Towards inferring language expertise using eye tracking

Kai Kunze
Graduate School of Engineering
Osaka Prefecture University
1-1 Gakuen-cho, Naka, Sakai,
Osaka, Japan
firstname.lastname@gmail.com

Hitoshi Kawaichi
Graduate School of Engineering
Osaka Prefecture University
1-1 Gakuen-cho, Naka, Sakai,
Osaka, Japan
kawaichi@m.cs.osakafu-u.ac.jp

Kazuyo Yoshimura
Graduate School of Engineering
Osaka Prefecture University
1-1 Gakuen-cho, Naka, Sakai,
Osaka, Japan
yoshimura@m.cs.osakafu-u.ac.jp

Koichi Kise
Graduate School of Engineering
Osaka Prefecture University
1-1 Gakuen-cho, Naka, Sakai,
Osaka, Japan
kise@cs.osakafu-u.ac.jp

CHI 2013 Extended Abstracts, April 27–May 2, 2013, Paris, France.
ACM 978-1-4503-1952-2/13/04.

Abstract
We present initial work towards recognizing reading activities. This paper describes our efforts detect the English skill level of a user and infer which words are difficult for them to understand. We present an initial study of 5 students and show our findings regarding the skill level assessment. We explain a method to spot difficult words. Eye tracking is a promising technology to examine and assess a user's skill level.

Author Keywords
language expertise, mobile eye tracker, eye movements

ACM Classification Keywords
H.5.2 [User Interfaces]: Input devices and strategies; H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous.

Introduction
As todays technologies and work environments are rapidly changing, continuous learning gains importance. With it, the question how to track the learning progress and assess the skill level becomes more and more critical. Today, a lot of people already monitor their daily physical activities, from jogging, over step counting to sleeping [1,2] to assess

[1]http://www.myzeo.com
[2]http://fitbit.com

and improve their physical life. We want to achieve the same for learning activities, with a special focus on reading. As reading is fundamental to knowledge acquisition, we would be able to improve the quality of our knowledge life.

Computer programs that can accurately assess our skill and expertise level could monitor our progress. Additionally, they could alter the information they provide us to be more custom tailored towards our needs, compensating for potential knowledge gaps. For example, a computer system recognizes that you don't understand a specific word while reading a text and it automatically alters the following sentences to incorporate the missing definition for you without you even realizing it. The research presented in this paper is driven by the question: Can we assess the expertise level and potential comprehension problems of a user during reading activities?

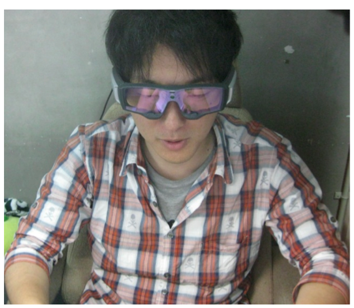

Figure 1: A user wearing the mobile eye tracker system from SMI used in the experiments.

The contributions of this paper can be summarized as follows:

- We present a study design and our initial findings assessing the English expertise of nonnative speakers while reading short texts and answering comprehension questions.
- We present initial findings on how to identify words that are difficult for the reader.

Approach

Expertise can be defined as a learned/trained skill towards a certain task. If we want to monitor the expertise level of a person and recognize potential improvements, it makes sense to look at the knowledge acquisition process. As we acquire a lot of our knowledge through reading, we decided to explore reading skills in greater detail.

We use eye tracking to analyze the reading process. It is not invasive and requires little or no augmentation on the user (depending on the type of eye tracker). Related research has already shown the usefulness of eye-tracking towards monitoring cognitive processes (e.g. distinguishing reading from not reading) [3, 5].

Experiments

Over the course of our reading studies, we want to get as close as possible to real life situations. Therefore, we chose paper documents instead of reading from a special, augmented screen. We also use a mobile eye tracker even for the initial experiments, as we want to explore skill improvements and assessments not only in a stationary setting (see figure 1). The SMI eye tracker used for our studies resembles a large pair of glasses and can be worn comfortably.

Setup

To assess the English skill level of the participants, we used texts and questions from the Cambridge English: Business Preliminary Exam [3]. It is a standard Business English examination for nonnative speakers.

The main reasons for choosing this examination over SAT, GRE, TOEIC or other standardized tests was that the text length, question size and difficulty seemed appropriate for the participants and our scenario. Furthermore, as the test focuses on Business English, we can be reasonably sure that the texts are not known to the participants of the study.

For the experiments, the students sit at a regular desk wearing the eye tracker with the test papers in front of

[3]http://www.cambridgeesol.org/exams/bec-preliminary/index.html

them (see Figure 2). Each student has 20 min to read each text and to answer the comprehension questions.

After each 20 minute experiment, the user is asked to mark words they did not understand (this is no longer recorded with the eye tracker).

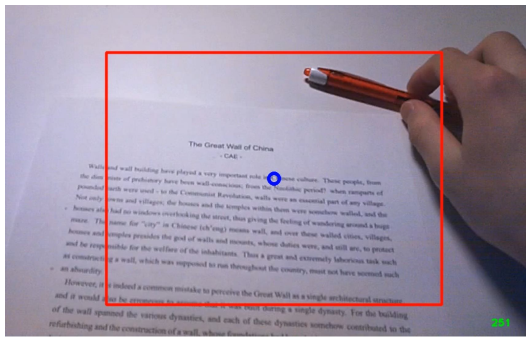

Figure 2: Picture form the experimental recordings. The blue circle shows the current eye focus of the user.

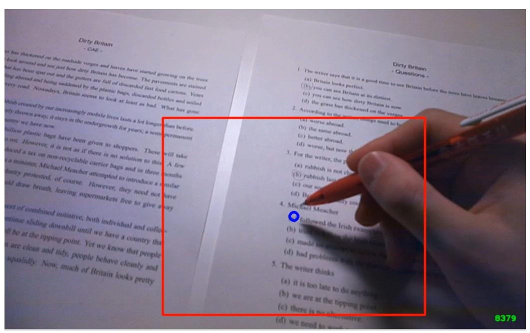

Figure 3: Picture form the experimental recordings. The blue circle shows the current eye focus of the user.

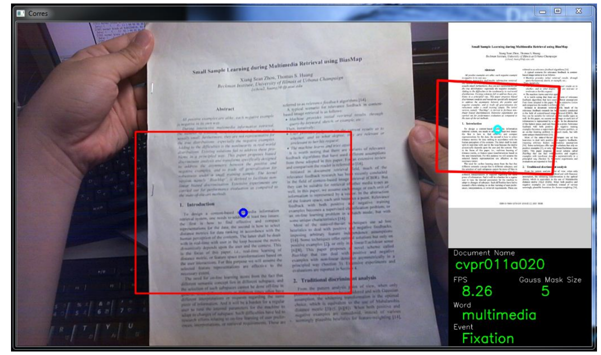

Figure 4: Eyetracking system with LLAH to match the paper document to the digital version.

Conduction

We record the eye-traces (gaze and fixations) at 30 Hz binocular as well as the scene and eye videos using the SMI mobile eye-tracker and the indexed digital document. We use a document image retrieval method based on "Locally Likely Arrangement Hashing" (LLAH) to associate the paper with the digital document [10]. We use the video feed from the eye tracker as input to LLAH. LLAH retrieves the corresponding pages from a document image database by comparing feature points. Combining the retrieved digital document with the eye gaze, we can get the approximate word a participant reads. The LLAH software [4] to identify and retrieve registered digital documents with their paper equivalent is publicly available and can be used by the community for non-commerical purposes.

We also gather basic student background information (age, sex, etc.) as well as answers to how tired they currently are, how they assess their current cognitive

[4] http://imlab.jp/LLAH/

abilities, and what their score range at the TOEIC was. So far, we record tiredness and cognitive load just informally for our personal information, we will use the NASA TLX for later assignments. So far, we recorded 5 students, all native Japanese speakers, reading 5 texts and answering the assigned questions (5-6 questions per text). We assign texts to students according to Graeco-Latin Square. The participants age ranges from 20-25 years, their average TOEIC scores are between 400-800.

Findings

As expected, the correctly answered questions correlated with the TOEIC scores.

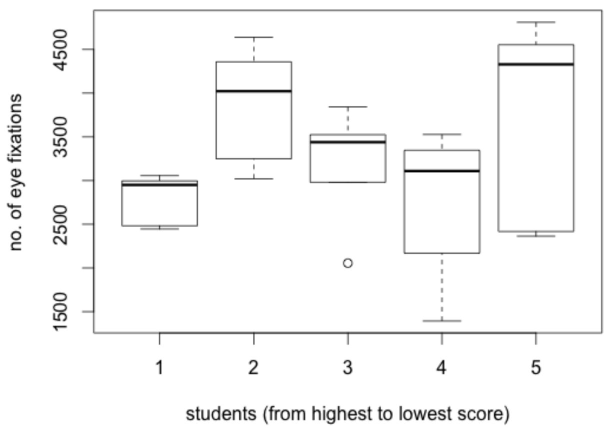

Figure 5: Average number of fixations with standard deviation for each student ordered by language skills according to the TOEIC and the correctly answered questions (starting with 1for the best student).

Let's look at the average eye fixations per document. We had the assumption that the eye fixations for readers with a higher skill would be lower (as they can read faster). Unfortunately, looking at our data this is not completely the case. Figure 5 shows a box and whiskers plot for the number of fixations (mean and standard deviation) for each student ordered by language skills according to TOEIC and the correctly answered questions (starting with 1for the best student). As expected the best student has the lowest mean and standard deviation. Student 3 and 4 have similar scores and it's possible they come from the same distribution (p-value of 0.4). Between the best and the worst student it's likely that the eye fixations come from different distributions(p-value of 0.15).

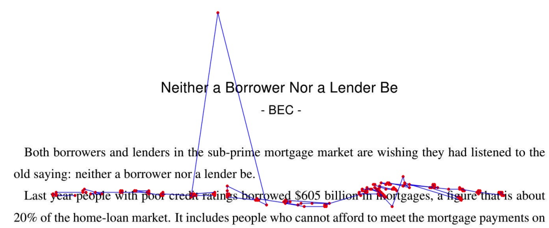

Figure 6: The raw eye trace as recorded by the MSI eye tracker.

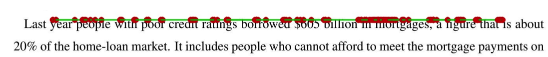

Figure 7: After identifying the line, we apply a horizontal projection of the fixations on the words of the line.

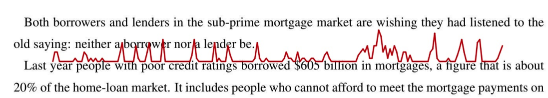

Figure 8: The histogram for eye fixations for one line. "Mortgages", the "difficult" word is easy to identify.

Interestingly, we can identify the difficult words a student marked on the text using the eye fixations. Figure 6 depicts raw eye trace recorded by the mobile eye-tracker overlaid over the digital document using LLAH and a manual line segmentation. Applying a horizontal projection of the read line results in the fixation points and gaze as shown in Figure 7. Now we can calculate the histogram of the eye fixations (seen in Figure 7). The participant had trouble understanding "mortgages" and marked it beforehand. This is easy recognizable on the fixation histogram in Figure 8. This method works for all words marked as difficult by the participants.

Discussion

Our initial findings seem promising, as there are differences between the different skill levels (especially regarding the standard deviation of fixation points). However, as mentioned the difference in eye fixations between the best and worst student are not statistically significant (p-value 0.15). Additionally, the number of eye fixations of student 2 do not fit our assumption. According to the interview after the experiments, Student 2 felt relatively tired and did the recordings after a long day of lectures. This could be a possible explanation for the diverging performance of student 2. The tiredness could affect his reading speed yet had no perceived effect on his level of understanding.

Due to the small sample size it is not possible for us to assess potential side-effects had on the results ((e.g. the time of day and tiredness of the students). For the follow-up experiments we already took these effects into account. Using the number of eye fixations alone, we cannot determine a fine grain level of the user's understanding. However, as seen from the experiments, there seem to be correlations. We have to search for better features to detect the skill level (e.g. looking at reading direction, as we assume people with lower language skill might re-read a passage/sentence multiple times).

The results regarding "Difficult" words are more encouraging. With manual line segmentation, we are able to spot the difficult words per user. The number of these words also correlates with the skill level. As a next step, we are in process of implementing a completely automated approach on recognizing lines and difficult words.

Related Work

There have been quite some studies looking at eye-tracking, cognitive load and reading comprehension. A lot of researchers focus on electroencephalography (EEG) and Functional magnetic resonance imaging (fMRI) [7, 4]. We see this work as complementary to ours. Also most experiments are performed in a lab setting. Mostly stationary eye trackers are used, as otherwise text cannot easily associated with eye gaze (without the use of a document image retrieval system).

Concerning reading, there are a few papers that try to classify (effective) skimming [6, 2]. Eye tracking while reading can also be used to provide summaries of documents or to find which words a user finds relevant [11, 9] An overview about eyer tracking research is given by Jakob et. al. [8].

Conclusions and Future Work

We have shown our initial studies towards assessing the language expertise using mobile eyeytracking. We can detect words that are hard to understand for the participants using the histogram of eye fixations.

We are in the process of recording more experimental runs to have a more representative sample. The next step is to explore the eye fixations more to find a better measure for the skill level of the participants and work on an automated line detection algorithm for the difficult word detection.

If we gain a better understanding about reading tasks, we might be able to alter the text after a difficult passage on a digital device to incorporate the definitions of difficult words (as proposed by Biedert et. al. [1]). Other applications include the long term skill improvement monitoring and automatic preference settings.

We will publish the data set and aggregated background information about the participants, however due to privacy reasons (the small sample size and the TOEIC score range as part of the data), we currently refrain from it.

Acknowledgements

This work was supported in part by the CREST project from Japan Society for the Promotion of Science (JSPS).

References

[1] Biedert, R., Buscher, G., Schwarz, S., Hees, J., and Dengel, A. Text 2.0. In *in Proc. CHI '10 Extended Abstracts* (2010), 4003–4008.

[2] Biedert, R., Hees, J., Dengel, A., and Buscher, G. A robust realtime reading-skimming classifier. In *Proc. of ETRA '12* (2012), 123–130.

[3] Bulling, A., Ward, J. A., Gellersen, H., and Tröster, G. Robust recognition of reading activity in transit using wearable electrooculography. In *Proc. of Pervasive '08*, Pervasive '08 (2008), 19–37.

[4] Clarke, A. R., Barry, R. J., McCarthy, R., and Selikowitz, M. Eeg analysis of children with attention-deficit/hyperactivity disorder and comorbid reading disabilities. *Journal of Learning Disabilities* (2002), 276–285.

[5] Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A., and Kliegl, R. Coregistration of eye movements and eeg in natural reading: analyses and review. *Journal of Experimental Psychology: General 140*, 4 (2011), 552.

[6] Duggan, G. B., and Payne, S. J. Skim reading by satisficing: evidence from eye tracking. In *Proc. of CHI 2011* (2011), 1141–1150.

[7] Ferstl, E. C., Neumann, J., Bogler, C., and von Cramon, D. Y. The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping* (2008), 581–593.

[8] Jacob, R., and Karn, K. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind 2*, 3 (2003), 4.

[9] Loboda, T. D., Brusilovsky, P., and Brunstein, J. Inferring word relevance from eye-movements of readers. In *Proc. of IUI '11* (2011), 175–184.

[10] Nakai, T., Kise, K., and Iwamura, M. Use of affine invariants in locally likely arrangement hashing for camera-based document image retrieval. *In Proc. of DAS 2006 3872* (Feb. 2006), 541–552.

[11] Xu, S., Jiang, H., and Lau, F. C. User-oriented document summarization through vision-based eye-tracking. In *Proc of IUI*, IUI '09 (2009), 7–16.