

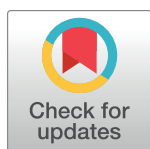
RESEARCH ARTICLE

Estimation of reading subjective understanding based on eye gaze analysis

Charles Lima Sanches , Olivier Augereau , Koichi Kise 

Osaka Prefecture University, Sakai, Osaka, Japan

* charles-ls@hotmail.fr



Abstract

The integration of ubiquitous technologies in the field of education has considerably enhanced our way of learning. Such technologies enable students to get a gradual feedback about their performance and to provide adapted learning materials. It is particularly important in the domain of foreign language learning which requires intense daily practice. One of the main inputs of adaptive learning systems is the user's understanding of a reading material. The reader's understanding can be divided into two parts: the objective understanding and the subjective understanding. The objective understanding can be measured by comprehension questions about the content of the text. The subjective understanding is the reader's perception of his own understanding. The subjective understanding plays an important role in the reader's motivation, self-esteem and confidence. However, its automatic estimation remains a challenging task. This paper is one of the first to propose a method to estimate the subjective understanding. We show that using the eye gaze to predict the subjective understanding improves the estimation by 13% as compared to using comprehension questions.

OPEN ACCESS

Citation: Lima Sanches C, Augereau O, Kise K (2018) Estimation of reading subjective understanding based on eye gaze analysis. PLoS ONE 13(10): e0206213. <https://doi.org/10.1371/journal.pone.0206213>

Editor: Phil Reed, Swansea University, UNITED KINGDOM

Received: March 2, 2018

Accepted: October 9, 2018

Published: October 25, 2018

Copyright: © 2018 Lima Sanches et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data used to compute the statistics in the paper have been made publicly available at the following address: <https://github.com/chalulu/Subjective-Understanding>.

Funding: This work is in part supported by Japanese Science and Technology Agency, Core Research for Evolutionary Science and Technology (JPMJCR16E1), <https://www.jst.go.jp/kisoken/crest/en/>; JSPS KAKENHI Grant Number 15K12712, <https://www.jsp.go.jp/english/e-grants/>; Japanese Society for the Promotion of Science, KAKENHI (16K16089), <https://www.jsp.go.jp/>.

Introduction

The recent development of small and inexpensive devices has led to a breakthrough in the education field. It allows students to learn from anywhere at any time and their learning activity can be recorded and analyzed. Such analysis is used to enhance their learning experience and provide new services such as context-aware learning systems [1] or adaptive teaching systems [2]. Moreover, the learning activity of a student can be used by a teacher as a feedback of the lesson comprehension [3], or by other students as a point of comparison. The idea is to set an environment where the students can be in a learning situation at any moments even if they are not aware of it. Such environment is called a "Ubiquitous Learning Environment" [4].

Ubiquitous learning environments have been widely explored in foreign language learning [5, 6] which demands a life-long practice to develop writing, listening, speaking and reading skills. Among these skills, reading has been proven to be a key activity [7] since it helps not only the development of conversation skills but also writing skills [8].

We distinguish two types of understanding: the objective understanding and the subjective understanding. The objective understanding can be measured by answering comprehension

go.jp/english/e-grants/; and Key Project Grant Program of Osaka Prefecture University.

Competing interests: The authors have declared that no competing interests exist.

questions about the content of the text. It is evaluated in most language proficiency tests such as the TOEIC [9]. The subjective understanding measures the reader's perception about his own understanding. It is linked to the reader's confidence, self-esteem or motivation [10, 11]. Thanks to the subjective understanding it is possible to detect when the reader thinks he understood the content of the text but actually did not. Usually, the subjective understanding is estimated by asking the reader to rate his own understanding. However, this type of evaluation is not compatible with a ubiquitous learning environment where an automatic estimation is preferable.

Researchers have shown that the understanding of a text is related to our eye movements while reading [12–14]. More generally, it has been proved that the reading behavior is correlated to the eye movements patterns [15–17]. Therefore, we make the hypothesis that it is possible to use the relation between understanding and eye movements to predict the text subjective understanding. Moreover, the recent development of affordable eye tracking technologies makes it possible to use it in an everyday life situation. Previous work using eye tracking technologies could only be made in laboratory settings with devices costing up to 25,000 USD, but nowadays some devices cost around 150 USD which make them usable in everyday life.

Our research goal is to use the eye tracking technology to build a system which automatically estimates the reader's subjective understanding. As a first step toward this goal, we propose in this paper a method to automatically estimate the overall subjective understanding of several Japanese texts based on eye tracking.

We show that we obtain 13% improvement in the estimation of the overall subjective understanding by using the eye gaze as compared to using comprehension questions. Moreover, because we aim to predict the understanding for each text independently, we analyze the effect of reducing the number of read texts on our algorithm. We show that not only analyzing the eye gaze is always more effective than using the comprehension questions, but also the smaller the number of read texts, the greater is the performance difference between both methods.

The contributions of the paper are twofold:

1. we provide a method to estimate subjective Japanese text understanding based on the eye gaze,
2. we show that the analysis of the eye gaze can predict the user's subjective understanding more efficiently than by analyzing the comprehension question answers.

The remainder of the paper is organized as follows. First, we present some related work on eye gaze while reading, and document understanding using eye movements. Next, we explain our algorithm and in particular which information from the eye gaze we use to deduce the subjective comprehension of a text. Then we present the database, the readers and the detailed results for our experiment. Finally, we draw a conclusion and discuss the future work.

1 Related work

The eye movements while reading is composed of two categories: saccades and fixations. A fixation is a short stop on a part of a word during reading which lasts about 250 ms and a saccade corresponds to the quick movement of the eyes between two fixations.

Rayner has shown that these two features of the eye movements pattern change in accordance with the comprehension of the reader [12]. For example, the number of fixations is increased if the reader does not understand the text he is reading. Also, the readers tend to produce more backward saccades (rereading regression) if they have difficulty to understand the text.

Kunze *et al.* have also tried to show that there is a link between the reading pattern and the reading skills of the reader [18]. In particular, they tried to show that a reader tends to focus

on a smaller number of words in the text if they have higher English reading comprehension skills. However, not only their database was small (5 participants reading 5 texts), but they also concluded that they did not find a statistically significant correlation between eye gaze and text understanding.

In their study Yoshimura *et al.* have tried to classify the English skill of readers in 3 classes: low, middle and high using eye gaze features [19]. The readers are assigned to a class depending on their TOEIC score which is a classic question-based language assessment test. As a consequence, the subjective understanding is not estimated in this study.

As an improvement of Yoshimura *et al.*'s work, Augereau *et al.* have proposed a method to estimate the TOEIC score of a student using the eye gaze information [20]. TOEIC is a test for assessing English proficiency. It includes a reading part and a listening part and the examinee is given a score, ranging from 10 to 990 points, which represents his English comprehension. Since the global TOEIC score and the reading part score are highly correlated [21], they have tried to estimate the TOEIC score using only the reading part. Each participant has been asked to read a text extracted from TOEIC test and to answer the questions related to it, while their eye movements were recorded. However, in order to predict the score, they use the correctness of the answers as an important feature. The estimation cannot be done without asking questions about the text to the participants. Moreover, no information about the participant's estimation of his own understanding has been asked during the experiment. Therefore, only the objective understanding is evaluated in this study.

Karolus *et al.* have proposed a method to detect whether a displayed sentence is in a language understandable to the reader [22]. In particular, they have shown that fixation and blink durations are impacted by the language proficiency of the reader. However, this study has been conducted by only analyzing the understanding of very simple questions such as "How many days are within a week?", on a very short period of time (4.5 seconds). Since the questions are very basic, the sentence difficulty is not taken into account in the study and could impact the reading behavior. Therefore, the degree of understanding of the sentence cannot be estimated.

Martinez-Gomez *et al.* have also proposed a method to recognize the language skill of a reader using eye gaze features [23]. Their method has been designed to retrieve both the subjective and objective understandings of the reader. The objective understanding, defined as the TOEIC or TOEFL score of the participants can be correctly estimated by their method. However, the reader's perception of their own English ability, defined as a self-reported measure (beginner, intermediate or advanced) cannot be retrieved by their algorithm. They have shown there is a considerable difference in the objective measure of the understanding, and the subjective participants' perception of their own ability.

The related work shows that it is possible to get some information about the comprehension of the reader by analyzing his eye movements. Moreover, estimating an objective measure of the language ability like the TOEIC score can be performed with a good accuracy. However, no work has been done for predicting the subjective understanding. In the following, we propose a method to predict the subjective understanding of Japanese texts by analyzing the eye movements of the reader.

This work is the continuation of the pre-study presented in [24]. It contains new results and an extended analysis.

2 From the understanding to the eye gaze

In this section, we describe the differences between objective and subjective understanding introduced in the Introduction section. In particular, we explain how important is the

subjective understanding in the context of education. We then describe the features which can be extracted from the reading pattern to predict the subjective understanding.

2.1 Subjective and objective understanding

The student's understanding of some reading material can be divided into two categories: the objective understanding and the subjective understanding. The objective understanding can be measured by using comprehension questions. It is usually measured via a score (the number of correctly answered questions) and is independent of the perception of the reader on his own understanding. In this study, we measure a different kind of understanding: the subjective understanding. The subjective understanding is measuring how well the student thinks he understood a text. The perception of a student on his own understanding is a key factor which will modify his self-esteem, his self-confidence, and his motivation [10, 11]. Moreover, it is an important information for the teacher (or the system, in a fully automated environment) who can adapt his teaching strategy according to it. For a student who answers a comprehension question, different scenarios can occur:

1. The student answered correctly to the comprehension question and he thinks that he understood the text
2. The student answered incorrectly to the comprehension question and he thinks that he did not understand the text
3. The student answered correctly to the comprehension question but he thinks that he did not understand the text
4. The student answered incorrectly to the comprehension question but he thinks that he understood the text

Although scenarios 1 and 2 are trivial, scenarios 3 and 4 require more attention from the teacher. These scenarios are accessible only if we can measure the reader's subjective understanding.

In our study, we want to focus on the subjective understanding prediction. We compare the performance of two estimators to estimate the reader's subjective understanding of one text: the correctness of a comprehension question and the eye gaze pattern of the reader. Objective and subjective understanding are different but correlated, and comprehension questions are the most common way to estimate the students' objective understanding. Therefore, it makes sense to try to use it as an estimator of the subjective understanding. Moreover, the texts of our study are extracted from the Japanese Language Proficiency Test, a standardized test to assess one's Japanese language ability. For each text, the associated comprehension question has been carefully designed by the test organizers to assess the student's objective understanding of the text. But our hypothesis is that the reader's eye gaze can be used to estimate the subjective understanding, and is a better estimator than the correctness of comprehension questions answers.

2.2 Features analysis

As a starting point of the 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 illustrate that the information of the understanding is included in the fixations and saccades pattern. Moreover,

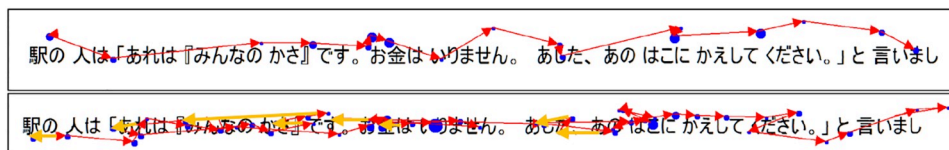


Fig 1. Eye gaze of 2 readers reading the same text. The fixations are represented by a blue dot which diameter depends on the duration: the more the reader looked at a specific region of the text, the bigger the dot is. The saccades between 2 fixations are represented by red strokes which direction is indicated by the arrow. The backward saccades (regressions) are represented in yellow. The first reader (upper part) understood the content of the text and the second did not understand the content of the text.

<https://doi.org/10.1371/journal.pone.0206213.g001>

according to the literature, the reader's fixations duration depends on his understanding [25]. Then, by analyzing several examples we can formulate hypotheses for two categories of readers:

1. Readers who understand the reading material:
 - **H1:** The reader will produce few, spread and quick fixations.
 - **H2:** The saccades produced by the reader are long and there is almost no backward saccades.
2. Readers who do not understand the reading material:
 - **H3:** The reader will produce numerous, grouped and long fixations.
 - **H4:** The saccades produced by the reader are short with many backward saccades.

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

2.2.1 Fixations and saccades features. According to the hypotheses **H1** and **H3**, the more the reader understands the text the less fixation he will produce. As a consequence, **f1 the number of fixations** is a good candidate as a feature to predict the understanding.

According to the hypothesis **H2**, a reader who understands the text will produce long forward saccades and a few numbers of backward saccades. Conversely, according to the hypothesis **H4** a reader who did not understand the reading material will produce long saccades and many forward saccades. We can select **f2 the number of forward saccades**, **f3 the number of backward saccades** to estimate the understanding. To know how fast the reader is reading we use the idea of "saccade velocity" which is defined as:

$$s_v = \frac{s_l}{s_d} \quad (1)$$

with s_v , s_b , s_d the saccade velocity, the saccade length and the saccade duration respectively. We can use **f4 the standard deviation of saccades velocities** as features to reflect the reading speed.

2.3 Features normalization

Some of the features previously described are highly dependent on the text length. For example, it is normal to produce more fixations on a long text because it contains more characters (or words). In the same way, it is normal to produce more forward saccades on a long text than a short text. Then, we can obtain variability on these features values because of the length of the text and not because of differences in the understanding. Therefore, it is essential to normalize the features in order to be independent from the length of the text. In order to do it,

Table 1. Features used to determine the reader comprehension of a text with their corresponding normalization.

Feature number	Feature	Normalization	Hypothesis
1	Number of fixations	Number of characters	H1,H3
2	Number of forward saccades	Number of characters	H2,H4
3	Number of backward saccades	Number of characters	H2,H4
4	Standard deviation of saccades velocities	Not needed	H2,H4

<https://doi.org/10.1371/journal.pone.0206213.t001>

some features are normalized by the number of characters in the text and for other features, we use the standard deviation over all the recording. The normalization is performed for each text.

The features relevant to estimate the subjective reading understanding, the corresponding normalizations, and the associated hypotheses are gathered in Table 1.

2.4 Regression

Once we have determined which features contain the subjective understanding information, we can see the retrieval of the reader's comprehension as a minimization problem. Let V be the vector we obtain after extracting the n features from the raw eye gaze: $V = [f_1 f_2 \dots f_n]$ with f_n the value of the feature n . And if we call U the reader's subjective understanding and U_E the corresponding estimated subjective understanding. We then look for a function $F(V)$ which minimizes the distance $|U - U_E|$.

To evaluate the performance of our algorithm, we can compute the average mean distance over all participants. If we call N the number of participants, the performance is computed as E , the mean absolute error of the understanding estimation:

$$E = \frac{1}{N} \sum_{i=1}^N |U - U_E| \quad (2)$$

2.5 Global average understanding

2.5.1 From the eye gaze. In our method, we compute the average subjective understanding of all the read texts for each user and try to predict this global average subjective understanding using the features extracted from the eye gaze. Because the average subjective understanding over all the texts read by a reader is a continuous value, our problem is not a classification problem. As a consequence, we use a Support Vector Regression (SVR) method to estimate this average subjective understanding. In order to find the global subjective understanding over all the texts read by each user, we compute the average of features over all the texts. Let $V_1 = [f_{11} f_{12} \dots f_{1n}]$ 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 $V_2 = [f_{21} f_{22} \dots f_{2n}]$ the vectors computed from the text 2. We can build V_1, \dots, V_p the p vectors computed from the p texts read by the user. Then we will build a vector $X = [a_1 a_2 \dots a_n]$ with

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

This vector X will be the input of our SVR.

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 SVR. We train

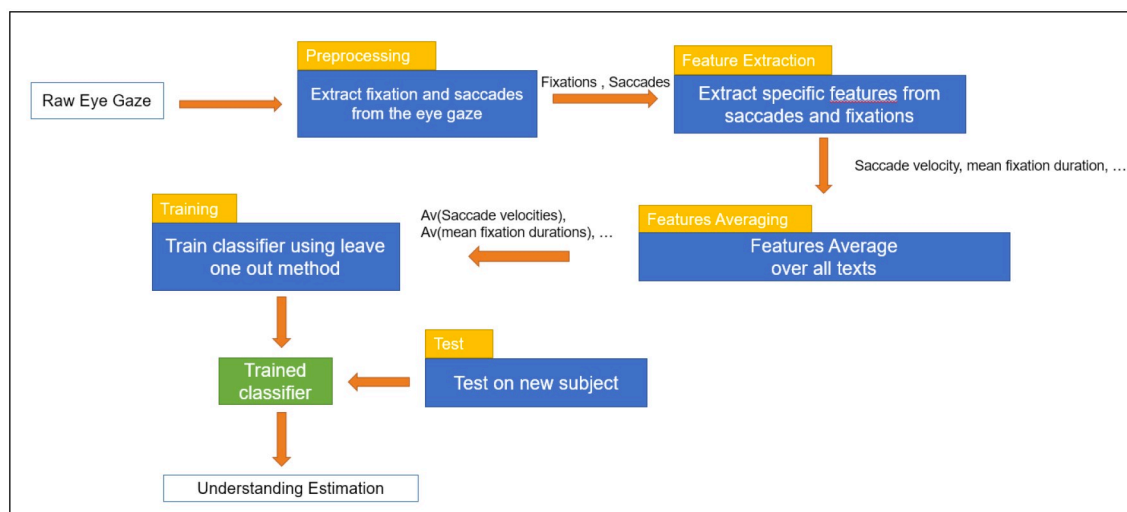


Fig 2. Algorithm overview from the raw eye gaze to the prediction of the understanding.

<https://doi.org/10.1371/journal.pone.0206213.g002>

our model using a leave-one-participant-out strategy. Therefore, our method is participant independent.

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

3 Experiment

This study has been approved by the ethics committee of the graduate school of engineering, Osaka Prefecture University. The authors received a written consent from all the participants. The participants have been recruited in February 2016 via e-mail or verbal invitation. All the participants were university students aged between 20 and 28 years old. All the data were anonymised.

In this section, we describe the experiment protocol. The eye tracker is used in this experiment is a nonprofessional Tobii eye tracker (<http://www.tobii.com/>). With an accuracy $<0.6^\circ$, a precision $<0.25^\circ$, and a latency $<50\text{ms}$, such eye trackers can be used to successfully measure fixations and saccades [26]. A research license is required to use it for research purposes. This kind of eye tracker is inexpensive and therefore can be spread to a large community. The eye tracker is attached to a screen and the documents are displayed on this screen. The head of the reader needs to be approximately between 45 to 80 cm apart from the screen. The reader is free to move his head, no bite bar nor head fixation have been used in order to be close to real reading conditions. At 70 cm distance, the allowed head movement is a plane of approximately 48 x 39 cm. If the participant moves more, the eye gaze will not be detected anymore by the eye tracker.

17 participants (7 females, 10 males) were asked to read 19 Japanese texts while their eye movements were recorded. The participants are from different nationalities: 4 Chinese, 10 French, 1 German, 1 Taiwanese and 1 Vietnamese. They are all university students (from 20

<p>「仕事ができる人」と「できない人」のいちばんの違いは何だろうか。私がいろいろな仕事で接してきた「できる人」に共通している特徴は、「自分の能力特性の分析」がうまいということだ。</p> <p>まず、自分は何が得意で、何が不得意なのかを分かっている。もちろんオールマイティ(どんなことも完全にできること)にあらゆる仕事をこなす(処理する)エリートもいるのだろうが、わたいてい人は、「仕事の得手、不得手」を持っている。そのうえで、「得意な仕事」を積極的になし、「不得意な仕事」からは上手に逃がっている人が、実は「仕事ができる人」と評価されるのだ。</p>	<p>中村さんの机の上に、先生からのメモが置いてある。</p> <p>中村さん</p> <p>おはようございます。</p> <p>昨日プリンターが故障したので、川名電気へ修理を頼みました。修理の人は、10時に来てくれるそうです。わたしは授業があるので、修理の人が来たら、プリンターの場所に案内をお願いします。</p> <p>故障の内容ですが、印刷するとき紙にインクの汚れがついてしまいます。実際に印刷した紙がプリンターのところに置いてありますから、それを修理の人に見せて説明してください。</p> <p>修理がすんだら、午後のゼミの資料を人数分用意しておいてください。</p> <p>下田</p>
<p>きのうの夜は、おそくまでしごとをしました。とてもつかれました。しごとのあと、電車で帰りました。</p> <p>家の近くの駅で電車をおりました。外は雨でしたが、わたしはかさがありませんでした。とてもこまりました。</p> <p>駅の人がわたしを見て、「あのはこの中のかさを貸してください。」と言いました。はこの中にはかさが3本ありました。わたしは「えっ、いいんですか。」と聞きました。</p> <p>駅の人は「あれは『みんなのかさ』です。お金はいりません。あした、あのはこにかえてください。」と言いました。</p> <p>わたしは「わかりました。ありがとうございます。」と言って、かさをかりて帰りました。</p>	<p>日本では、むかしから引っ越しをしたとき、近所の家へあいさつに行く習慣があります。「これからいろいろお世話になります。どうぞよろしくお願いします。」という意味です。</p> <p>アパートやマンションでは、自分の部屋の隣に住んでいる人や、上の部屋と下の部屋に住んでいる人などにあいさつをします。引っ越しをしたら、すぐにあいさつに行きましょう。あいさつに行くときは、小さな品物を持って行くことが多いです。例えば、タオルやせっけん、おかしななどです。しかし、大事なものはあいさつをすることなので、どんな物を持って行くかはあまり心配しなくてもいいです。あいさつに行ったら、留守だったときは、あいさつをくれた手紙などを玄関のポストに入れておくのがいいです。</p>

Fig 3. Example of JLPT texts. Level N2 (top left), level N3 (top left), level N4 (bottom right), level N5 (bottom left).

<https://doi.org/10.1371/journal.pone.0206213.g003>

years old to 28 years old), with different majors and Japanese abilities. The texts are extracted from some Japanese Language Proficiency Test textbooks and correspond to the different levels of the JLPT: from N2 to N5, N5 being the easiest. Because of the lack of people capable of understanding the N1 level (most difficult level) we did not use any N1 level texts in this experiment. Samples of the used texts are shown in Fig 3. Each participant had to read four texts N2, five texts N3, five texts N4, five texts N5. The average number of characters per text is 210. The texts were written using the font MS UI gothic, and the font size was 30 pt. For each text, the participant was asked to answer one comprehension question designed by the test organizers.

Then, the participant was asked to rate his understanding of each text from 1 to 5. Such self-assessed methods induce bias in the measures which can be caused by four factors [27]:

- Failure in memory: we tried to minimize this factor by asking the participant to rate his understanding immediately after reading the text.
- Lack of appropriate motivation: We minimized this factor by paying the participants for the experiment.
- Failure in communication or lack of knowledge: We minimized this factor by explaining carefully which marks corresponds to which level of understanding (equivalence).

For each value of the subjective understanding, the corresponding equivalence has been given.

- 1: Aborted the text
- 2: Understood almost nothing
- 3: Understood part of the content
- 4: Understood the content but some parts were unclear
- 5: Understood everything

In the evaluation of our algorithm, we did not take into account the case 1 as the participant did not read a large part of the corresponding text. Then, the average understanding we want to predict is a continuous value within the interval [2;5]. The distribution of each recording in

Table 2. Understanding distribution in the database.

Understanding	Number of Recordings
1	23
2	45
3	23
4	74
5	137

<https://doi.org/10.1371/journal.pone.0206213.t002>

the database is shown in Table 2. The descriptive statistics of eye movements data for this experiment is shown in Table 3. The data used to compute the statistics is publicly available (<https://github.com/chalulu/Subjective-Understanding>).

3.1 Results and analysis

This section gathers the results of our experiment. As a first analysis of the results, we compute the average value of the features for the different subjective understanding values. Table 4 shows that for most of the features, there is a correlation between the value of the feature and the self-assessed understanding. The p-values were also computed using the Pearson method to show the significance of this correlation. Since the feature 2 did not show a strong significance of the correlation between its value and the self-assessed understanding, the model was capable of finding a non-linear relationship between the two values.

The statistical method used here assumes that the data is normally distributed. The normality has been assumed after examination of the Quantile-Quantile (QQ) plots for each feature. Because of the number and the size of the plots, they have not been included in the paper. The QQ plots and corresponding computed values have been made publicly available (<https://github.com/chalulu/Subjective-Understanding>).

For our main result, we compare the precision of the estimation using the eye gaze features with the one using the answer feature. We want to know which estimation is the closest to the real average understanding value in the interval [2;5]. The corresponding results are shown in Table 5. In our experiment, we did not use the recordings corresponding to aborted texts. Therefore, the number of texts is not the same for all the participants.

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 answers to the questions cannot relate perfectly to the general comprehension of a text. For example, in Fig 4 we can see that the reader struggled with two specific parts of the text (green

Table 3. Descriptive statistics of eye movements data (average and standard deviation for each feature).

Eye Movements Data	Statistics	
	Average	STD
Fixation count	111.53	65.89
Fixation duration (ms)	291.48	70.11
Forward saccade count	82.13	49.17
Backward saccade count	28.38	19.43
Ratio forward/backward saccade	3.15	1.38
Saccade velocity (px/ms)	0.32	0.10
Forward saccade length (px)	114.99	39.64
Backward saccade length (px)	-327.46	128.86

<https://doi.org/10.1371/journal.pone.0206213.t003>

Table 4. Average features values for the different subjective understandings. The p-values were computed to show the significance of the correlation between the subjective understanding and the features value.

Subjective Understanding	Feature Number (see Table 1)			
	1	2	3	4
2	0.74	0.45	0.08	0.20
3	0.71	0.40	0.09	0.18
4	0.61	0.46	0.11	0.16
5	0.46	0.49	0.14	0.12
p-value	1.10^{-12}	7.10^{-3}	2.10^{-9}	1.10^{-7}

<https://doi.org/10.1371/journal.pone.0206213.t004>

Table 5. Error in the understanding estimation using eye gaze features or the answer feature.

Features	Mean absolute error
Eye Gaze	0.26
Answers	0.30

<https://doi.org/10.1371/journal.pone.0206213.t005>

squares): the fixations are condensed and the saccades are very short. Therefore, the eye gaze indicates which parts were difficult for the reader. However, the reader will be able to answer correctly if the information given in the difficult parts is irrelevant for the right answer. 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 part of the text is included in the eye gaze and then the global understanding is easier to predict.

However, the algorithm can be “fooled” by the reader’s reading behavior. For example, if the reader keeps reading smoothly even if he does not understand the text. In that case, the lack of understanding will not be included in the eye gaze. This case is represented in Fig 5. There is no fixation agglomeration in this recording and the small fixation duration indicates that the reader read smoothly the text. However, he did not answer correctly the comprehension question, so he clearly misunderstood some part or all the text. This can explain that answering the question can be more accurate than analyzing the eye gaze for some specific reading behaviors. A solution to correct this error will be discussed in the section 4.

The comparison of the mean absolute error participant by participant between the answer features and the eye gaze features is shown in Fig 6. In this figure, we can see that using the eye gaze is globally more effective than using the comprehension question, but for some specific

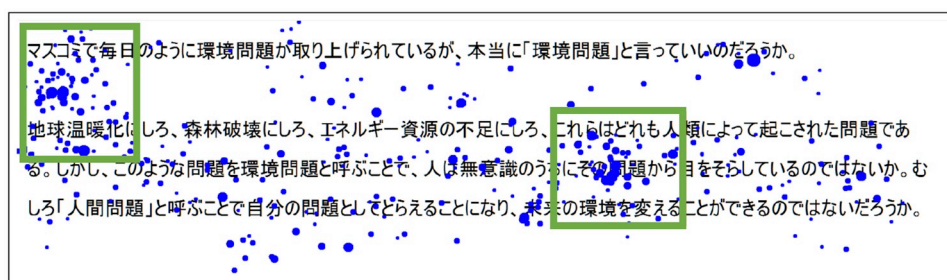


Fig 4. Example of recording: The reader struggled with two parts of the text but read the rest more smoothly. The understanding information of these two parts of the text is included in the eye gaze but not in the question.

<https://doi.org/10.1371/journal.pone.0206213.g004>

アパートやマンションでは、自分の部屋の隣に住んでいる人や、上の部屋と下の部屋に住んでいる人などにあいさつをします。引っ越しをしたら、すぐにあいさつに行きましょう。あいさつに行くときは、小さな品物を持って行くことが多いです。例えば、タオルやせっけん、おかしなどです。しかし、大事なのはあいさつをすることなので、どんな物を持って行くかはあまり心配しなくてもいいです。あいさつに行ったら、留守だったときは、あいさつのことを書いた手紙などを玄関のポストに入れておくのがいいです。

Fig 5. Example of recording where the reader keeps reading smoothly: There is no fixation agglomeration and the fixation duration are small (dot diameter). However, the reader did not understand the text very well and gave a wrong answer for the comprehension question.

<https://doi.org/10.1371/journal.pone.0206213.g005>

reading behavior (for example reader 1) answering the question is more efficient than analyzing the eye gaze.

3.2 Impact of the number of texts on the understanding estimation

After analyzing the recording of 19 texts read by a user and estimating the global understanding of these texts, we analyze the impact of reducing the number of texts on the algorithm performance. Less information will be available to predict the understanding if we reduce the number of texts, which makes the estimation more difficult. The estimation is then more and more difficult. In the limit case, we try to estimate the understanding for each text individually. In this limit case, we cannot compare the performance of our eye gaze method with the question feature method because only one question has been asked for each text. The correctness of the answer is a binary value, and this information is not precise enough to predict an understanding which ranges in the interval [2;5].

The results are gathered in Fig 7. Because we removed the aborted texts, all participants have not read the same number of documents. As a consequence for some participants, the number of available texts is smaller than the number of texts tested. When the number of texts tested is greater than the number of texts available for a participant, we still include this

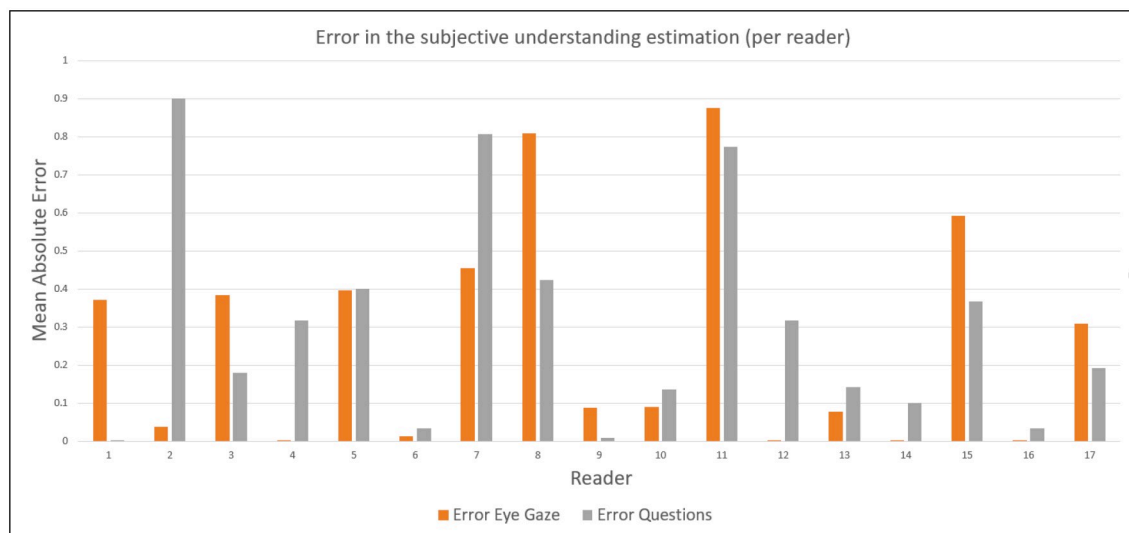


Fig 6. Mean absolute error in the understanding estimation using two different types of information: The eye gaze or the comprehension question.

<https://doi.org/10.1371/journal.pone.0206213.g006>

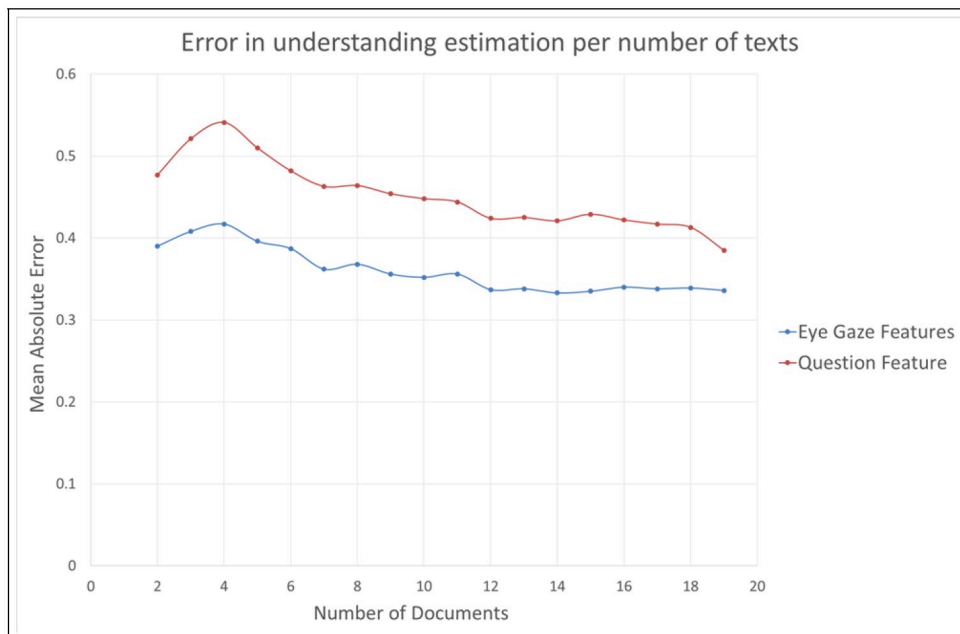


Fig 7. Error in the understanding estimation per number of texts. The eye gaze features method is more accurate than the comprehension question method for any number of documents. For a large number of documents (>15) the performance of both method tends to be similar.

<https://doi.org/10.1371/journal.pone.0206213.g007>

participant in our computation by using all the texts he read. We can see that the fewer number of texts, the lower the precision for both the eye gaze features method and the comprehension question method. This can be explained because for a fewer number of texts the algorithm is more sensitive to unusual reading behaviors. For example, Fig 8 shows a recording with numerous backward saccades and fixations. This means the participant reread a lot some parts of the text and focused on some words. Therefore, the algorithm estimated that he did not understand the text so well (estimated understanding = 2.6). However, when asked about his understanding, the participant answered 5, which corresponds to a complete comprehension of the text. This behavior can occur when the participant reread some parts of the text to check again his own understanding. Knowing that the participant will be asked one comprehension question, he can be tempted to check again the text in order to be sure he will be able to answer the question.

We also notice that analyzing the eye gaze is more accurate than using the comprehension question for any number of documents. However, for a number of texts ranging from 15 to 19, the comprehension question method tends to improve its accuracy although the eye gaze

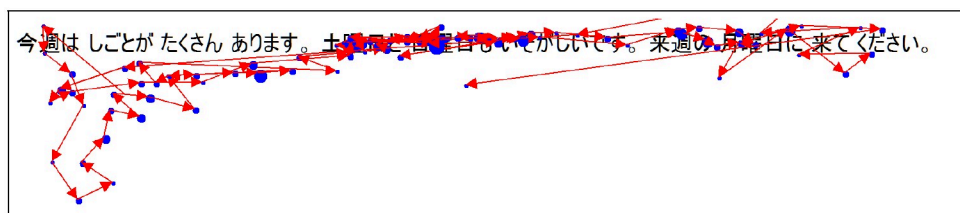


Fig 8. Unusual reading behavior: Even if the participant perfectly understood the text, he reread and focused a lot on some parts of the text.

<https://doi.org/10.1371/journal.pone.0206213.g008>

method performance is stable. This indicates that for a large number of texts and a large number of questions, there is as much information brought by the question as information brought by the eye gaze.

4 Discussion on the sources of error and limitations

In the results above, we have shown that using the eye gaze is an efficient way to predict the understanding of a text for a participant. Analyzing the fixations and the saccades information helps to distinguish a reader who did not understand the text from a reader who did understand, confirming our hypotheses **H1**, **H2**, **H3** and **H4**. However, we have underlined some limitations which affect the algorithm performance. One of the first error is linked to the eye tracker accuracy. Because of the vertical error of the eye tracker, some fixations are not on the text characters. However, because we do not use the position of the fixations relatively to the text in our algorithm (only the fixations number and their duration), this misplacement does not affect the algorithm. As a consequence, no fixations were excluded from the analysis. Nevertheless, the recordings can contain some noise which will be interpreted as a specific reading behavior by the algorithm. For example, in the Fig 8 we can see a drift at the beginning of the recording which is an error of the eye tracker. Moreover, with the vertical error of the eye tracker, some fixations are not on the text characters. This kind of errors can be avoided by improving the calibration step or by using post-processing algorithms [28].

Another source of errors can be unusual reading behaviors. If the reader reads the text smoothly although he did not understand it, or if he rereads many parts of the texts although he perfectly understood them, the algorithm will not be able to estimate accurately the reading understanding. This could be detected by using the content of the text as an additional feature: studies have shown that the word length and the visual complexity can affect the reading behavior of Japanese sentences even for native speakers [29]. In particular, the visual complexity of Japanese characters (made from more or fewer strokes) directly influences the reading time and the text areas fixed by the reader. Therefore, in a text containing areas of simple complexity and difficult complexity, the reading behavior of the reader should change. If it does not change because the reader read smoothly forward the text, it can be detected as an unusual reading behavior and processed differently. It would also be possible to compare the reading behaviors of participants with similar Japanese skill to detect an unusual reading behavior. However, this would require to know the Japanese skill of the reader in advance. This solution cannot be used in the context of a test but can be considered in the context of a self-understanding quantification. Also, in the context of our study, recording only people with similar profiles might lead to similar subjective understanding and the algorithm would not generalize well. Therefore, we recruited participants from various nationalities and various Japanese reading abilities since this variation helps the algorithm to generalize with different readers' profiles.

Our study is limited to the analysis of the reader's subjective understanding. Since no questions have been asked about it, no conclusion can be drawn about the reading ease or reading fluency. Similarly, we use the comprehension questions as a point of comparison to evaluate our eye gaze based method but they are not used by the algorithm. Therefore, the objective understanding cannot be directly predicted from this method. The prediction of the reader's objective understanding with the eye gaze has been studied in the past and requires the extraction of different eye gaze features [30].

5 Conclusion

In this paper, we presented a method to estimate the reading subjective understanding based on the eye gaze. Our method is based on the analysis of average features to predict a subjective

understanding after reading several short texts. We showed that our method 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 even if we reduce the number of texts, analyzing the eye gaze is still more efficient than using the comprehension question. Moreover, the smaller the number of read documents, the greater is the difference in performance. However, there are still some errors in the subjective understanding estimation. This can be explained by the small size of the texts and by some specific reading behaviors such as repeated 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.

Nevertheless, our method can be used in a ubiquitous learning environment as an input so that the learning material is automatically adjusted to the user's subjective understanding. In the future, by using this understanding on texts of different levels, and by comparing with different users we plan to estimate the reader's confidence about his Japanese ability. Also, because our method is not only meant to estimate the reader's subjective understanding of only Japanese texts, we plan in the future to apply our method to other languages.

Acknowledgments

This work was supported in part by JST CREST Grant Number JPMJCR16E1, JSPS KAKENHI Grant Numbers 15K12712 and 16K16089, and Key Project Grant Program of Osaka Prefecture University.

Author Contributions

Formal analysis: Charles Lima Sanches.

Methodology: Charles Lima Sanches.

Software: Charles Lima Sanches.

Supervision: Olivier Augereau, Koichi Kise.

Writing – original draft: Charles Lima Sanches.

Writing – review & editing: Charles Lima Sanches.

References

1. Chen CC, Huang TC. Learning in a u-Museum: Developing a context-aware ubiquitous learning environment. *Computers & Education*. 2012; 59(3):873–883. <https://doi.org/10.1016/j.compedu.2012.04.003>
2. Jones V, Jo JH. Ubiquitous learning environment: An adaptive teaching system using ubiquitous technology. In: *Beyond the comfort zone: Proceedings of the 21st ASCILITE Conference*. vol. 468. Perth, Western Australia; 2004. p. 474.
3. Munoz-Cristobal JA, Jorin-Abellan IM, Asensio-Perez JI, Martinez-Mones A, Prieto LP, Dimitriadis Y. Supporting teacher orchestration in ubiquitous learning environments: a study in primary education. *IEEE Transactions on Learning Technologies*. 2015; 8(1):83–97. <https://doi.org/10.1109/TLT.2014.2370634>
4. Chougale P, Desai S. UBIQUITOUS LEARNING ENVIRONMENT. *International Journal of Emerging Trends and Technology in Computer Science*. 2016; 5(2):unknown.
5. Ogata H, Yano Y. How ubiquitous computing can support language learning. *Proceedings of KEST*. 2003; p. 1–6.
6. Liu TY, Chu YL. Using ubiquitous games in an English listening and speaking course: Impact on learning outcomes and motivation. *Computers & Education*. 2010; 55(2):630–643. <https://doi.org/10.1016/j.compedu.2010.02.023>

7. Krashen SD. The power of reading: Insights from the research: Insights from the research. ABC-CLIO; 2004.
8. Mason B, Krashen S. Extensive reading in English as a foreign language. *System*. 1997; 25(1):91–102. [https://doi.org/10.1016/S0346-251X\(96\)00063-2](https://doi.org/10.1016/S0346-251X(96)00063-2)
9. Woodford PE. The Test of English for International Communication (TOEIC). 1980;.
10. Bénabou R, Tirole J. Self-confidence and personal motivation. *The Quarterly Journal of Economics*. 2002; 117(3):871–915. <https://doi.org/10.1162/003355302760193913>
11. Christie H, Tett L, Cree VE, Hounsell J, McCune V. A real rollercoaster of confidence and emotions: Learning to be a university student. *Studies in Higher Education*. 2008; 33(5):567–581. <https://doi.org/10.1080/03075070802373040>
12. Rayner K. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*. 1998; 124(3):372. <https://doi.org/10.1037/0033-2909.124.3.372> PMID: 9849112
13. Biedert R, Dengel A, Elshamy M, Buscher G. Towards robust gaze-based objective quality measures for text. In: *Proceedings of the Symposium on Eye Tracking Research and Applications*. ACM; 2012. p. 201–204.
14. McConkie GW, Kerr PW, Reddix MD, Zola D. Eye movement control during reading: I. The location of initial eye fixations on words. *Vision research*. 1988; 28(10):1107–1118. [https://doi.org/10.1016/0042-6989\(88\)90137-X](https://doi.org/10.1016/0042-6989(88)90137-X) PMID: 3257013
15. Beymer D, Russell DM. WebGazeAnalyzer: a system for capturing and analyzing web reading behavior using eye gaze. In: *CHI'05 Extended Abstracts on Human Factors in Computing Systems*. ACM; 2005. p. 1913–1916.
16. Kudo R, Augereau O, Rou T, Kise K. Reading similarity measure based on comparison of fixation sequences. In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*. ACM; 2015. p. 1221–1226.
17. Biedert R, Buscher G, Dengel A. The eyebook—using eye tracking to enhance the reading experience. *Informatik-Spektrum*. 2010; 33(3):272–281. <https://doi.org/10.1007/s00287-009-0381-2>
18. Kunze K, Kawaichi H, Yoshimura K, Kise K. Towards inferring language expertise using eye tracking. In: *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. ACM; 2013. p. 217–222.
19. Yoshimura K, Kunze K, Kise K. 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*; 2015. p. 251–255.
20. Augereau O, Fujiyoshi H, Kise K. Towards an automated estimation of English skill via TOEIC score based on reading analysis. In: *Pattern Recognition (ICPR), 2016 23rd International Conference on*. IEEE; 2016. p. 1285–1290.
21. TOEIC. 2014 Report on Test Takers Worldwide: The TOEIC Listening and Reading Test; http://www.ets.org/s/toEIC/pdf/www_data_report_unlweb.pdf.
22. Karolus J, Wozniak PW, Chuang LL, Schmidt A. Robust Gaze Features for Enabling Language Proficiency Awareness. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM; 2017. p. 2998–3010.
23. Martínez-Gómez P, Aizawa A. Recognition of understanding level and language skill using measurements of reading behavior. In: *Proceedings of the 19th international conference on Intelligent User Interfaces*. ACM; 2014. p. 95–104.
24. Lima Sanches C, Augereau O, Kise K. Using the Eye Gaze to Predict Document Reading Subjective Understanding. In: *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. IEEE; 2017. p. 28–31.
25. Rayner K, Chace KH, Slattery TJ, Ashby J. Eye movements as reflections of comprehension processes in reading. *Scientific studies of reading*. 2006; 10(3):241–255. https://doi.org/10.1207/s1532799xssr1003_3
26. Gibaldi A, Vanegas M, Bex PJ, Maiello G. Evaluation of the Tobii EyeX Eye tracking controller and Matlab toolkit for research. *Behavior research methods*. 2017; 49(3):923–946. <https://doi.org/10.3758/s13428-016-0762-9> PMID: 27401169
27. Initiative OBL, et al. OECD guidelines on measuring subjective well-being; 2013.
28. Lima Sanches C, Augereau O, Kise K. Vertical error correction of eye trackers in nonrestrictive reading condition. *IPSJ Transactions on Computer Vision and Applications*. 2016; 8(1):7. <https://doi.org/10.1186/s41074-016-0008-x>

29. White SJ, Hirotani M, Liversedge SP. Eye movement behaviour during reading of Japanese sentences: Effects of word length and visual complexity. *Reading and Writing*. 2012; 25(5):981–1006. <https://doi.org/10.1007/s11145-010-9289-0>
30. Lima Sanches C, Augereau O, Kise K. Japanese Reading Objective Understanding Estimation by Eye Gaze Analysis. *Adjunct Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. 2017; p. to be published.