Task Specific Eye Movements Understanding for a Gaze-Sensitive Dictionary

Abdelaziz Khiat, Yoshio Matsumoto and Tsukasa Ogasawara
Nara Institute of Science and Technology - NAIST
Graduate School of Information Science
630-0192 Nara, JAPAN
abdela-k@is.aist-nara.ac.jp

ABSTRACT

In this paper, we study the relation between the user's degree of understanding and his/her eye movements; in an effort to realize a proactive interface that monitors the user and provides a contextual support. The application is a gaze sensitive dictionary that helps the user when reading a text in a browser's window. Not only is the user's gaze analyzed but also the context and thus the difficulty degree of the text being read. The experiment results suggest using regressions as an indicator to trigger the help process along with a context grounding approach.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems—*Human Information Processing*; I.3.6 [**Computing Methodologies**]: Methodology and Techniques—*Interaction Techniques*

General Terms

Human Factors, Measurement

Keywords

Contextual support, eye movements, proactive interface.

1. INTRODUCTION

Gaze-based research has gained growing interest in the last decade [1]. Within this particular field many applications were built for object selection, object moving, text scrolling, menu commands and more [2]. All of which are using gaze as a new kind of pointing device. On another hand some researchers support the fact that we can infer the cognitive process of a person from the eyes or the gaze [3, 4]. Tacit information plays an important role in our lives, to the extent that some people can communicate without uttering a single word. This extraordinary capability is attracting and is revealing another source of information that has not yet been fully explored.

Our claim is that we could improve the flow of information between the computer and the user if we succeed to decipher the hidden information available in gaze patterns. Some works in this direction are the adaptive interface in [5] and iDict in [6]. Recent findings state that the vision process is not only a passive process working on the retina level but a cognitive process is also taking place at the same time [7] leading to the necessity to model the vision process.

Copyright is held by the author/owner(s).
IUI'04, January 13–16, 2004, Madeira, Funchal, Portugal.
ACM 1-58113-815-6/04/0001.

2. PROACTIVENESS AND USER'S STATE PREDICTION

The noncommand interface concept introduced by Nielsen [8] has influenced a lot of research to develop the kind of interfaces that evolved into AUI (Attentive User Interface). But until recently not many realizations have seen light due to the difficulty of the task and to the subtlety of the concept itself. It says that noncommand interfaces are based on an implicit interchange based on a passive monitoring of the user's actions rather than waiting for his explicit commands. The reaction of the computer should be based on distinguishing whether the user thinks he/she is issuing a command or not. Another problem, which might make the interface and the interaction technique very annoying to the user, is when the computer reacts to all or most of the actions. A compromise should be found and a correct interpretation of the user's actions is required.

The final goal of this study is to be able to discover if the user is facing an understanding trouble when reading a text on a navigator window in a different language than his/hers. When this is achieved the system then proposes to help the user by providing a contextual translation of the word that has likely caused the trouble. Experience has shown that one can understand when someone is facing a problem by looking at his face and thus his eyes. We suggest that if the facial expressions are omitted, it can still be possible to figure out that a person if facing an understanding problem just by monitoring his/her gaze patterns. This possibility is shown clearly in the gaze pattern examples here below (the text was extracted from the online edition of The Japan Times newspaper):

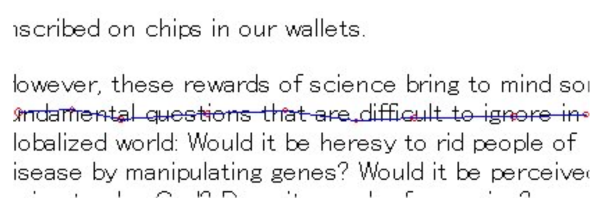

Figure 1: An example of a normal reading pattern

It is known that good English readers do read almost word by word from left to right and line by line, the eyes fixate on each word and sometimes more than once but do skip familiar words (see gaze path in Figure 1). When it comes to unfamiliar words, the reader may fixate the word many times and often regress to grab the meaning from the context (see gaze path in Figure 2).

There exists a widely accepted assumption, which says that attention is highly correlated to the user's gaze direction (fovea), even though this is not always the case. In this paper we are consid-

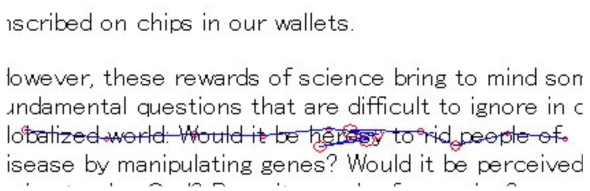

Figure 2: An example of a reading pattern when facing a problem

ering the selective attention which is resulting from the top-down cognitive process in the vision act, as well as to the directing attention or the attention resulting from the bottom-up process in the vision, that attracts a person's sight due to some salient features in the viewed scene. The directing attention is concerned by finding "where" to look at, by analyzing, in parallel, the entire scene using the peripheral vision (parafoveal) which is more sensitive to dim stimulus. Once a salient feature, which attracts the attention, is detected the eyes are directed to inspect that interesting feature more clearly with the higher resolution vision (foveal) and find out "what" is so interesting. Clearly, this combination of "where" and "what" is governing our eye movements. In order to study this relation a large amount of gaze data was recorded with a number of non-natives reading texts online. The analysis of this data revealed less than what was expected but enough to realize the application.

3. EXPERIMENTAL STUDY

3.1 Experiment Setup

An EMR-NC (Eye Mark Recorder-Non Contact) system produced by NAC Corporation and developed at NTT-BRL was used for this experiment. The EMR-NC uses two cameras to detect a triangular frame that the user sticks into the frame of his/her glasses, and then an active mirror tracks the eye position. The gaze can be tracked with a precision of $\pm 0.3^{\circ}$ and a sampling rate of 30[Hz]. Placing the subjects at a distance of approximately 700[mm] from a monitor (of 1024×768[pixels]) then the covered region on the screen is about 12×12[pixels], with the fore mentioned precision. Knowing that text characters usually cover a region of about 8×12[pixels] on the screen the current coverage is a little big, so the characters need to be enlarged for the experiments. But this depends on the utilized font and the user's preferences, which can vary largely. In this experiment the text font was doubled in size and the interline space was set to a larger distance.

3.2 Experiment Conditions

Nine (9) non-native subjects (25∼48 years old) with normal vision participated in this experiment. They had three (3) different levels of English proficiency according to their TOEIC (Test Of English for International Communication) scores (Namely beginner, intermediate, and advanced). They were asked to read three (3) different texts each. The texts were of two (2) different levels of difficulty (normal and difficult). The texts used in this experiment were all extracted from the online issues of The Japan Times newspaper covering several subjects of interest.

The reading order of the texts was done arbitrarily to avoid any biasing to the results and to exclude any expectation from the subjects. They were also asked to indicate when they face an understanding problem. Namely when they do not understand a word in the text they would notice that with a mouse click without considering the pointer's position. The click has been used later as a time stamp around which the gaze data should be analyzed.

Level	Advanced	Intermediate	Beginner
Normal text	1.25	"5.5"	2
Difficult text	4.5	5.2	6.25

Table 1: The average of unknown words for different levels of English proficiency and texts with different levels of difficulty.

3.3 Experiment Results

Table 1 shows how the average of unknown words, as indicated by the users, changes according to the level of proficiency and the difficulty of the texts. As it can be expected, the average is bigger for difficult texts and increases as the level of proficiency decreases, except the average for the intermediate level with normal texts that is inexplainably big.

An extracted portion from the gaze path is shown in Figure 3, where $\mathbf{X}$ represents the horizontal axis of the screen (from left to right), and $\mathbf{Y}$ represents the vertical axis of the screen (from up to down). You can notice that the value of $\mathbf{Y}$ is almost constant, which means that the user was reading on a single line. At the time P where the user has clicked on the mouse indicating an understanding problem, was inserted a zero value as a distinguishing mark of the event (wider highlight).

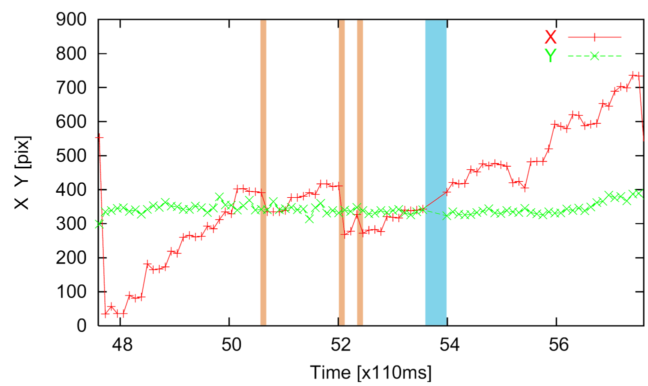

Figure 3: Gaze pattern on a line length with highlighted features.

In the same way, all scans of single lines where the users indicated having problems were extracted and analyzed to distinguish similarities. Some observations could be discerned and are summarized in the following:

- Presence of regressions before P (thinner highlights).
- Variable distances from the regressions to P.
- Average of the number of regressions was about 1.3[units].
- Sometimes no regressions were noticed before P, in 10% of the cases.

From the above we can infer that the presence of regressions is important for indicating when the user is being troubled. This is an aspect in the eye movements that can be used in deciding when to provide assistance to the user. This allows considering the spatial dimension of the gaze information contrary to the approach in [3] that considers only the temporal dimension of the gaze information or in other words the fixation's duration.

Nevertheless, the regression information cannot be used all alone, thus supplementary information is clearly necessary. This can be explained by the fact that eye movements taken without their environment will not be of great help. We need a grounding of the gaze information within the current context and a closer look has to be made on the user's task. In the next section is described a solution to cope with this problem.

4. CONTEXT GROUNDING FOR AN EFFECTIVE SUPPORT

To demonstrate the feasibility of using gaze information, a preliminary implementation was realized of a translator that can be imbedded in a Microsoft Internet Explorer as a dynamic link library[9]. Using the high level interpretation information from the gaze tracker, the application proposes a contextual assistance to the user provided that it can detect when this one is facing a problem or when he/she is likely facing a problem.

The gaze information is used as a context indicator and then that context is explored to find the word that does trouble the user. Naturally, a good candidate is the word with the most of fixations on it, but sometimes the fixations lay between two words or on a wrong position due to some inaccuracies. To solve this problem we have to add a supplementary source of information that involves a preprocessing of the viewed text. When the text is first loaded in the browser window a preprocessing, which associates a difficulty coefficient (rate) to each word in the text, is performed. This coefficient, reflecting the degree of difficulty of the word, can be used later to discern which word (hot spot) is more likely to cause a problem in the case of an ambiguity in the decision. We have chosen to utilize the words frequency of usage calculated from the British National Corpus, which contains more than 100 million words.

5. CONCLUSION AND FUTURE WORK

In this paper, we presented the design idea and the first result of a new proactive gaze-based interaction technique that will bring an additional benefit to non-native users browsing the net. The system monitors the user's actions (gaze patterns) to detect when an understanding problem is faced by the user. Then the proper action is taken by figuring out the word that is causing the trouble and translating it to the user's native language.

An experimental study was conducted to extract the common features that reflect the user's state or help to infer about it. The importance of grounding the user's gaze within its context has been discussed and a difficulty rate has been defined and used as additional information that blends the gaze information for better detection results. The proposed framework can be extended later to handle other languages if we can find a common behavior for the users in difficult understanding situations.

In the future work prospects we are going to evaluate the system's performance by conducting usability and accuracy tests. The final goal is to use the user's natural eyes movements to understand and to detect his intention and his state of mind. This will allow exploiting a new information channel between the user and the computer for better interactive systems.

Acknowledgments

This work is supported in part by the MEXT (Ministry of Education, Culture, Sports, Science and Technology) and in part by the CREST, JST (Japan Science and Technology Agency).

6. REFERENCES

[1] Duchowski, A.T. "Eye Tracking Methodology: Theory and Practice", *Springer-Verlag* , London, 2003.

[2] Jacob, R.K. "Eye Movement-Based Human-Computer Interaction Techniques: Toward Non-Command Interfaces", *In Advances in Human-Computer Interaction*, pp. 150-190, Vol. 4, 1993.

[3] Just, M.A. and Carpenter, P.A. "A Theory of Reading: From Eye Fixations to Comprehension", *In Psychological Review, Vol.87, No.4*, pp. 329-354, July 1980.

[4] Salvucci, D.D. "Inferring Intent in Eye-Based Interfaces: Tracing Eye Movements with Process Models", *In proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'99)*, pp. 254-261, USA, May 1999.

[5] Edwards, G. "A Tool for Creating Eye-Aware Applications that Adapt to Changes in User Behavior", *In proceedings of The third International ACM SIGCAPH Conference on Assistive Technologies (ASSETS'98)*, pp. 67-74, USA, April 1998.

[6] Hyrskykari, A., Majaranta, P. and Räihä, K.-J. "Proactive Response to Eye Movements", *In proceedings of The* 9^{th} *IFIP International Conference on Human-Computer Interaction (INTERACT2003)*, Switzerland, September 2003.

[7] Harris, L.R. and Jenkin, M. "Vision and Attention", *In M.Jenkin and L.Harris(Eds.) Vision and Attention, Springer-Verlag*, pp. 1-17, New York, 2001.

[8] Nielsen, J. "Noncommand User Interfaces", *In Communications of the ACM*, pp. 83-99, 36(4), April 1993.

[9] Khiat, A., Matsumoto, Y. and Ogasawara, T. "Towards Gaze-based Proactive Support for Web Readers", *In proceedings of the* 12^{th} *IEEE Workshop on Robot and Human Communication (RO-MAN2003)*, USA, pp.121-127, October-November 2003.