Analyzing the eye gaze to estimate Japanese reading understanding

Charles Lima Sanches[1,a)] Olivier Augereau[1,b)] Koichi Kise[1,c)]

Abstract: Analyzing the eye gaze to estimate the text understanding is a good way to overcome the drawbacks of classic subjective/objective understanding assessment tests. In this paper we propose a method to estimate the understanding of a learner by analyzing his eye movements while reading. We conduct our experiment on Japanese texts and try to predict, by analyzing the eye gaze, the number of questions about the text the reader can answer (objective understanding) and the reader's self assessment understanding (subjective understanding). We show that we obtain 5.27% of error in the number of correct answers estimation and 0.33 of mean absolute error in the self assessment understanding estimation by using eye gaze features.

1. Introduction

In the education field, assessing the student's understanding is an important step of the learning process.

In particular, foreign language learning is a particularly challenging field since it demands to develop writing, listening, speaking and reading skills. Among these skills, text reading has been proven to be a key activity [4] since it helps not only on the development of conversation skill but also writing skill [5]. In order to measure and improve the reading skill it is necessary to have access to the understanding of the reader about the reading material. This understanding can usually be obtained by 2 different ways:

(1) objectively by asking comprehension questions to the reader
(2) subjectively by asking the subject to rate his own understanding

However, the first method to get the understanding objectively has some major drawbacks. There is no analysis on the way the candidate answered the question: he could have given a correct answer by chance, or misunderstood the question. Plus, using comprehension questions to estimate the understanding leads to an incomplete estimation: the questions cannot focus on all the parts of the text.

The second method to get the subject's understanding of a text is to ask him to rate his own understanding. However this subjective understanding highly depends on the honesty and the awareness of the reader.

Researchers have shown that the reading understanding is related to the way we move our eyes while reading[2]. Therefore, we can use the relation between understanding and eye movements to predict the text understanding. Plus, the recent development of affordable eye tracking technologies makes it possible to use it in everyday life. Previous work using eye tracking technologies could only be made in laboratory settings with devices costing up to 25,000 USD, but it exists now devices costing around 150 USD.

Our research goal is to use the eye tracking technology to build a system which automatically estimates the subject's understanding. In this paper, we try to estimate the global average subject's understanding (subjective and objective) of Japanese texts only by using the eye gaze information. We show that the global subjective understanding of the texts can be estimated with a mean absolute error of 0.33 by analyzing the eye movements while reading. Plus, we show that the objective understanding can be estimated with 5.27% of estimation error.

The contributions of the paper are: (1) we provide a method to make objective and subjective understanding estimations based on the eye gaze, (2) we show that analyzing the eye gaze is more effective than using self assessment understanding to predict the number of correct answers and (3) we show that analyzing the eye gaze is more effective than using the number of correct answers to predict the reader's self assessment understanding.

2. Related Work

Augereau *et al.* have proposed a method to estimate the TOEIC score (which ranges from 10 to 990) of a student by analyzing the eye gaze information [1]. However, they rely on the correctness of the answers to perform the estimation.

Karolus *et al.* have proposed a method to detect if a sentence is displayed in a language the reader can understand [3]. They have shown that the blink durations and the time spent on the words depends on the language proficiency of the reader. However, the reading material consists of very simple questions displayed during a short period of time (4.5 seconds on average). Since the questions are simple, the difficulty is not taken into account with

1 Osaka Prefecture University
a) charles-ls@hotmail.fr
b) augereau.o@gmail.com
c) kise@cs.osakafu-u.ac.jp

their method but it could have a major impact on the reading behavior.

The related work shows that it is possible to get some information about the comprehension of the reader by analyzing his eye movements. However, estimating this comprehension without using the answers of some comprehension questions remains a challenging task. Plus, this analysis has not been performed for Japanese texts yet.

3. From the eye gaze to the understanding

When the reader reads a text, we want to analyze his eye gaze pattern to determine his understanding. In order to do so, we can analyze the distribution of the fixations and the saccades. We have to find some features related to the fixations and the saccades which contain the understanding information. In the following we describe the features analysis.

3.1 Features analysis

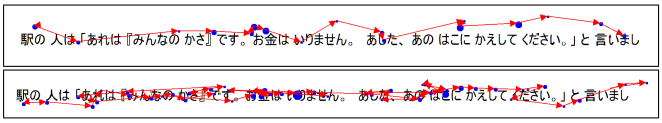

Fig. 1 Eye gaze of 2 subjects reading the same text. The fixations are represented by a blue dot. The saccades between 2 fixations are represented by red arrows. The first subject (upper part) understood the content of the text and the second did not understand the content of the text.

As a starting point of analysis let's compare the recording of a reader who did not understand what he has just read and a reader who almost completely understood the text he read. **Fig. 1** illustrates these two different scenarios. If we compare the saccades and the fixations of the two recordings we can find several differences: direction and length differences for the saccades, numbers and position differences for the fixations. These differences illustrates that the information of the understanding is included in the fixations and saccades pattern. Then, by analyzing several examples we can formulate hypotheses for two categories of readers:

(1) Readers who understand the reading material:
- **H1:** The subject will produce few and spread fixations.
- **H2:** The saccades produced by the subject are long and there is almost no backward saccades.

(2) Readers who does not understand the reading material:
- **H3:** The subject will produce numerous and grouped fixations.
- **H4:** The saccades produced by the subject are short with many backward saccades.

Therefore, to estimate the understanding we can look for features related to fixations or related to saccades.

3.2 Global average understanding

3.2.1 Subjective and objective understanding from the eye gaze

In order to find the global understanding over all the texts read by each user we compute the average of eye gaze features over all the texts. Let $V_1 = \begin{bmatrix} f_{11} & f_{12} & \dots & f_{1n} \end{bmatrix}$ be the vector we obtain after extracting the n features from the raw eye gaze of the text 1. In the same way we have $V_2 = \begin{bmatrix} f_{2n} & f_{2n} & \dots & f_{2n} \end{bmatrix}$ the vectors computed from the text 2. We can build $V_1, ..., V_p$ the p vectors computed from the p texts read by the user. Then we will build a vector $X = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}$ with

$$a_y = \frac{f_{1y} + f_{2y} + f_{3y} + ... + f_{py}}{p} \quad (1)$$

This vector X will be the input of our classifier.

The overview of the algorithm is described in **Fig. 2**. First, we extract the fixations and the saccades from the eye gaze, then we extract some features from it. We average the features for each text read by the user and use these average features as an input for the classifier. We train our model using a leave one out strategy (subject independent) and test on a new subject using the trained model. The output will be different depending on the understanding we want to estimate:

- the self assessment understanding estimation (subjective understanding)
- the percentage of correctly answered questions (objective understanding)

3.2.2 From the answers

In order to evaluate the subjective understanding estimated by using the eye gaze features, we compare it with the same method using the correctness of the answers given by the subject. The question are text comprehension questions directly extracted from JLPT text books. The JLPT test is a standardized test for assessing Japanese ability *1. The answer (*wrong=0 or right=1*) is used as a feature for the classifier. The answers of all the questions answered by one subject are summed and the sum is divided by the number of questions answered by the subject. The result is an average score S which will be the input of the classifier.

3.3 From the self assessment understanding

In order to evaluate the objective understanding estimated by using eye gaze features, we compare it with the same method using the self assessment understanding provided by the user. This self assessment (*from 1: not understood the text to 4: perfectly understood the text*) is used as a feature for the classifier

4. Experiment

In this section we describe the experiment protocol. The eye tracker used in this experiment is a non professional eye tracker like the Tobii EyeX Controller *2. This kind of eye tracker is inexpensive and therefore can be spread to a large community.

17 subjects were asked to read 19 Japanese texts while their eye movements were recorded. The subjects are from different nationalities: 4 Chinese, 10 French, 1 German, 1 Taiwanese and 1 Vietnamese. They are all university students with different majors and Japanese abilities. The texts are extracted from some Japanese Language Proficiency Test text books, and correspond to different level of JLPT: from N2 to N5, N5 being the easiest. For each text, the subject was asked to answer one comprehension question.

Then, the subject was asked to rate his understanding of each

*1 [http://www.jlpt.jp/e/]
*2 [http://www.tobii.com/]

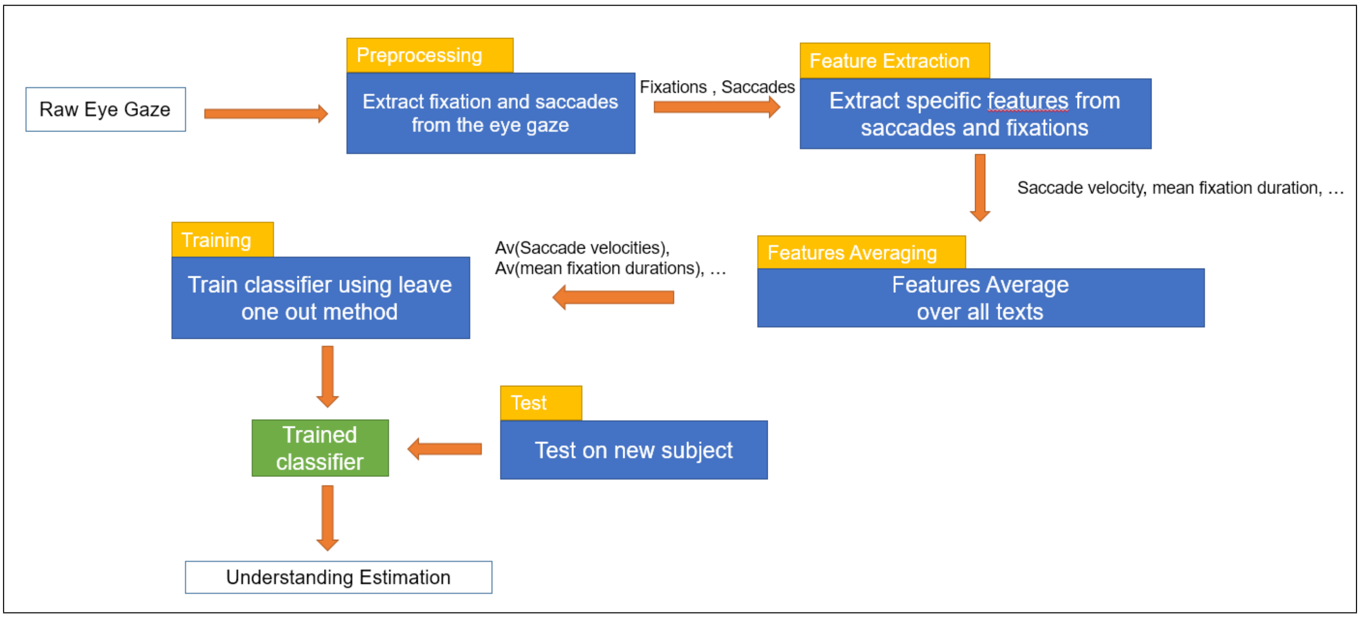

Fig. 2 Algorithm overview from the raw eye gaze to the prediction of the understanding.

text (*from 2: not understood the text to 5: perfectly understood the text*). Among all participants 217 questions were correctly answered, 81 were not correctly answered. The recordings with a score understanding 1 which corresponds to a text reading abortion have been removed. Among all participants, the distribution of the understandings are, from 1 to 5: 23, 45, 74, 137 respectively.

4.1 Results and analysis

In this section are gathered the results of our experiment. We first evaluate the estimation of the subjective understanding, then the precision of the objective understanding estimation.

4.2 Subjective understanding estimation

The number of fixations, the standard deviation of fixation durations, the number of forward saccades, the number of backward saccades, the average saccade duration, the average saccade velocities and the standard deviation of saccades velocities were used as features in this part.

We compare the precision of the estimation using the eye gaze features with the estimation using the answer feature. We want to know which estimation is the closest to the real average understanding value in the interval [2;5]. As a reference we compute the baseline regression performance which is defined as the mean absolute error if all decisions were equal to the mean value.

Table 1 Error in the subjective understanding estimation using eye gaze features, the answer feature or the baseline classifier.

Features	Mean absolute error
Eye Gaze	0.33
Answers	0.38
Baseline	0.62

The corresponding results are shown in Table 1.

In this table we can see that using the eye gaze to predict the subjective understanding is more effective than using the correctness of the answers. This can be explained because the answer

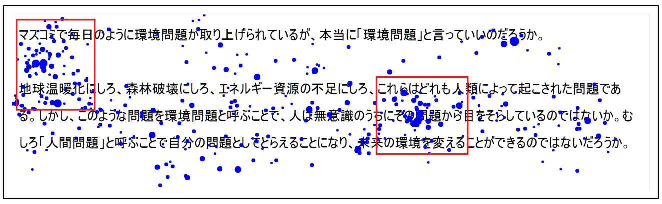

Fig. 3 Example of recording: the reader struggled with two parts of the text but read the rest more smoothly. The understanding information of these two parts of the text is included in the eye gaze but not in the question.

of the question cannot relate the general comprehension of a text. For example, in **Fig. 3** we can see that the reader struggled with two specific parts of the text (red square): the fixations are condensed and the saccades very short. Therefore, by looking at the eye gaze we can know that these part were difficult for the reader. However, if the information included in these two parts are not needed for answering the question, the subject will answer correctly. The information of the understanding is not accessible by using the question because it cannot cover all the text parts. In order to be more efficient it would be necessary to ask questions about all the parts of the text, which is impossible. Conversely, the understanding of every parts of the text is included in the eye gaze and then the global understanding is easier to predict.

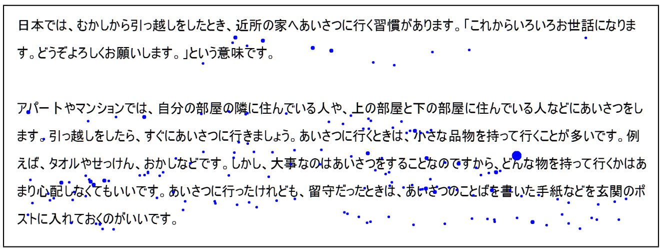

Fig. 4 Example of recording where the subject keeps reading smoothly: there is no fixation agglomerations and the fixation duration are small (dot diameter). However, the subject did not understand the text very well and gave a wrong answer for the comprehension question.

However, the algorithm can be "fooled" by the subject's reading behavior. For example, if the subject keeps reading smoothly even if he does not understand the text. In that case, the non understanding will not be included in the eye gaze. This case is represented in **Fig. 4**. There is no fixation agglomerations in this recording and the fixation duration are small, indicating that the reader read smoothly the text. However, he did not answer correctly the comprehension question, so he clearly misunderstood some parts or all the text. This can explain why answering the question can be more accurate than analyzing the eye gaze for some specific reading behaviors.

5. Objective understanding estimation

Table 2 Error in the objective understanding estimation using eye gaze features, the understanding feature or the baseline classifier.

Features	Error(%)
Eye Gaze	5.27%
Understanding	9.04%
Baseline	16.42%

The reading time, the standard deviation of fixation durations, the number of backward saccades and the average saccade duration were used as features in this part. We compare the precision of the estimation using the eye gaze features with the estimation using the self assessment understanding feature. We want to know which estimation is the closest to the real percentage of correctly answered questions. As a reference we compute the baseline regression performance which is defined as the mean absolute error if all decisions were equal to the mean value. The corresponding results are shown in Table **Table 2**.

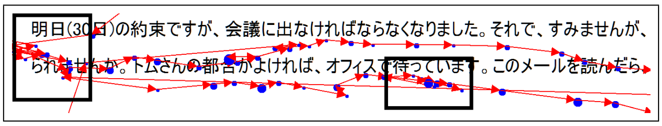

Fig. 5 Example of recording where the subject claimed he perfectly understood the text but gave a wrong answer. By analyzing the eye gaze, we can actually detect he struggled with some parts of the text (more fixations/regressions).

In this table we can see that we can accurately predict the number of correctly answered questions by using the eye gaze. Plus, using the eye gaze is more efficient than using the self assessment understanding indicated by the user. This can be explained because the reader has usually a wrong perception of his understanding. For example in **Fig. 5** the reader claimed he perfectly understood the text, yet has given a wrong answer. However, with the eye gaze recording, it is possible to detect he actually struggled on several parts of the text (more regressions, and more fixations). In this case, using the eye gaze is more accurate than using the self assessment understanding because it does not depends on the subject awareness of his understanding.

On the other hand, the estimation can be wrong for some specific behaviors. Because the subjects knew they would have to answer comprehension questions, some of them over checked the texts even if they understood it the first time they read it. Such behavior is displayed in **Fig. 6**. The subjects produces many fixations and many backward saccades which will be interpreted by

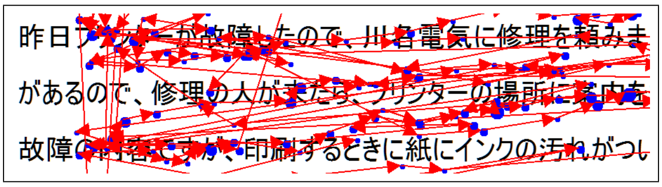

Fig. 6 Example of recording where the subject overcheck the text even if he understood it.

the algorithm as a non understanding behavior. However, the subject correctly understood the question and claimed he perfectly understood the text.

6. Conclusion

In this paper we have presented a method to estimate both the objective and subjective understandings of Japanese texts using the eye gaze. Our method is based on the analysis of average features to predict a global understanding over several texts.

We have shown that our method for estimating the subjective understanding is more accurate than asking questions about the texts. With our method the mean absolute error is 0.33 against 0.38 by using the comprehension question.

We have also shown that our method for estimating the objective understanding is more accurate than using the self assessment understanding provided by the reader. With our method we can estimate the percentage of correctly answered questions with an error of 5.17% using the eye gaze against 9.04% using the self assessment understanding.

However there are still some error in the understanding estimation. This can be explain by the small size of the text and by some specific reading behaviors such as understanding over checking. We plan to use the content of the text, to detect unusual reading behavior which does not match with the visual complexity of the sentences. Plus, in the future we plan to combine the question features and other eye gaze features to reach the best accuracy in the text understanding estimation.

References

[1] Augereau, O., Fujiyoshi, H. and Kise, K.: Towards an automated estimation of English skill via TOEIC score based on reading analysis, *Pattern Recognition (ICPR), 2016 23rd International Conference on*, IEEE, pp. 1285–1290 (2016).

[2] Biedert, R., Dengel, A., Elshamy, M. and Buscher, G.: Towards robust gaze-based objective quality measures for text, *Proceedings of the Symposium on Eye Tracking Research and Applications*, ACM, pp. 201–204 (2012).

[3] Karolus, J., Wozniak, P. W., Chuang, L. L. and Schmidt, A.: Robust Gaze Features for Enabling Language Proficiency Awareness, *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ACM, pp. 2998–3010 (2017).

[4] Krashen, S. D.: *The power of reading: Insights from the research: Insights from the research*, ABC-CLIO (2004).

[5] Mason, B. and Krashen, S.: Extensive reading in English as a foreign language, *System*, Vol. 25, No. 1, pp. 91–102 (1997).