Towards Extraction of Subjective Reading Incomprehension: Analysis of Eye Gaze Features

Ayano Okoso
Osaka Prefecture Univ.
1-1 Gakuen-cho, Naka, Sakai
Osaka, Japan
okoso@m.cs.osakafu-u.ac.jp

Joachim Folz
DFKI GmbH
Trippstadter Strasse 122
Kaiserslautern, Germany
joachim.folz@dfki.de

Takumi Toyama
DFKI GmbH
Trippstadter Strasse 122
Kaiserslautern, Germany
takumi.toyama@dfki.de

Marcus Liwicki
DFKI GmbH
Trippstadter Strasse 122
Kaiserslautern, Germany
marcus.liwicki@dfki.de

Kai Kunze
Keio Media Design
Keio University
Yokohama, Japan
kai.kunze@gmail.com

Koichi Kise
Osaka Prefecture Univ.
1-1 Gakuen-cho, Naka, Sakai
Osaka, Japan
kise@cs.osakafu-u.ac.jp

CHI 2015, April 18-23, 2015, Seoul, Republic of Korea.

http://dx.doi.org/10.1145/2702613.2732896

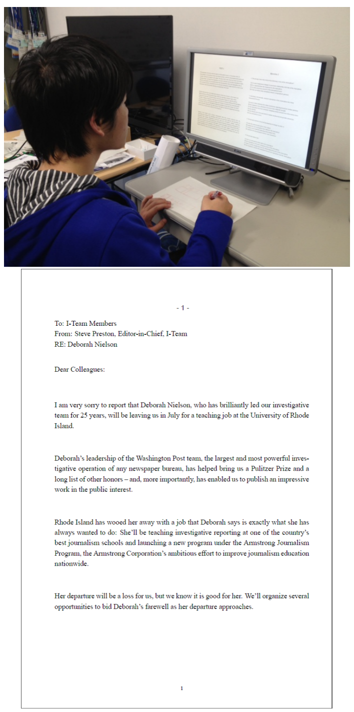
Abstract
One way to optimize learning processes is to clearly inform the learner about problematic areas. Recent work on gaze-based CHI showed that a reader's language skill can be inferred by gaze analysis. However, only few approaches have been proposed to identify those document parts a reader finds problematic. Our goal is to develop a computational method for reading incomprehension extraction. As initial work, we analyze which eye gaze features are useful for such part-based reading incomprehension extraction at three levels of document structure: paragraphs, segments and words.

Author Keywords
Eye tracking, subjective reading comprehension, gaze features, reading analysis

ACM Classification Keywords
H.5.2 [User Interfaces]: Input devices and strategies.

Introduction
The eyes play a central role for information processing in reading: humans effectively control them to collect data to gradually build an understanding of the text. Considerable effort has been made to reveal the nature of human eye movements while reading text [3, 4, 9]. In smooth reading, it is said that eye fixations last about

In classical eye movement research, there are two classes of typical eye movements: *fixations* and *saccades*. Saccades are rapid eye movements that occur when a viewer switches the target of focus. Fixations are the states that the eyes remain relatively still and occur between saccades.



- 1 -

To: I-Team Members
From: Steve Preston, Editor-in-Chief, I-Team
RE: Deborah Nielson

Dear Colleagues:

I am very sorry to report that Deborah Nielson, who has brilliantly led our investigative team for 25 years, will be leaving us in July for a teaching job at the University of Rhode Island.

Deborah's leadership of the Washington Post team, the largest and most powerful investigative operation of any newspaper bureau, has helped bring us a Pulitzer Prize and a long list of other honors – and, more importantly, has enabled us to publish an impressive work in the public interest.

Rhode Island has wooed her away with a job that Deborah says is exactly what she has always wanted to do: She'll be teaching investigative reporting at one of the country's best journalism schools and launching a new program under the Armstrong Journalism Program, the Armstrong Corporation's ambitious effort to improve journalism education nationwide.

Her departure will be a loss for us, but we know it is good for her. We'll organize several opportunities to bid Deborah's farewell as her departure approaches.

1

Figure 1: Top: data recording setting. Bottom: a sample image of the document we used in the analysis.

200-250 ms and the mean saccade size is 7-9 letter spaces [9]. When the reader finds it difficult to understand the text, such smooth eye movements may be disrupted.

Based on prior studies for eye movements during reading, research revealed the power of eye gaze-based approaches for assessing reading comprehension, language skill, and text quality. Biedert et al. presented a method for measuring objective quality of written text [1]. They aggregated gaze features during reading of an entire text area to find out which text passages are comprehensible and which are not. Although the exact degree of comprehension of text is in part subjective, they showed that gaze-based objective measurement is also reliable. Martinez-Gomez and Aizawa showed the potential of eye gaze analysis for subjective understanding level recognition [6]. Their work mainly focused on recognition of a reader's language level, where the comprehension of individual text parts is dismissed. The authors also discussed linguistic features compared to eye gaze ones. However, those linguistic features were not found discriminative in their experiments. Similarly, Kunze et al. presented a method for inferring language expertise from reading behavior [5]. Furthermore, they proposed an approach for spotting difficult words for the reader.

Inspired by the above mentioned prior work, we prototype an eye gaze-based reading comprehension assessment approach. Particularly, we aim to extract subjective part-based reading incomprehension (i.e., detecting parts of a document that are not understood by the reader) on multiple document structure levels: does the reader have difficulty understanding a paragraph, sentence or word. In this paper, we analyze eye gaze features as a preliminary for formulation of such an extraction method.

We record eye gaze data of participants reading English documents. Afterwards they also answer test questions regarding the content. Then we ask the participants to mark which paragraphs, sentences/clauses, and words they find difficult. Finally we extract eye gaze features from the recording and analyze them with respect to difficulty as annotated by the reader.

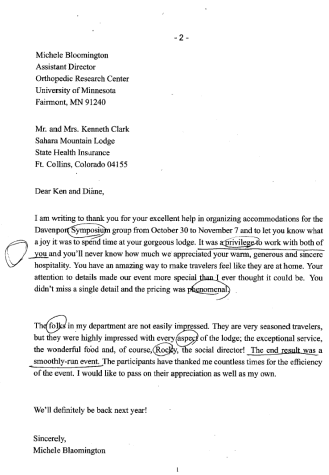
Data Collection

Our data collection procedure is as follows: As shown in Figure 1, we set a stationary eye tracker (SMI RED 250) and a display on the desk.First, the eye tracker is calibrated for the participant. Since our eye tracker does not have a head motion compensation function, we ask the participant to keep his/her head as still as possible.After checking the calibration, we start a recording session. For each recording, the participant reads a single page document presented in full-screen mode. Subsequently, he/she answers several questions regarding the content. In total, the participant completes 10 English language documents. Documents and questions were sourced from the *Test of English for International Communication* (TOEIC). A sample image of a document page is shown in Figure 1.

Every time a recording is completed (document is read and questions are answered), we also ask the participant to mark difficult parts in the document. We asked for three types of annotations for different document structure levels: 1) difficult (e.g. unknown) words, 2) difficult sentences or clauses, and 3) the most difficult paragraph in the document. Note that 3) was chosen instead of labeling individual paragraphs, because some participants had problems to tell whether a specific paragraph was difficult or not. A sample document annotated like this is shown in Figure 2.

Table 1: Extracted eye gaze features at three document structure levels.

Level	Feature	Description
Fixation	fixation duration (f_{fd})	fixation duration (microsec.)
	pupil size (f_{ps})	diameter of the pupils (pixel)
Segment	number of fixations (s_{nf})	number of fixations in a segment
	avg. fixation duration (s_{fd})	average of fixation duration in a segment (microsec.)
	avg. saccade length (s_{sl})	average of saccade length in a segment (pixel)
Paragraph	number of fixations (p_{nf})	number of fixations in a paragraph
	avg. of fixation duration (p_{fd})	average of fixation duration in a paragraph (microsec.)
	avg. saccade length (p_{sl})	average of saccade length in a paragraph (pixel)
	total reading duration (p_{td})	total fixation duration (microsec.)
	avg. number of fixations per word (p_{fw})	—
	avg. reading duration per word (p_{dw})	—
	number of regressions (p_{nr})	number of backward saccades (excluding line break)
	total number of words (p_{nw})	*total number of words in a paragraph*


- 2 -

Michele Bloomington
Assistant Director
Orthopedic Research Center
University of Minnesota
Fairmont, MN 91240

Mr. and Mrs. Kenneth Clark
Sahara Mountain Lodge
State Health Insurance
Ft. Collins, Colorado 04155

Dear Ken and Diane,

I am writing to thank you for your excellent help in organizing accommodations for the Davenport Symposium group from October 30 to November 7 and to let you know what a joy it was to spend time at your gorgeous lodge. It was a privilege to work with both of you and you'll never know how much we appreciated your warm, generous and sincere hospitality. You have an amazing way to make travelers feel like they are at home. Your attention to details made our event more special than I ever thought it could be. You didn't miss a single detail and the pricing was phenomenal.

The folks in my department are not easily impressed. They are very seasoned travelers, but they were highly impressed with every aspect of the lodge; the exceptional service, the wonderful food and, of course, Rocky, the social director! The end result was a smoothly-run event. The participants have thanked me countless times for the efficiency of the event. I would like to pass on their appreciation as well as my own.

We'll definitely be back next year!

Sincerely,
Michele Blaomington

1

Figure 2: Annotations of difficult parts in a document. Difficult words are circled. Difficult sentences (clauses) are underlined. The most difficult paragraph is marked by a circle next to it.

We recorded gaze data from seven participants. However, we only focus one representative reader in this preliminary analysis as work-in-progress.

Feature Extraction

We use the following three gaze structure levels to represent the different document parts: fixations (i.e. words), gaze segments (i.e. lists of words, clauses, sentences), and paragraphs. Gaze segments are used since they are similar to sentences in terms of document structure level, but allow more flexibility in our approach. For each level, we extract gaze features and analyze the distributions of individual feature values with respect to difficult/non-difficult document parts. Table 1 shows a summary of the features we extract in this analysis.

First, we detect fixations and saccades from the recorded gaze data using a dispersion approach [2]. Then, individual features are calculated based on detected fixations and saccades.

Fixation Level Gaze Feature

The smallest unit of gaze feature in this work is a fixation. During reading, a reader continuously fixates on a word to process information. When a problem for word comprehension occurs, a fixation duration may be longer [3]. We also extract pupil size during fixation. We suspect that it corresponds to the reader's stress level, so it may change if he/she is confronted with an unknown word.

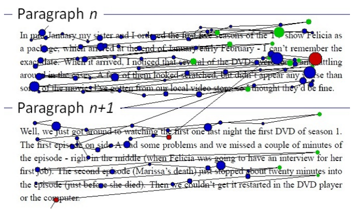


Figure 3: Detected fixations (circles) are not always located exactly on a word. However, we can see which paragraph the fixations belong to.

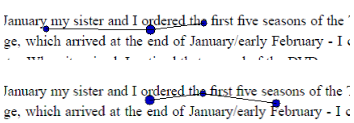


Figure 4: Fixation labeling. Fixations at time t (top) and $t+1$ (bottom). Although the new fixation at $t+1$ is closer to *February* than *seasons*, prior fixations suggest it should be associated with *seasons* by the human labeler.

Segment Level Gaze Feature
On this level, we want to extract reading incomprehension for an arbitrary document part. Thus, instead of considering "sentence" or "clause" as our unit, we extract features based on a *gaze segment*. A gaze segment is a sequence of consecutive fixations within W_t seconds. Based on a preliminary test, we set $W_t = 7.5$ in this analysis.

Paragraph Level Gaze Feature
Recorded raw gaze data can be noisy, because of e.g. the inaccuracy of the eye tracker or the natural eye gaze behavior of the participant. For instance the gaze is not necessarily fixated on the center of a word region [7]. Thus, it can be hard to tell which word a fixation belongs to (see also Figure 3). Previous work showed that gaze data can be robustly associated with a paragraph, since spaces between paragraphs are relatively large in a normal document [8]. Therefore, we consider that a paragraph-level is the smallest unit which can contain document features. The *total number of words* p_{nw} is one of such document feature.

Analysis

Gaze Data Correction
As previously discussed, gaze samples are too noisy to find the exact word the reader fixates on. However, for our analysis, we need to know which gaze samples are associated with which words, segments, or paragraphs as ground truth. If we have such associations between gaze and text, we can analyze the extracted features with regard to the annotations from the participants. We create such association data by manually labeling individual fixations. Using a custom labeling tool, a human labeler monitors each fixation (and its immediate predecessors) and assigns the most probable word the reader is fixating as shown in Figure 4. Although we could also use a filter to automate this manual association process, the recorded data is too noisy to assure the quality of the resulting ground truth.

Consequently, we create fixation data in which each of them is associated with one word. If a fixation is associated with a difficult word, we consider that the fixation is on the difficult word. However, note that just because it is associated with a difficult word (or segment) does not necessarily mean that the extracted feature values differ from those associated with non-difficult words [10]. A reader often ignores or skips a difficult (unknown) word if he/she can understand its context. We discuss this point in the analysis.

If the ith fixation F_i is associated with a difficult word, we label that $F_i = +1$ and otherwise $F_i = -1$. Similarly, if a gaze segment S_i contains a difficult word, clause, or sentence, $S_i = +1$. The most difficult paragraph is also labeled as $P_i = +1$.

Analysis of Variance
For each feature, we calculated the average value and the standard deviation (SD) on both difficult and non-difficult parts. Figure 5 shows all averages and SDs from one representative participant. For visualization, we normalize the values using the maximum of each feature. In this figure, we also show *p-values* of individual one-way ANOVA tests. If the p-value is small (normally, $P < 0.05$), we could infer that the samples are drawn from different distributions, which means that the value is likely be different when the reader has a comprehension problem. For example, the p-value for the pupil size in a fixation-level f_{ps} is 0.71. Thus, we infer that for this feature there is no significant difference between difficult and non-difficult. On the other hand, for the fixation

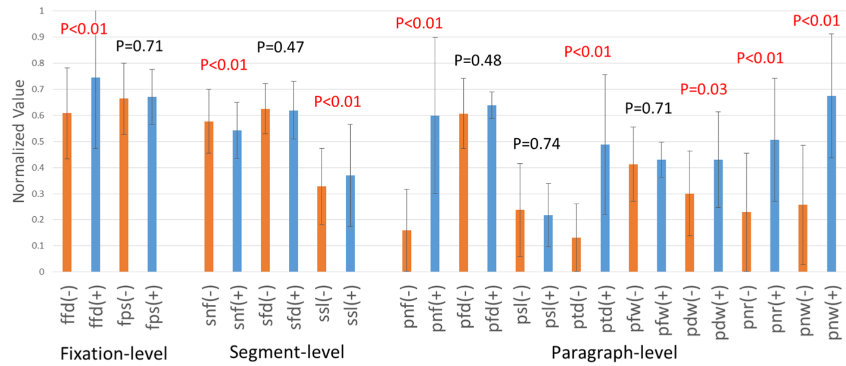


Figure 5: Average and standard deviation (SD) of each feature on both difficult $(+)$ and non-difficult (-) document parts. Error bars represent for SDs. P-values for each feature are calculated by ANOVA.

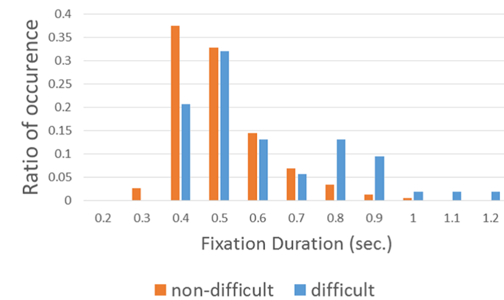


Figure 6: Histogram of ratio of occurrences for fixation duration in f_{fd}. Orange is non-difficult one and blue is difficult one.

duration f_{fd}, $P < 0.01$. It statistically shows that there is a significant difference. The p-value for the average fixation duration of gaze segments is $s_{fd} = 0.47$. Opposite to the fixation-level, there is no significant difference in the average of fixation duration.

On the other hand, the number of fixations s_{nf} and the average saccade length s_{sl} may likely be different between difficult and non-difficult gaze segments. Not surprisingly, the number of words in a paragraph p_{nw} has a low p-value $(P < 0.01)$. This result shows that a reader is likely to find a paragraph difficult if it is very long. The total reading duration of a paragraph p_{td} also has a low p-value, likely because a longer paragraph usually takes longer to read.

However, a short paragraph is sometimes marked as the most difficult by the reader when it has to be re-read. The number of fixations per word p_{fw}, average fixation duration p_{fd} and average saccade length p_{sl} have high p-values. As such, features of individual fixations appear to be less important to assess the difficulty on a paragraph level.

Histogram Analysis

As previously mentioned, a reader sometimes ignores words that he/she finds difficult. Figure 6 shows the histogram distribution of fixations associated with different durations f_{fd}. We can see that longer fixations (longer than 0.7 sec.) are associated with difficult words, whereas the difficulty is not clear with medium-length durations (about 0.4 - 0.7). Therefore, we should consider that the subjective part difficulty is not always inferable from eye gaze only when we evaluate the incomprehension extraction system. It could happen that the reader has comprehension problems even though eye movements are almost normal.

Initial Classification Test
In addition to the analysis, we attempted to recognize difficult parts using a Random Forest (RF) classifier trained on the extracted features as an initial test. However in this initial test, the classification was not successful. We believe that this was caused by insufficient or unfit training data. For future work in this direction we see two options: collect more data or develop an unsupervised approach, which does not require training data.

Conclusion and Outlook
From the analysis, we found several features that can be effective for extracting reading incomprehension. On low level document structures (words or segments), features of fixations (duration and number of fixations) or saccades (avg. of saccade length) showed significant differences between difficult and non-difficult ones. On the other hand, on high level structures (paragraphs), comprehensive features such as total reading duration or number of words in a paragraph are most promising, while the previously mentioned values are less discriminative.

Based on this analysis, we are able to select effective features for each level of document structure. Using our findings, we will develop a method for subjective reading incomprehension extraction using the features.

Acknowledgements
This work is partly supported by Kakenhi Grant-in-Aid B No. 26730095 and the CREST project "Development of Fundamental Technologies for Innovative Use of Character/Document Media".

References
[1] Ralf Biedert, Andreas Dengel, Mostafa Elshamy, and Georg Buscher. Towards robust gaze-based objective quality measures for text. *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA '12*, p. 201, 2012.

[2] Georg Buscher, Andreas Dengel, and Ludger van Elst. Eye movements as implicit relevance feedback. *Proceeding of CHI extended abstracts on Human factors in computing systems - CHI '08*, p. 2991, 2008.

[3] Charles Clifton, Adrian Staub, and Keith Rayner. Eye movements in reading words and sentences. In *Eye Movements*, pp. 341–371. 2007.

[4] M A Just and P A Carpenter. A theory of reading: from eye fixations to comprehension. *Psychological review*, Vol. 87, pp. 329–354, 1980.

[5] Kai Kunze, Hitoshi Kawaichi, Kazuyo Yoshimura, and Koichi Kise. Towards inferring language expertise using eye tracking. *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*, p. 217, 2013.

[6] Pascual Martínez-Gómez and Akiko Aizawa. Recognition of understanding level and language skill using measurements of reading behavior. In *Proceedings of the 19th international conference on Intelligent User Interfaces - IUI '14*, pp. 95–104, New York, New York, USA, 2014. ACM Press.

[7] G.W. McConkie, P.W. Kerr, M.D. Reddix, and D. Zola. Eye movement control during reading: I. The location of initial eye fixations on words. *Vision Research*, Vol. 28, No. 10, pp. 1107–1118, January 1988.

[8] Ayano Okoso, Kai Kunze, and Koichi Kise. Implicit gaze based annotations to support second language learning. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct Publication - UbiComp '14 Adjunct*, pp. 143–146, 2014.

[9] Keith Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, Vol. 124, No. 3, pp. 372–422, November 1998.

[10] W Schroyens, F Vitu, M Brysbaert, and G D'Ydewalle. Eye movement control during reading: foveal load and parafoveal processing. *The Quarterly journal of experimental psychology. A, Human experimental psychology*, Vol. 52, pp. 1021–1046, 1999.