The Eye as the Window of the Language Ability: Estimation of English Skills by Analyzing Eye Movement While Reading Documents

Kazuyo Yoshimura, Koichi Kise
Graduate School of Engineering,
Osaka Prefecture University, Japan
yoshimura@m.cs.osakafu-u.ac.jp, kise@cs.osakafu-u.ac.jp

Kai Kunze
Graduate School of Media Design
Keio University, Japan
kai.kunze@gmail.com

***Abstract*—Reading-life log is a research field of analyzing our activities of reading documents to know more about readers and documents. In this paper we propose an implementation of reading-life log which is to estimate the English language skill by analyzing the activities of reading English documents. As input for the analysis, we employ eye movement information, because we consider the eye movement of skillful readers is far different from that of novices. From the experiments, we have found that the following two features are informative: (1) the sum of fixation duration, and (2) the sum of the velocity of saccades. By using these features the proposed method is to estimate the class of English skill from among low, middle and high, which are defined based on the scores of English standardized test called TOEIC. From the experimental results with 11 subjects and 10 documents, we have been successful to estimate the class with the accuracy of 90.9%.**

I. Introduction

People acquire a lot of information daily through reading. Thus by recording and analyzing reading activities, we are able to obtain various types of information about the reader including his / her interests and preferences. In addition, once the reading activities of many readers are recorded in relation to the contents of documents, we are able to consider the analysis of documents based on reading activities such as finding interesting and difficult parts for many readers. In other words we can open up a new research field of document analysis through the mutual analysis of reading behavior and document contents.

Our research on reading-life log has started with the motivation mentioned above. In our research the main source of information about the reading activities is the eye movement, because we think it contains rich information about reading activities. By analyzing eye movement, we can obtain various types of information about his/her reading. The simplest is the amount of reading: how many words are read by a reader in a specific time frame such as a day [17]. In addition, the analysis of eye movement allows us to segment reading activities from others, to estimate a type of documents the reader is working on such as novels, magazines, newspapers and mangas [15]. The contents of read documents can also be recorded as a log of reading [13].

In this paper we are concerned with a higher level of reading-life log as compared to the previous approaches. As such a high-level log we have selected the estimation of a level of language ability by analyzing reading activities.

A simplest and traditional method for obtaining the language abilities is based on paper tests. For the case of English, for example, standardized tests such as TOEFL and TOEIC are well accepted for this purpose. However, such tests require time to take. Another problem would be that the test is available not so often, say once a month, and thus it is not possible to have a “real-time” monitoring of the ability.

Our research attempts to build a method for easy and real-time monitoring of the language ability with a special focus on English. The proposed method analyzes eye movement obtained through an eye tracker to learn/classify the English ability of readers. To be precise features about the fixations and saccades of eye movement while reading documents are employed for the analysis. The main contribution of this paper is that, though the experiments, we have found that the two features, the sum of duration of fixations and the sum of angular velocity of an eye, enable us to estimate the class (low, middle, high) of English ability with a high accuracy of more than 90%. All of the work described here was done under the permission of the research ethics committee of the graduate school of engineering, Osaka Prefecture University.

II. Related Work

The strong relationship between reading and eye movements is well explored in cognitive science and psychology [21], [11]. For example, Kligel et al. investigate correlations of eye fixations with cognitive tasks related to reading [14]. Most of the reading research in psychology however emphasizes on older adults or disabled [8], [7].There are only a few research publications centering around reading detection in mobile and stationary settings [5], [4]. As such detecting reading can be used as a very simple word counting mechanism, as there’s a relation between time read and the read volume. Biedert et al. look into how people read text. They presented a method to discriminate skimming from reading using a novel set of eye movement features [3]. Their algorithm works in real-time, deals with distorted eye tracking data and provides robust classification with 86% accuracy. They also showed a method to recognize text comprehensibility with an accuracy of 62% from gaze data recorded from multiple readers [2].

In a series of works, Biedert et al. studied ways to enhance the reading experience of the user. They presented Text 2.0 [1] as a reading interface that observes which part of the text is currently being read by the user and that generate appropriate effects (e.g. playing sounds). However, they do not evaluate what suitable interventions are to increase users enjoyment, comprehension or attention. Xu et al. apply eye movement analysis for document summaries, yet the enviroment is very controlled, e.g. the users need to rest their chin on a support when performing the reading task[22].

Concerning reading habits, there are some questionnaires based evaluations giving advice about effective reading techniques to second language learners, as well as for children with reading disabilities [9], [19].Hansen [10] reports on a series of studies on reading comprehension with rapid readers trained in the Evelyn Wood method. Several mention rigorous practice and steady increase in reading volume as one of the key factors to success [12].

There are also some efforts to infer the users expertise, language skill and other higher level cognitive activities using eye tracking [16], [18], [6].

III. Proposed method

A. Basic concepts

It is often said that the eye is the window of the mind. We consider that this holds for our specific purpose of estimation of English ability. Eye movement of persons with a high skill of English should be different from that of low skill persons. Thus we consider in this paper to establish a method of estimating the English skill by analyzing eye movement.

As measurement of English skill, we employ a standardized test called TOEIC. TOEIC is one of the well-known English tests, which consists of two parts: hearing and reading. The score ranges from 10 to 990 as a result of the process of standardization of the raw scores. Thus the score is considered as a general measure of the English ability of a person.

One may ask the reason why we need to "estimate" the TOEIC score, because the easiest way is to take the paper test and receive the score. The answer is quite simple. Taking the test requires long time (more than two hours). Thus it is not possible for us to employ the test for frequent monitoring the English ability of a person. In other words, if we are able to estimate the score by the analysis of eye movement, it is possible for us to obtain the information about the English skill in real time.

Our ultimate goal is to estimate the score itself by calculating the function from features of eye movement to the score. However this is not so easy due mainly to the lack of the number of samples for learning the function. Thus in this paper we attempt to learn a classifier that takes as input features of eye movement to estimate the subject's class as a three class classification problem: low (less than 600), middle (between 600 and 800) and high (higher than 800).

B. Outline of processing

Figure 1 illustrates processing steps of the proposed method. First, eye movement data are obtained. For this purpose, we utilize a stationary eye tracker when the subject reads a document on the display. Then these raw data are analyzed to segment eye activities into three states: blink, fixation and saccade, where the fixation and the saccade indicate a short stop of the eye and a rapid movement between short stops, respectively. Human reads text by repeating the fixation and saccade. At the third step, various features are obtained from these two states. At the last step, extracted features are employed for learning or classifying reader's English ability.

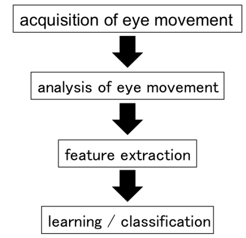

Fig. 1. Overview of the processing.

In the following, we explain the details of steps 2–4.

C. Analysis of eye movement

The raw data of eye movement is a sequence of eye gazes represented as the 2D coordinates in the coordinate system of the display. The first task is to analyze the data to convert them to a sequence of states: blink, fixation and saccade. The eye gazes located close with each other are combined into one fixation with the coordinates as the average of its members. Once the fixations are recognized, saccades are identified as the eye movement that connects two consecutive fixations. Blinks are recognized based, for example, on irregular eye movements. In our system, the above states are identified by the software called BeGaze, which is a product of SMI.

D. Feature extraction

Features are extracted from the sequence of states. Table I shows the list of features we employed in our method. These features are employed to select the best combination to classify the subject's English ability at the next step.

E. Learning and classification

This step consists of two phases: learning of the classifier and its application. As a classifier, we employ the support vector machine (SVM).

At the phase of learning, we first apply feature selection from a pool of features in Table I. To be more precise, we apply the backward stepwise selection. We start from learning with all the features. Then the feature with the lowest correlation to the TOEIC score is removed if learning without it gives us a better or the same result. Otherwise the feature is kept and the feature with the next lowest correlation is tested. This process continues until no more features are removed. Note that we employ multiple documents read by a single subject to obtain enough number of blinks, saccades and fixations to calculate features.

At the phase of classification, we apply the learned SVM to the features in order to estimate the class of subject's English ability. Note again that the features are extracted from the eye movement for multiple documents for a better estimation.

TABLE I. FEATURES.

feature		explanation
End_Time[ms]		elapsed time of reading
Blink_Count		the number of blinks
Blink_Frequency[count/s]		the number of blinks per second
Blink_Duration [ms]	_Total	sum, average, maximum and minimum of the time of blinks
	_Average	
	_Maximum	
	_Minimum	
Fixation_Count		the number of fixations
Fixation_Frequency[count/s]		the number of fixations per second
Fixation_Duration [ms]	_Total	sum, average, maximum and minimum of duration of fixations
	_Average	
	_Maximum	
	_Minimum	
Fixation_Dispersion [px]	_Total	sum, average, maximum and minimum dispersion of fixations
	_Average	
	_Maximum	
	_Minimum	
Scanpath_Length[px]		sum of the distance of saccades
Saccade_Count		the number of saccades
Saccade_Frequency[count/s]		the number of saccades per second
Saccade_Duration [ms]	_Total	sum, average, maximum and minimum of duration of saccades
	_Average	
	_Maximum	
	_Minimum	
Saccade_Amplitude [°]	_Total	sum, average, maximum and minimum of the rotation angle of an eye ball during saccades
	_Average	
	_Maximum	
	_Minimum	
Saccade_Velocity [°/s]	_Total	sum, average, maximum and minimum of the angular velocity of an eye ball during saccades
	_Average	
	_Maximum	
	_Minimum	

TABLE II. CLASSES OF TOEIC SCORES.

class	toeic score	# subjects
1 (low)	400 - 600	3
2 (middle)	600 - 800	6
3 (high)	800 - 990	2

IV. EXPERIMENTS

A. Conditions and goals

We employed the dataset of eye movement obtained from 11 subjects who are students of graduate schools. All of them are Japanese and thus non-native of English. The distribution of TOEIC scores is shown in Table II.

As the eye tracker, SMI RED250 was used. This is a stationary eye tracker with a sampling rate of 250 Hz, which is high enough to catch eye movement while reading.

In the experiments, we asked the subjects to perform the task of reading parts (PART7) of TOEIC tests. Each part consists of a document with several paragraphs and four questions about it. Each question is to select from among four choices.

Figure 2 shows the procedure of experiments. First, the calibration of the eye tracker was done every time just before measuring eye movement. Next, a subject was asked to start reading the text shown on the display. At the time of reading, the subject was not allowed to read the questions. After reading, the text disappears from the display and the subject was asked to answer to the questions. Finally the subject also answered to some questionnaires. This process was repeated until first five sheets of text are read. After 15 min. break, the subject repeated the process for the remaining five sheets.

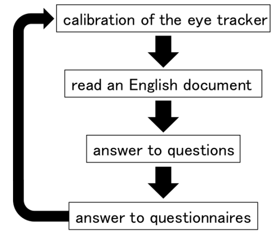

Fig. 2. Procedure of experiments.

TABLE III. RESULTS OF ESTIMATION.

subject	A	B	C	D	E	F	G	H	I	J	K	ave.
estimated	2	1	1	3	2	2	2	3	2	1	1	
correct	2	1	1	3	2	2	2	3	2	1	2	
error	0	0	0	0	0	0	0	0	0	0	1	0.09

Using the above data of eye movement the following experiments were done: The first is to estimate the English abilities of subjects using all the data. At the learning phase, leave-one-subject-out was applied to make the learning result independent from subjects. The second experiment is to evaluate the dependency on documents. Different documents were employed to verify whether the method is influenced by documents. The third and the fourth are to know the relationships with the number of documents and classes, respectively.

B. Estimation of the English ability

The first task here was the feature selection. Thanks to the backward stepwise selection, we selected the following two features: Fixation_Duration_Total and Saccade_Velocity_Total. The former is the sum of the duration of fixation and the latter is the sum of angular velocity of an eye ball. Note that these sums are not only for a single document but for all documents employed in the learning.

Table III shows the results of test. Estimated and correct mean the estimated and correct classes, and the error is the difference between them. The average indicates the average of all errors. As shown in the table, only one subject K was with the error. The accuracy of estimation is 90.9% ($= 10/11$).

Figure 3 shows the distribution of features. One outlier with a high value of Saccade_Velocity_Total corresponds to the subject K. Figures 4 and 5 show the distributions of Fixation_Duration_Total and Saccade_Velocity_Total, respectively.

The distribution in Fig. 4 shows that the class 1 (low) can be distinguished by the feature Fixation_Duration_Total. This is understandable since it has been known that the feature Fixation_Duration_Total is relevant to the difficulty of a document for a subject [20]. However it is also shown that the feature is not effective to separate the class 2 (middle) from the class 3 (high).

As shown in Fig. 5 the feature Saccade_Velocity_Total plays this role. For the class 2 subjects, their saccade velocity is smaller than that of the class 1 and class 3, because it is necessary for them to read carefully to understand the contents. For the class 1 subjects, they read faster simply because of a lot of local re-reading; this could be due to the difference of grammatical structure between English and Japanese. Low skill subjects are not able to understand English in its word order

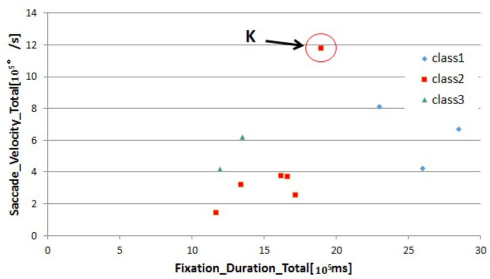

Fig. 3. Distribution of features.

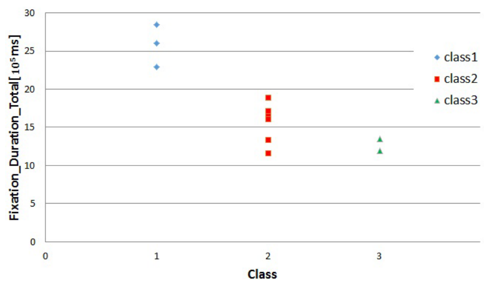

Fig. 4. Distribution of Fixation_Duration_Total.

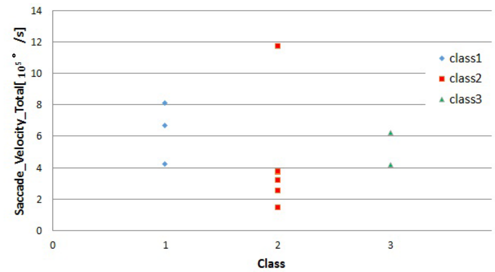

Fig. 5. Distribution of Saccade_Velocity_Total.

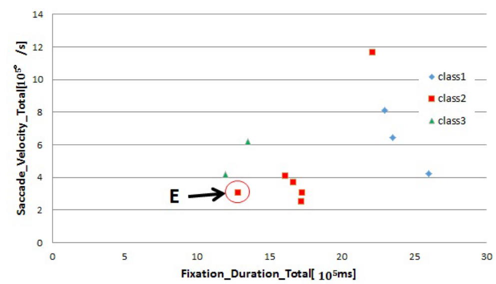

Fig. 6. Feature distribution with new documents.

and thus must go back and forth, which makes the velocity higher. The reason of high velocity for the class 3 subjects is that they require shorter time to understand the contents as compared to the class 2 subjects.

C. Dependency on documents

The above result should be independent of the documents we employed for the experiment. In order to verify this point, we changed all of the ten documents for five subjects. The new documents were as difficult as the originals from the viewpoint of subjective difficulties.

The result is shown in Table IV. Only one additional error, which was with the new documents, was observed for the subject E. From this result, we have confirmed that the estimation is well independent of documents. Figure 6 illustrates the distribution of features for this experiment. The new error was caused by the position of the subject E closer to a class 3 subject. However the distribution is similar to that of the previous experiment in Fig 3, from which we can also confirm the independence.

D. The number of documents

The next is about the number of documents necessary for the estimation. It is better to employ a smaller number of documents if the same accuracy of estimation is obtained. Generally speaking, the accuracy can drop as the number of documents is reduced. Thus in this experiment we clarify the relationship between the number and the accuracy. In the experiment, reduced sets of documents were prepared by a random sampling without replacement for 30 times and the accuracy was obtained as their average.

The result is shown in Fig. 7, where the horizontal axis represents the number of documents employed for learning and the vertical axis is the accuracy of classification. It is observed that the accuracy improves as the number of documents increases. This means that if the user of this method is interested in accuracy he/she should use documents as many as possible. For those who would like to minimize the efforts at the subject side, at least two documents should be used. The use of a single document cannot be recommended.

E. The number of classes

The final question here is to verify the setting of classes. The accuracy of classification could be different depending on the classes. We defined the different classes as shown in Table V. In addition to the original setting of three classes, we employed four and five classes. Table VI shows the accuracy obtained by using different numbers of classes. As shown in this figure, the accuracy drops when the number of classes increase from three. We consider this is mainly because of the lack of the number of subjects in the same class.

F. Discussions

From the above experimental results, we have confirmed that the proposed method is capable of estimating the subject's English ability from among the basic three classes: low, middle and high. It has also been shown that the method is user- and document-independent. If the number of documents for evaluation increases, we can obtain a more accurate estimation. Note that even with ten documents, the time required for the estimation is much less than taking the full TOEIC test, which requires more than two hours.

On the other hand, we consider that the experimental results have posed some limitations. An important one is about the ”resolution” of classes. The accuracy of estimation dropped when we increased the number of classes. However this should not be considered as the limitation of the proposed method but caused by a small number of learning samples.

TABLE IV. DEPENDENCY ON DOCUMENTS. IN THE "DOCUMENT SET" ROW, "O" AND "**n**" INDICATES THE ORIGINAL AND THE NEW DOCUMENTS, RESPECTIVELY.

subject	A	B	C	D	E	F	G	H	I	J	K	ave.
document set	o	o	o	o	**n**	o	**n**	o	**n**	**n**	**n**	
estimated	2	1	1	3	3	2	2	3	2	1	1	
correct	2	1	1	3	2	2	2	3	2	1	2	
error	0	0	0	0	1	0	0	0	0	0	1	0.18

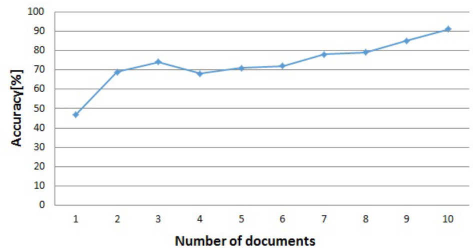

Fig. 7. The number of documents and accuracy.

V. CONCLUSION AND FUTURE WORK

It is often said that the eye is a window of the mind. We claim that this holds for the language ability. The subject's language ability can be estimated by analyzing his/her eye movement. Based on this concept we have proposed a method of estimating a class (low, middle, high) of English ability defined based on the TOEIC score. The key contribution of this paper is that in addition to the known feature of the sum of fixation duration, the sum of saccade velocity plays an important role to distinguish the classes. From the experimental results with 11 subjects and 10 documents, we have confirmed that the proposed method is capable of estimating the class with the accuracy of 90.9% in a user- and document-independent way.

The future work includes experiments with a larger number of subjects and documents, as well as building attractive services based upon the functionality of the estimation of language abilities.

ACKNOWLEDGMENT

This work was supported in part by JST CREST and JSPS KAKENHI (25240028).

REFERENCES

[1] Ralf Biedert, Georg Buscher, Sven Schwarz, Jörn Hees, and Andreas Dengel. Text 2.0. In *Ext. Abs. CHI 2010*, pages 4003–4008, 2010.

[2] Ralf Biedert, Andreas Dengel, Mostafa Elshamy, and Georg Buscher. Towards robust gaze-based objective quality measures for text. In *Proc. ETRA 2012*, pages 201–204, 2012.

[3] Ralf Biedert, Jörn Hees, Andreas Dengel, and Georg Buscher. A robust realtime reading-skimming classifier. In *Proc. ETRA 2012*, pages 123–130, 2012.

[4] Andreas Bulling, Jamie A. Ward, and Hans Gellersen. Multimodal Recognition of Reading Activity in Transit Using Body-Worn Sensors. *ACM Trans. on Applied Perception*, 9(1):2:1–2:21, 2012.

[5] Andreas Bulling, Jamie A. Ward, Hans Gellersen, and Gerhard Tröster. Eye Movement Analysis for Activity Recognition Using Electrooculography. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 33(4):741–753, April 2011.

TABLE V. TOEIC SCORES AND CLASSES.

subjects	C	J	B	E	K	F	A	G	I	D	H
TOEIC score	465	480	550	605	665	685	730	750	775	930	945
3 classes	1	1	1	2	2	2	2	2	2	3	3
4 classes	1	1	1	2	2	2	3	3	3	4	4
5 classes	1	1	2	2	3	3	4	4	4	5	5

TABLE VI. CLASSES OF ENGLISH ABILITIES AND ACCURACY.

# classes	3	4	5
accuracy [%]	90.9	54.5	45.5
ave. error	0.09	0.45	0.91

[6] Andreas Bulling and Thorsten O Zander. Cognition-aware computing. *Pervasive Computing, IEEE*, 13(3):80–83, 2014.

[7] Adam R. Clarke, Robert J. Barry, Rory McCarthy, and Mark Selikowitz. Eeg analysis of children with attention-deficit/hyperactivity disorder and comorbid reading disabilities. *Journal of Learning Disabilities*, pages 276–285, 2002.

[8] Evelyn C. Ferstl, Jane Neumann, Carsten Bogler, and D. Yves von Cramon. The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping*, pages 581–593, 2008.

[9] Fred Genesee, Kathryn Lindholm-Leary, William Saunders, and Donna Christian. English language learners in us schools: An overview of research findings. *Journal of Education for Students Placed at Risk*, 10(4):363–385, 2005.

[10] Dorothy Morrison Hansen. A discourse structure analysis of the comprehension of rapid readers. 1975.

[11] David E Irwin. Fixation location and fixation duration as indices of cognitive processing. *The interface of language, vision, and action: Eye movements and the visual world*, pages 105–134, 2004.

[12] Marcel Adam Just and Patricia Ann Carpenter. *The psychology of reading and language comprehension.* Allyn & Bacon, 1987.

[13] Takashi Kimura, Rong Huang, Seiichi Uchida, Masakazu Iwamura, Shinichiro Omachi, and Koichi Kise. The reading-life log — technologies to recognize texts that we read. In *Proceedings of ICDAR2013*, pages 91–95. IEEE, 2013.

[14] Reinhold Kliegl, Antje Nuthmann, and Ralf Engbert. Tracking the mind during reading: the influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General; Journal of Experimental Psychology: General*, 135(1):12, 2006.

[15] K. Kunze, A. Bulling, Y. Utsumi, S. Yuki, and K. Kise. I know what you are reading – recognition of document types using mobile eye tracking. In *ISWC 2013*, 2013.

[16] K. Kunze, H. Kawaichi, K. Yoshimura, and K. Kise. Towards inferring language expertise using eye tracking. In *Ext. Abs. CHI 2013*, pages 4015–4021, 2013.

[17] K. Kunze, H. Kawaichi, K. Yoshimura, and K. Kise. The wordmeter – estimating the number of words read using document image retrieval and mobile eye tracking. In *Proc. of ICDAR 2013*, pages 25–29. IEEE, 2013.

[18] 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*, pages 95–104. ACM, 2014.

[19] Carol A Rashotte and Joseph K Torgesen. Repeated reading and reading fluency in learning disabled children. *Reading Research Quarterly*, pages 180–188, 1985.

[20] Keith Rayner. Eye movements and attention in reading, scene perception, and visual search. *The quarterly journal of experimental psycology*, 62(8):1457–1506, 2009.

[21] George D Spache. Is this a breakthrough in reading? *The Reading Teacher*, 15(4):258–263, 1962.

[22] Songhua Xu, Hao Jiang, and Francis C.M. Lau. User-oriented document summarization through vision-based eye-tracking. In *Proc of IUI*, IUI '09, pages 7–16, 2009.