Estimation of English skill with a mobile eye tracker

Olivier Augereau
Osaka Prefecture University
1-1 Gakuen-cho, Naka, Sakai,
Osaka , JAPAN
augereau.o@gmail.com

Hiroki Fujiyoshi
Osaka Prefecture University
1-1 Gakuen-cho, Naka, Sakai,
Osaka , JAPAN
fujiyoshi@m.cs.osakafu-u.ac.jp

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

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

UbiComp/ISWC '16 Adjunct, September 12-16, 2016, Heidelberg, Germany
ACM 978-1-4503-4462-3/16/09.
http://dx.doi.org/10.1145/2968219.2968275

Abstract
Learning a foreign language such as English is an important task for many people. The process of learning takes time and it is important to have a simple way to evaluate the progress of the skill. We propose a method to evaluate the reader's English skill based on a mobile eye tracking system. The eye tracker captures the reader's behavior while reading a document. The front camera of the eye tracker records the scene image that contains the read document. By using a retrieval algorithm we can recognize the read document and project the eye gaze data from the scene image to the document space. Then, some features related to the reading and solving behavior on several documents are computed. As a first result, we show that the TOEIC score can be estimated with an error of 36.3 points.

Author Keywords
eye tracking; language estimation; LLAH.

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

Introduction
Due to globalization, it is more and more important for non-native English speakers to study English language. During the learning process, it is necessary for the learners to know their current skill by taking English tests. One of

Figure 1: The mobile eye tracker can be worn as glasses. One camera records the scene and two others record the eyes.

the most famous English test in some Asian and European countries is the TOEIC (Test of English for International Communication) [2]. For example, in 2014, around 2.4 millions Japanese people took this test to estimate their language skill [1]

Unfortunately, taking the TOEIC has several constraints. The participants must go to a specific place at a specific time to take the exam, pay the fee and wait around 1 month after taking the test to get the results.

This motivates us to propose a new method for estimating the English skill by simply wearing a mobile eye tracker. The eye tracker records the eye movements which reflect the reader's skill. We compute some features based on the eye movements and combine them with the number of good answers given by the subject to estimate the TOEIC score. Figure 1 is an illustration of a mobile eye tracker.

In our experiment, 9 Japanese students read 10 documents and solve the associated questions. The first result shows that, on average, the TOEIC score can be estimated by our method with an error of 36.3 points, which is 3.7 % of the whole range of the possible scores (from 10 to 990 points).

Related Work

In 2013, Kunze et al. [4] proposed a method for estimating the English skill of a reader. The authors show that the average number of fixations and the standard deviation of the number of fixations is different depending on the reader's skill.

In 2014, Martinez-Gomez et al. proposed another method to analyze the language skill (TOEIC or TOEFL score) of a reader [6]. Unfortunately, the performance obtained by the authors is around 82% of error for the understanding prediction.

In the same year, Copeland et al. proposed to estimate the reading comprehension based on eye gaze movements by using an artificial neural network [3]. The misclassification rate is quite low, but the dataset is heavily unbalanced (most answers are good) and the result are equal or under the baseline classifier.

In 2015, Yoshimura et al. proposed a method to classify the TOEIC score in three classes: low, middle and high [9]. By computing 33 features related to blinks, saccades, fixations, etc. the authors succeeded to classify the English skill of 11 readers in 3 classes with an accuracy of 90.9%. But if the number of classes if augmented from 3 to 4, the classification accuracy drop to 54.5%. So the skill can be estimated only roughly.

The main difference of our proposal is the use of a mobile eye tracker. Furthermore, our analysis of the English skill is based on the reading behavior when the subject read a text and solve the related questions which allow us to estimate the TOEIC score more accurately.

Proposed method

Our proposed method consists of the 5 following consecutive steps:

1. Recording the eye gaze
2. Converting the eye gaze to the document coordinates
3. Computing fixations and saccades
4. Extracting the features
5. Estimating the TOEIC score

[1] http://www.toeic.or.jp/library/toeic_data/toeic_en/pdf/data/TOEIC_Program_DAA.pdf

Fixations
Average duration
Total number
Sum of duration
Proportion of duration
Saccades
Average length
Average speed
Average time
Sum of time
Sum of the speed
Sum of the distance
Inverse of the average time

Table 1: 11 features computed for the 3 areas (main text, questions, and the overall document). The “proportion of duration” represents the ratio of the sum of duration divided by the total time spend. To these 33 features we add 2 other features: the number of transitions between the document and the questions, and the number of good answers. In total, 35 features are used.

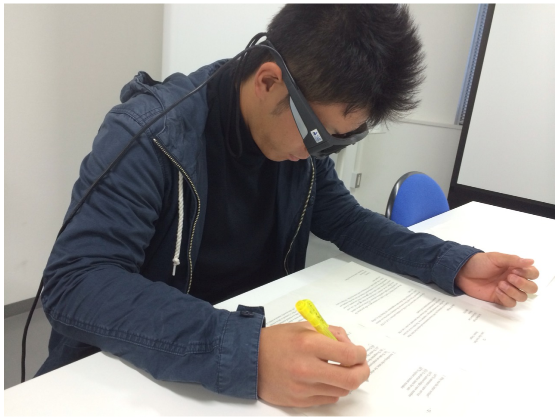

Figure 2: The subject is wearing a mobile eye tracker while reading a document.

Recording the eye gaze

The mobile eye tracker contains 3 cameras. Two are located inside the glasses and record each eye. The last one is a front camera recording the scene image. The eye gaze positions are computed using the data from the two eye cameras, and the position is displayed on the scene image captured by the front camera. Figure 2 represents a subject wearing a mobile eye tracker while reading a document.

Converting the eye gaze to the document coordinates

By using a mobile eye tracker, the subject can read freely a paper document. The position of the eye gaze is given for the scene image but the document in the scene image can have a perspective distortion. Before analyzing the eye gaze, we need to compute the geometrical transformation of the document in order to project the eye gaze position from the scene to the document coordinates. Figure 3 illustrates this principle.

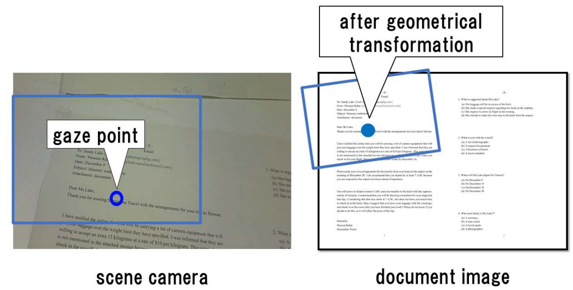

Figure 3: The eye gaze is recorded in the scene image. We apply LLAH to retrieve the document image and compute the geometrical transformation.

In order to solve this problem, we use the LLAH (Locally Likely Arrangement Hashing) algorithm [8]. It is a document image retrieval algorithm. An example of each document image to retrieve is stored in a database. Then, if one of these documents appears on the camera, the document will be retrieved and the geometrical transformation computed. After this step, the eye gaze positions are converted in the document coordinates.

Computing fixations and saccades

The eye gaze movement of the reading behavior is composed of fixations (the position where the reader stop briefly to acquire information) and saccades (the jumps between the fixations) [7]. The Buscher algorithm [1] is used to compute the fixations and saccades from the raw eye gaze data. The characteristics of fixations and saccades are different depending on the reader’s skill.

Extracting the features

As displayed in the Fig. 4 the document is divided in three areas corresponding to the right page, the left page and the

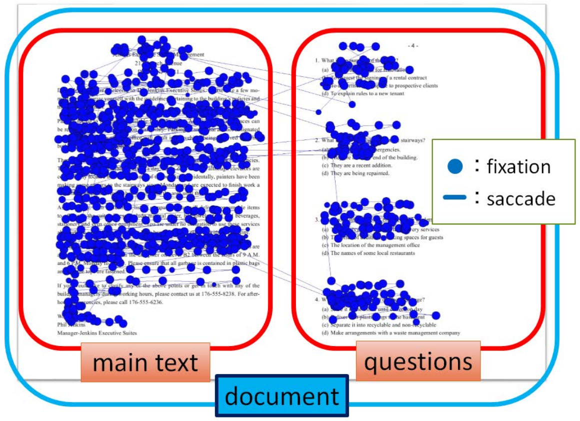

Figure 4: The document is divided into 3 areas: the main text area, the question area and the overall document area. Some features are computed for each area.

combination of both areas. The left page contains the main text and the right page contains the questions. We extract four features about fixations and seven features about saccades (see table 1 for details). These 11 features are computed for the three areas of the document. The document contains fixations and saccades from the main text plus the ones from the questions, plus the saccades going from one to the other. Two other features are added to these 33 features: the number of correct answers given by the reader, and the number of saccades between the main text and the questions. A total of 35 features are used.

Estimating the TOEIC score

Before estimating the TOEIC score, a backward stepwise feature selection is used to remove the features which are not relevant. The feature selection is not applied for each subject independently but for all the subjects based on a leave-one-out subject cross validation.

Then, the ridge regression method [5] is used to estimate the TOEIC score of each subject as a function of 35 features.

Experiment

We asked nine subjects to read 10 documents. Each document consists of two pages: the left one contains the main text and the right one contains four questions. The subject is free to read and solve the question in the order he wants. After processing five documents, the subject has a short break. The mobile eye tracker used for the experiment is an "Eye tracking Glasses 2.0 (ETG2.0)" made by SensoMotoric Instruments (SMI). The eye tracker was calibrated one time for each participant before starting the experiment.

The average age of the subjects is 22.4 (maximum: 24, minimum: 20, and standard deviation: 1.16). All subjects are Japanese males and university students. Seven are undergraduate students and 2 are master students. The average of the TOEIC score of the subjects is 660 (maximum: 930, minimum: 435, and standard deviation: 154.1).

Result

First, the feature selection was applied. The backward stepwise feature selection selected 20 features among the 35 ones. For the evaluation of the selected features, we computed the correlation coefficient between the feature and the TOEIC score. The features with the highest correlation coefficient are: the number of correct answers (0.74), the sum of the saccade time of the question area (0.70), the average fixation duration of the overall area (0.66), and the sum of the saccade length of the overall area (0.66).

The table 2 shows the result of the TOEIC score estimation. For each participant, we compute the absolute er-

Subject	Abs. Error
A	87.0
B	86.9
C	1.52
D	28.1
E	42.9
F	24.1
G	9.38
H	24.5
I	21.9
Average	36.3

Table 2: Estimation of the TOEIC score.

ror between the real TOEIC score and the estimated one. On average, the absolute mean error is 36.3 points, with a standard deviation of 46.5 points.

Conclusion and future work

We presented a new method for estimating the language skill of a reader by using a mobile eye tracker. After retrieving the document in the scene image recorded in the front camera, the eye gaze of the user is projected in the document space. Some features related to the reading behavior are extracted and then, the TOEIC score of the subject can be estimated.

Our main outlook is to try to reduce the number of needed documents for estimating the English skill. For now, reducing the number of document greatly reduce the performance.

Acknowledgment

This work is supported in part by JST CREST and JSPS KAKENHI Grant Numbers 25240028, 15K12172, and 1681207700.

REFERENCES

1. Georg Buscher and Andreas Dengel. 2009. Gaze-based filtering of relevant document segments. In *International World Wide Web Conference (WWW)*. 20–24.
2. Kiyomi Chujo and Kathryn Oghigian. 2009. How many words do you need to know to understand TOEIC, TOEFL & EIKEN? An examination of text coverage and high frequency vocabulary. *Journal of Asia TEFL* 6, 2 (2009), 121–148.
3. Leana Copeland, Tom Gedeon, and Sumudu Mendis. 2014. Predicting reading comprehension scores from eye movements using artificial neural networks and fuzzy output error. *Artificial Intelligence Research* 3, 3 (2014), p35.
4. Kai Kunze, Hitoshi Kawaichi, Kazuyo Yoshimura, and Koichi Kise. 2013. Towards inferring language expertise using eye tracking. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. ACM, 217–222.
5. Saskia Le Cessie and Johannes C Van Houwelingen. 1992. Ridge estimators in logistic regression. *Applied statistics* (1992), 191–201.
6. Pascual Martínez-Gómez and Akiko Aizawa. 2014. Recognition of understanding level and language skill using measurements of reading behavior. In *Proceedings of the 19th international conference on Intelligent User Interfaces*. ACM, 95–104.
7. Keith Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin* 124, 3 (1998), 372–422.
8. Kazutaka Takeda, Koichi Kise, and Masakazu Iwamura. 2011. Real-time document image retrieval for a 10 million pages database with a memory efficient and stability improved llah. In *Document Analysis and Recognition (ICDAR), 2011 International Conference on*. IEEE, 1054–1058.
9. Kazuyo Yoshimura, Kai Kunze, and Koichi Kise. 2015. The Eye as the Window of the Language Ability: Estimation of English Skills by Analyzing Eye Movement While Reading Documents. In *Document Analysis and Recognition (ICDAR), 2015 13th International Conference on*. 251–255.