what we think is significant.

(vi) The subject was not feeling stressed during the experiment, which indicates that the LF/HF ratios can be used as an indicator for subconscious stress.

(vii) Higher arousal levels were noticed at roads of higher traffic volume.

As our research is still in progress, a lot of issues are still open and should be covered in the future. Our focus of research will be segmented into two directions. One part is aimed to continue the recording of ECG/GPS data on different driving routes with a larger number of recordings each (e. g. ≥ 10). For these tests it is planned to integrate, apart from ECG and GPS, other biosensors to improve data set quality. We will then repeat the conducted on-the-road studies for a certain driving route with at least one different driver in order to provide evidence for person-related differences. In addition to the "real" driving studies we will conduct tests on a predefined simulated track, e. g. by using a trace-driven experiment as described in [27] or a driving simulator. On the other hand, but concurrently in time, we will use more effort in the mapping of data and selection of algorithms with respect to improving the computational model for emotion representation and interpretation.

Acknowledgements

This work is supported under the FP7 ICT Future Enabling Technologies programme of the European Commission under grant agreement No 231288 (SOCIONICAL) and under grant FACT (Flexible Autonomic Context Technologies) of SIEMENS CT SE-2, Siemens AG, Munich.

5. REFERENCES

[1] A. Aleksandrowicz and S. Leonhardt. Wireless and Non-contact ECG Measurement System - the "Aachen SmartChair". *Acta Polytechnica, Journal of Advanced Engineering*, 47(4–5):68–71 (4), 2007.

[2] L. Barrett. Discrete emotions or dimensions? The role of valence focus and arousal focus. *Cognition & Emotion*, 12(4):579–599, 1998.

[3] M. Benovoy, J. R. Cooperstock, and J. Deitcher. Biosignals analysis and its application in a performance setting - towards the development of an emotional-imaging generator. In P. Encarnação and A. Veloso, editors, *BIOSIGNALS (1)*, pages 253–258. INSTICC - Institute for Systems and Technologies of Information, Control and Communication, 2008.

[4] N. O. Bernsen. Defining a taxonomy of output modalities from an HCI perspective. *Computer Standards & Interfaces*, 18(6-7):537–553, December 1997.

[5] P. Blobel, L. Zvanmi, and B. AKA. Estimation of Mental Stress Levels Based on Heart Rate Variability and Stress Factor in Mobile Settings. In *EHealth: Combining Health Telematics, Telemedicine, Biomedical Engineering and Bioinformatics to the Edge: CeHR Conference Proceedings 2007*. Akademische Verlagsgesellschaft, 2008.

[6] G. Clifford, F. Azuaje, and P. McSharry. *Advanced methods and tools for ECG data analysis.* Artech House, 2006.

[7] A. Ferscha and A. Riener. Pervasive Adaptation in Car Crowds. In *First International Workshop on User-Centric Pervasive Adaptation (UCPA) at MOBILWARE 2009, Berlin, Germany*, pages 1–6 (6). Springer Berlin, Heidelberg, April 2009.

[8] A. Gluhak, M. Presser, L. Zhu, S. Esfandiyari, and S. Kupschick. Towards mood based mobile services and applications. In *Proceedings of 2nd European*

Conference on Smart Sensing and Context , (EuroSSC 2007), Kendal, England, pages 159–174 (16). Springer, October 2007.

[9] A. Goldberger and E. Goldberger. Clinical Electrocardiography, A Simplified Approach. *Critical Care Medicine*, 9(12):891, 1981.

[10] P. Green. Crashes induced by driver information systems and what can be done to reduce them. In *SAE CONFERENCE PROCEEDINGS P*, pages 27–36. SAE; 1999, 2000.

[11] M. Grimm, K. Kroschel, H. Harris, C. Nass, B. Schuller, G. Rigoll, and T. Moosmayr. On the Necessity and Feasibility of Detecting a Driver's Emotional State While Driving. In *Proceedings of the 2nd Intern. Conference on Affective Computing and Intelligent Interaction (ACII '07)*, pages 126–138, Berlin, Heidelberg, 2007. Springer-Verlag.

[12] A. Haag, S. Goronzy, P. Schaich, and J. Williams. Emotion recognition using bio-sensors: First steps towards an automatic system. *Lecture Notes in Computer Science*, 3068:36–48, 2004.

[13] C. Jones and M. Jonsson. Automatic recognition of affective cues in the speech of car drivers to allow appropriate responses. In *Proceedings of the 17th Australia Conference on Computer-Human Interaction*, pages 1–10. CHISIG of Australia, Narrabundah, Australia, Australia, 2005.

[14] I.-M. Jonsson, C. Nass, H. Harris, and L. Takayama. Matching in-car voice with driver state: Impact on attitude and driving performance. In *Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, pages 173–180, Stevenson, Washington, 2005.

[15] D. Kim, Y. Seo, S. Kim, and S. Jung. Short Term Analysis of Long Term Patterns of Heart Rate Variability in Subjects under Mental Stress. In *BioMedical Engineering and Informatics, 2008. BMEI 2008. International Conference on*, volume 2, 2008.

[16] R. Lane, K. McRae, E. Reiman, K. Chen, G. Ahern, and J. Thayer. Neural correlates of heart rate variability during emotion. *Neuroimage*, 44(1):213–222, 2009.

[17] L. Li and J. Chen. Emotion Recognition Using Physiological Signals. *Lecture Notes in Computer Science*, 4282:437, 2006.

[18] I. B. Mauss and M. D. Robinson. Measures of emotion: A review. *Cognition & Emotion*, 23(2):209–237, February 2009.

[19] G. Mauter and S. Katzki. The Application of Operational Haptics in Automotive Engineering. Business Briefing: Global Automotive Manufacturing & Technology 2003 pp. 78–80, Team for Operational Haptics, Audi AG, 2003.

[20] W. B. Mendes. Autonomic nervous system. In *Methods in the Neurobiology of Social and Personality Psychology*. E. Harmon-Jones & J. Beer (Eds.) Guilford Press, in press.

[21] C. Nass, M. Jonsson, H. Harris, B. Reaves, J. Endo, S. Brave, and L. Takayama. Improving automotive safety by pairing driver emotion and car voice emotion. In *Conference on Human Factors in Computing Systems*, pages 1973–1976. ACM New York, NY, USA, 2005.

[22] M. Nygårds and L. Sörnmo. Delineation of the QRS complex using the envelope of the ECG. *Medical and Biological Engineering and Computing*, 21(5):538–547, 1983.

[23] R. Picard. Affective Computing. Technical Report 321, Cambridge, Massachusetts, MIT Media Laboratory,Perceptual Computing Section, Cambridge, Massachusetts, 1995.

[24] P. Rainville, A. Bechara, N. Naqvi, and A. Damasio. Basic emotions are associated with distinct patterns of cardiorespiratory activity. *International Journal of Psychophysiology*, 61(1):5–18, 2006.

[25] T. A. Ranney, E. Mazzae, R. Garrott, and M. J. Goodman. NHTSA driver distraction research: Past, present, and future. In *40th Annual Proceedings of the Association for the Advancement of Automotive Medicine, Des Plaines, IL*, 2000.

[26] A. Riener. *Sensor-Actuator Supported Implicit Interaction in Driver Assistance Systems.* PhD Thesis, Johannes Kepler University, Linz, Austria, January 2009.

[27] A. Riener and A. Ferscha. Simulation Driven Experiment Control in Driver Assistance Assessment. In D. Roberts, editor, *Proceedings of the 12th IEEE International Symposium on Distributed Simulation and Real Time Applications (DS-RT 2008)*, number 3425, pages 217–226 (10), Vancouver, BC, Canada, October 2008. IEEE Computer Society Press.

[28] G. Rigas, C. Katsis, G. Ganiatsas, and D. Fotiadis. A user independent, biosignal based, emotion recognition method. *Lecture Notes in Computer Science*, 4511:314, 2007.

[29] S. Rohrmann and H. Hopp. Cardiovascular indicators of disgust. *International Journal of Psychophysiology*, 68(3):201–208, 2008.

[30] J. Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161–1178, 1980.

[31] A. Rydström, P. Bengtsson, C. Grane, R. Broström, J. Agardh, and J. Nielsson. Multifunctional systems in vehicles: a usability evaluation. In A. J. Thatcher and A. Todd, editors, *Proceedings of the Fourth International Cyberspace Conference on Ergonomics (CybErg 2005), Johannesburg*, pages 768–775 (8). Ergonomics Association Press, 2005.

[32] N. Sebe, I. Cohen, and T. Huang. *Handbook of Pattern Recognition and Computer Vision*, chapter Multimodal Emotion Recognition. World Scientific, January 2005. ISBN: 981-256-105-6.

[33] J. Wagner, J. Kim, and E. Andre. From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification. In *IEEE International Conference on Multimedia and Expo, 2005. ICME 2005*, pages 940–943, 2005.

[34] J. Wang and Y. Gong. Recognition of multiple drivers' emotional state. In *ICPR'08: The 19th International Conference on Pattern Recognition*, December 2008.

[35] A. Zinck and A. Newen. Classifying emotion: a developmental account. *Synthese*, 161(1):1–25, March 2008.

Drivers' quality ratings for switches in cars: Assessing the role of the vision, hearing and touch senses

Gary Burnett
Human Factors Research Group
University of Nottingham
Nottingham, UK.
NG7 2RD
+44 (0)115 9514030
gary.burnett@nottingham.ac.uk

Ainojie Irune
Mixed Reality Lab (MRL)
School of Computer Science
University of Nottingham, UK.
NG8 1BB
+44 (0)115 9514226
aai@cs.nott.ac.uk

ABSTRACT

Sensory integration is critical to the perception of quality in automobile interior design. To investigate the relative contribution of the senses of vision, touch and hearing to the perception of quality for in-car switches, 30 participants rated eight switches taken from two vehicles when all senses were available and under various conditions of sensory deprivation: no hearing; no vision; no touch. Results indicated that touch had the greatest role to play in judgements of quality, enabling participants more easily to differentiate between the two vehicle designs. Furthermore, correlation and regression analyses for specific switches indicated that touch contributed up to three times as much to quality ratings compared to either the vision or hearing senses. Future research should aim to verify such findings and to establish which aspects of touch have particular influence.

Categories and Subject Descriptors

H.5.1 [Multimedia Information Systems]: Evaluation, Methodology; H.5.2 [Information interfaces and presentation]: User Interfaces, User-centered design, Interaction styles, Haptic I/O

General Terms

Design; Human Factors

Keywords

Quality perception; In-car switches; Affective design; Usability

1. INTRODUCTION

Designing consumer products to account for a human's affective needs is now widely recognised to be an important and growing research area within Human Factors and Human-Computer Interaction (HCI). As is commonly the case for emerging topics (particularly those which are interdisciplinary), a variety of overlapping terms exist in the literature as labels for the concept, including: affective human factors design [8]; emotional design [12]; hedonomics [4]; engineering aesthetics [11]; hedonic quality [3]; pleasure-based design [7]; and emotional usability [9].

AutomotiveUI'09, September 21-22, 2009, Essen, Germany
ACM 978-1-60558-571-0/09/009

The prevailing view relevant to all terms is that designers should consider a broader perspective of the user-product experience, given that products are increasingly associated with users' lifestyles. In particular, it is noted that official definitions of usability [5], with their emphasis on task completion measures, do not account for the full scope of human-focused qualities that a product must possess to be successful in the marketplace [7].

Such a shift in emphasis takes Human Factors and HCI into the domains traditionally considered by those working in marketing and consumer behaviour, in particular, the area of quality perception. In assessing the quality of a product, such as an automobile, users typically assimilate and synthesise information from across the senses. For cars, a common scenario in which critical quality ratings are made concerns the 'show room' experience, specifically, encounters made with the vehicle interior including interactions with the range of switches on offer. Ultimately, this multi-sensory 'contact' with a product, and the subsequent quality judgements made, will have a significant effect on overall purchasing decisions [6, 16]. For vehicles, such interactions are of particular importance given the rapid rise in the adoption of new technologies (e.g. satellite navigation) with their potentially complex user-interfaces [1].

As researchers trying to understand this situation scientifically, it is clear that a wide range of factors (relating to switch design, and task, individual and environmental issues) will have an impact on overall ratings of quality for switches in a car. Figure 1 attempts to highlight this complexity, by providing a non-exhaustive listing of factors expected to contribute to quality perception according to different categories.

Whilst useful in developing an appreciation of the problem, such an analysis provides little assistance for vehicle manufacturers attempting to design switches to maximise quality ratings. Moreover, it is clear that there are too many variables to be sensibly considered in an experimental research programme. Accordingly, there is a need to generate knowledge which enables switch designers to restrict the design space associated with this problem, that is, to focus on the specific design characteristics in further work/development which are most likely to impact on quality ratings. An understanding of the comparative role of the three key senses of vision, hearing and touch to the perception of in-car switch quality would provide such information. Designers would then be able to concentrate their efforts on the limited range of design factors relevant within specific senses.

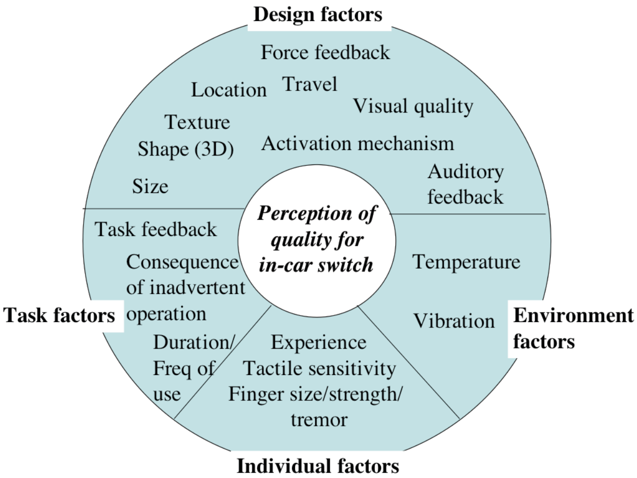


Figure 1. Range of factors relevant to perception of in-car switch quality

The aim of the present study was to establish the relative contribution of the three primary senses (vision, touch and hearing) to the perception of quality for in-car switches. In searching the literature prior to the start of the study, it was clear that there was no research reported that considered this specific issue. Therefore, the design of the study drew initially on the model of the quality perception process proposed by Steenkamp [15] where it is argued that quality perception is primarily affected by five key variables:

1. Quality cues – either physical characteristics of the product (intrinsic) such as its shape, size, and so on, or features associated with the product (extrinsic) such as its brand name, pricing, etc.
2. Quality attributes – perceived benefits of the product, such as the functions and potential social advantages it offers, either based on actual experience or expert viewpoints.
3. Interactions – the context in which the human engages with the product prior to making quality ratings, including the physical and social environment, and whether it is possible to make comparative judgements.
4. Timing – whether ratings are made pre or post consumption, that is, before or after extensive use of the product.
5. Personal perspective – individual differences, such as level of education, product knowledge, motivation, and so on.

In considering the 'show room' scenario, it is clear that quality perception in this situation is largely pre-consumption and involves static consumer-vehicle interactions with non-operational switches, often in a comparative fashion. Furthermore, both quality attributes and quality cues (intrinsic and extrinsic) are utilised in making judgements. In developing a study methodology which could investigate the relative contribution of the senses to quality perception in a design context, it was clear that intrinsic quality cues were of greatest relevance. Moreover, extrinsic quality cues (particularly branding) were likely to confound any results relevant to intrinsic cues. Therefore, a critical aspect of the methodology concerned the exclusion of extrinsic quality cues.

A further consideration in the design of the study was whether to take a survey or experimental based approach. Within the marketing domain, researchers have conducted large-scale surveys in order to ascertain the relative impact of the senses in the development of brand loyalty [10]. However, whilst the use of surveys may be appropriate when considering consumers' opinions for generic qualities such as brand image, they are not suitable as a method for investigating users' direct sensory experience with a product. Consequently, an experiment was conceived in which participants' multi-sensory encounters with in-car switches were manipulated in a systematic fashion. In this regard, it was anticipated that the presence (or conversely, the absence) of a sense would provide information regarding its relative contribution to quality perception.

2. METHOD

2.1 Participants

Thirty participants took part in the study (18 male and 12 female); the majority were aged between 18 and 35. Participants were generally experienced and regular drivers – on average, they had possessed a full UK driving licence for 11 years (SD=7.00, Range 3-30) and drove 4.3 days per week (SD=2.45, Range 2-7). None of the participants had experience with either of the two cars associated with the study.

2.2 Equipment

Two control panels were constructed for the study (Panel 1 and Panel 2) – see Figure 2. Each contained an array of switches taken from the central dashboard and driver door areas of a commercially available 'medium' (C) class car. The two arrays were chosen because it was felt they, a) represented typical examples of current switch design, and b) they provided a range of switch designs with varying sensory qualities. The panels were built to be solid (so that they did not move when switches were pressed) and portable (so that they could easily be moved to alter the presentation to the left/right of the participant). Furthermore, the panels were anonymised (i.e. by placing stickers over company names/logos) so that participants could not readily associate them with specific manufacturers. The switches were non-operational, that is, operating them did not lead to the execution of a function, such as turning the audio system on.

2.3 Experimental Design

In a repeated measures design, all 30 participants pressed a range of switches from both of the panels under each of the following conditions:

A. NO HEARING, i.e. touch and vision only. In this condition, whilst interacting with the panels, participants wore headphones through which classical music was played at a constant volume.

B. NO VISION, i.e. touch and hearing only. Participants wore a blindfold comprising blacked out goggles.

C. NO TOUCH, i.e. vision and hearing only. In this case, the switches were operated by the experimenter, whilst the participant watched and listened.

D. ALL SENSES, i.e. touch, vision and hearing.

Figure 2. Panels 1 and 2 (with examples of labelling used)

Participants experienced each of the restricted sense conditions in a counterbalanced order (i.e. five did A-B-C, five A-C-B, five B-A-C, and so on). All participants experienced the 'ALL SENSES' condition at the end of the experiment, that is, as the final condition. Figure 3 highlights the four conditions experienced in the study.

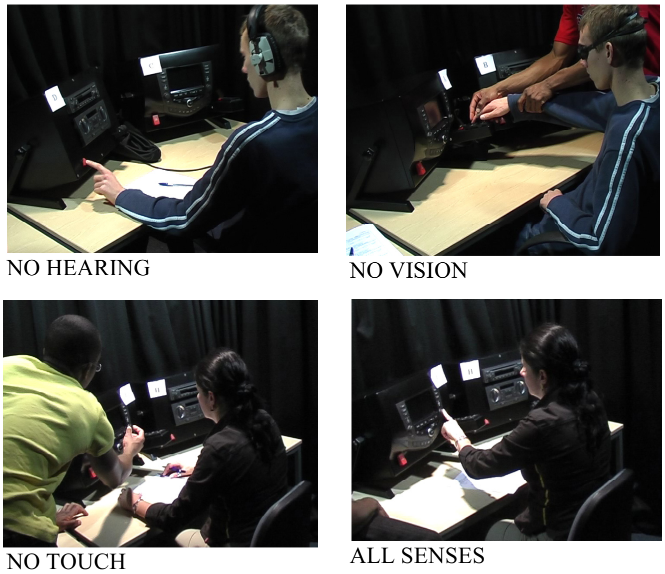


Figure 3. Conditions used in the study

2.4 Tasks

For each panel, the following eight tasks (and their associated switches) were used: hazard on/off; audio on/off; increase/decrease audio volume; seek up/down radio station; eject CD; demist the rear window; recirculate the air within the car; and raise/lower driver window. The tasks were chosen for two primary reasons: they were all common in-car secondary tasks; and the switches needed to execute the tasks were associated with a range of visual, auditory and touch characteristics.

2.5 Dependent variables

The two principal dependent variables captured in the study were quality ratings and preferences, for both individual switches and panels as a whole. Quality ratings were made using a simple five-point numerical scale with semantic anchors (very poor quality; very high quality) in which the following question was set: What sense of quality does using this switch/panel provide? In addition, participants were encouraged to speak aloud during the study and sessions were videoed in order to provide qualitative supporting data regarding drivers' opinions.

2.6 Procedure

Initially, participants completed a consent form and a questionnaire regarding their driving experience. They were then provided with an overview of the study's aims and informed in general terms of what would occur during the course of the study.

In planning the study, it was felt that participants might develop an overall view of a panel based on their initial experiences which would affect subsequent ratings. To counteract this possibility, participants were led to believe that eight panels were being rated. This was achieved by:

- Informing participants at the beginning of the study that there were eight panels to be rated, stressing that whilst they would look similar, they might differ with respect to a range of visual, sound and/or touch characteristics.
- Keeping panels hidden behind curtains until the experimenter was ready to commence a condition. During this time an assistant moved the panels around as if different panels were being introduced.
- Placing different labels onto the panels (letters A to H).

For each condition, participants were presented with the panels in pairs on a desk in front of them. For each of the eight tasks described in section 2.4 (taken in turn in a fixed order), the switches for both panels were then operated using the appropriate hand/finger for a right-hand drive vehicle (i.e. left hand and index finger for all switches apart from the driver door control). For all conditions (apart from the 'NO TOUCH' condition), participants were instructed to operate the switches 'a few times'. In the 'NO TOUCH' condition, the experimenter operated the switches (typically, two to three times) according to the participants' instructions. In the 'NO VISION' condition, participants' hands were guided towards the switches. Following an interaction with the two switches for a given task, participants were instructed to make a quality rating for the individual switch for each of the panels and to state an overall preference.

When all the eight tasks had been covered within a condition, participants rated the quality of both panels as a whole and gave a panel preference. The experimenter then moved onto the next condition. The study lasted approximately one hour in total.

3. RESULTS

3.1 Overall panel preferences and ratings

Figure 4 reports the number of participants who gave an overall preference for each of the two panels according to each of the four experimental conditions. The graph shows clearly that there was a significant preference for panel 1 over panel 2 throughout the study, apart from the 'NO TOUCH' condition where there was a marginal preference for panel 1.

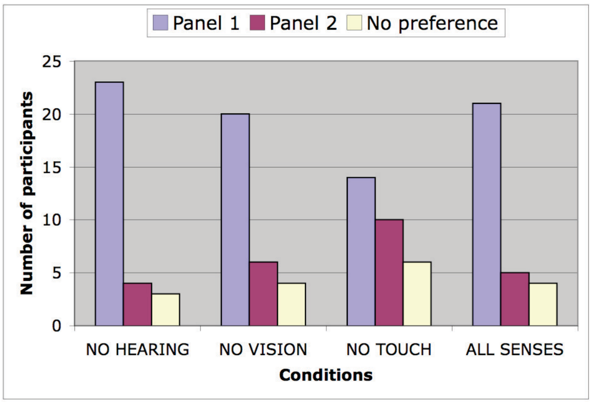


Figure 4. Responses to question, 'Which panel did you prefer for quality?'

Table 1 shows the ratings (means with standard deviations in brackets) for the two panels according to each of the four experimental conditions. The table highlights the fact that panel 1 was generally rated to be of higher quality than panel 2 for all conditions, apart from the 'NO TOUCH' condition in which there were no apparent differences in ratings. Two-tailed paired t-tests confirmed such an observation.

Table 1. Responses to question 'What sense of quality did the panel provide?' – means, standard deviations (in brackets) and p-values: where 1=Very Poor Quality; and 5=Very High Quality

Condition	Panel 1	Panel 2	Paired t-test
NO HEARING (n=30)	3.5 (0.860)	2.8 (0.711)	p<0.005
NO VISION (n=30)	3.5 (0.682)	2.6 (0.809)	p<0.0005
NO TOUCH (n=30)	3.2 (0.785)	3.1 (0.860)	p=0.27
ALL SENSES (n=30)	3.6 (0.621)	2.8 (0.714)	p<0.0001

A bivariate correlation analysis was then conducted in which the presence or absence of a sense was indicated in a spreadsheet utilising '1' or '0' respectively. Table 2 shows the Pearson correlations between the presence/absence of each of the senses and the ratings for each of panels as a whole. The table reveals that the presence of touch was significantly related to ratings for panel 1, whereas the absence of touch was significantly related to ratings for panel 2.

Table 2. Pearson correlations between absence/presence of touch and quality ratings for panels

Sense	Panel 1	Panel 2
Touch	0.16*	-0.21*
Vision	-0.02	0.16
Hearing	-0.05	0.01

* p<0.05 (two-tailed)

3.2 Individual switch preferences and ratings

With respect to the preference data for individual switches, Table 3 provides a summary of the results showing the switches where:

- There was a strong preference for a switch from panel 1 over the equivalent switch from panel 2, defined as occurring when at least two thirds of participants (20 from 30) indicated that they preferred panel 1
- There was a strong preference for a switch from panel 2 over the equivalent switch from panel 1, occurring when at least two thirds of participants preferred panel 2
- There were no strong preferences for a switch from either of the two panels, that is, when neither of the above criteria could be applied

Table 3. Responses to question 'Which switch did you prefer for quality?' – listing of switches

Condition	Panel 1 strongly preferred c.f. Panel 2	Panel 1 switch similar prefs c.f. panel 2	Panel 2 strongly preferred c.f. Panel 1
NO HEARING	Hazard, Window, Seek, Audio on/off, Eject CD, Recirculate	Volume, Demist	
NO VISION	Hazard, Window, Seek, Eject CD, Recirculate	Audio on/off, Volume, Demist	
NO TOUCH	Hazard, Window, Seek	Audio on/off, Volume	Recirculate, Demist, Eject CD
ALL SENSES	Hazard, Window, Seek, Eject CD, Recirculate, Volume	Audio on/off, Demist	

A similar analysis was conducted for the rating data utilising two-tailed paired t-tests and the results are shown in table 4.

Table 4. Responses to question 'What sense of quality did the switch provide?' – listing of switches

Condition	Panel 1 switch rated > than Panel 2 switch*	Panel 1 switch rated the same as Panel 2 switch	Panel 2 switch rated > than Panel 1 switch*
NO HEARING	Hazard, Window, Seek, Eject CD, Recirculate	Audio on/off, Volume, Demist	
NO VISION	Hazard, Window, Seek, Eject CD, Recirculate	Audio on/off, Volume, Demist	
NO TOUCH	Hazard, Window, Seek	Audio on/off, Volume, Recirculate, Eject CD	Demist
ALL SENSES	Hazard, Window, Seek, Eject CD, Recirculate, Volume	Audio on/off, Demist	

* $p<0.05$ (two-tailed)

A bivariate correlation analysis was conducted to assess the relationship between the presence/absence of each of the three senses and quality ratings for each of the individual switches (from both panels). Table 5 shows the results of this analysis for those switches where at least one significant correlation occurred (according to a two-tailed test).

Table 5. Pearson correlations between absence/presence of touch and quality ratings for key switches

	Panel 1		Panel 2		
Sense	Eject CD	Re-circulate	Seek	Eject CD	Demist
Touch	0.32**	0.17*	-0.35**	-0.33**	-0.19*
Vision	0.05	0.02	0.15	0.24**	0.16
Hearing	-0.22*	-0.07	0.10	0.05	0.08

* $p<0.05$ (two-tailed); ** $p<0.01$ (two-tailed)

As the "Eject CD" switch was associated with significant correlations for at least two of the senses, it was decided to conduct a linear multiple regression analysis for this switch to assess the relative contribution of the different senses to overall ratings. This revealed that the three senses accounted for a significant amount of the variance in quality ratings for the "Eject CD" switch for both panel 1: $F(3, 116)=5.69$, $p<0.001$ and for panel 2: $F(3,116)=5.66$, $p<0.001$. In both cases, the senses accounted for 13% of the variance in ratings. With respect to the specific contribution of each sense, touch was the only sense in which the contribution was significant, and table 6 shows the standardised coefficients for each of the senses for the "Eject CD" switch for each panel.

Table 6. Standardised coefficients (Beta) for each sense for Eject CD switch for each panel

	Panel 1		Panel 2	
Sense	Beta	Sig level	Beta	Sig level
Touch	0.34	p<0.005	-0.27	<0.05
Vision	0.11	p=0.19	0.16	p=0.14
Hearing	-0.06	p=0.56	0.01	p=0.89

3.3 Order analysis

During the informal scanning of the data, it was clear that an order effect existed in the results, focused specifically on the 'NO VISION' condition. A more detailed analysis revealed that preferences and ratings for the 'NO VISION' condition were significantly different dependent on whether this condition was experienced first or as the second/third condition. Figure 5 and Table 7 summarise this finding.

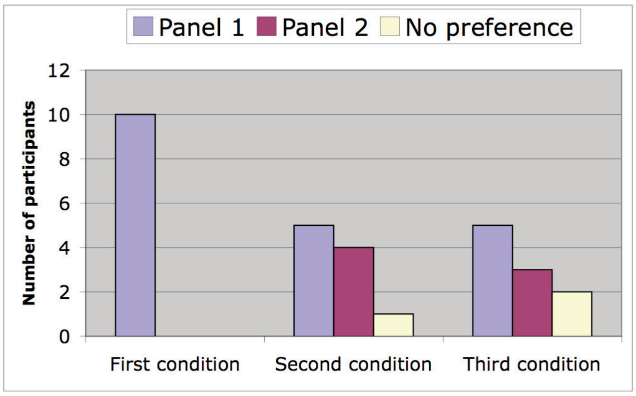


Figure 5. Responses to question 'Which Panel did you prefer for quality?'- NO VISION condition only, split by order

Table 7. Responses to question 'what sense of quality did the panel provide?' NO VISION condition only, split by order – means, standard deviations (in brackets) and p-values: where 1=very poor quality; and 5=very high quality

Condition	Panel 1	Panel 2	Paired t-test
1st condition (n=10)	3.8 (0.421)	1.8 (0.632)	P=0
2nd condition (n=10)	3.3 (0.622)	3.0 (0.632)	P=0.08
3rd condition (n=10)	3.5 (0.926)	3.1 (0.333)	No difference

4. DISCUSSION

4.1 Which sense provides the greatest contribution?

Taking the results as a whole, it is argued that touch provides the greatest contribution to drivers' ratings of switch quality in a static situation. Indirect evidence is apparent from participants' preferences and ratings of each of the two panels, which were similar for all conditions, apart from the situation in which participants were deprived of the ability to use their sense of touch. When touch was removed, participants did not generally differentiate between the panels with respect to the quality of the switches. This is despite the fact that switches from the two panels differed considerably in relation to their visual (e.g. size, shape) and auditory (e.g. amplitude/frequency of feedback sounds) characteristics.

More direct evidence was established from the correlation analyses where the factor of touch had a significant positive relationship with quality ratings for panel 1, and a significant negative relationship with ratings for panel 2. In other words, the touch-related characteristics of panel 1 led to increased ratings of quality, whereas the opposite was true for panel 2. No such relationship was observed for the other senses. More detailed assessments for individual switches found that specific switches were associated with a strong touch 'sensitivity', that is, touch has a considerable positive or negative relationship with quality ratings. For instance, according to a regression analysis conducted for the "Eject CD" switch, the sense of touch provided approximately three times as great a contribution to ratings, as compared with the other senses.

Final support for the significance of touch was revealed in an unexpected order effect. Those participants who experienced the panels first without vision were noticeably more differentiating in their preferences and ratings for this condition, as compared to those who had seen the panels earlier in the experiment. With a 'blind' initial experience, it is argued that participants were acutely sensitive to differences in the touch-related characteristics of the switches (also noted by Burnett and Porter [1]). In contrast, when the non-vision condition occurred at a later stage, a visual mental model of the panels might have already been developed, which was used in subsequent quality judgements.

4.2 Why is touch so important?

Returning to Steenkamp's model of the quality perception process [15] it was apparent that perhaps the most significant reason for the importance of touch concerns the nature of the human-switch interaction. Touch is evidently the only sense from the three investigated in this study which necessitates a 'close' physical interaction, that is, the human must be near to the object (in this case a switch) in order to use the sense. Burnett and Porter [1] have made this point in stressing the need for utilising a greater range of touch and kinaesthetic cues in in-car control design, in particular as a means of enabling drivers access to new technology systems (e.g. navigation, email, Internet services). In making the argument, Burnett and Porter cite research from the Virtual Reality domain in which haptic interfaces have been shown to increase users' sense of emotional involvement in collaborative tasks, in relation to traditional visual and auditory interfaces [14].

Furthermore, it was evident in many of the comments made by participants during the study that the intrinsic quality cues associated with the sense of touch were fundamental to drivers' ratings of in-car switches. Interestingly, in making a quality judgement some participants were clearly using intrinsic touch-related cues, but were also concerned with absent extrinsic quality cues, notably regarding pricing. The following comments provided by two participants highlight the importance of intrinsic touch cues as well as the significance of extrinsic quality cues for an image-related product such as an automobile [15].

> "This one [Panel 2] felt cheaper to me. Some of the buttons seem like they would end up breaking after not too long"

> "[The two panels] are quite similar in different things, but there are a couple of buttons which really deteriorate the whole thing [for Panel 2]. So, for example the Seek button here feels very cheap… very cheap and very tacky"

Further issues concerned the context in which the ratings were being made. Despite the fact that ratings were being made in a static situation, some participants considered the importance of touch when using in-car switches whilst driving. As commented by one participant:

> "Once I know my radio I want to be able to do it with no visual input at all, just feel the thing. So, I suppose… size of buttons will be a lot more important. I don't normally stare right at my radio… unless I'm stationary obviously. Most of the time I'm… fiddling around hoping I'm pressing the right button"

The findings in relation to specific switches are also of interest, as they highlight key influences for the overall panel results. The preferences and ratings for switches associated with three functions ("Demist", "Recirculate" and "Eject CD") were largely the same for all conditions, apart from the situation in which touch cues were absent. Without touch, participants' views altered from a general preference for panel 1 to a preference for panel 2. It was unsurprising that these three switches were associated with similar trends in the data as they had comparable designs (within a panel). Nevertheless, across the two panels there were considerable differences between the three switches, in particular in relation to their touch characteristics (e.g. the surface texture, force/travel relationships). In relation to the participants' experiences, the three switches for panel 1 had what was commonly referred to as a 'soft' feel, whereas the switches for panel 2 were often referred to as 'clicky' or 'harsh'.

4.3 What future work is required?

The significance of touch in the judgement of the quality of in-car switches raises the question: Which characteristics of touch are most important? In addressing this question, it must first be noted that a number of design characteristics will be of relevance, for instance, force feedback and travel distributions, switch lateral stability, texture, size, shape, and so on. Related work within vehicles has taken a Kansei Engineering approach in order to identify variables considered to be of relevance to the perception of quality for the touch factor, either of seat fabrics [2], or of surface materials on components such as the steering wheel [16]. Clearly, research is required which focuses on the relative priority of touch characteristics for in-car switch panels.

A further key issue concerns individual differences in the perception of quality for in-car switches. Whilst the results of this study indicate that touch had the greatest role across participants, it was apparent from the spread of data that this was not a universal truth. In this respect, research in the marketing area is of particular interest. Peck and Childers [13] have developed a questionnaire which aims to establish the 'need for touch', that is, the extent to which people require touch-related information when interacting with a potential product. Evidently, there is a need to understand how consumers with varying 'need for touch' preferences are likely to respond to in-car switch designs with differing degrees of tactile features.

Whilst important as an initial study in this area, it must also be recognised that a range of limitations exist in the present study. Consequently, there is a requirement to verify results. Three key considerations for future work include:

1. The need for a wider range of switch characteristics. A concern with this study is that the results are unique to the panels and tasks utilised.
2. The need to utilise switches fitted in representative locations within a vehicle. Whilst the use of panels enabled easy experimental manipulation, it was not possible to arrange them in the same orientations in which they would be operated in a vehicle.
3. The need to consider the relationship between the consistency of switch characteristics and the effect on subsequent ratings. That is, do ratings for an overall design improve or reduce if switches look, feel and/or sound the same? In this study, consistency of switch design was not manipulated as an independent variable, yet it was evident from some participants' comments that it had an influence on ratings.
4. The need to consider the specific influences of the driving task on the perception of quality for in-car switches. Whilst this paper has argued the importance of the initial showroom experience for vehicle purchases, comments made by participants highlight the impact that driving conditions will have ultimately on quality perception (e.g. due to varying vibration, noise, illumination).

5. CONCLUSIONS

On the basis of the results described in this paper, it is argued that the most important sense to get right in designing in-car switches is touch. If the touch-related characteristics of the switches of a vehicle are positively received, this can significantly enhance quality ratings. Perhaps more importantly, if the touch aspects of switches are viewed in a negative light, quality ratings can be considerably reduced. Moreover, quality ratings for an overall design can be strongly affected by the judgements made on specific switches - particularly those with well or poorly regarded touch characteristics.

Touch is considered to be of particular significance in this context for two key reasons. Firstly, there is an intimacy in the use of touch which is congruent with the nature of making a subjective quality rating. Secondly, touch-related feedback can be critical in the visually demanding driving situation, a fact which consciously or subconsciously affects drivers' views on the quality of in-car switches.

It is worth noting that the results of this study are of particular significance given the recent trend for the use of touchscreen technology within vehicles. Traditional touchscreens provide minimal touch cues to users (only the feedback of pressing against a solid object) and instead place an emphasis on visual and auditory information during interaction. The results of the present study suggest that such devices fail to provide cues to drivers that would be considered important in this context. Haptic touchscreens now exist though (see [17]) and it would be of interest to examine their use in a vehicle environment. Such research could investigate the design variables that influence quality perception with such technology. Moreover, it would be extremely worthwhile to consider whether haptic touchscreens reduce visual distraction in relation to traditional touchscreens.

As a final point, it is also important to note that vision and hearing clearly also impact on quality ratings for in-car switches and should not be ignored in the design process. Whilst it is argued that touch provides the greatest relative contribution, the experience of using an in-car switch is clearly multi-sensory and switch designs with inappropriate visual and/or auditory characteristics (e.g. garish colours, high frequency sounds) will inevitably be considered to be of poor quality.

6. ACKNOWLEDGEMENTS

The authors would like to acknowledge the support of Honda R&D Europe Ltd (in particular, Katharine Prynne and Steve Watson) in the funding and development of the methodology for this study. Furthermore, we would like to thank Michael Pettitt for assistance in the running of the study and all the participants for providing their time.

7. REFERENCES

[1] Burnett, G.E., Porter, J.M., 2001. Ubiquitous computing in cars: designing controls for non-visual use. International Journal of Human-Computer Studies. 55(4), 221-231.

[2] Giboreau, A., Navarro, S., Faye, P., Dumortier, J., 2000. Sensory evaluation of auomotive fabrics: the contribution of categorising tasks and no-verbal information to set up a descriptive method of tactile properties. Food Quality and Preference. 12, 311-322.

[3] Hassenzahl, M. 2001. The effect of perceived hedonic quality on product appealingness. International Journal of Human-Computer Interaction. 13(4), 481-499.

[4] Helander, M.G., Tham, M.P., 2003. Editorial – Hedonomics-affective human factors design. Ergonomics. 46(13/14), 1269-1272.

[5] ISO. 1999. ISO 9241: Ergonomic requirements for office work with visual display terminals. Part 11: Guidance on usability. International Organization for Standardization. (1999).

[6] Jindo, T., Hirasago, K., 1997. Application studies to car interior of Kansei engineering. International Journal of Industrial Ergonomics. 19, 105-114.

[7] Jordan, P.W. 2000. Designing pleasurable products. Taylor and Francis: London.

[8] Khalid, H.M. 2004. Guest editorial: Conceptualizing affective human factors design. Theoretical Issues in Ergonomics Science. 5(1), 1-3.

[9] Kim, J., Moon, J.Y., 1998. Designing towards emotional usability in customer interfaces - trustworthiness of cyber-banking system interfaces. Interacting with Computers. 10, 1-29.

[10] Lindstrom, M. 2005. Brand Sense – How to build powerful brands through touch, taste, smell, sight and sound. Kogan Page: London.

[11] Liu, Y. 2003. Engineering aesthetics and aesthetics ergonomics: Theoretical foundations and a dual-process research methodology. Ergonomics. 46(13/14), 1273-1292.

[12] Norman, D. A. 2004. Emotional Design: Why we love (or hate) everyday things. Basic Books: New York.

[13] Peck, J., Childers, T.L., 2003. Individual differences in haptic information processing: The 'need for touch' scale. Journal of Consumer Research. 30, 430-442.

[14] Sallnas, E. 2000. Supporting collaboration in distributed environments by haptic force feedback. Proceedings of 1st International workshop on Haptic Human-Computer Interaction, pp. 12-16. Glasgow, Scotland: Glasgow Interactive Systems Group (http://www.dcs.gla. ac.uk/haptic).

[15] Steenkamp, J.E.M. 1990. Conceptual model of the quality perception process. Journal of Business Research. 21, 309-333.

[16] Yun, M.H., You, H., Geum, W., Kong, D., 2004. Affective evaluation of vehicle interior craftsmanship: systematic checklists for touch/feel quality of surface-covering material. Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting.

[17] Hoggan, E. Brewster, S. Johnston, J, 2008. Investigating the Effectiveness of Tactile Feedback for Mobile Touchscreens Proceedings of Computer-Human Interaction conference (CHI), pp. 1573-1582. ACM Press.

An On-Road Assessment of the Impact of Cognitive Workload on Physiological Arousal in Young Adult Drivers

Bryan Reimer*, Bruce Mehler, Joseph F. Coughlin, Kathryn M. Godfrey and Chuanzhong Tan
MIT AgeLab & The New England University Transportation Center
77 Massachusetts Avenue, E40-291
Cambridge, MA 02139
*Corresponding author: reimer@mit.edu (617) 452-2177

ABSTRACT
In this paper, we describe changes in heart rate and skin conductance that result from an artificial manipulation of driver cognitive workload during an on-road driving study. Cognitive workload was increased systematically through three levels of an auditory delayed digit recall (n-back) task. Results show that changes in heart rate and skin conductance with increasing levels of workload are similar to those observed in an earlier simulation study. Heart rate increased in a step-wise fashion through the first two increases in load and then showed a less marked increase at the highest task level. Skin conductance increased most dramatically during the first level of the cognitive task and then appeared to more rapidly approach a ceiling (leveling) than heart rate. Findings further demonstrate the applicability of physiological indices for detecting changes in driver workload.

Categories and Subject Descriptors
J4 Social and Behavioral Sciences: Psychology; J.7 Computers in Other Systems: Real time; H5.m Information Interfaces and Presentation: Miscellaneous

General Terms
Measurement, Performance, Experimentation, Human Factors.

Keywords
Detecting driver state, cognitive workload, driving performance, physiology, driver distraction.

1. INTRODUCTION
The increase in the complexity of information available to drivers through in-vehicle interfaces and handheld devices has the potential to increase driver workload to a level at which driving performance begins to decline, thus increasing accident risk [3, 16]. As the percentage of older drivers increases, the advent of autonomous driving systems, and continued surge of secondary demands, active workload managements systems will increasingly need to monitor and manage driver state [1, 12]. Examples of early generation systems include Volvo's Intelligent Driver Information System [5] and Saab's ComSense [13]. These systems primarily help drivers prioritize information and delay the presentation of dynamic content based on the driving situation. Future workload management systems may better adapt to changing demands by incorporating real time measures of individual drivers' capacity. Physiological measures are a noninvasive method of characterizing relative cognitive workload [3] and have been suggested as being complementary to driving performance based measures in a more comprehensive assessment of driver workload [16]. Further, physiological measures have been frequently described as being potentially more sensitive to initial changes in workload then performance based measures [2, 7, 8, 15] although only limited published data is currently available to support this assertion [8]. In this research, we further explore the capability of physiological indices to discriminate subtle changes in driver workload.

2. BACKGROUND
Theoretically, performance and arousal have an inverted U relationship (Yerkes-Dodson law); performance increases with arousal up to an optimal point and then declines as workload and the arousal associated with it continues to build. This point or, more broadly, optimal operating range, will vary with differences in operator capacity and other individual and situational factors. Drivers too often function in under aroused or distracted states that result in suboptimal arousal levels and a higher potential for accidents [16]. Real-time detection of an operator's workload might be used to provide feedback that attempts to help the driver self-manage the demands of operating the vehicle and associated tasks, thereby optimizing workload and enhancing automotive safety [12].

In a series of driving studies, heart rate and other physiological indices have been shown to be responsive to increased cognitive demand [2, 4, 7]. However, the literature on the relationship between heart rate, other physiological indices, and driving performance appears to be mixed. Lennenman, Shelley & Black [7] found that simulated driving performance did not degrade with increasing workload and heart rate. Engström, Johansson & Östlund [6] reported changes in physiological measures that were inconsistent across dual task conditions, visual vs. cognitive, and between locations, simulation vs. field. In a simulation study in our lab, Mehler et al. [8], heart rate and skin conductance detected changes in workload prior to a discernable decrement in driving performance. A general decrease in driving performance was only apparent at the highest level of secondary task demand. In a field study using the same secondary task presentation, Reimer [11] showed that changes in visual attention appear to be another method of detecting levels of secondary cognitive workload. Consistent with the earlier simulation work [8], a decrease in driving performance was only found at the highest level of workload. In the present paper, we report on physiological data collected during this study.

AutomotiveUI'09, September 21-22, 2009, Essen, Germany
ACM 978-1-60558-571-0/09/009

3. METHODS

3.1 Participants

Twenty-six drivers between 22 and 27 years of age participated in the study (M = 23.85, SD = 1.57). Participants were required to read and sign an approved informed consent form, to present a valid driver's license, and attest to having their license for more than three years, driving more than three times per week and being in good health. A research assistant ensured that participants clearly understand and spoke English. Participants were excluded if they had been involved in a police reported accident in the past year, needed glasses to drive, or were taking a medication that caused drowsiness. Participants were recruited in the greater Boston area using online and newspaper advertisements.

3.2 Apparatus

The experiment was conducted in the MIT AgeLab "Aware Car", an instrumented Volvo XC 90 equipped with a customized data acquisition system designed for time synchronized measurement of vehicle, driver and environmental factors. Data capture was facilitated through a number of embedded sensing systems including a MEDAC System/3 instrumentation unit and NeuGraph software (NeuroDyne Medical Corporation, Cambridge, MA) for physiological measurement. In addition to capturing data, the system included functionality for manual and time based triggering that was used for the presentation of an auditory prompt / verbal response task.

Three levels of difficulty of a delayed digit recall task (n-back) were employed to present drivers with a low, moderate and high level of secondary cognitive workload (0-, 1-, & 2-back respectively). The presentation order of the levels was "ramped-up" to systematically increase the demands on the driver. Further details on the development and presentation of the task can be found in Reimer [11].

3.3 Procedure

The experiment consisted of two counterbalanced assessments of identical presentations of the secondary task. One presentation took place in a driving simulator and the second, reported on here, in a field vehicle.

A modified lead II configuration was employed for EKG recording; the negative lead was placed just under the right clavical (collar bone), ground just under the left clavical, and the positive lead on the left side over the lower rib. The skin was cleaned with isopropyl alcohol and standard pre-gelled silver/silver chloride disposable electrodes (Vermed A10005, 7% chloride wet gel) were applied. Skin conductance level was measured utilizing a constant current configuration and non-polarizing, low impedance gold plated electrodes that allowed electrodermal recording without the use of conductive gel. Sensors were placed on the underside of the outer flange of the middle fingers of the non-dominant hand and held in place with medical grade paper tape.

During the field portion of the experiment, a research associate seated in the back of the vehicle observed the participant for signs of fatigue and other unsafe driving behaviors that could compromise safety. The research associate also operated the data collection equipment and provided driving direction. After leaving MIT, participants drove for approximately 30 minutes before stopping for a short break. Following the break, the experimental section of the protocol began. Instructions and other components of the experiment were prerecorded and played automatically over the vehicle sound system. The eight segments of the experiment and the time allotted to each appear in Table 1. The posted speed limit was 65 mph (104.60 km/h).

Table 1. Experimental Protocol Overview

Segment	Duration (min:sec)	notes
Baseline	10:00	Minutes 7:22 – 9:30 used for the analysis period.
0-back training	1:04	Instructions and one 10 digit practice trial
0-back test	2:08	Four 10 digit evaluation trials
1-back training	2:08	Two repetitions of instructions and a 10 digit practice trial
1-back test	2:08	Four 10 digit evaluation trials
2-back training	3:12	Three repetitions of instructions and a 10 digit practice trial
2-back test	2:08	Four 10 digit evaluation trials
Recovery	7:00	Minutes 5:00 – 7:08 used for the analysis period.

3.4 Data Analysis

Due to recording issues, one participant's heart rate data and a second participant's skin conductance data were unavailable. A variety of algorithms were applied to the raw EKG and skin conductance data to remove noise and identify heart rate. Heart beats were detected using EKG Wave Editor release 1.8 (NeuroDyne Medical Corporation, Cambridge, MA), a software package that identifies R-wave peaks in the raw EKG signal and provides editing functionality. Processed records were reviewed by trained research associates and skipped and double beats were edited to provide a normalized heart rate record following general guidelines recommended for heart rate variability analysis [10]. To ensure accuracy and consistency, a second review of all records was performed by the second author.

Skin conductance is a "smooth" physiological signal characterized by relatively slow changing tonic levels and phasic sign wave like peaks associated with discrete arousal events. Artifacts associated with deformation of the sensor skin interface due to movement of the fingers or contact with the steering wheel typically appear as abrupt signal changes that are easily visualized. Artifact removal was carried out by first filtering skin conductance data using a wavelet decomposition with a coif5 mother wavelet decomposed to level 3 using MatLab [9]. The signal was reconstructed using the level 3 approximation and detail coefficients using a similar coif5 mother wavelet. This multi-level decomposition of the raw SCL signal into successive approximation and detail coefficients allows the breakdown of the SCL data into various lower resolution components of the original signal. This procedure allows the removal of high frequency artifacts without compromising the low frequency information in the SCL data. The second author directed manual removal of remaining low frequency movement artifacts. During this review, two cases were classified as outliers and dropped from the analysis.

Statistical comparisons were computed with SPSS 16 using a repeated measures general linear model (GLM) procedure and a Greenhouse-Geisser correction for models that violated the assumption of sphericity. Gender was evaluated and later dropped from the analyses presented here after being established as a non-significant factor (p>.05). Pairwise t-tests were computed for significant results with a least significant difference (LSD) adjustment for multiple comparisons. Means and standard deviations reported in the figures are computed across participants where data exist for all levels of the comparison.

4. RESULTS

4.1 Heart Rate

Mean and standard deviation values by period are listed in Table 2 and mean heart rate is displayed graphically in Figure 1. Period significantly impacted heart rate, $F(1.97, 47.21) = 30.95$, $p<.001$, in a manner consistent with the position that heart rate varies with the level of workload. Overall, heart rate increased by 3.1 beats per minute (bpm) from baseline to the low secondary workload level (0-back task), by an additional 4.5 bpm at the moderate workload level (1-back task) and finally by another 1.1 bpm at the high workload level (2-back task). This represents an 8.7 bpm increase in heart rate from baseline to the highest level of secondary task engagement. All pairwise comparisons across the workload periods, except the change between the moderate and high workload periods, are significantly different (p<.01). Five minutes following the highest workload period heart rate recovers to within 0.6 bpm of the baseline value (p>.05). As would be expected with the variations in how participants respond to the secondary cognitive task, the standard deviation of heart rate increases with each level of task demand.

Table 2. Summary of Physiological Response Measures

	Heart Rate (beats / min)	Skin Conductance (micromhos)
Baseline	75.4 (9.2)	13.7 (4.8)
0-Back	78.4 (10.7)	15.0 (5.5)
1-Back	82.9 (11.9)	15.3 (5.7)
2-Back	84.0 (12.8)	15.3 (5.4)
Recovery	76.0 (9.2)	14.6 (5.2)

Note: Means with the standard deviations in parenthesis.

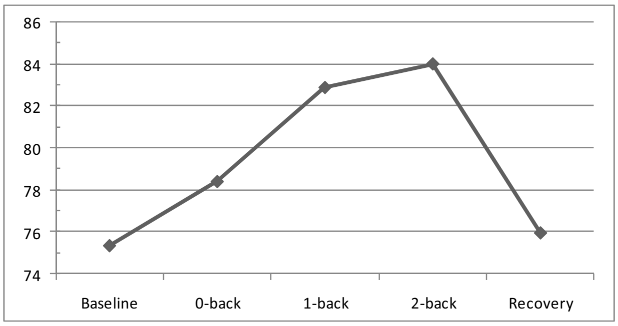


Figure 1. Mean heart rate (bpm) by period

Changes in mean heart rate from baseline to the highest level of task demand vary across subjects from -1.9 to 21.3 bpm. In two cases heart rate decreased. In one of these, the participant clearly had difficulty remaining engaged in the 2-back task, committing errors in 11 of 32 potential responses. This error rate was the largest exhibited in the sample. In the second case, the individual did not appear to be physiological responsive to the overt changes in demand, with heart rate varying only 1.9 bpm across the five periods. In five cases heart rate increased by less than three bpm and by 3 to 10 bpm in 8 cases. In the 10 remaining cases heart rate increased by over 10 bpm between the baseline and high demand periods.

Comparing the data generated in this field study to results presented in Mehler et al. [8], incremental increases in heart rate across the demand levels appear strikingly similar (3.1, 4.5, and 1.1 vs. 3.1, 4,7, and 1.1 bpm respectively). Overall heart rate was higher in real vs. simulated driving, averaging 4.9, 4.8, 4.6 and 4.6 bpm higher over the first four periods respectively.

4.2 Skin Conductance

Mean skin conductance values by period are presented in Figure 2. As with heart rate, there is a significant effect of period, $F(2.32, 50.92) = 7.98$, $p=.001$. Continuing the pattern seen in the simulation study [8], the largest percentage increase in skin conductance occurs between the baseline and the initial / lowest secondary workload period (0-back). This mean change of 1.3 micromhos represents over 80% of the total rise in electrodermal activity observed across the three cognitive task periods. The incremental increase from the 0-back to the 1-back period was a more modest 0.3 micromhos and there was no additional increase in the mean during the 2-back. Pairwise comparisons show that baseline skin conductance differs from all other periods (p<.005 or less). However, skin conductance levels across the three tasks are not significantly different (p<.05) from each other. In contrast with heart rate, where there was essentially a return to baseline during the recovery period, skin conductance shows a more modest decrease. In fact, the change in mean values for skin conductance level during the 2-back and recovery period show a trend toward recovery but are not significantly different (p=.07).

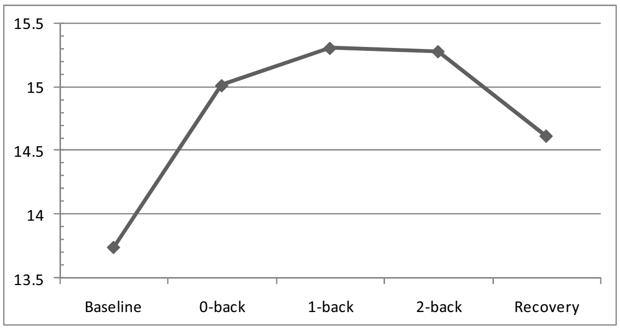


Figure 2. Mean skin conductance (micromhos) by period

5. CONCLUSION

Heart rate and skin conductance measures are presented as indicators of changes in driver workload associated with a secondary cognitive task. Consistent with Mehler et al. [8], heart rate appears to be a robust measure of incremental changes in real time workload associated with the secondary task. The current study replicates our simulation work by showing a remarkably consistent pattern of heart rate change in an on-road setting using the same protocol. Further, it extends this work by presenting data showing good recovery of heart rate to baseline levels following the cessation of the secondary load. Skin conductance data collected in the field also show a pattern similar to that seen in the simulation data. Skin conductance is quite sensitive to an initial increase in cognitive demand but appears to reach a ceiling effect more quickly. It is interesting to observe that the increase from baseline to the initial / low demand 0-back task was 2.4 micromhos in the simulator versus 1.3 on-road. One explanation might be that the base level of arousal during actual driving is greater than in the simulator which would reduce the magnitude of the skin conductance level change if it tends to reach a ceiling level more readily. Skin conductance also differs from heart rate in that it appears to show a less rapid recovery following the completion of the secondary tasks. It may be that this slower recovery represents some remaining emotional engagement with the secondary tasks that is more evident in electrodermal activity than in heart rate.

The results presented here, in combination with task performance and vehicle performance data collected in this study [11], provide additional clarity as to the potential role of physiology in the detection of cognitive workload. Reimer [11] shows that drivers remained engaged in all levels the secondary tasks, while driving performance, characterized through lane keeping and forward velocity, only exhibits signs of deterioration at the highest level of secondary workload. A significant decrease in the standard deviation of gaze was also observed at this time. The physiological data presented here show that only a modest additional increase in heart rate occurred between the medium and high workload conditions, suggesting that drivers were nearing their available cognitive capacity for the combined load of driving and secondary tasks.

Aging and the corresponding health related declines in physical and cognitive function play an obvious role in reducing spare cognitive capacity [14]. These same factors may also potentially impact the sensitivity of real-time detection of workload through physiological measures. For example, various cardiac conditions and medications commonly used by older adults might reduce the effectiveness of using heart rate as a means of detecting driver workload by reducing beta-adrenergic reactivity.

The sequential presentation of the task demands from low to medium to high was carried out to allow a direct comparison with the earlier simulation [8] study. It is reasonable to question the extent to which the pattern of results obtained was influenced by presentation order and further investigation looking at the various load levels presented in a random order is clearly warranted. A consideration of presentation order effects and further assessment of age and health associated characteristics is currently underway. Continued exploration of these factors and other individual characteristics is important for determining the usefulness of physiological indices as a component of an algorithm for the real-time detection of driver workload.

6. ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of Ford Motor Company, the Santos Family Foundation and the United States Department of Transportation's Region I New England University Transportation Center at the Massachusetts Institute of Technology. We also wish to acknowledge the contributions of Jonathon Long, Vincent Lammers and Alea Mehler for their efforts in the collection and processing of data.

7. REFERENCES

[1] Andreone, L., A. Amditis, E. Deregibus, S. Damiani, D. Morreale, and F. Bellotti. Beyond Context-Awareness: Driver-Vehicle-Environment Adaptively. From the COMUNICAR Project to the AIDE Concept. Proc. IFAC 16th World Congress, (2005).

[2] Brookhuis, K.A. and D. De Waard, The use of psychophysiology to assess driver status. *Ergonomics*, Vol. 36, 9, (1993), 1099-1110.

[3] Brookhuis, K.A. and D. De Waard, *Assessment of drivers' workload: Performance and subjective and physiological indexes*, in *Stress, Workload, and Fatigue*, P.A. Hancock and P.A. Desmond, Editors. Lawrence Erlbaum Associates, Mahwah, NJ, pp. 321-333, 2001.

[4] Brookhuis, K.A., G. De Vries, and D. De Waard, The effects of mobile telephoning on driving performance. *Accident Analysis and Prevention*, Vol. 24, 3, (1991), 309-316.

[5] Broström, R., J. Engström, A. Agnvall, and G. Markkula. Towards the next generation intelligent driver information system (IDIS): The Volvo car interaction manager concept. Proc. 2006 ITS World Congress, (2006).

[6] Engström, J., E. Johansson, and J. Östlund, Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F*, Vol. 8, 2, (2005), 97-120.

[7] Lenneman, J.K., J.R. Shelley, and R.W. Backs. Deciphering psychological-physiological mappings while driving and performing a secondary memory task. Proc. Third International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, (2005), 493-498.

[8] Mehler, B., B. Reimer, J.F. Coughlin, and J.A. Dusek. The impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. Proc. Transportation Research Board of The National Academies, (2009).

[9] Misiti, M., Y. Misiti, G. Oppenheim, and J. Poggi. *Wavelet Toolbox 4 User's Guide*. 2009 [cited 2009 May 29]; Available from: http://www.mathworks.com/access/helpdesk/help/ pdf_doc/wavelet/wavelet_ug.pdf.

[10] Mulder, L.J., Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biological Psychology*, Vol. 34, 2-3, (1992), 205-236.

[11] Reimer, B. Cognitive Task Complexity and the Impact on Drivers' Visual Tunneling. Proc. Transportation Research Board of The National Academies, (2009).

[12] Reimer, B., J.F. Coughlin, and B. Mehler. Development of a Driver Aware Vehicle for Monitoring, Managing & Motivating Older Operator Behavior. Proc. ITS-America, (2009).

[13] Saab Innovations. 2009 [cited 2009 May 27]; Available from: http://www.saabsunited.com/2005/12/saab-innovations.html.

[14] Salthouse, T.A., The processing-speed theory of adult age differences in cognition. *Psychological Review*, Vol. 103, 3, (1996), 403-428.

[15] Wilson, G.F., *Psychophysiological test methods and procedures*, in *Handbook of Human Factors Testing and Evaluation*, S.G. Charlton and T.G. O'Brien, Editors. Lawrence Erlbaum Associates, Mahwah, NJ, pp. 127-156, 2002.

[16] Wu, C. and Y. Liu, Queuing network modeling of driver workload and performance. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 8, 3, (2007), 528-537.

INTERACTING WHILE DRIVING

Driver behaviour during haptic and visual secondary tasks

Annie Rydström
Dept. of Human Work Sciences
Luleå University of Technology
SE-971 87 Luleå, Sweden
annryd@ltu.se

Camilla Grane
Dept. of Human Work Sciences
Luleå University of Technology
SE-971 87 Luleå, Sweden
camgra@ltu.se

Peter Bengtsson
Dept. of Human Work Sciences
Luleå University of Technology
SE-971 87 Luleå, Sweden
pebn@ltu.se

ABSTRACT
There has been an increasing interest for in-vehicle interfaces that make use of haptic information. A simulator study was conducted to investigate whether haptic information can facilitate the interaction with an interface while driving. The conceptual in-car interface consisted of a visual menu of four textures displayed on a screen and corresponding haptic information displayed through the interaction device – a rotary device. The experimental conditions included either visual or haptic or both visual and haptic information. One advantage of the condition including only haptic information was that the participants' eyes remained on the road during the interaction. However, since the haptic interaction necessitated serial processing, the experimental task took longer when using only haptic information. Therefore the participants seem to have relied more on the visual information when it was available. The degradation in driving performance and mental workload assessment did not differ between the conditions.

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *haptic I/O*

General Term
Human Factors

Keywords
In-car interfaces; Haptic information; Driver distraction

1. INTRODUCTION
The information input needed for driving a car is predominantly visual [25] and the major output is generally manual by the hands (steering wheel and gear shift) and by the feet (accelerator, break pedal and clutch). Visual in-vehicle displays require the eyes to be taken off the road and manual controls often require the hands to be taken off the steering wheel. While gaze has to be moved from the road to an in-vehicle display to gather detailed information, the two hands can perform different manual tasks simultaneously [31]. According to Wierwille [31] it is easier to time-share driving and manual in-vehicle tasks than driving and visual in-vehicle tasks, as long as large corrections of the steering wheel that require the use of both hands are not needed. It is a challenge for vehicle manufacturers to design in-vehicle systems that can be operated safely while driving. It has been suggested that haptic cues available through the interaction devices may have the potential to make the interaction with in-car interfaces safer since the interaction may be less visually demanding [3, 30].

Humans can passively perceive vibrations, applied forces and motions. In the automotive domain the use of vibrotactile stimuli has been most investigated and has been shown to be useful as warnings [11, 16], as well as to provide navigation information [27]. Humans can actively, through hand movements, perceive shapes, sizes, textures and locations. The perceptions resulting from these active movements are sequential and form what is called haptic perception [9]. Haptic interaction implies the ability to both sense and manipulate an interface [12]. Hence, haptic cues can be used to create haptically discriminable buttons and switches [19]. Haptic cues can also be provided through one single interaction device, which can change its mechanical properties just as a graphic display can change its optical properties [10]. In the domain of desktop interaction, augmenting a visual interface with usable haptic information through the interaction device (e.g. a computer mouse), such as texture, friction, gravity and force, has been shown to result in a decrease in task completion time [1, 4] and decrease in workload and the number of errors made [7, 20]. Although there are in-car interfaces available on the market providing haptic cues through single multifunctional interaction devices [24], there has not yet been sufficient research on the use of haptic cues while driving [2]. In a study by Porter et al. [21] an in-car interface was designed in which the interface devices (three pods) were coded in terms of the haptic properties size, shape and location. It was shown that the number and duration of glances made to the display and controls were reduced compared to a standard interface. However, in the applied study by Porter et al. the comparison between the interfaces was not entirely systematic since the haptically coded interface and the standard interface differed in many aspects.

Studies carried out in real and simulated driving environments have shown that interaction with equipment within the car while driving cause changes in visual behaviour and driving performance. It has been shown that visual-manual tasks, such as interacting with a mobile phone or manipulating a car radio, leads to frequent and long periods of visual time off road and impaired lane-keeping performance, harmed detection ability and increased brake reaction time [15, 26, 32, 33]. It has also been revealed that non-visual withdrawal of attention, for example a phone

AutomotiveUI'09, September 21–22, 2009, Essen, Germany
ACM 978-1-60558-571-0/09/009

conversation, leads to impaired driving [15]. There are several different ways to measure drivers' visual behaviour and driving performance. One limitation of the ISO metrics glance frequency and glance duration [13] is that these measures only can be used to measure visual behaviour during visual secondary tasks. It has been shown that non-visual secondary tasks can lead to gaze concentration towards the centre of the road [22, 29]. Hence, an alternative to glance based measures is the measure called Percent Road Centre (PRC) [29]. This measure focuses on how much time is spent looking at the centre area of the road and can be used to evaluate visual behaviour during both visual and non-visual secondary tasks. Several driving simulation software programs are available for evaluating the distraction caused by the interaction with in-vehicle systems. The simulations range from advanced ones that record numerous measures on both lateral and longitudinal performance to more hands-on ones that record one or a few measures. An applied driving simulation software currently under investigation to be an ISO standard is the Lane Change Test (LCT) [14]. In this PC simulated environment drivers are requested to change lanes while interacting with an in-car interface. The perception and reaction to lane-change signs shows the driver's awareness of the environment and the lane-keeping shows the driver's ability to control the vehicle [18]. The LCT derives a single measure of driving performance – the mean deviation from a normative path. Mattes [18] showed that results from the LCT correlates with results from a high end moving base driving simulator. Since not only visual and manual, but also mental workload, leads to impaired driving [15] it is central to also consider the mental workload imposed by different secondary tasks. Mental workload can be measured by using subjective assessment techniques.

The objective of the present simulator study was to investigate whether haptic information can ease the interaction with an interface while driving. The conceptual interface used in the study provided visual or haptic or both visual and haptic information. Secondary task performance, eye movement behaviour, driving performance and subjective assessment of mental workload were measured.

2. METHOD

2.1 Participants

Forty participants (six women and 34 men) were recruited to take part in the study. The majority of the participants were students recruited via e-mail at Chalmers University of Technology. Their ages ranged from 20 to 46 (*M* = 26.8, *SD* = 5.2). The criteria for participation were possession of a driving license and no need of eyeglasses (to ensure eye-tracking data of good quality). The participants were to be able to wear contact lenses if vision correction was needed.

2.2 Equipment

The study was conducted using a fixed base Volvo XC90 driving simulator (Fig. 1). A 140 cm wide and 110 cm high driving scene was projected on a screen approximately 200 cm in front of the participants. The driving simulation used was the Lane Change Test (LCT) [14]. In the LCT simulation a participant drives at a constant speed of 60 km/h on a straight, three-lane road on which no other cars are present. Signs on both sides of the road instruct the participant to change lanes. The information appears 40 metres ahead of the sign. In the LCT, the deviation from a normative path is recorded. Consequently, late perception of signs (or missed signs), slow lane change and poor lane-keeping result in greater deviation [18]. An LCT track takes three minutes to complete, and 18 lane changes are made during a track. An LCT analysis software was used to compute the mean deviation for the participants.

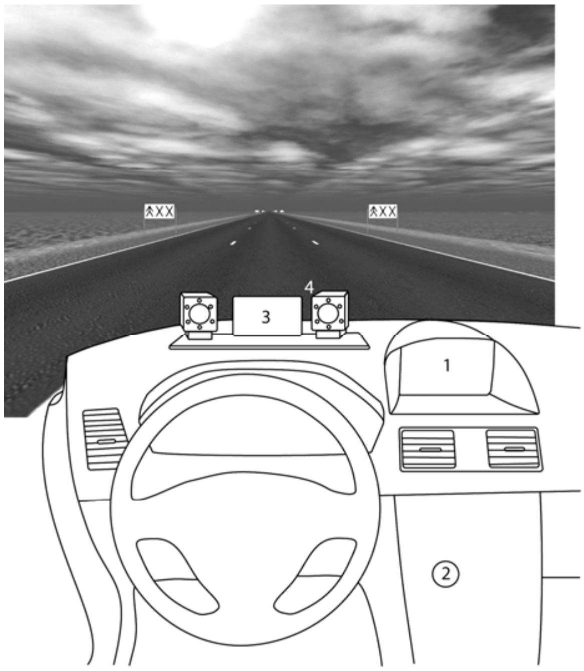


Figure 1. The LCT driving scene and Volvo XC90 simulator, including a secondary task interface display (1) , a haptic rotary device (2), a display presenting the secondary tasks to be completed (3) and eye-tracking cameras (4).

The secondary task interface was implemented in Macromedia Director 8.5 (Adobe Systems Inc., USA). A visual menu was displayed on an 8.5" screen in the centre panel of the simulator (Fig. 1). The interaction with the interface was made with a haptic rotary device (ALPS Haptic Commander, ALPS Automotive Products Division, Japan) mounted on the centre panel. The device had a knob diameter of 3.5 cm. The interface program managed the sensations provided through the rotary device and the graphical scenes displayed on the centre panel display. The program also managed the task presentation – orally in headphones and written on a 6.4" display placed on the dashboard in front of the participants (Fig. 1). faceLAB 4.1 (Seeing Machines, Australia) was used to record eye movements. By means of video signals from two cameras, the faceLAB system measures 3D head position and gaze direction at a rate of 60 Hz. The two eye-tracking cameras were mounted on the dashboard in front of the participants (Fig. 1). To improve data quality face markers were placed on the participants' face. An analysis software, Visual Demand Measurement (VDM) Tool [28], was used to analyse the eye movement data.

2.3 Secondary Task Interface

The secondary task interface was designed to contribute to the central theme of haptic and visual information and did not consist of real vehicle functions. The interface was designed to display congruent information in the haptic and visual modalities. Since textures can be effectively perceived both haptically and visually [6] a menu with texture items, or more specifically items with different roughness, was designed.

The four secondary task conditions were named: *visual information* (V), *visual information and haptic ridges* (VHr), *visual information and haptic ridges and textures* (VHrt), and *haptic ridges and textures* (Hrt). As the rotary device was turned in the conditions including visual information, a transparent blue cursor moved in the menu displayed in the centre panel display (Fig. 2). The four menu items, A, B, C and D, were arranged horizontally, and each texture had a height and width of 25 mm. The graphical representations of the textures were identical for the three conditions including visual information. The visual menu was not displayed in the Hrt condition.

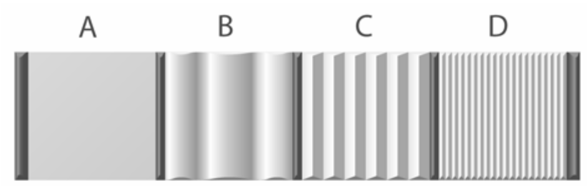


Figure 2. The visual menu displayed in the centre panel display in the secondary task conditions including visual information (V, VHr and VHrt). The textures are here presented in alphabetical order from A to D. In the experiment, the active texture was marked with a transparent blue cursor.

The haptic sensations provided through the rotary device varied between the conditions (Fig. 3). The total angle of operation was 150° for all conditions. Hence, a participant was able to comfortably rotate the device through the menu without changing the grasp. Restricting walls were incorporated outside the scale limits on each end of the menu, and a damper torque made forces increase and decrease with device speed. A smooth sensation was provided in the V condition as the device was turned. In the VHr condition salient ridges were incorporated between every texture in the menu to indicate borders. The angle of the ridges was 10° and the amplitude of the elastic torque was 50 mN·m. The salient ridges indicated borders in the VHrt and Hrt conditions and, in addition, representations of the textures were provided through the device. The haptic textures were rendered as repeated and evenly distributed ridges, i.e. alternated high and low torque. The peak torque of the textures was 10 mN·m and textures A, B, C and D had 0, 3, 6 and 30 ridges, respectively.

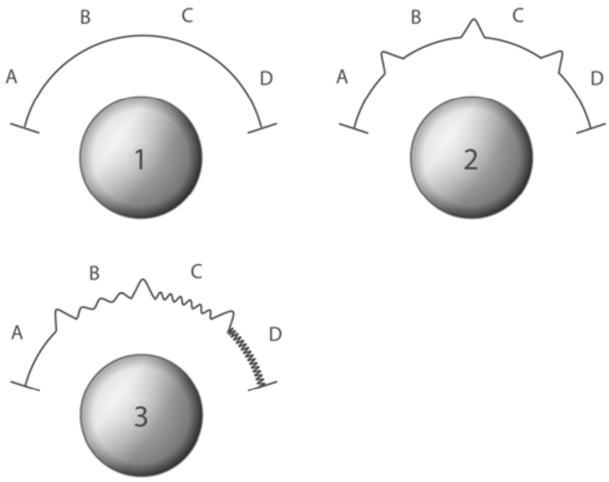


Figure 3. A representation of the haptic sensations provided through the rotary device in the four secondary task conditions: V (1), VHr (2), VHrt and Hrt (3).

Task

The tasks to be completed with the secondary task interface, e.g. "Locate C", were automatically presented to the participants orally in the headphones and provided in written form on the 6.4" display in front of the participants. The participants located and selected the requested item in the menu by turning and pushing the rotary device. As soon as one task was completed, the next was initiated. The target texture and the positions of the textures changed for every trial. If the wrong texture was selected, the textures stayed in the same order until the right texture was selected. A beep was given as feedback after a task was completed successfully. The device was programmed to start at the leftmost texture for every new trial, and the participants were to initially turn the device clockwise. The reset was not felt, and it was therefore not necessary to let go of the device.

2.4 Procedure

The experiment had a between-subjects design and the participants were randomly assigned to the four secondary task conditions, V, VHr, VHrt and Hrt. During the test the test leader sat in the front passenger seat of the simulator and controlled the equipment and read test instructions aloud from a manuscript. A brief description of the experiment as a whole was given at the beginning of a session. The participants were then instructed to adjust the seat, markers were placed on the participants' face, and the eye-tracking cameras were calibrated. The participants were given instructions about the LCT and were specifically informed to change lanes quickly, as soon as the information appeared on the signs. Each participant then drove three LCT tracks, of which the first two were training tracks and the third was a baseline driving track, i.e. driving without secondary task. Eye movements and driving performance were recorded for the baseline track.

Subsequent to baseline driving the participants practiced the secondary task in isolation in two training series. In the first series the participants were to learn which letter represented which texture and were free to explore a menu in which the menu items were provided in alphabetical order. As the rotary device was turned, the name of the active item (A, B, C or D) was displayed on the centre panel display. In the second training series the participants practiced the secondary task as it would be displayed while driving. The participants had to successfully complete 12 tasks in a row to pass the training.

Following this training the participants completed two dual-task driving tracks, i.e. driving with secondary task, of which the first was a training track. For the second track, data were recorded on secondary task performance, eye movements and driving performance. Since it may be tempting to adopt a strategy where secondary tasks are completed on the straight sections between the lane changes, the participants were instructed to perform as well as they could on both the primary and secondary tasks. The participants finally drove a second baseline track, for which data on eye movements and driving were recorded. After the test the participants were asked to fill in a participant characteristics form (concerning gender, age, handedness etc.) and mental workload (NASA-TLX) forms. A whole session took altogether about one hour.

2.5 Dependent Measures

Secondary task performance was measured in terms of the number of tasks completed and the number of push and turn errors made. When the wrong item was selected, the action was recorded as a push error, and the action was recorded as a turn error when the participants passed the right item without selecting. The eye movement data were analysed in terms of percent road centre (PRC) [29]. In the PRC analysis the road centre area was defined as a circle with a radius of 10°. LCT driving performance was measured in terms of mean deviation [14]. The NASA-TLX rating method was used to measure subjective mental workload. NASA-TLX is a multidimensional rating method that gives an overall workload score based on the weighted average of six workload-related factors (mental demand, physical demand, temporal demand, performance, effort and frustration level) [8].

3. RRESULTS

An alpha level of .05 was used for all statistical tests.

3.1 Secondary Task Performance

Figure 4 shows the number of tasks completed in the four secondary task conditions. A between-subjects ANOVA with secondary task condition (V, VHr, VHrt and Hrt) as the factor was used to test the statistical significance of differences. The number of tasks completed was found to be significantly different between the conditions, $F(3, 36) = 21.1$, $p<.001$. The Tukey HSD procedure, used for post hoc pairwise comparisons of means, showed that there were no significant differences between the V, VHr and VHrt conditions. However, significantly fewer tasks were completed with the Hrt condition as compared to the other conditions (all $p<.001$).

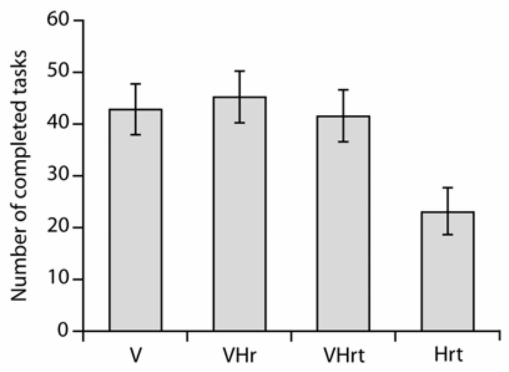


Figure 4. The number of tasks completed for the four secondary task conditions. The error bars represent 95% confidence intervals for the means.

In terms of push error, floor effects were present for the V, VHr and VHrt conditions, whereas 10% of the tasks in the Hrt condition included a push error. Figure 5 shows the percentage of tasks that included a turn error. To rectify the differences between group variances the turn error data was transformed by taking the square roots of the values. A between-subjects ANOVA with secondary task condition (V, VHr, VHrt and Hrt) as the factor was used to test the statistical significance of differences. The number of turn errors made was found to be statistically different between the conditions, $F(3, 36) = 14.5$, $p<.001$. The Tukey HSD procedure showed that there were no significant differences between the V, VHr and VHrt conditions. However, significantly more turn errors were made with the Hrt condition as compared to the other conditions (all $p<.001$).

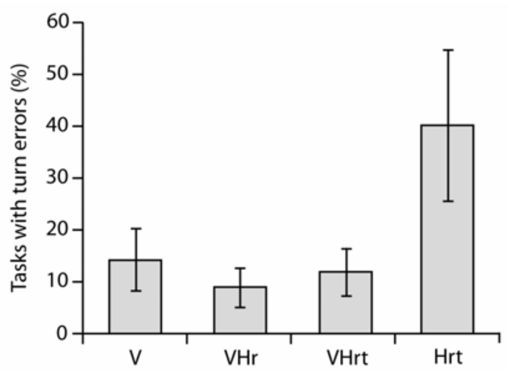


Figure 5. The percentage of tasks including a turn error in the four secondary task conditions. The error bars represent 95% confidence intervals for the means.

3.2 Eye Movements

Figure 6 shows the PRC values for baseline and dual-task driving for the four conditions. Since each participant conducted two baseline tracks, mean PRC values were calculated. Within-groups t-tests (two-tailed) were used to compare the PRC values between baseline and dual-task driving. The PRC values decreased drastically when tasks including visual information were performed: $t(9) = 14.77$, $p<.001$, for the V condition, $t(9) = 16.50$, *p*<.001, for the VHr condition and *t*(9) = 9.19, *p*<.001, for the VHrt condition. No significant difference was found between baseline and dual-task driving for the Hrt condition.

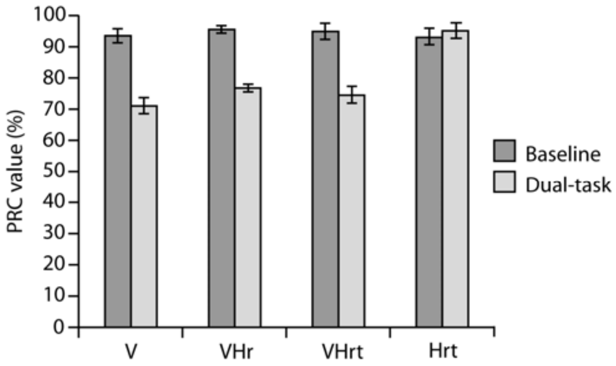


Figure 6. PRC values for baseline and dual-task driving for the four secondary task conditions. The error bars representing 95% confidence intervals for the means have been adjusted in the figure to suit within-subject comparisons [17].

To compare the PRC values between the conditions, the relative decrease or increase in PRC was calculated for each participant by dividing the dual-task value with the baseline value. More time spent looking at the centre area of the road in the dual-task condition as compared to the baseline leads to a higher quotient (>1) (V: *M* = 0.62, *SD* = 0.07; VHr: *M* = 0.64, *SD* = 0.07; VHrt: *M* = 0.69, *SD* = 0.10; Hrt: *M* = 1.03, *SD* = 0.06). A between-subjects ANOVA with secondary task condition (V, VHr, VHrt and Hrt) as the factor was used to test the statistical significance of differences. The quotient was found to be statistically different between the conditions, *F*(3, 36) = 60.9, *p*<.0.001. The Tukey HSD procedure showed that there were no significant differences between the V, VHr and VHrt conditions. However, the time spent looking at the centre area of the road as compared to the baseline was significantly higher for the Hrt condition as compared to the other conditions (all *p*<.001).

3.3 Driving Performance

Figure 7 shows the mean deviation for baseline and dual-task driving tracks for the four conditions. Since each participant conducted two baseline tracks, mean baseline deviation values were calculated. Within-groups *t*-tests (two-tailed) were conducted to compare the mean deviation between baseline and dual-task driving. There was a significant increase in mean deviation from baseline driving to dual-task driving for all conditions: *t*(9) = -3.1, *p*<.05, for the V condition, *t*(9) = -2.8, *p*<.05, for the VHr condition, *t*(9) = -2.8, *p*<.05, for the VHrt condition and *t*(9) = -5.06, *p*<.01, for the Hrt condition.

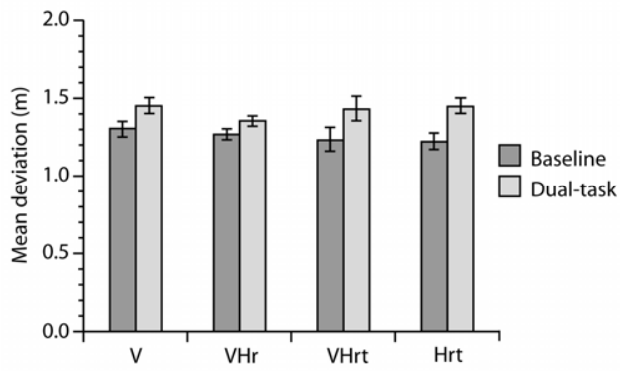


Figure 7. Mean deviations for baseline and dual-task driving for the four secondary task conditions. The error bars representing 95% confidence intervals for the means have been adjusted in the figure to suit within-subject comparisons [17].

To compare the decrease in performance between the conditions, the relative decrease in performance was calculated for each participant by dividing the dual-task value with the baseline value (Mattes, 2003). Performing poorly in the dual-task condition as compared to the baseline leads to a higher quotient (>1) (V: *M* = 1.11, *SD* = 0.12; VHr: *M* = 1.06, *SD* = 0.08; VHrt: *M* = 1.16, *SD* = 0.18; Hrt: *M* = 1.20, *SD* = 0.13). A between-subjects ANOVA with secondary task condition (V, VHr, VHrt and Hrt) as the factor was used to test the statistical significance of differences. Degradation in driving performance did not differ significantly between the conditions.

3.4 Subjective Mental Workload

The mean weighted workload scores (V: *M* = 58.3, *SD* = 11.0; VHr: *M* = 62.4, *SD* = 14.7; VHrt: *M* = 54.0, *SD* = 18.9; Hrt: *M* = 62.9, *SD* = 6.9) were not significantly different between the secondary task conditions.

4. DISCUSSION

Fewer tasks were completed in the Hrt condition in comparison with the other three conditions. When using solely haptic information the menu had to be sequentially processed texture-by-texture, which was time-consuming, whereas the menu items in the V, VHr and VHrt conditions could be visually processed more simultaneously. Analogously, more turn errors were made in the Hrt condition. A comparison between textures could be made visually without turning the device (V, VHr and VHrt), while the device had to be turned in order to compare textures haptically (Hrt), which resulted in an increased number of turn errors. Furthermore, and interestingly, it can be concluded that there were no significant differences between the three conditions including visual information, irrespectively of whether redundant haptic information was provided or not. Studies have shown that vision often dominates an integrated percept [5]. Thus, visual dominance in combination with the fact that the visual information in this experiment could be processed faster and more precise than the haptic seem to have led to a reliance on visual information when it was available. Rydström and Bengtsson [23] also observed in a desktop experiment that haptic texture information was often ignored by the participants if corresponding visual information was provided. Even though the present experiment included a concurrent visual task (driving) the participants do not seem to have used the redundant haptic information instead of or even as a complement to the visual. Nonetheless, in the domain of desktop interaction, augmenting a visual interface with usable haptic information through the interaction device has been shown to enhance the interaction [1, 4]. However, one difference in these interfaces is that the haptics is designed to enhance the visual interaction rather than replace it. Research in human-computer interaction has also shown that workload is decreased when a visual interface is augmented with usable haptic information [7, 20]. However, the reported mental workload did not differ between the secondary task conditions in this study. In view of the fact that the haptic interaction was more time consuming and less precise it is notable that it was not demonstrated that haptic interaction was more mentally demanding.

The PRC measure showed that the eyes were kept on the road during the non-visual Hrt condition. From this view, it is apparent that there is a potential for haptic interaction. In contrast, the PRC values were significantly lower for the V, VHr and Vhr conditions compared to baseline driving. Further, in terms of the PRC values, the visual behaviour did not differ between the V, VHr and VHrt conditions. Hence, complementing visual information with corresponding haptic information did not increase the time the participants spent looking at the road. Interestingly, from a traffic safety point of view the non-visual haptic condition might be advantageous, even though it is more time consuming and less precise, as concluded above. A practical implication of this would be that when visual information is not needed it should not be provided. For example volume controls sometimes has a graphical representation in a display when the volume is increased or decreased. The visual information may make the driver look at the display instead of only rely on the auditory and haptic information.

When using the LCT, which is a PC simulated environment, the road scene is projected only in front of the participants, and no rear-view mirrors are used. In addition, no glances at the speedometer are necessary since the driving speed is system-controlled. In this experiment there were therefore relatively high values in the PRC value for baseline driving, over 90%. Real driving normally gives a PRC value of about 70% [29]. This may have made it infeasible to calculate an effect of gaze concentration caused by attention to a non-visual secondary task [22, 29]. It should also be mentioned that glances at the display presenting the tasks to be completed may have induced some noise in the data.

Compared to baseline driving, all four secondary task conditions caused an increase in mean deviation. This finding supports research showing that both visual and non-visual secondary tasks have a negative influence on driving performance [15, 33]. Furthermore, it was shown that the participants performed similarly in all conditions. It could, however, perhaps be expected that the degradation in driving performance should be less for the Hrt condition since the eyes were kept on the road. One possible explanation for the absence of this effect could be that the lane-change signs in the LCT are highly expected. A participant can therefore adopt a strategy in which the tasks are solved using visual information at the straight sections between the signs. Alternatively, since the tasks were provided one after another during the dual-task track, most of the driving was spent with only one hand on the steering wheel. In view of the fact that LCT requires a great deal of manipulation of the steering wheel, and large corrections of the steering wheel often actually require the use of both hands [31], there was substantial manual time-sharing for all conditions. However, since the LCT derives a single measure of driving performance, different characteristics of driving are not considered in isolation. Perhaps other measures, such as response to an unexpected external event, could capture any differences.

Porter et al. [21] found that an in-car interface designed with consideration to haptics was preferable to a conventional one in terms of the number and duration of eye glances. The study of Porter et al. indicates that haptic interaction has the potential to facilitate the interaction with in-vehicle equipment. The present study showed that during the condition including only haptic information the participants' eyes remained on the road during the interaction. Even if the experimental task took longer when using only haptic information, the degradation in performance and mental workload assessment did not differ from the conditions including visual information, which is a result of vital importance for future implementations of haptics. However, the haptic information needs to be improved. Multifunctional, menu-based systems are common in cars today [24]. These systems include a wide range of different functions, and the interaction often requires several steps of visual and manual interaction. This experiment serves as a basis for investigating the use of haptic and visual information in the interaction with such systems. The next step will be to implement haptic and visual information in combination with a more genuine in-car interface.

5. ACKNOWLEDGMENTS

This study was supported by the Swedish Program Board for Automotive Research.

6. REFERENCES

[1] Akamatsu, M., MacKenzie, I.S., 1996. Movement characteristics using a mouse with tactile and force feedback. Int. J. Human-Computer Studies 45, 483-493.

[2] Bengtsson, P., Grane, C., Isaksson, J., 2003. Haptic/graphic interface for in-vehicle comfort functions – a simulator study and an experimental study, in: Proc. of the IEEE International Workshop on Haptic, Audio and Visual Environments and Their Applications. Ottawa, Canada, pp. 25-29.

[3] Burnett, G.E., Porter J.M., 2001. Ubiquitous computing within cars: designing controls for non-visual use. Int. J. Human-Computer Studies 55, 521-531.

[4] Campbell, S.C., Shumin, Z., May, K.W., Maglio, P.P., 1999. What you feel must be what you see: adding tactile feedback to the Trackpoint, in: Sasse, M.A., Johnson, C. (Eds.), Human-Computer Interaction – Proceedings of INTERACT'99. IOS Press, Edinburgh, pp. 383-390.

[5] Ernst, M.O., Bülthoff, H.H., 2004. Merging the senses into a robust percept. Trends. Cogn. Sci. 4, 162-169.

[6] Gentaz, E., Hatwell, Y., 2003. Haptic processing of spatial and material object properties, in: Hatwell, Y., Strieri, A., Gentaz, E. (Eds.), Touching for Knowing: Cognitive Psychology of Haptic Manual Perception. J. Benjamins, Amsterdam, pp. 123-159.

[7] Grane, C., Bengtsson, P., 2005. Menu selection with a rotary device founded on haptic and/or graphic information, in: Proc. of World Haptics 2005. Piza, Italy, pp. 475-476.

[8] Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research, in: Hancock, P.A., Meshkati, N. (Eds.), Human Mental Workload. Elsevier, Amsterdam, pp. 139-182.

[9] Hatwell, Y., 2003. Introduction, in: Hatwell, Y., Strieri, A., Gentaz, E. (Eds.), Touching for Knowing: Cognitive Psychology of Haptic Manual Perception. J. Benjamins, Amsterdam, pp. 1-14.

[10] Hayward, V., Astley, O.R., Cruz-Hernandez, M., Grant, D., Robles-De-La-Torre, G., 2004. Haptic interfaces and devices. Sensor Review 24, 16-29.

[11] Ho, C., Tan, H.Z., Spence, C., 2005. Using spatial vibrotactile cues to direct visual attention in driving scenes. Transport. Res. F 8, 397-412.

[12] Hollis, R., 2004. Haptics, in: Bainbridge, W.S. (Ed.), Berkshire Encyclopedia of Human-Computer Interaction. Berkshire Publishing Group, Great Barrington, pp. 311-316.

[13] ISO 15007-1. Road vehicles – Measurement of driver visual behaviour with respect to transport information and control systems – Part 1: Definitions and parameters.

[14] ISO 26022 (standard under development). Road vehicles – Ergonomic aspects of transport information and control systems – Simulated lane change test to assess driver distraction.

[15] Lamble, D., Kauranen, T., Laakso, M., Summala, H., 1999. Cognitive load and detection thresholds in car following situations: safety implications for using mobile (cellular) telephones while driving. Accid. Anal. Prev. 31, 617-623.

[16] Lee, J.D., Hoffman, J.D., Hayes, E., 2004. Collision warning design to mitigate driver distraction, in: Proc. of CHI 2004. ACM Press, New York, pp. 65-72.

[17] Loftus, G. R., Masson, M. E., 1994. Using confidence intervals in within-subject designs. Psychon. B. Rev. 1, 476-490.

[18] Mattes, S., 2003. The lane-change-task as a tool for driver distraction evaluation, in: Strasser, H., Kluth, K., Rausch, H., Bubb, H. (Eds.), Quality of Work and Products in Enterprises of the Future. Ergonomia Verlag, Stuttgart, pp. 57-60.

[19] Moore, 1974. Tactile and kinaesthetic aspects of push-buttons. Appl. Ergon. 5, 66-71.

[20] Oakley, I., McGee, M. R., Brewster, S., Gray, P., 2000. Putting the feel in 'look and feel', in: Proc. of CHI 2000. ACM Press, New York, pp. 415-422.

[21] Porter, J.M., Summerskill, S.J., Burnett, G.E., Prynne, K., 2005. BIONIC – 'eyes-free' design of secondary driving controls, in: Proc. of the Accessible Design in the Digital World Conference. Dundee, UK.

[22] Recarte, M.A, Nunes, L.M., 2003. Mental workload while driving: effects of visual search, discrimination, and decision making. J. Exp. Psychol. Appl. 9, 119-137.

[23] Rydström, A., Bengtsson, P., 2007. Haptic, visual and cross-modal perception of interface information, in: de Waard, D., Hockey, G.R.J., Nickel, P., Brookhuis, K.A. (Eds.), Human Factors Issues in Complex System Performance. Shaker Publishing, Maastricht, pp. 399-409.

[24] Rydström, A., Bengtsson, P., Grane, C., Broström, R., Agardh, J, Nilsson, J., 2005. Multifunctional systems in vehicles: a usability evaluation, in: Thatcher, A., James, J., Todd A. (Eds.), Proceedings of CybErg 2005, the Fourth International Cyberspace Conference on Ergonomics. International Ergonomics Association Press: Johannesburg.

[25] Sivak, M., 1996. The information that drivers use: is it indeed 90% visual? Perception 25, 1081-1089.

[26] Strayer, D.L., Johnston, E.A., 2001. Driven to distraction: dual-task studies of simulated driving and conversing on a cellular phone. Psychol. Sci. 12, 462-466.

[27] Van Erp, J., Van Veen, H., 2004. Vibrotactile in-vehicle navigation system. Transport. Res. F 7, 247-256.

[28] Victor, T.W., Blomberg, O., Zelinsky, A., in press. Automating driver visual behavior measurement. Vision in Vehicles 9.

[29] Victor, T.W., Harbluk, J.L., Engström, J.A., 2005. Sensitivity of eye-movement measures to in-vehicle task difficulty. Transport. Res. F 8, 167-190.

[30] Vilimek, R., Hempel, T., Otto, B., 2007. Multimodal interfaces for in-vehicle applications. Lect. Notes. Comput. Sc. 4552, 216-224.

[31] Wierwille, W., 1993. Demands on driver resources associated with introducing advanced technology into the vehicle. Transport. Res. C 1, 133-142.

[32] Wikman, A.S., Nieminen, T., Summala, H., 1998. Driving experience and time-sharing during in-car tasks on roads of different width. Ergonomics 41, 358-372.

[33] Zwahlen, H.T., Adams, C.C., Jr., DeBald, D.P., 1988. Safety aspects of CRT touch panel controls in automobiles, in: Gale, A.G., Freeman, M.H., Haslegrave, C.M., Smith, P.A., Taylor, S.P. (Eds.), Vision in Vehicles II. Elsevier, Amsterdam, pp. 335-344.

Glancing at Personal Navigation Devices Can Affect Driving: Experimental Results and Design Implications

Andrew L. Kun[1], Tim Paek[2], Željko Medenica[1], Nemanja Memarović[1], Oskar Palinko[1]

[1] University of New Hampshire
Electrical and Computer Engineering Department
Kingsbury Hall
Durham, NH 03824, USA
603.862.1357
{andrew.kun, zeljko.medenica, nemanja.memarovic, oskar.palinko}@unh.edu

[2] Microsoft Research
One Microsoft Way
Redmond, WA 98052 USA
425.703.8647
timpaek@microsoft.com

ABSTRACT
Nowadays, personal navigation devices (PNDs) that provide GPS-based directions are widespread in vehicles. These devices typically display the real-time location of the vehicle on a map and play spoken prompts when drivers need to turn. While such devices are less distracting than paper directions, their graphical display may distract users from their primary task of driving. In experiments conducted with a high fidelity driving simulator, we found that drivers using a navigation system with a graphical display indeed spent less time looking at the road compared to those using a navigation system with spoken directions only. Furthermore, glancing at the display was correlated with higher variance in driving performance measures. We discuss the implications of these findings on PND design for vehicles.

Categories and Subject Descriptors
H5.2. User Interfaces: Evaluation/methodology.

General Terms
Measurement, Design, Reliability, Experimentation, Human Factors.

Keywords
In-car navigation, user interfaces, driving performance.

1. INTRODUCTION
As computer form factors shrink and communication bandwidth and networks expand, ubiquitous computing is starting to play an increasingly important role in our lives. This prospect is particularly exciting with regards to interaction with users while they are engaged in the manual-visual task of driving. In some countries, driving is the primary mode of commuting. For example, according to the U.S. Census Bureau [1], Americans spend more than 100 hours a year commuting on the road. Given the large amount of time that some people spend behind the wheel, and the increasing availability of computational resources that can now operate inside a vehicle, many companies have been introducing a myriad of mobile services and functionalities into the consumer market just for drivers. A few notable examples are hands-free voice dialing, GPS navigation, live traffic reports, automated directory assistance, and infotainment systems. Unfortunately, the question of how these in-car services impact driving performance remains largely unanswered.

This paper addresses the effect of in-car personal navigation devices (PNDs) on driving. In order to guide drivers, a PND usually combines a map-based visual display of the GPS location of the vehicle with spoken directions. However, any visual output to the driver may constitute a potentially dangerous source of distraction. As such, we sought to answer two important research questions:

1. *Does a PND with combined visual and spoken output cause drivers to spend less time looking at the road ahead than a PND that provides spoken output only?*
2. *What is the effect of glancing at the PND visual display on driving performance?*

These two questions are motivated by an industry trend towards PNDs with increasingly sophisticated graphical user interfaces (GUI), such as 3D views of the terrain [2][3]. At the same time, many mobile phones with much smaller screens offer driving directions that rely primarily on spoken output to guide users, such as the Verizon VZ Navigator [4].

This paper is organized as follows. After surveying related research in Section 2, we describe the experiment we conducted using a high fidelity driving simulator to address the two research questions above in Section 3. We report our results in Section 4 and discuss the implications of our results on PND design in Section 5. Finally, we conclude with directions for future research.

2. RELATED RESEARCH
Although many researchers have worked on evaluating the visual and cognitive load of driving as well as that of participating in concurrent activities such as talking on a cell phone [5], no research to date has specifically explored the effects of interacting with a PND on driving performance. We now describe the most

AutomotiveUI'09, September 21–22, 2009, Essen, Germany
ACM 978-1-60558-571-0/09/009

Figure 1. 180º field of view driving simulator.

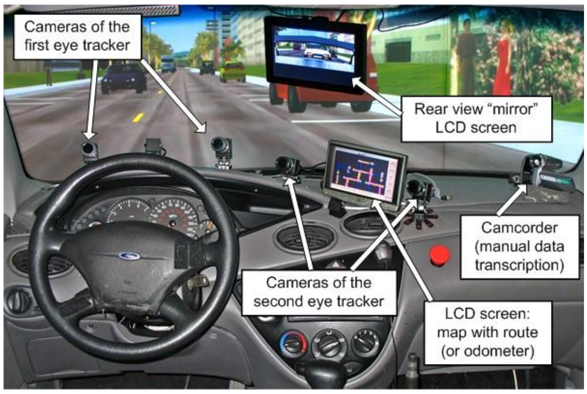


Figure 2. Experimental setup inside the vehicle.

relevant prior work that either sets a precedent for our experimental methodology or discusses similar in-car interfaces.

Because assessing driving performance in real vehicles can be impractical and hazardous, simulator studies are a common way to evaluate driving performance as well as visual attention while interacting with in-car devices. The work of Lew et al. supports the validity of this approach [6] and researchers often make design recommendations based on simulator studies. In fact, Lew et al. explored how well simulator performance could predict driving performance among participants recovering from traumatic brain injury. The authors reviewed a number of studies on this topic and found it difficult to compare results, due to a lack of standard driving simulator scenarios. In their study, they used driving performance measures from the simulator, such as lane position variance and steering wheel angle variance, in conjunction with human observation data, to predict driving performance at a future date (when participants have hopefully recovered some of their abilities lost to the injury). They found that simulator performance measures were good predictors of future driving performance in the real-world.

Besides simulator studies, a few large-scale naturalistic studies have also been conducted. In order to assist the development of crash countermeasures, Neale et al. [7] collected data about the driving habits, performance and other factors of 100 drivers over a period of one year. Their study provides useful data on the causes of crashes and near-crashes; for example, the most common causes involved a lead vehicle braking. Indeed, the unexpected events we generated in our simulator experiment were informed by their study.

While our experiment assesses the effects of PND output on driving performance, Tsimhoni et al. investigated the effects of entering addresses while driving using word-based speech recognition, character-based speech recognition and typing on a touch-screen keyboard [8]. They found that employing speech recognition allowed for shorter and safer address entry than using a keyboard.

Prior research has examined a variety of other in-car devices, and even cognitive architectures for predicting the effect of in-car interfaces on driving performance [9]. In a simulator experiment, Chisholm et al. [10] looked at manual-visual interactions with mp3 players while driving. They found that complicated interactions with the mp3 player increased reaction time to road hazards. Using an eye gaze tracker, the study also concluded that the interactions re-directed driver attention from the road to the mp3 player, increasing the chance of crashes. Medenica and Kun [11] compared the driving performance of participants when using a police radio's manual user interface versus a speech user interface. They found that using the manual user interface degraded driving performance significantly whereas using the speech interface did not.

Using a simulator experiment, Horrey et al. investigated the influence of in-car devices in general on the visual attention of drivers and driving performance [12]. They found that as the amount of time drivers spent observing the outside world (or the percent dwell time on the outside world) decreased, the variability in lane position increased. In other words, their experiments showed that visual distractions negatively influenced driving performance. While general findings provide critically important guidance, they need to be validated for specific domains. Our simulator experiment validates their finding specifically for PNDs.

3. EXPERIMENT

Before we delve into the details of our simulator experiment, it is worth noting that we conducted a preliminary study comparing paper directions against a PND with and without a visual display [13]. In examining the ways in which a PND in general was better than paper directions, and observing how drivers with a visual display spent less time looking at the road than those with spoken directions only, we decided to conduct a follow-up experiment that could more thoroughly inspect the relationship between glancing and driving performance. We did this by making the simulation more typical of a city route, with short and long road segments, ambient traffic conditions characteristic of city driving, and pedestrians walking here and there. In other words, we developed a more "realistic" simulation populated with things to look at – primarily, other cars and people. We now describe how we conducted the simulator experiment and collected data.

3.1 Equipment

Our experiment was conducted in a high-fidelity driving simulator with a 180º field of view. As shown in Figure 1, the simulator provides a full-width automobile cab on top of a motion base that allows drivers to feel bumps in the road as well as braking. Figure 2 displays the equipment inside the vehicle. Because we were interested in visual attention, we equipped the simulator with two eye trackers that provide gaze information from two cameras

Figure 3. LCD screen displaying real-time location of the vehicle on a map.

each. Figure 2 also shows where we mounted a 7'' LCD screen for displaying map information. PNDs are typically mounted either on the windshield, on top of the dashboard, or are built into the dashboard. We decided to place the LCD screen on top of the dashboard because the gaze angle generally has to change less if the PND is located higher than if the PND is built into the dashboard. Although a 7'' screen is typically larger than most portable PNDs, our larger screen ensures that users can clearly see the map and read the street names. Indeed, the consumer market has exhibited a steady trend toward larger screen PNDs with greater multimedia functionality.

3.2 Method

3.2.1 Participants

We collected data from 8 male participants. All were university students between the ages of 21 to 29 (the average age was 22.4). They received a $15 gift card to a popular store chain for their participation.

3.2.2 Procedure

Participants in the experiment interacted with two types of navigation aids:

1. Standard PND directions: Standard PNDs provide real-time map location as well as turn-by-turn spoken directions. Likewise, our LCD screen presented users with real-time location of the vehicle in the simulator world along with spoken prompts for impending turns. Figure 3 shows the LCD screen with map information. The map was presented in a dynamic, exocentric, forward-up view, where the car remains at the center of the screen while the road moves. In order to eliminate problems associated with the comprehension of synthesized speech while driving [14], we used spoken prompts recorded by a female voice talent.

2. Spoken directions only: Here, we utilized the same spoken prompts as in the standard PND and displayed no map information on the LCD. The spoken directions provided distances to the next turn (e.g., "In 75 yards turn right onto Fifth Avenue."). Because the simulator does not provide an odometer, we displayed odometer information on the LCD.

The experimental protocol proceeded as follows. Participants were given an overview of the simulator and the driving and navigation tasks, and were then trained in the driving simulator.

Figure 4. Simulated two-lane city road with lane markings and ambient traffic. The image also illustrates an unexpected event: a pedestrian walking into the roadway from behind a vehicle parked on the side of the road.

Training consisted of driving in a city environment as shown in Figure 4. Participants were instructed to drive as they normally would and to obey all traffic laws. They first drove for about 5 minutes following directions from a standard PND and then another 5 minutes following directions from a PND with spoken directions only. During training, participants were exposed to two unexpected events, one for each navigation aid. In one event, a pedestrian walked out from behind a vehicle parked on the side of the road (see Figure 4), and in the other, a parked vehicle pulled out and cut off the participant. Participants were warned that they may encounter such events before they started the driving portion of their training.

After training, participants completed two routes, one for each of the navigation aids. Two routes were used to prevent participants from learning the directions over the course of the experiment. In order to keep the driving task complexity equal across routes, the two routes were identical, and participants simply traversed them in different directions for the two PNDs. Figure 5 displays the route used in the experiments (bottom left side). Roads were presented in daylight with ambient traffic characteristic of city driving. Each route consisted of two-lane (one lane in each direction) city roads, with lane markings, all with 3.6 m wide lanes. The total route lengths were 10 km and each took about 15 minutes to complete. Each route also exposed participants to three unexpected events, as listed in the legend of Figure 5.

In this paper, we concentrate on two-lane city roads with lane markings, ambient vehicle traffic and pedestrian traffic (e.g., Figure 4). We focus on these roads because this type of road demands constant visual attention from drivers. This, in turn, means that driving performance measures and visual attention are likely to be affected by differences in the visual demands of the two navigation aids.

3.2.3 Design

We conducted a within-subjects factorial design experiment with the two navigation aids as our primary independent variable, *Nav*. The order of *Nav* was counter-balanced among the participants. We measured the following dependent variables.

Standard driving performance measures. We recorded the variances of lane position, steering wheel angle and velocity. In

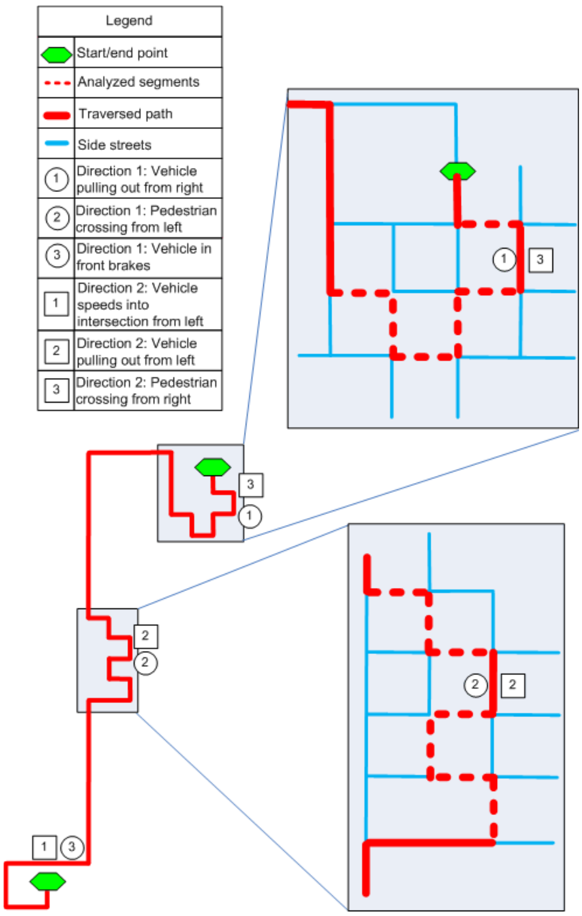


Figure 5. The simulated route (bottom left) and the short segments (right) used for analysis. Use the legend to locate the traversed path, the analyzed short segments, and the location of unexpected events.

each case, a higher variance represents worse driving performance. We also analyzed the mean velocity of travel for each participant. A lower mean velocity may indicate harder perceived driving conditions.

Lane position constitutes the position of the center of the simulated car and is measured in meters. Clearly, large variances in lane position are the most serious sign of poor driving performance, since they indicate that the participant has weaved in his/her lane, and perhaps even departed from the lane.

Steering wheel angle is measured in degrees. In the case of curvy roads, large steering wheel angle variance is not in itself a sign of poor driving performance. After all, just following a curvy road requires varying the steering wheel angle constantly. However, steering wheel angle variance can be used as a relative measure of driving performance when comparing the performance of multiple participants on road segments of similar driving difficulty. A higher variance is an indication of increased effort expended by a driver to remain in his/her lane.

The *velocity* of the vehicle is measured in meters/second. A relatively large variance in the velocity of a car does not necessarily indicate unsafe driving. However, drivers often reduce speed when they are concerned about safety or when they are distracted. For example, a driver may slow down on a narrow road or when talking to a passenger. Similarly a low mean velocity for a portion of the road may indicate that the driver was concerned about safety or otherwise distracted.

Number of collisions. We counted the number of instances when the participant's vehicle touched another object, such as a parked or moving vehicle, a pedestrian, etc. Based on our experience with simulator studies, we did not expect collisions to happen during normal driving, but thought they might occur when drivers were confronted with unexpected events. Since the unexpected events were designed to be avoidable by an alert driver, any collision during such an event may indicate distraction.

Percent dwell time (PDT) on the outside world. The *PDT* is the percentage of time that the participant spent looking at items displayed on the three simulator screens (most importantly the roadway). A low value may indicate that the driver was distracted, which in turn could lead to collisions. In addition to total PDT, we also tracked changes in PTD as participants traveled between intersections. Changes in PDT that depend on proximity to a given intersection may shed light on what causes distractions and hopefully lead to better PND designs that can avoid these dips.

Cross-correlation peaks. We performed cross-correlation analyses to identify time lags in increased variance for lane position and steering wheel angle (if any) in response to decreased PDT on the outside world. Peaks in the cross-correlation of the PDT on the outside world and the variance of a driving performance measure may indicate a causal relationship between decreased PDT and increased variance. If a peak exists for a given lag between the PDT and the variance of a driving performance measure, the lag (expressed in seconds) may indicate the time lag between the onset of decreased PDT on the outside world (e.g. due to a participant looking at the standard PND) and the increase in the variance of the driving performance measure.

3.2.4 Measurement

Raw data for the four driving performance measures were provided by the simulator and sampled at a 10 Hz rate. Using the eye tracker, we also recorded gaze angles throughout the experiment. Eye tracker data was sampled at a 60 Hz rate. Using the eye tracker, we automatically classified gazes as being directed at the outside world if the participant was looking at any of the simulator's front projection screens.

For the rare cases in which the eye tracker could not track a participant's gaze (e.g. when the participant's hand blocked the eye tracker's view of his/her eyes for an instant), we reviewed video footage obtained from the eye tracker cameras as well as from the camcorder in the simulator (Figure 2) and hand-transcribed dwell times.

3.2.5 Calculation

As discussed in section 3.2.2, our experiment presented participants with city driving routes. The routes can be broken up into segments by treating roads between two intersections as separate segments. Figure 5 displays the route used in the experiments (bottom left side) and zooms in on the short segments of the routes used in the experiment (right). We calculated all of our results, such as the variances and mean velocity, using data from 13 segments. These segments all had the same characteristics, thereby controlling factors that could potentially confound our results. In particular, the segments were short, with 200 meters separating the centers of adjacent intersections. Although longer segments were utilized in the routes to make the driving task feel more realistic, we expected that participant driving patterns (e.g. the frequency content of the vehicle velocity reflecting the acceleration and deceleration over a segment) and visual attention patterns (how often and where people look) would be different for segments of different lengths, making comparisons between them difficult.

Furthermore, at both the beginning and end of each segment, there was a four-way intersection where participants made either a right or left turn. Although routes had short (200 m) segments that did not meet this criterion (e.g. when participants entered some of the short segments by driving straight through a four-way intersection), we did not include them. Driving performance and visual attention are likely to be different on these segments than on segments where one or both of the turns may be missing.

Finally, participants did not encounter an unexpected event in the segments we analyzed. Unexpected events may require sudden braking and steering wheel motion, which in turn can result in very large variances for these measures, again making comparisons with other segments difficult.

In analyzing all of the segments, we excluded data collected close to the intersections. This was done because driving performance data at the beginning of a segment is typically dominated by the turning maneuver that is necessary to get through the intersection, and data collected at the end of a segment is dominated by deceleration before turning. Variances resulting from the effects of turning maneuvers and deceleration close to intersections are much larger than variances encountered in data generated away from the intersection, which of course makes it difficult to compare intersection and straight segment data. In particular, we excluded data generated 60 meters after exiting the previous intersection and 40 meters before an upcoming intersection, and analyzed data generated over (200 – 60 – 40) m = 100 meters.

3.2.5.1 Driving Performance

For each participant and navigation type, the variances of the driving performance measures (lane position, steering wheel angle and velocity) were calculated for each short segment. The same was done for average velocity. We then calculated the average of the variances and velocities for the segments.

We also searched the simulator log files for signs of collisions between the simulated vehicle and surrounding objects.

3.2.5.2 Visual Attention

For each participant *p* and navigation aid *nav*, we also calculated the average percent dwell time, $APDT_{p,nav}$, on the outside world by finding the ratio of the sum of dwell times for all 13 segments and the sum of the total time spent traversing all 13 segments. We used the same approach in calculating the APDT at the standard PND for parts of the experiment when this PND was in use. Finally, we used an analogous approach to calculate how the APDT at the road ahead changed as participant vehicles traveled through five 20 meter segments between consecutive intersections (from 60 m after the preceding intersection to 40 m before the upcoming intersection).

3.2.5.3 Cross-correlation

We calculated the cross-correlation between the instantaneous percent dwell time, IPDT, on the outside world and the short-term variance of two driving performance measures: lane position and steering wheel angle. The IPDT was calculated at a 10 Hz rate by calculating a separate PDT for each consecutive 100 ms window of eye tracker data. Since the eye tracker data is recorded at 60 Hz, we calculated instantaneous PDTs using six eye tracker data samples at a time. For cross-correlation calculations, the IPDT was transformed into the transformed IPDT (TIPDT) such that a TIPDT value of 0 represented 100% IPDT (attention fully on the outside world), while a TIPDT value of 1 represented 0% IPDT (e.g. when the participant is looking at the LCD screen). Thus peaks in the cross-correlation indicate worse driving performance (larger variance values) correlated with reduced visual attention on the outside world (larger transformed IPDT values).

The short-term lane position and steering wheel angle variances, were calculated at a 10 Hz rate for 1 second long windows (i.e., for 10 samples of the given driving performance measure at a time). The choice of 1 second for the window length reflects our expectation that on straight roads the corrections to lane position, accomplished by relatively large changes in the steering wheel angle, will take less than 1 second.

We calculated two cross-correlations. $Rlp_{nav}[lag]$ is the cross-correlation between lane position variance and the TIPDT on the outside world for navigation aid *nav*. Rlp_{nav} was calculated as the average of cross-correlations for each of the 13 segments and each of the 8 participants. $Rstw_{nav}[lag]$ is the cross-correlation between the steering wheel angle variance and the TIPDT and it was calculated analogously to $Rlp_{nav}[lag]$. Both calculations were implemented using Matlab's xcorr function. The *lag* variable indicates the number of samples by which the variance measure lags behind the PDT measure. Thus, for positive values of lag, a peak in the cross-correlation indicates that there is an increase in the variance following an increase in the time the participant spent not looking at the outside world.

3.3 Results

3.3.1 Driving Performance

We performed a one-way ANOVA for each of the driving performance measures with *nav* as the independent variable. We found no significant effects for any of the three variances of driving performance measures or for average velocity. This result mirrors our findings in our preliminary study [13]. We also found no collisions in any of the experiments. Hence, participants were able to pay sufficient attention to the road to avoid contact with other objects or pedestrians.

3.3.2 Visual Attention

To assess the effect of different navigation aids on visual attention, we performed a one-way ANOVA using PTD as the dependent variable. As expected, the time spent looking at the outside world was significantly higher when using spoken directions as compared to the standard PND directions, $p<.01$. Specifically, for spoken directions only, the average PDT was 96.9%, while it was 90.4% for the standard PND.

To assess the effect of distance from the previous intersection on PDT on the outside world for the two navigation aids, we performed one-way ANOVAs for each of the navigation aids using PDT as the dependent variable. For the standard PND, we found a significant main effect, $p<.01$, while the effect was less significant for the PND with spoken output only, $p<.05$. Figure 6 shows the differences in PDT on the outside world. For the standard PND, we also assessed how the PDT on the PND screen changes with the distance from the previous intersection. Using a

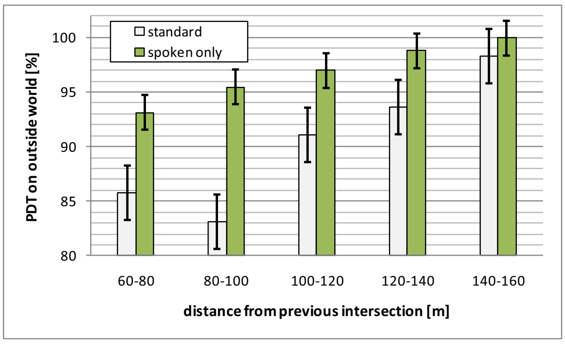


Figure 6. PDT on the outside world (with standard error), changing as vehicles travel between intersections.

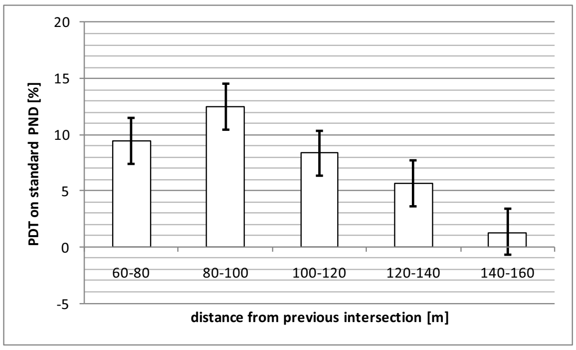


Figure 7. PDT on LCD screen of the standard PND (with standard error), changing as vehicles travel between intersections.

one-way ANOVA we again found a significant main effect, $p<.01$. Figure 7 shows the differences in PDT on the LCD screen for the standard PND. These results indicate that on short road segments, when drivers are expecting to possibly turn at the upcoming intersection, they are likely to look at the display of a standard PND. However, they are less likely to do so as they approach the next intersection.

3.3.3 Cross-correlation

Our cross-correlation analysis indicates that there is a relationship between the IPDT on the outside world and the two short-term variances. This relationship is evident from peaks in the two cross-correlation functions, $Rlp_{nav}[lag]$ and $Rstw_{nav}[lag]$, shown in Figure 8. In order to evaluate whether the peaks arose due to chance, we conducted a randomization test in a manner similar to the one used by Veit et al. [15]. Specifically, while we used pairs of sequences of TIPDT and variance values from the same segment in our cross-correlation calculations (section 3.2.5.3), in our randomization test, we found the cross-correlation between the TIPDT from one segment and variances from a different segment. We created 1000 random arrangements of TIPDT values with respect to the variances. Thus, for each value of *lag* we had 1000 cross-correlation results. For each value of *lag* we then found the bottom $(1\text{-}p)\cdot 1000$ cross-correlation values. We estimated statistical significance by comparing cross-correlation values for the original data with these values. If the cross-correlation for the original data was larger, then the result was considered statistically significant with probability less than p. E.g. to estimate the $p<.05$ significance level, we found the bottom 1000 - 50 = 950 cross-correlation values for each value of *lag*. If, for a given value of *lag*, the cross-correlation value from the original data was larger than these values, the result was statistically significant with $p<.05$.

The cross-correlation results are shown in Figure 8. As the graph in the top part of Figure 8 indicates, for the standard PND, the cross-correlation between transformed instantaneous PDT on the outside world and short-term lane position variance, $R_{lp,standard}$, has several statistically significant peaks. For the most prominent of these peaks, the lag is about 0.8 seconds, indicating that an increase in the lane position variance follows reduced attention to the outside world. The graph at the bottom of Figure 8 indicates that similar peaks exist for the steering wheel angle variance ($R_{stw,standard}$). The two graphs also show that statistically significant peaks exist for PND with spoken directions as well. In tracing the source of the peaks, we found that when drivers were not looking at the roadway, they were looking at either the speedometer,

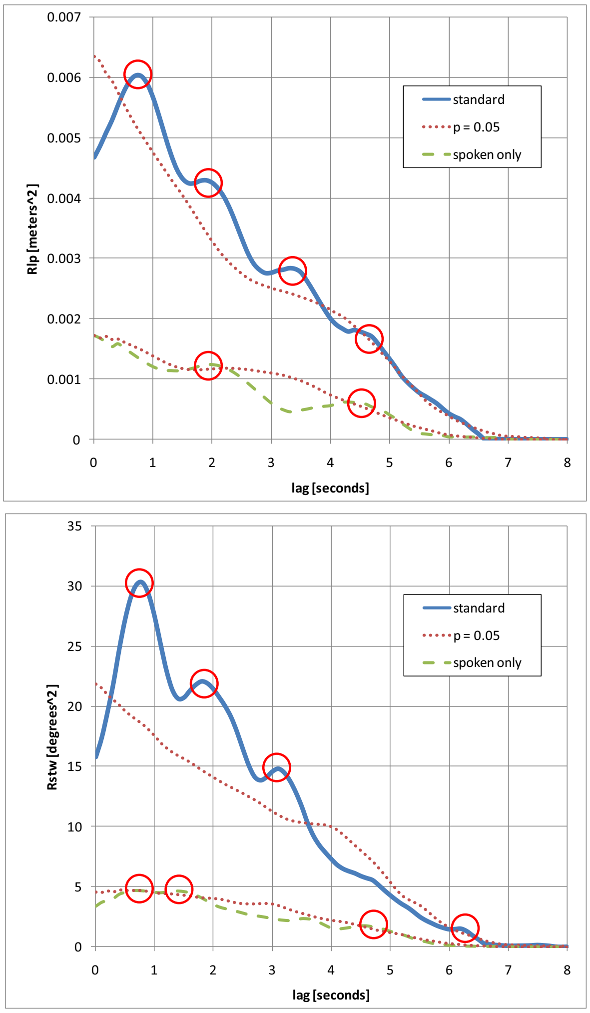


Figure 8. Cross-correlation between TIPDT on the outside world and lane position variance (top) and steering wheel variance (bottom). Circled peaks indicate statistically significant increases in variance occurring after decreases in the IPDT, with the delay indicated by the value of lag.

dashboard, or steering wheel. This is to be expected. However, the peaks for the spoken directions are about six times smaller than for the standard PND.

Why is there such a difference in the magnitude of the effects? Our data indicates that the answer is in the length of gazes drivers use to view the standard PND. Figure 9 again shows cross-correlation values for the two navigation aids, however in this case the cross-correlations were calculated using gazes away from the outside world that are 200 ms or more in length. Clearly, there is a striking resemblance between the graphs in Figure 8 and Figure 9, respectively: peaks are located in practically the same locations and the magnitudes are almost the same. We can conclude that gazes away from the outside world lasting 200 ms or longer are the major contributors to peaks in the cross-correlations. And, as Figure 10 shows, about 60% of all fixations (gazes at the same location lasting at least 100 ms) at the standard PND are in fact at least 200 ms long.

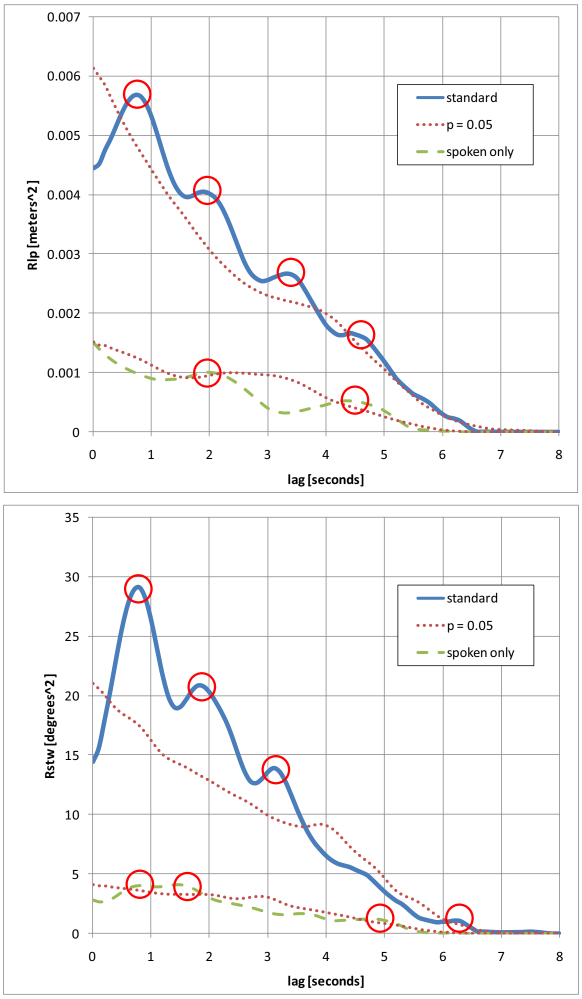


Figure 9. Cross-correlation between transformed instantaneous PDT on the outside world and lane position variance (top) and steering wheel variance (bottom). Calculated only using gazes away from the outside world of 200 ms or longer. Circled peaks indicate statistically significant increases in variance occurring after decreases in the IPDT, with the delay indicated by the value of lag.

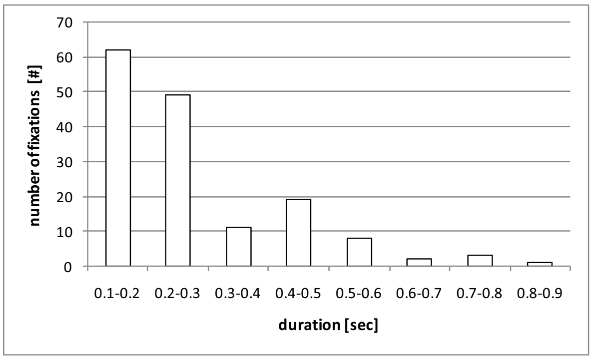


Figure 10. Fixations at the standard PND by duration.

In summary, whenever drivers look away from the road in such a way that it causes higher variance in lane position or steering wheel angle, it is because they are spending at least 200 ms doing so. When a visual display is present, the magnitude of the effect on driving performance is about six times greater. This is probably due to the fact that unlike looking at the dashboard, looking at a map that is changing in real-time requires a fair amount of cognitive effort. Drivers need to mentally parse the information in the display, and that is more distracting.

4. DISCUSSION

In our introduction, we started out by asking two questions.

1. *Does a PND with combined visual and spoken output cause drivers to spend less time looking at the road ahead than a PND that provides spoken output only?*

Because we found a significant difference in visual attention directed at the outside world for the two navigation aids, with drivers spending less time looking at the road ahead when they had a visual display, the answer to this question is affirmative. Note that glancing at the visual display was not necessary to complete the navigation task. In fact, there were no cases of missed directions for any of the navigation aids. For the city route and traffic conditions utilized, spoken directions provided sufficient information without introducing a visual distraction.

2. *What is the effect of glancing at the PND visual display on driving performance?*

Despite the fact that we did not find significant differences in driving performance measures when averaging over all segments, we did find statistically significant peaks in the cross-correlation between the TIPDT on the outside world and the short-term lane position and steering wheel variances. These peaks indicate that there may be a causal relationship between looking away from the outside world (e.g. to look at the PND), and an increase in the variance of lane position and steering wheel angle. We also found that the cross-correlation peaks are larger for gazes away from the outside world lasting 200 ms or longer. This is important since about 60% of all fixations at the standard PND were at least 200 ms long. In other words, the way in which users interact with standard PNDs very often results in looking away from the outside world for more than 200ms at a time. This in turn is correlated with increased short-term lane position and steering wheel variances. Although any increase in the risk of accidents due to these increased variances still needs to be quantified, our results provide designers of in-car navigation aids with reason for caution and a framework for assessing any negative impact on driving due to visual displays.

4.1 Design Implications

With respect to designing in-car navigation aids, our results seem to suggest that if users can trust a PND enough to follow whatever spoken directions they are given, even when they are lost, a navigation system with no visual display may be the most favorable option since visual attention and consequently driving performance will likely be improved. This finding is important for two reasons. First, any sophistical GUI that could hold a driver's attention even more than the simple 2D view we presented, such as 3D terrain maps [2][3], is likely to affect driving performance in an even worse way. Second, small PND devices that rely primarily on speech present viable alternatives to the typical GPS form factor. For example, Verizon VZ Navigator [4] provides spoken turn-by-turn directions along with a map, but on some phones (e.g., flip phones), the map and text are too small to read. Our research suggests that, if the map is intentionally turned off, using these devices may not result in worse driving performance than using PNDs with larger displays, and may even result in better visual attention and consequently better driving performance.

The key to a successful PND interface may be to earn the trust of the users. At the end of our experiment, we asked participants to rate their experiences with the three navigational aids. Five of the eight participants strongly agreed or agreed with the following statement: "I prefer to have a GPS screen for navigation." We hypothesize that this sentiment will be especially strong on roads where users may seek reassurance that they are on the right path. For example, on long road segments, drivers may get anxious that they have missed a turn and may want to get feedback from the navigation aid. These may be times when drivers cast a glance at the visual output of a navigation aid.

5. Conclusion & Future Directions

In this paper, we describe the experimental evaluation of the influence of two navigation aid types on driving performance and visual attention while driving a simulated car in a city environment. We found that participants spent significantly more time looking at the outside world when using a spoken output-only PND compared to using a standard PND with an LCD screen and spoken output. In fact, participants on average spent about 6.5% more time looking at the road ahead when using the spoken output-only PND – a difference of about 4 seconds for every minute of driving. We also found evidence that this difference negatively impacted two driving performance measures: lane position variance and steering wheel angle variance. Specifically, we found statistically significant cross-correlation peaks between the increases in these variances and decreases in the time spent looking at the outside world.

In our next investigation we intend to explore a larger variety of PND displays. We plan to explore interactions with displays that provide egocentric maps, as such maps have been shown to improve user performance on navigation tasks [16], as well as augmented reality navigation aids. We are also exploring building predictive models of when users are likely to look at the PND display for reassurance. Such models could assist the development of spoken only navigation aids that deliver prompts reassuring drivers that they are on the right track.

6. ACKNOWLEDGMENTS

This work was supported in part by the US DOJ under grants 2005CKWX0426 and 2006DDBXK099 and Microsoft Research.

7. REFERENCES

[1] Lowe, S. Many Workers Have Long Commutes to Work. US Census Bureau Press Release. http://www.census.gov/Press-Release/www/releases/archives/american_community_survey_acs/004279.html. March 29, 2005. Retrieved 5/30/2009.

[2] Whenham, T.O. 3D GPS Mapping coming to North America with Nav N Go. http://www.mobilemag.com/2008/01/07/3d-gps-mapping-coming-to-north-america-with-nav-n-go/. Jan. 7, 2008. Retrieved 5/30/2009.

[3] Eaton, K. Navigon's 8100T GPS Has 3D Terrain-View Maps. http://gizmodo.com/5081866/navigons-8100t-gps-has-3d-terrain+view-maps. Nov. 10, 2008. Retrieved 5/30/2009.

[4] Verizon VZ Navigator: http://www.verizonwireless.com/b2c/splash/turnbyturn.jsp

[5] Strayer, D. and Johnston, W. Driven to Distraction: Dual-Task Studies of Simulated Driving and Conversing on a Cellular Telephone. Psych. Science, 12(5), 462-466, 2002.

[6] Lew, H.L., Poole, J.H., Lee, E.H., Jaffe, D.L., Huang, H., and Brodd, E. Predictive Validity of Driving-Simulator Assessments Following Traumatic Brain Injury: A Preliminary Study. Brain Injury, vol. 19, no. 3, pp. 177-188, 2005

[7] Neale, V.L., Dingus, T.A., Klauer, S.G., Sudweeks, J. and Goodman, M. An Overview of the 100-Car Naturalistic Study and Findings. National Highway Traffic Safety Administration, Paper No. 05-0400, 2005.

[8] Tsimhoni, O., Smith, D., Green, P. Address Entry while Driving: Speech Recognition versus a Touch-screen Keyboard. Human Factors, 46, 6 (2004), 600-610.

[9] Salvucci, D. Predicting the Effects of In-Car Interfaces on Driver Behavior Using a Cognitive Architecture. CHI, 2001.

[10] Chisholm, S.L., Caird, J.K., Lockhart, J., Fern, L. and Teteris, E. Driving Performance while Engaged in MP-3 Player Interaction: Effects of Practice and Task Difficulty on PRT and Eye Movements. Driving Assessment, 2007.

[11] Medenica, Z. and Kun, A.L. Comparing the Influence of Two User Interfaces for Mobile Radios on Driving Performance. Driving Assessment, 2007.

[12] Horrey, W.J, Wickens, C.D., Consalus, K.P. Modeling Drivers' Visual Attention Allocation While Interacting With In-Vehicle Technologies. Journal of Experimental Psychology: Applied 12, 2 (2006), 67-78.

[13] Kun, A.L., Paek, T., Medenica, Z., Oppelaar, J.E., Palinko, O. The Effects of In-Car Navigation Aids on Driving Performance and Visual Attention. Technical Report ECE.P54.2009.3, Electrical and Computer Engineering Department, University of New Hampshire, 2009.

[14] Lai, J., Cheng, K., Green, P. and Tsimhoni, O. On the Road and on the Web? Comprehension of Synthetic and Human Speech While Driving. CHI 2001.

[15] Veit, R.R., Pyle, P., McGowan, J.A. Ocean Warming and Long-Term Change in Pelagic Bird Abundance within the California Current System. Marine Ecology Progress Series, 139 (1996), 11-18.

[16] Porathe, T., Prison, J. Design of Human-Map System Interaction. CHI, 2008.

Usability Evaluation of a Volkswagen Group In-Vehicle Speech System

Jackie C. Chang, Annie Lien, Brian Lathrop, Holger Hees
Electronics Research Lab
Volkswagen Group of America, Inc.
4005 Miranda Avenue, Suite 100
Palo Alto, CA 94304
+1 (650) 496-7000
{jackie.chang,annie.lien,brian.lathrop,holger.hees}@vw.com

ABSTRACT
Driving behavior has been trending towards more time in the car and longer commutes. This has fueled the demand for an increasing number of in-vehicle infotainment features, at the cost of the driver splitting attention between the primary task of driving and other secondary tasks. To demonstrate one process we use for generating continuous improvements to the usability of our infotainment systems, we discuss a study where 30 participants were asked to interact with the speech dialogue system of a Volkswagen Group in-vehicle speech system. Participants performed tasks in telephone, navigation, and map contexts. Tasks were timed and videotaped for analysis of three performance measures: 1) Task Completion, 2) Task Time, and 3) participant rating of Task Difficulty. From this analysis, we identified issues that are especially important to the interaction between the system and the driver, which we categorized into a few broad areas: System Organization, Push-To-Talk Functionality, Data Entry, and Speech Commands. Analysis of the issues specific to each category and usability recommendations for each are discussed.

Categories and Subject Descriptors
H.5.2 [Information Systems]: Information interfaces and presentation – *user interfaces, evaluation/methodology.*

General Terms
Measurement, Performance, Design, Experimentation, Human Factors.

Author Keywords
Driver user interfaces, driver safety, voice user interfaces, speech interface, speech technology, speech dialogue systems.

1. INTRODUCTION
With the ever-increasing features available in today's in-vehicle infotainment systems, the need for a simple, easy-to-use interface has become a necessity [2]. In the last 40 years, automotive infotainment displays and control elements have more than doubled [6]. As drivers spend more time in their vehicles, and with the trend towards longer commutes [2], the demand for infotainment features will not subside. Especially with the focus of current events and legislation on hands-free devices and the impact of these secondary tasks on driver distraction, the need for a usable interface that does not distract from the primary task of driving is ever more important. Of the control methods currently available, hands-free speech recognition is one of the most promising methods, resulting in better driving performance, less mental taxation, and less glances off the road compared to manual data entry [1], and better driving quality especially in more complex tasks such as navigation and phone dialing [3]. In fact, it has been said that the level of distraction involved in entering a voice command and listening to the vehicle's subsequent response is so low that it is comparable to that of listening to the car radio [12].

Although there have been attempts at developing natural language speech systems [4, 9, 10, 11], there is yet to be a viable product for the mass market, due to a number of challenges that must still be overcome when communicating between human and machine. These challenges range from issues with the sophistication of the speech technology to the user interaction schemes used to guide task performance. With respect to the speech technology itself, issues such as car noise interference and a limited vocabulary of speech commands that the system can recognize [5] constrain the ways that humans can interact with the system. On the usability side, many speech interfaces do not have a clear and transparent menu structure [5], which leads to confusion about why a certain command is misrecognized in certain contexts but not others. Also, the pace of speech system dialogues is far from the natural pace of human conversation [8]. Oftentimes, information needs to be entered in pieces, such as entering an address with house number, street, and city as separate utterances [4].

In light of the issues mentioned above, the goal of this paper is to illustrate a process of evaluating the usability of a speech interface system, and discuss the method of analysis that led to suggestions and rationale for future improvements.

2. EXPERIMENT DESIGN

We conducted an in-vehicle user study to evaluate the usability performance of a Volkswagen Group speech dialogue system. The goals of the user study were to: 1) Observe participants' interactions with the voice-activated phone, navigation, and map contexts of the system, and 2) Generate system specific recommendations as well as general suggestions and principles for how such a system can best interact with a user. All tasks were timed and performed only through voice operation.

2.1 Participants

Thirty native English speaking adults (17 male and 13 female), between the ages of 21 – 56 years old (*M*=37, *SD*=9), were recruited to participate in the study. Of the 30 participants, all were computer literate and all possessed a valid driver's license with 27 participants having had a license for 10 or more years; 24 participants drove 10,000 or more miles per year; and 26 participants had previous experience with navigation systems. In terms of level of familiarity with navigation systems, on a scale from 1 to 5 (1=Not Familiar, 5=Very Familiar), 25 participants possessed a familiarity level of 3 or more. Participants were recruited from the local community and compensated with a payment of an $80 check for experiment participation. Each experiment session lasted approximately 1.5 hours.

2.2 Apparatus

Throughout the experiment, participants were seated in the driver's seat of a stationary Volkswagen Group vehicle equipped with a speech recognition system for infotainment control, which allowed users to operate contexts such as navigation, phone, media, and setup. Voice commands were available for operating a subset of the functions of the system.

2.3 Procedure

Participants began by signing a consent form and filling out a pre-experiment questionnaire in order to collect their demographic information and driving habits. After completing the pre-experiment questionnaire, each participant received a brief training session so that they could become familiar with the basic operation of the speech user interface and know how to access the system's help feature. Training included instruction and practice on operating the Push-To-Talk button on the steering wheel, listening for the beep as a cue to speak to the system, and accessing the help menu by saying the command "Help." Additionally, participants practiced interrupting the system while the system was speaking and canceling an action by saying "Cancel," or holding down the Push-To-Talk button (long press). No additional training was provided, as for the purposes of this experiment, it was important to keep participants naïve about which commands to use.

After training, participants were given a paper packet that listed 21 tasks in phone, navigation, and map contexts of the speech system (see below, Tasks). All tasks were performed by voice operation only. Tasks were videotaped and timed, with a maximum allowed time of 3 minutes to complete each task. After 3 minutes had passed, if the participant was not done with the task, the experimenter stepped in to finish the task. This was necessary from a practical and logistical standpoint.

Participants were encouraged to use the vehicle's "Help" menu if they were stuck and needed help. Beyond answering any questions the participant had prior to timing began, once the stop watch had started, the experiment administrator only stepped in to clarify a task. Clarification was required if participants seemed confused about a task to the point where they were obviously off task and unaware of it, or thought they had completed a task but in actuality had not. Experimenter intervention occurred in approximately 19% of the tasks.

2.4 Tasks

The following is an example of the sequence of a task. The participant reads the task aloud, for example, "You need to find a house located at *NNN* S Blaney Ave in Cupertino, CA. Please enter this house address as a destination into the navigation system." The participant then asks the experimenter for any clarification if the task was not understood. Once any and all questions have been answered, the participant begins performing the task, and the experimenter starts timing. In the case of correct completion of this task, the participant begins by pressing the Push-To-Talk button and, upon hearing the subsequent beep of the microphone turning on, says the command "Enter destination." The system would then prompt the participant by asking "Please enter the city" and beep when it is ready to receive the next command. The participant states the city name. The system then repeats the city name that it registered, prompts the participant for the street name, and beeps when ready to receive the response. The task is complete once the participant has finished entering a destination in its entirety and activated route guidance. At this point, task timing stops.

The 21 tasks (in order of appearance in the experiment) were as follows:

2.4.1 Telephone Tasks

(1) "Please navigate to the directory using voice command."

(2) "Please call Harrison by finding his phone number in the directory through voice command and dialing the number."

(3) "Please find Jackie's contact information, this time by using voice command to scroll down the list of contacts in the directory until you find Jackie."

(4) "Please navigate to the telephone menu."

(5) "Please call the number (*NNN*) *NNN-NNNN*." (Note: a real phone number was used; it is masked here for security reasons.)

(6) "Please call the international phone number 011 *NN NNN NNNNNNNN*." (Note: a real phone number was used; it is masked here for security reasons.)

(7) "Please redial the last number that you have just called."

2.4.2 Navigation Tasks

(8) "Please switch over to the Navigation menu."

(9) "You need to find a house located at *NNN* S Blaney Ave in Cupertino, CA. Please enter this house address as a destination into the navigation system." (Note: a real address was used; it is masked here for security reasons.)

(10) "You have just realized that the house is actually located at *NNNN* Alma St. in Palo Alto, CA, and you need to correct the address. Please change the destination." (Note: a real address was used; it is masked here for security reasons.)

(11) "You no longer need to go to this house. Cancel the route guidance of the navigation system."

(12) "Instead, you would like to go to Daniel Jones' address in Palo Alto, which has already been saved in your address

book. Find this address from your address book and navigate to it."

(13) "You need to go to the previously entered destination of *NNNN* Castro Dr in San Jose. Please select this address from a list of previous destinations." (Note: a real address was used; it is masked here for security reasons.)

(14) "You would also like to visit the corner of Prospect Rd and Miller Ave in Saratoga. Enter this intersection of Prospect Rd and Miller Ave as a new destination."

(15) "Select Gas Stations as a Point of Interest and use the navigation system to locate and display the addresses of the nearest gas stations. Select the 3rd gas station listed on the second page of the list of gas stations as your Point of Interest."

2.4.3 Map Tasks

(16) "You want to view a map of your vehicle and the surrounding area where you are currently located. Bring up a map of your current location onto the display screen."

(17) "Proceed to zoom out from your vehicle's location on the navigation map display."

(18) "Zoom in on the navigation map to a scale of 50 yards."

(19) "Change the orientation of the map so that the display is oriented northward."

(20) "Change the map display from daytime to nighttime display."

(21) "Switch from a 2D map image to a 3D map display."

In choosing the wording to describe each task, emphasis was placed on colloquial usage, in other words, tasks were not worded to provide clues to the participant as to which speech command to use. Any similarities between task wording and the system's actual speech commands were not intentional.

2.5 Performance Measures

For each task, the following performance measures were recorded:

2.5.1 Task Completion

A task was considered "Complete" if the participant was able to finish the task successfully within the 3-minute time limit. If at 3 minutes the participant was not able to finish the task, the task was considered "Not Complete" and the experiment administrator intervened and finished the task for the participant.

2.5.2 Task Time

Total task time needed to complete a task was recorded with a stop watch. Both the task time including experimenter intervention (which means total task time recorded was more than 3 minutes), and excluding experimenter intervention (which means the maximum task time was cut off at 3 minutes), were recorded.

2.5.3 Task Difficulty

After performing each task, on the same paper packet containing the list of tasks, the participants rated task difficulty on a scale from 1 to 10, with 1 indicating "Very Easy" and 10 indicating "Very Difficult."

3. RESULTS

We utilize the results of our analysis in two main ways: 1) to make system specific recommendations, and 2) to provide general guidelines for enhancing the usability of the voice or graphical interface.

System Specific Recommendations. Through our analysis of individual tasks (see Results by Task), we use the most problematic tasks to identify areas for system specific recommendations. These recommendations can be as detailed as examining the exact steps required to complete that particular task.

General Guidelines for Voice and Graphical Interface. Additionally, we analyze the data for more general problem areas (see Results by Problem Area) to provide guidelines in the overall design of the voice and graphical interface. These guidelines can be applied system wide, and are oftentimes general enough to be relevant to any similar speech interface system.

3.1 Results by Task

The following graphs show results for each task in terms of the three performance measures described above. (Two graphs are shown for the performance measure of Task Time, one which includes experimenter intervention time, and one which excludes experimenter intervention time).

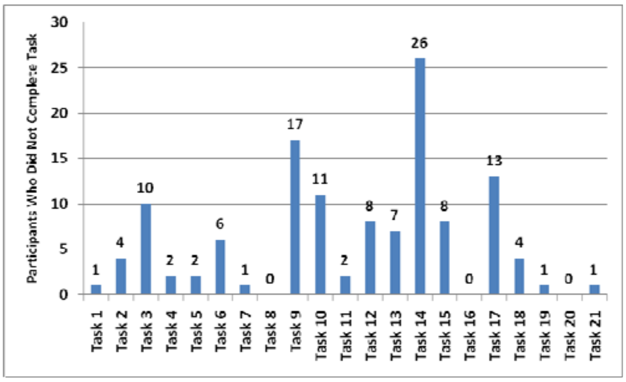


Figure 1. Task Completion (number of participants who could not complete the task)

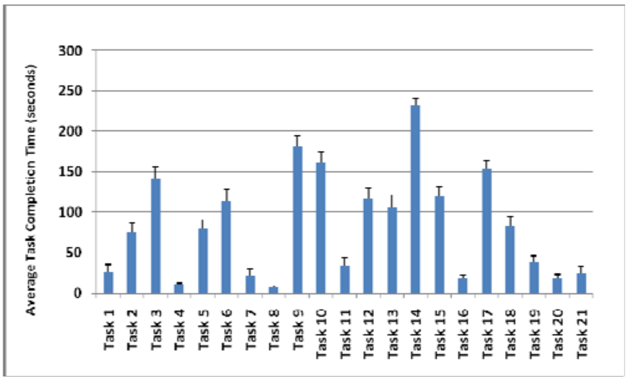


Figure 2a. Average Task Time, *Including* Experimenter Intervention (across 30 participants)

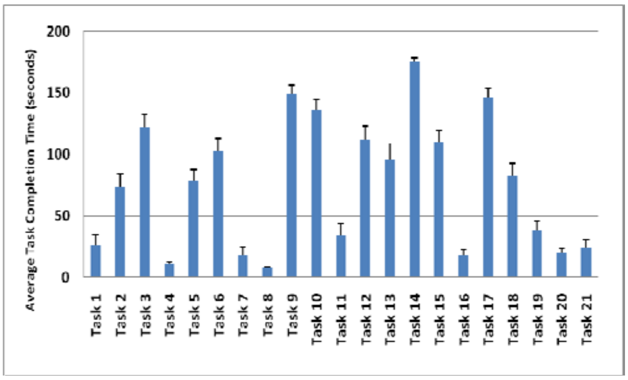


Figure 2b. Average Task Time, *Excluding* Experimenter Intervention (across 30 participants)

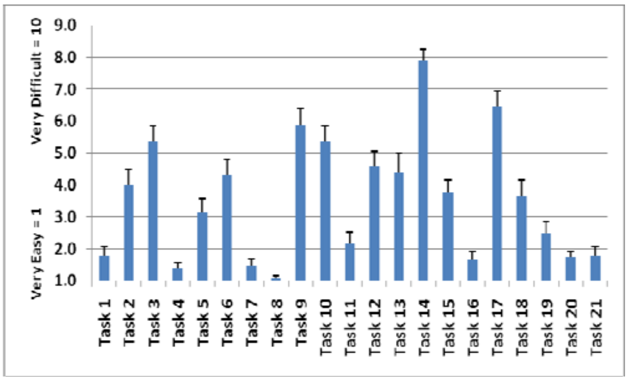


Figure 3. Average Participants' Reported Difficulty Rating (across 30 participants)

The table below (see Table 1) shows the correlations between Task Completion and the other dependent variables.

Table 1. Correlations between Task Completion and Other Dependent Variables

Dependent Variable	Correlation Coefficient
Task Time (*Incl.* Experimenter Intervention)	r = 0.95
Task Time (*Excl.* Experimenter Intervention)	r = 0.91
Difficulty Rating	r = 0.92

Since Task Completion was highly correlated ($r \geq 0.90$) with all other dependent variables, we used Task Completion as the marker in identifying areas for making system specific improvements. In particular, we focused on those tasks where one third or more of the participants could not complete the task. Additional conclusions relating to the other dependent variables will not be discussed here due to their high correlation to Task Completion. Given these criteria, the problematic tasks are listed below (see Table 2).

Table 2. Top 5 Problematic Tasks for Participants

Task 3: "Please find Jackie's contact information, this time by using voice command to scroll down the list of contacts in the directory until you find Jackie."

Task 9: "You need to find a house located at *NNN* S Blaney Ave in Cupertino, CA. Please enter this house address as a destination into the navigation system."

Task 10: "You have just realized that the house is actually located at *NNNN* Alma St. in Palo Alto, CA, and you need to correct the address. Please change the destination."

Task 14: "You would also like to visit the corner of Prospect Rd and Miller Ave in Saratoga. Enter this intersection of Prospect Rd and Miller Ave as a new destination."

Task 17: "Proceed to zoom out from your vehicle's location on the navigation map display."

3.2 Results by Problem Area

After tabulating the task performance data, we did a more thorough analysis of all the videotaped sessions for all tasks to understand the causes of confusion and find improvements that are not necessarily task-specific. Through the experimenter's observation of the participants, the assumptions made from these observations, and the participants' self-reported comments, we identified a number of repetitive and consistent problem areas which disrupted the interaction between the user and the system and hindered participants' ability to perform tasks. The problem areas are system-wide and not specific to any particular task, and often spanned across multiple tasks. Problem areas we identified are comparable to those reported in other studies of speech system usability [3, 4, 5]. Those problem areas that were observed in 5 or more participants are listed in the table below (see Table 3).

Table 3. Problem Areas Identified ($\geq$5 participants with problem)

Problem Area	# Participants with problem
System Organization	
- Global vs. Local Commands	21 out of 30
- Undo or Back	21 out of 30
Push-To-Talk Functionality	
- System Playback Interruption	13 out of 30
- System Misrecognition Not Conveyed	10 out of 30
- Microphone On/Off Not Apparent	23 out of 30
- Timing of Microphone On Indication	15 out of 30
Data Entry	
- Pace of Data Entry	5 out of 30
- Order of Data Entry	17 out of 30
- Format of Data Entry	19 out of 30
Speech Commands	
- Misleading Help Commands	22 out of 30
- Selection of Wrong Command	28 out of 30

To illustrate how problem areas were identified through observation of a task, we use the example of entering an intersection (Task 14). During the course of performing this task, there are a number of problems a participant may encounter.

Many of these problems (for example, "Microphone On/Off Not Apparent" or "Global vs. Local Commands") can probably be overcome in the 3-minute time frame and successful completion of the task is still achievable. Other problems may be more difficult to overcome and thus will ultimately be the main factor in preventing a participant from successfully completing the task under 3 minutes. In our analysis (see Table 3), we have provided a general overview of how many particpants experienced the problem area, which could have occurred in any of the 21 tasks.

Though one of the purposes of this experiment was to find system-specific improvements, we feel the problem areas identified could be generalized to be applicable to any similar speech system design. We will now discuss each problem area in more detail, using examples from the particular system evaluated.

3.2.1 System Organization

In using any system, users feel more comfortable when they have a mental map of how the system is organized. Having an unclear or overly complicated structure leaves the user feeling lost and unable to repeat an action that they have just completed. Having an unclear idea of the system's structure creates further usability issues that manifest themselves in different ways.

3.2.1.1 Global vs. Local Commands

21 out of 30 participants had issues identifying which commands could be used globally (throughout the entire system) vs. locally (in certain contexts, such as navigation only or radio only). Global commands included commands such as "Navigate to [*destination*]," "Call [*name*]," or "Enter Destination." Local commands included such commands as "Enter Street" which is only functional when the user is already in the navigation context.

A related issue that exacerbates the problem is not having a clear delineation between the various contexts. Many systems, understandably, try to link "Map" and "Navigation" contexts, or link "Radio" and "CD/Media" contexts. The problem arises when the system attempts to partially link as well as partially keep these contexts separate. As with the test system used in this case study, while "Map" and "Navigation" were separate screens, they both shared the same color in terms of text font and graphics. This creates problems due to the global and local commands discussed earlier. Users are confused as to which command they can use and when. Contexts, therefore, should be clearly differentiated both in terms of the speech and graphical user interface to minimize user confusion.

3.2.1.2 Undo or Back

21 out of 30 participants had issues making a correction to their task. This happened in two main instances: 1) in making a correction during data entry, and 2) in returning to a previous state or menu. The system we evaluated employed two different commands that were similar but varied slightly in functionality. "Cancel" allowed the user to stop the current dialogue. (For example, when the user says "Navigation", the system repeats "Navigation" once it arrives at the navigation context, then opens the microphone. At this point, saying "Cancel" would stop the dialogue and close the microphone). "Correction" allowed the user to make a correction to a data entry field. To complicate the matter, however, saying "Cancel" during a data entry field (instance 1) does not work because the system attempts to match "Cancel" to the closest sounding phonetic representation, which could be a street name, city name, or person name. For example, during street address entry, saying "Cancel" when the system is expecting a street name causes the system to find the street name that sounds most like "Cancel." In this particular system, there was no way to return to a previous state or menu (instance 2).

An example of an interface where this is not a problem is the computer. Word processing applications have one command (CTRL-Z) which works all the time, either in instance 1 or 2, and simply undoes the last user action. It is a feature that almost all users are familiar with and is easy to understand.

3.2.2 Push-To-Talk Functionality

Although this case study is an evaluation of a speech system interface, we would like to emphasize that we are not evaluating the accuracy of the speech recognition technology. Rather, we are evaluating the usability of the speech interface, in terms of interaction with the user, pace and timing of dialogue, etc.

3.2.2.1 System Playback Interruption

The particular system used in this evaluation repeated the command that it heard back to participants, as a method of feedback and verification. Many speech systems incorporate some form of this. In the example of the system used for this evaluation, when the user states "Navigation," the system repeats the command "Navigation" as confirmation before changing to the navigation context. This worked for users most of the time, but this method of confirmation did not work very well in the telephone context. 13 out of 30 participants expressed confusion when the system interrupted them by repeating back the telephone number digits that they had just entered, prior to their completion of the entire telephone number entry. While the user can continue with the number entry where they left off once this happens, 13 out of 30 users did not know this, and instead thought they had to delete the entire number and start over from the beginning.

A few minor adjustments can be made to this interface in order to provide more clarity to the user. First, instead of repeatedly telling the participant to enter a number when a number has already been partially entered in the telephone number field, the system could give a command such as "Please *finish* the number entry" or "Please *continue* the number entry" which gives the user a cue that the previously entered digits are still acknowledged. A visual cue that can accompany the vocal dialogue is having the phone number field split into three separate smaller data entry fields, rather than having one large data field, which would subtly suggest to the user that the verbal entry can also be broken up into three separate parts. The system can also wait longer before interrupting with the playback, as most people cannot verbally recite an entire phone number without at least one short pause.

3.2.2.2 System Misrecognition Not Conveyed

With any technology that receives input and interaction from a user, such as speech recognition, it is necessary to convey to the user whether a misrecognized command is the fault of the system (speech recognition) or the fault of the user (wrong command used). The system sometimes asks for verification or asks the user to repeat the command when it does not recognize a command. However, above a certain threshold of certainty, the system does not ask the user to repeat, but rather simply executes the command. We noticed that in 10 out of 30 cases, the participant had in fact used the correct command, but due to system misrecognition and the fact that the system did not ask for verification, the user thought the wrong command was used and thus never tried that particular command again. One way to alleviate this problem is for the system to repeat what it thought it heard every time, while executing the command which allows the user to go back and repeat the command if necessary.

3.2.2.3 Microphone On/Off Not Apparent

Given the noisy environment of the vehicle and the fact that speech recognition systems cannot yet detect the difference between human-to-human speech and commands directed at the system, it necessitates that there is a state when the microphone of the system is turned on in order to receive commands from the user, and turned off in order to ignore speech not intended for the system. In the system that we evaluated, the differentiation between the microphone on and off states was not always apparent to the user (in 23 out of 30 cases). The system indicates that it is ready to receive a command by beeping as well as showing a microphone on icon on the screen. In order to improve upon the current interface, we suggest that the icons for microphone on and off be more visually differentiated. Some systems make it apparent for the user when they can talk because the user needs to hold down a button the entire time that speech is being inputted. However due to the fact that both hands are occupied while driving, instead we suggest that when the microphone closes, the system can say "microphone off" to make it obvious to the user.

3.2.2.4 Timing of Microphone On Indication

As stated previously, the system evaluated uses a short beep to indicate to the user every time the microphone is on and ready to accept speech commands. The timing of the beep is such that for very short commands in rapid succession, the user's command is in many cases partially not registered by the system because the user spoke before the beep has occurred. In such instances, which happened in 15 out of 30 cases, the pace of the dialogue can be improved such that it more closely mimics the pace of natural conversation. Additionally, there can be a buffer so that the system is listening for a command already, shortly before the beep occurs.

3.2.3 Data Entry

Given the linear nature of speech input, data entry is a special consideration. The main use cases of data entry that were tested are: 1) telephone number entry, and 2) destination entry for navigation.

3.2.3.1 Pace of Data Entry

In the telephone context, when dialing a phone number, 5 out of 30 participants did not know that digits can be said in groups, for example, the user saying "650" [wait for system to register] versus the user saying "6" [wait for system to register], "5" [wait for system to register], then "0" [wait for the system to register]. This caused considerable frustration on the users' part because data entry could easily take twice as long when it is done digit by digit. One easy way that this could be solved is by providing a visual cue to users, for example (as stated previously), breaking up the telephone number field into three separate smaller fields as opposed to having one long field.

3.2.3.2 Order of Data Entry

Another unique aspect of navigation systems is that sometimes data entry must occur in a specific order before the system can move forward. This can be difficult for non-experienced users to understand, because it is counterintuitive to how a computer works. It is especially problematic when entering a destination for navigation. Given the nature of navigation systems, oftentimes the database of addresses needs to be narrowed down before the system can find the proper data. For instance, most systems require the user to specify the state in the U.S. where the destination is located before it can find a street. This is so that the system can narrow down the possible matching streets, since it does not have the vast memory and processing power of a desktop computer. When users are not aware of this mannerism of in-vehicle navigation systems, this can cause some problems. In 17 out of 30 cases, participants did not know that in order to find the particular street that they are interested in, they need to have the correct city first. For example, when the city field is prefilled to "San Francisco" (from the previous destination entry), unless the user changes the city to "San Jose" prior to entering the street name, the system will only look for streets in "San Francisco."

To alleviate this problem, systems should clearly gray out unavailable fields, in order to guide the user into the correct order of entry. The order of the destination entry fields on the screen itself can also be rearranged. The other option is for the street entry prompt to clue the user in to the fact that only streets in the displayed city will be found. For example, changing the prompt from "Please enter a street" to "Please enter a street in the displayed city" will alert the user to this fact. If this is not enough, the prompt can say something more discrete such as "The city field needs to be updated first before a street can be entered."

3.2.3.3 Format of Data Entry

Because the system does not possess human understanding of speech, in addition to the data input that is needed, another piece of information that needs to be conveyed to the user is the format of the data input that is needed. This is most apparent in destination entry tasks. Many participants will enter an address by saying street and house number all as one string, for example, "*NNN* South Blaney Avenue" all as one string for the street field, rather than saying "South Blaney Avenue" for the street field and "*NNN*" for the house number field. Another issue occurs when entering an intersection. The system asks "Please enter the intersection." It is unclear what format the data for an intersection is supposed to be. Is it the first street, then the second street? Or both street names at once, separated by an "and"? When designing a system, these considerations need to be put in place, because when speaking to another human being, the particular format does not matter—all formats can be understood.

3.2.4 Speech Commands

The particular wording of the speech commands themselves has a large effect on the usability of the system. First, many real life users might not consult the system's instruction manual, and second, commands that more closely match users' natural predilection will be more memorable, and thus perceived as easier to use.

3.2.4.1 Misleading Help Commands

In using any unfamiliar system, the user looks to the instructions given by the system as guidance. When the instructions are misleading, it is almost impossible for the user to disregard the misleading information and do what is intuitive to him or her. 22 out of 30 participants came across this problem in the particular instance of attempting to enter in a destination as an intersection into the system. In the destination entry process, once participants specify the city and street, they have the option to enter either a house number or a second, intersecting street. The voice guidance at this point tells the participant to enter a house number. It does not mention that an intersecting street can also be entered at this time. While a minority of participants were able to intuitively guess that an intersecting street might also be acceptable by the system at this juncture, the majority did not figure this out.

3.2.4.2 Selection of Wrong Commands

Commands that most naturally correspond to local dialect are those that will be most accessible and memorable to the user. In a few instances of our evaluation, having a poorly chosen command for a particular feature in the system rendered that feature almost impossible for the user to access, even with the aid of the help menu list of possible commands. The most poignant example was the task of zooming out on the navigation display map. While users are familiar with the terms "zoom in" and "zoom out," the system's commands for this feature were "Map Smaller" and "Map Larger." Since this is not the vocabulary that users commonly use to refer to this feature, even after they had viewed a list of possible commands, of which "Map Larger/Smaller" was one of them, 28 out of 30 participants still could not activate the function. To further confound the issue was the fact that another, unrelated feature used similar language as what was thought to be appropriate for the zoom in/zoom out feature. The feature of "Intersection Zoom," where the vehicle zooms in on the map display whenever the vehicle approaches the intersection, was commonly mistaken for zooming in and out of the map.

4. DISCUSSION

The current study discusses the methodology used and the results found in our evaluation of the Volkswagen Group speech system. Questionnaire and videotape data were collected across 30 participants. Statistics were compiled for task performance and by problem area. Major problem areas in System Organization, Push-To-Talk Functionality, Data Entry, and Speech Commands were identified. System specific suggestions as well as general recommendations for addressing these common speech interface usability issues were discussed.

An obvious limitation of this study was the fact that it was conducted while the vehicle was stationary. (Though it has been reported that speed of performing a speech task is relatively unaffected by whether the participant was driving or stationary [7]). However, in order to gain a more complete picture of the impact of driving on speech system use performance, the experiment should be replicated in a closed driving course during real driving, or during simulated driving using a driving simulator.

Although we touched on some visual cues in the context of how they could have supported the speech interface, a much deeper analysis into the system in its entirety, speech and visual cues (and maybe even tactile cues) in conjunction, could provide some more elegant solutions to improving the usability of the system.

Additionally, the task list could be broadened to cover some other areas of functionality. The current task list was developed to research the most common use cases according to our own personal experiences.

Also, to have a more accurate evaluation of different implementation details, the same experiment could be replicated with different vehicle infotainment systems of various brands.

The current study discusses findings and results in terms of usability from a human perspective. Although current limitations to speech recognition technology (difficulty distinguishing between similar sounding words, processing power, and limited database of commands) constrains much of what the system can do, the goal of this paper is not to offer a definitive solution that can be technically realized, but rather to explain why some current ways of implementation can be confusing to a user. In this way, we are influencing the future design of our systems by offering some explanations for the mismatch between the mapping of the system design and the human brain.

5. ACKNOWLEDGMENTS

We thank Kristin Dettmer and Dr. Stefan Christ for their support of this project. The statements contained in this document represent claims of the authors only, and are not necessarily representative of the views of Volkswagen AG, or their associates. We would also like to thank Jessica Yuan for her assistance in data collection, and Jerry LeBlanc of Volkswagen Group of America, Inc. for sharing his expertise regarding the system's functionality.

6. REFERENCES

[1] Barón, A., Green, P. Safety and Usability of Speech Interfaces for In-Vehicle Tasks while Driving: A Brief Literature Review. Technical Report UMTRI-2006-5. University of Michigan Transportation Research Institute (2006).

[2] Driver Distraction: trends and issues. Computing & Control Engineering, IEE (2005), 28-30.

[3] Gärtner, U., König, W., Wittig, T. Evaluation of Manual vs. Speech Input When Using a Driver Information System in Real Traffic. SENECA Evaluation. DOI=http://ppc.uiowa.edu/driving-assessment/2001/Summaries/Driving%20Assessment%20Papers/02_Gartner_Wittig.htm

[4] Gruenstein, A., Reimer, B., Orszulak, J., Mehler, B., Liu, S., Seneff, S., Roberts, S., Glass, J., Zabel, J., Coughlin, J. City Browser: Developing a Conversational Automotive HMI. In Proceedings of CHI 2009, ACM Press (2009), 4291-4296.

[5] Hataoka, N., Araki, M., Matsuda, T., Takahashi, M., Ohtaki, R., Obuchi, Y. Evaluation of Interface and In-Car Speech – Many Undesirable Utterances and Sever Noisy Speech on Car Navigation Application. In Proceedings of IEEE Workshop on MMSP (2008), 956-959.

[6] Reissner, U. Gestures and Speech in Cars. DOI= http://www14.informatik.tu-muenchen.de/konferenzen/Jass07/courses/3/3_Gestures-and-Speech-in-Cars_Report.pdf

[7] Tsimhoni, O., Smith, D., Green, P. Address Entry While Driving: Speech Recognition Versus a Touch-Screen Keyboard. Human Factors. Human Factors and Ergonomics Society (2004), 600-610.

[8] Villing, J., Holtelius, C., Larsson, S., Lindström, A., Seward, A., Åberg, N. Interruption, Resumption and Domain Switching in In-Vehicle Dialogue. In Proceedings of GoTAL, Springer LNCS/LNAI, 5221 (2008), 488-499.

[9] Weng, F., Varges, S., Raghunathan, B., Ratiu, F., Pon-Barry, H., Lathrop, B., Zhang, Q., Bratt, H., Scheideck, T., Xu, K., Purver, M., Mishra, R., Lien, A., Raya, M., Peters, S., Meng, Y., Russell, J., Cavedon, L., Shriberg, E., Schmidt, H., & Prieto, R. CHAT: A Conversational Helper for Automotive Tasks. In Proceedings of Interspeech: ICSLP (Pittsburgh, Pennsylvania, September 17-21, 2006).

[10] Weng, F., Yan, B., Feng, Z., Ratiu, F., Raya, M., Lathrop, B., Lien, A., Mishra, R., Varges, S., Lin, F., Purver, M., Meng, Y., Bratt, H., Scheideck, T., Zhang, Z., Raghunathan, B., Peters, S. CHAT To Your Destination. In Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue (Antwerp, Belgium, September 2007).

[11] Yan, B., Weng, F., Feng, Z., Ratiu, F., Raya, M., Meng, Y., Varges, S., Purver, M., Lin, F., Lien, A., Scheideck, T., Raghunathan, B., Mishra, R., Lathrop, B., Bratt, H., Peters, S. A Conversational In-car Dialog System. In Proceedings of NAACL/HLT (Rochester, New York, April 22-27, 2007).

[12] Yoshitsugu, N., Miki, Y., Ito, T., Matsunaga, M. Study of Driver Distraction Due to Voice Interaction. SAE International (2003), 71-75.

Enhanced Turning Point Displays Facilitate Drivers' Interaction with Navigation Devices

Myounghoon Jeon[1], Junho Park[2], Ubeom Heo[2], & Jongmin Yun[2]

[1]Sonification Lab, School of Psychology
Georgia Institute of Technology
654 Cherry Street
Atlanta, GA 30332 USA
+1 404 894 8265
mh.jeon@gatech.edu

[2]Corporate Design Center
LG Electronics
Seocho R&D Center
Seoul, Korea
+82 2 2005 3319
{juncrom, playsoul, ericyun}@lge.com

ABSTRACT

Recently, the use of in-vehicle navigation devices, such as PNDs (Personal or Portable Navigation Devices) has become pervasive, and the device functions have been rapidly expanded and updated. Unfortunately, drivers often have considerable difficulty using these complex technologies. To improve and optimize PND user interfaces, the present study suggested several display improvements for the turning point, which is one of the critical usability issues. *Advanced Turn-By-Turn Display* and *Spatial Turning Sound* were suggested to facilitate the preparation of the next turns. *Leading Tones for Turning* was also presented to help drivers tune the timing of their turns. We evaluated these new concepts with domain experts in three countries, and improved the details of the functions. We are currently implementing those features and looking forward to demonstrating new displays on the real product in our presentation at the Automotive User Interface conference.

Categories and Subject Descriptors

H.5.2. **[Information Interfaces And Presentation (e.g., HCI)]**: User Interfaces – graphical user interfaces (GUI), interaction styles (e.g., commands, menus, forms, direct manipulation), user-centered design

General Terms

Design, Human Factors, Performance

Keywords

Advanced Turn-By-Turn Display, AUI, GUI, IVTs, Leading Tones for Turning, PND, Spatial Turning Sound

1. INTRODUCTION

Driving is one of the most attention-demanding tasks in modern everyday life, with dangerous contexts and complex human-system interactions. Driving is even more challenging when the driver simultaneously uses in-vehicle navigation devices such as Personal or Portable Navigation Devices (PNDs), because it requires multi-tasking and can result in additional distraction from the primary driving task. Not surprisingly, this inattention to the driving task has been identified as one of the leading causes of car accidents. Research has pointed out that the increasing provision of a range of types of complex in-vehicle technologies (IVTs) means that the problem of driver inattention is likely to become even worse [1, 5, 9].

To compensate for this risk, the latest generation of PNDs has adopted more sophisticated navigation features including 3 dimensional maps, a quick spelling, and voice recognition [17, 18]. On the other hand, these new devices also have extended non-navigation functions involving music, movies and telephone. Despite the pervasive use of PNDs (which should make users more familiar), and updated technology (which has been done in an attempt to make the interaction better), users still complain about the difficulty of using PNDs. Even the most basic functions (e.g., entering an address, or learning when to make the next turn) are still in need of considerable research and enhancement. For example, various vendors have begun to support a 3 dimensional display as well as a bird's eye viewing angle, but this does not seem to help users identify the precise time or place to make a turn. Rather, it causes information pollution by conflicting 3D image with text on it. Previous research has shown that a visually optimized navigation system can decrease the total map fixation time and the number of glances needed to interpret the display [8]. This type of benefit using abstracted information properly illustrates how we can overcome the naïve realism in display design [13], but it often remains to be implemented effectively in real devices.

In order to provide a more effective display and safer use of PNDs, we focused on improving the way to present information pertinent to turning points, which is the most fundamental display problem of navigation devices.

2. ISSUES WITH CURRENT TURNING POINT DISPLAYS

Once driving starts, the PND provides visual and auditory information regarding turning points. The use of both visual and auditory cues makes a lot of sense, since it allows the driver to listen to cues while driving, when necessary. From a more theoretical perspective, models of multimodal information processing, such as Wickens' Multiple Resources Theory [16], have led many researchers to study this multimodal approach, particularly in terms of the use of spoken turning commands from the navigation system [3, 6, 14, 15]. Typically, better results are obtained with the multimodal navigation system than with visual-only PNDs. However, there remain considerable issues still to overcome.

2.1 Turning Point Planning and Preview

The first category of usability problems with turning point displays relates to the planning of routes, and the planning and previewing of upcoming turns.

Memory Capacity Issues.

Before getting started to drive, drivers can check all of the routes to their destination on the PND. They can trace the route with using a simulation function. They can also get an overview of important turning points with turn-by-turn list. These functions are clearly helpful in preparing for driving because they can form a schema on the entire route. Nevertheless, they cannot memorize all the directions where they have to go in every single road. What they can memorize are just overall destination direction and a few intersections. It is necessary to provide directions in sequence, and preferably in such a way that the driver need not look down at the list of turns, or navigate from map view to list view.

Advanced Planning.

One of the most important reasons why drivers need more information for further directions while driving is that they should prepare for turns in advance. Although the current turning arrow display can make drivers expect the next direction and prepare for it, it is not sufficient. What if they have to turn again just after the next turn? If a driver needs to turn left just after right turning, she must change lanes immediately after the right turn. Drivers have to decide which lane they will turn into, depending on the next turning direction after the current one. The importance of advanced planning in the dynamic context has been stressed in various fields. For instance, expert musicians play even unknown scores well, because they read the next several notes in advance, which allow them to prepare for the next whole sequence of movements [12]. For drivers, multi-turn planning needs to be part of the instructions, and presented before the first turn, in order for adequate sequencing of sub-goals.

Decision Making.

Even with a PND, drivers in an unfamiliar locale have a high possibility of missing the correct turning point. Even though they listen to the voice guidance, they might not have confidence to turn when directed. Part of the problem is trust in the technology, and part of the issue is a mismatch between the instructions and the view out the window. Visual displays on the small screen are confusing and distracting, and do not have realistic images. Improving the context of the instructions can help enhance the match between system and street, and thus increase the driver's recognition of the correct turning location, and therefore trust in the system. As an example, Reagan and Baldwin [11] suggested that when voice instructions included a *salient landmark*, driving performance was significantly improved. For example, a voice prompt that says, "Turn right in five miles *at the police station*" should lead to better results than a prompt that does not include the police station landmark.

2.2 Cue Sound Location

Typical turning point instructions include a series of prompts, progressively closer to the turn. For instance, listeners may hear a voice prompt at 3km, 1km, and 500m before the turn. While this may help planning to some degree, to date all of these sounds are recorded (or synthesized) and played in mono, via a single speaker on the PND, or via both stereo channels of the car stereo system. This has the effect that the sound cues appear to originate either very near to, or in front of the listener (driver). While this may not be detrimental, per se, it is generally regarded as more congruent to have the sound cue originate from the same side as the required action. That is, a right turn could be cued by a sound coming more from the right side.

2.3 Turning Synchronization

At the first stage of using the PND, many people complain that they cannot know exactly when or where they are to make a turn. Even experienced users experience the same problem because the tuning of their timing is different from the system timing. Some people simply give up using a PND before they become familiar with it. Though PNDs currently present various ways to inform the user of the precise turning point, individual differences between users will always be a big obstacle to overcome. Clearly, there is a need for some way for the PND to overcome the timing-synchronization issue, in order for users to achieve the fast acclimation and adjustment to the timing of the various turn prompts.

3. REDESIGN OF TURNING POINT DISPLAYS AND BENEFITS

Based on these issues, this paper presents several solutions in terms of visual and auditory displays. Solutions involve two separate display timing points. One is preparing for the current turn in advance. The other one is just within a measurable distance of turning i.e., just before turning. For this purpose, visual and auditory components are added to each context.

3.1 Advanced Turn-By-Turn Display

First, to predict and prepare for turnings more properly, we created *Advanced Turn-By-Turn Display*, which could display the next several turning directions on the map screen. If the route requires a second turn soon after the first turn, the PND automatically displays a piece of additional turning information beside the current turning arrow. Moreover, if users touch the arrow, they can check additional turning points (see Figure 1). Once users touch it again, it will disappear. Otherwise, it automatically disappears in a few seconds.

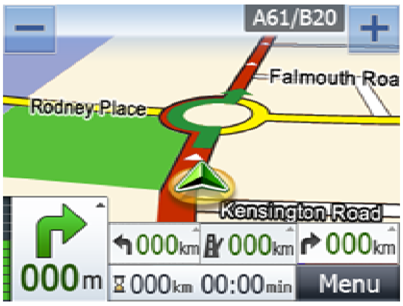


Figure 1. Screen capture of *Advanced Turn-By-Turn Display*. It shows several next turning points from left to right.

Suppose that the drivers stop at the crosswalk on the red light, they might want to check the next few turning points. Previously, for this, drivers had to enter the menu and navigate several depths more in order to reach the Turn-by-Turn list. After checking the list, they had to return to the current map display. In contrast, by using the *Advanced Turn-By-Turn Display*, they can check it by only one touch of the map screen and they can also leave it on the screen. This means drivers might feel that the *Advanced Turn-By-Turn Display* requires navigating a shorter physical and psychological distance and is more approachable than the current Turn-by-Turn list in the menu. Therefore, it can provide drivers with advanced awareness of future required lane changes and further turns, and can allow them to be free from our typically limited memory capacity that might otherwise be a problem when driving a route.

3.2 Spatial Turning Sound

To enhance any potential benefits of cue-response compatibility, we devised a *Spatial Turning Sound* (see Figure 2). If the next turn is right in a mile, the PND may say, "Turn right in one mile." To date, it has been generated in mono. In this newer version, the sound is provided in stereo. That is, if the next turn is to the right, the sound generates from the right speaker.

Spatial Turning Sound uses the basic perception principle of spatial sound. It would affect users' anticipation of the turning direction. Even if users cannot know it consciously, it might render a type of subliminal perception like a framing effect. Users can obtain additional information from the acoustic properties of sound (such as spatial location) before they interpret the meaning of the words. This can clearly lessen the information processing load for drivers. Even if users miss the message of the voice prompt due to a dialogue with passengers or radios, they could identify the next turn direction from the spatialized location of the audio cue. According to Ho and Spence [7], spatial attention is attracted more efficiently when information presented to multiple senses originates from approximately the same spatial region. Thus, *Spatial Turning Sound* may play a role in terms of attracting drivers' attention.

3.3 Leading Tones for Turning

Finally, the sound presented just before turning was redesigned. This *Leading Tones for Turning*, generates tones of increasing duration and pitch, like "Pip.. Pip.. Pip.. PiiiiP" (see Figure 3). Adding contextual sounds before the exact moment might help users sense the appropriate timing.

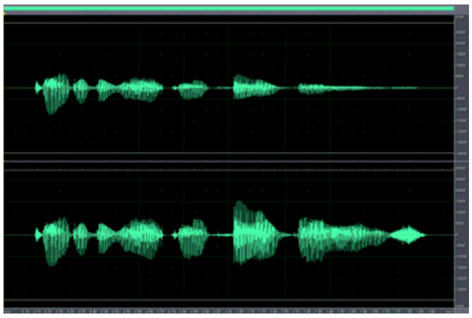


Figure 2. Screen capture of redesigned voice prompt. *Spatial Turning Sound* pans from the center to the right.

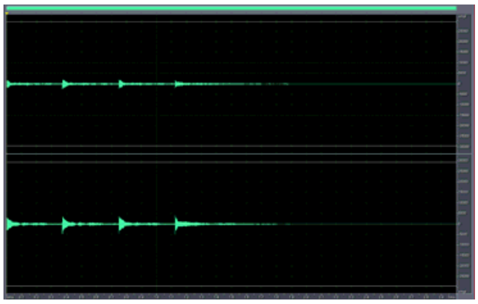


Figure 3. Screen capture of redesigned *Leading Tones* just before the turning. It consists of a series of leading sounds.

Even though the current concept of presenting a short sound like an earcon [2] just before turning has been recently added to many PNDs in order to help turn timing, it still tends to make users miss the correct timing because the processing from 'perception' to 'behavioral reaction' requires a certain time. Drivers still need anticipation and preparation for the precise turning timing. *Leading Tones for Turning* can let users perceive where the next turn is more accurately, and be ready to turn appropriately. Furthermore, through *Leading Tones*, users are able to compute the accurate turning timing by using all of the series of sounds.

Both these two auditory displays are easy to make and implement because the device can always use the same files in each situation. Vision is the most heavily taxed sense in driving, even though driving requires integration of information coming from multiple modalities [4]. Thus, the workload of the visual modality could be lessened by using the additional auditory modality. Further, all of these new features might help users in terms of decision making. Based on this additional information for turning point display, drivers can more conveniently decide whether to make a turn at a particular time or not.

4. FGI & DESIGN IMPROVEMENTS

We conducted several Focused Group Interview sessions for gathering experts' feedback and improving these newer display design concepts, in the U.S.A., Hungary, and Germany.

4.1 Participants

Seventeen participants (all male) participated in the FGI sessions. They ranged from telematics system providers and car audio specialists to salesmen at electronic goods stores.

4.2 Materials & Procedure

For the *Advanced Turn-By-Turn*, a simple movie clip was created in Flash 8.0. For the auditory display features, we composed wave files using Cubase SX 3.0 and played them via Microsoft Power Point 2003. Total seven FGI sessions (the U.S.A. and Hungary = 2, Germany = 3) were conducted including one to four participants in each Focused Group. FGI sessions were held at our office or the participants' office in each country. At first, a coordinator introduced the new display design concepts, using the Power Point slides for visuals, and playing sounds via stereo desktop speakers. Another interviewer simultaneously took notes of the participants' comments using a laptop computer.

4.3 Design Improvements

As a result of subsequent FGI sessions, we gained a couple of critical improvements as well as the preference of the most of participants (domain experts). Among them, the present paper describes two major points pertinent to each display design. The first one was related to the compatibility issue of the *Advanced Turn-By-Turn Display*. Some said that the top to bottom order of the turning point display is congruent with typical reading flow, but some preferred the bottom to top because it is compatible with the moving direction of the vehicle. This meant that regardless of which design we implemented, about half the users would have an incongruent display. To solve this compatibility issue, we changed it into the left to right order. The leftmost arrow means the nearest turning point and the rightmost arrow denotes the farthest turn. Since in most of countries except some at the Middle East, people read from left to right, we could expect that it would work well.

The next suggestion enhanced the *Spatial Turning Sound* presentation. It was suggested that if the sound moves to either side, the dynamic sound should be more compelling and more commanding of attention [10]. For these reasons, we developed new dynamic turning cues to move from the center out to directed side.

5. CONCLUSION & FUTURE WORKS

This paper presented the visual and auditory display concepts for facilitating drivers' interaction with a navigation device and potential users' benefits. Subsequent FGI results showed that experts favored those features and improved the details. These optimized turning point displays might dramatically decrease the driver's perceptual and cognitive load during navigation tasks which would lead to increased safety for drivers with use of IVTs. Despite this promising expectation, work is still needed to further validate those concepts in the context of real driving with normal traffic sounds. Therefore, future research is planned to evaluate one of our new models which incorporates those features.

6. REFERENCES

[1] Ashley, S. (2001). Driving the info highway, *Scientific American, 285(4)*, 52-58.

[2] Blattner, M. M., Sumikawa, D. A., & Greenberg, R. M. (1989). Earcons and icons: Their structure and common design principles. *Human-Computer Interaction, 4,* 11-44.

[3] Burnett, G., & Joyner, S. (1997). An assessment of moving map and symbol-based route guidance systems, *Ergonomics and safety of intelligent driver interfaces*, 115-136.

[4] Deblasio, J. M., Jeon, M., Walker, B. N., & Nees, M. (2009). Designing an auditory display: Report on a case study for in-vehicle use. Georgia Institute of Technology Sonification Lab Technical Report, 2009.

[5] Dukic, T., Hanson, L., & Falkmer, T. (2006). Effect of drivers' age and push button locations on visual time off road, steering wheel deviation and safety perception, *Ergonomics, 49(1),* 78-92.

[6] Gish, K., Staplin, L., Stewart, J., & Perel, M. (1999). Sensory and cognitive factors affecting automotive head-up display effectiveness, *Transportation Research Record: Journal of the Transportation Research Board, 1694(1)*, 10-19.

[7] Ho, C., & Spence, C. (2005). Assessing the effectiveness of carious auditory cues in a capturing a driver's visual attention. *Journal of Experimental Psychology: Applied, 11(3)*, 157-174.

[8] Lee, J., Forlizzi, J., & Hudson, S. E. (2005). Studying the effectiveness of MOVE: A contextually optimized in-vehicle navigation system. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 571-580. NY: ACM Press.

[9] Patten, C., Kircher, A., Ostlund, J., & Nilsson, L. (2004) Using mobile telephones: Cognitive workload and attention resource allocation," *Accident Analysis and Prevention, 36(3),* 341-350.

[10] Polkosky M. D., & Lewis, J. R. (2001). The function of nonspeech audio in speech recognition applications: A review of the literature, IBM Voice Systems Technical Report TR 29.3405, 2001.

[11] Reagan, I., & Baldwin, C. L. (2006). Facilitating route memory with auditory route guidance systems. *Journal of Environment Psychology 26*, 146-155.

[12] Sloboda, J. A. (1982*). The musical mind.* Englewood Cliffs. NJ: Prentice-Hall.

[13] Smallman, H.S. & St. John, M. (2005). Naïve realism: Limits of realism as a display principle. *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting,* 1564-1568.

[14] Srinivasan, R., & Jovanis, P. (1997). Effect of selected in-vehicle route guidance systems on driver reaction times, *Human Factors, 39(2),* 200-215.

[15] Streeter, L., Vitello, D., & Wonsiewicz, S. (1985). How to tell people where to go: Comparing navigational aids, *International Journal of Man-Machine Studies, 22(5),* 549-562.

[16] Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3(2),* 159-177.

[17] www.mybecker.com

[18] www.tomtom.com

AUTHOR INDEX