SubMe

An Interactive Subtitle System with English Skill Estimation Using Eye Tracking

Katsuya Fujii
The University of Tokyo
Bunkyo-ku, Tokyo, Japan
katsuya.f0515@gmail.com

Jun Rekimoto*
The University of Tokyo
Sony Computer Science Laboratories, Inc.
Bunkyo-ku, Tokyo, Japan
rekimoto@acm.org

ABSTRACT

Owing to the improvement in accuracy of eye tracking devices, eye gaze movements occurring while conducting tasks are now a part of physical activities that can be monitored just like other life-logging data. Analyzing eye gaze movement data to predict reading comprehension has been widely explored and researchers have proven the potential of utilizing computers to estimate the skills and expertise level of users in various categories, including language skills. However, though many researchers have worked specifically on written texts to improve the reading skills of users, little research has been conducted to analyze eye gaze movements in correlation to watching movies, a medium which is known to be a popular and successful method of studying English as it includes reading, listening, and even speaking, the later of which is attributed to language shadowing. In this research, we focus on movies with subtitles due to the fact that they are very useful in order to grasp what is occurring on screen, and therefore, overall understanding of the content. We realized that the viewers' eye gaze movements are distinct depending on their English level. After retrieving the viewers' eye gaze movement data, we implemented a machine learning algorithm to detect their English levels and created a smart subtitle system called SubMe. The goal of this research is to estimate English levels through tracking eye movement. This was conducted by allowing the users to view a movie with subtitles. Our aim is create a system that can give the user certain feedback that can help improve their English studying methods.

CCS CONCEPTS

• **Human-centered computing** → *Interactive systems and tools*; *Empirical studies in interaction design*;

KEYWORDS

Human Computer Interaction, Learning, User Interface

*This author is the one who did all the really hard work.

AH2019, March 11–12, 2019, Reims, France

ACM ISBN 978-1-4503-6547-5/19/03…$15.00
https://doi.org/10.1145/3311823.3311865

Figure 1: Watching a movie with subtitles is a popular way to learn English. In this paper, we introduce a system called SubMe, an Interactive subtitle system that detects English level through tracking eye movement and, depending on their english levels, it helps users to comprehend the content better by using interactive help.

ACM Reference Format:
Katsuya Fujii and Jun Rekimoto. 2019. SubMe: An Interactive Subtitle System with English Skill Estimation Using Eye Tracking. In *Augmented Human International Conference 2019 (AH2019), March 11–12, 2019, Reims, France.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3311823.3311865

1 INTRODUCTION

Many people studying English in the classroom, or through some other means, will frequently find themselves reciting language drills in an attempt to memorize newly acquired vocabulary. While this traditional method may be effective in some cases, hearing vocabulary used naturally in some form of conversation or narration with a visual aid can assist in its understanding and practical knowledge of that vocabulary's usage. Therefore, using movies or television series as a tool for studying language can be very beneficial.

Due to its relative ease of access and entertainment value, this has become a very popular and successful method of studying language, as it is not just limited to English [1, 2]. (Figure 1) Watching movies or series in a target language with subtitles in that language has become an ever expanding trend. In fact, it has been demonstrated that many language learners prefer to have subtitles in the original language as it is beneficial to their understanding, but also because of its originality [3]. The original content in its original form is something that many loyal fans are interested in watching. An example of this would be Japanese anime [4]. These fans develop the motivation to learn Japanese by the desire to understand the original

content. As this can lead to a style of active learning in which the viewer can gain quite an impressive amount of vocabulary while also enjoying some of their favorite content, there is a prime opportunity for development and advancement of this trend.

Although watching media in its original form may serve as a type of motivation, most people who try this method of watching movies or series with subtitles may find it takes too much time to learn new words. When they come across difficult vocabulary, or a new word which they would like to confirm, they must pause the video, check the definition, and then unpause. Typically, this becomes so troublesome that potential new words are skipped in favor of simply continuing the movie or series at the expense of total comprehension. In this way, what could have become an efficient learning experience is compromised. On the other hand, if people do actively pause the movie, they may not enjoy it as much, and as such, this may also dissuade them from continuing this beneficial practice, regardless of their motivation. Another possibility is that people prefer watching something on a TV rather than on a laptop or a smartphone. In this situation, it may not be as easy or simple to pause what the viewer is watching as it would be on some other device. If there was a way to passively watch a movie or series with feedback automatically given to the viewer afterwards, the problems above could be solved.

Researchers have explored eye gaze movement and its correlation to comprehension. One example of this was to determine English level. In a related experiment regarding reading, researchers could accurately estimate someone's English level just by measuring their eye gaze movement [5]. With regards to videos, computer technology can also use our eye gaze movements to record our progress in the same way [6]. However, since, the visuals on screen are always in motion, and the subtitles change at a certain speed, the dynamics are different. The participants in the original reading experiment were able to take their time reading and they had the freedom to return to parts of the texts they were not sure about. As stated above, those using videos to study English may encounter problems. Therefore, other solutions or algorithms are necessary in order to create a system which could be beneficial in the same way or more as the reading system. If this method can be harnessed, customized options that target the individual's needs can therefore be developed and create a possibility for deeper understanding even with moving screens and subtitles. For our research, we want to take advantage of these modern movie trends as well as the potential of eye gaze movement technology to make a system related to language learning that would not be a hassle for the viewer. Therefore, we seek to create an effective and entertaining system that can estimate English levels through tracking eye gaze movements using a movie with English subtitles. In addition, we aim to give the viewer certain feedback that will be reflective of how well they are able to follow the movie and subtitles, and in turn, how well they are able to comprehend the content. This feedback includes information on how they can improve their English studying through different optional methods recommended by our application. Some of these methods are slowing down the speed of the movie, highlighting difficult words, or giving the user a list of words that they possibly did not understand. The system can learn over time about the viewer and estimate which words will be difficult or that the user will not understand.

2 RELATED WORK

2.1 Estimation of English Skills by Analyzing Eye Movement

Researchers have explored the correlation between reading and eye movements.Fixations and saccades are the elements which compose eye gaze movement in regards to reading. Fixations are the points at which people reading pause in order to absorb information. Saccades are the lengths in between these fixations [7]. Buscher implemented an algorithm which computes the fixations and saccades from the collected eye gaze movement data [8]. Related research has already proved the usefulness of utilizing eye gaze movements for monitoring physical activities such as distinguishing reading from not reading [9, 10].

Analyzing these features enables us to retrieve characteristics of the users' eye gaze movements that can potentially be utilized to estimate English skill [11, 12]. Researchers et al has widely explored methods to recognize reading activities and one of their findings was that the average number of fixations and the standard deviation of the number of fixations differs depending on the reader's skill. Using this factor, they proposed a method to detect the English skill level of a user and infer which words are difficult for them to understand and also applied their method for estimating their TOEIC (Test of English for International Communication) levels [13–16]. In 2013, Copeland et al. took advantage of machine learning methods to estimate the reading comprehension by tracking eye gaze movements [17].

2.2 Smart Subtitle

Even though subtitles are helpful to learn a language, they can be optimized for skill acquisition. Conventionally, when users find an unknown word, they stop a movie, look it up in a dictionary, go back, and then rewind the movie or series as needed. This is time consuming and sometimes prevents them from purely enjoying the content. Attempts to tackle this problem have been made in a variety of ways including through research and even commercial products [18, 19]. Kovacs presented Smart Subtitles, a system of interactive subtitles tailored towards vocabulary acquisition [20, 21]. It provides features such as exposing vocabulary definitions upon hovering over an unknown word and dialog-based video navigation. Ma proposed InteractiveSubtitles, a video language learning method specifically focusing on UI design by comparing multiple ways of showing subtitles with a method of using pop-ups to show the meanings of keywords [22]. Kurzhals payed attention to the positions of subtitles and claimed that traditionally displayed text centered at the bottom of the screen can cause eye strain because of the large distances between the text and the relevant image content. They proposed a new approach called speaker-following subtitles, which results in a better distribution of focus and shorted saccades for a less exhausting viewing experience [23].

As the researchers above demonstrated, analyzing eye gaze movement can be used to estimate English level. We aim to create a proof-of-concept prototype to estimate a viewer's English level by tracking their eye gaze movement using a movie with subtitles. In addition, we will provide certain feedback so as to provide them with certain methods of improving their study habits.

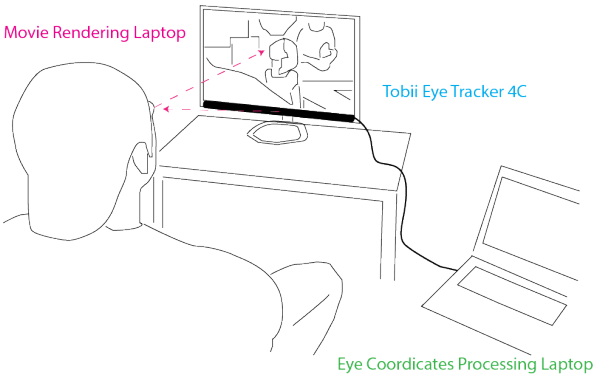


Figure 2: The eye tracking devise captures X and Y coordinates with timestamps. From there, we sent the data to another laptop that renders movies.

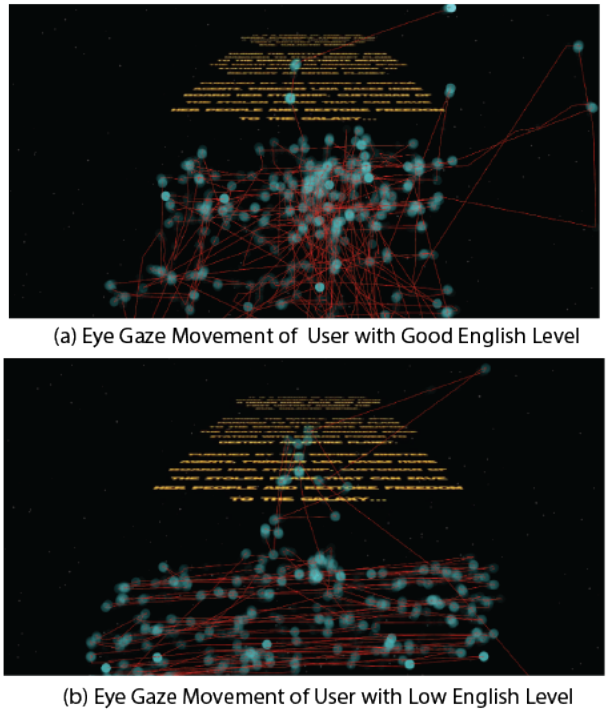
3 IMPLEMENTATION

3.1 Exploratory Findings

The idea of this research arose during the process of recording viewers' eye gaze movements while they were watching the introduction of Star Wars: Episode IV - A New Hope. In this sequence, sentences scroll up from the bottom of the screen towards the top. Many non-native level speakers have difficulty reading these sentences since they continue scrolling up and there may not be enough time to read or absorb everything they are reading. We recorded eye movements of a native level English speaker, a participant with a fairly fluent grasp of English, and an English learner. Figure 3 shows the visualization of fixations and saccades of their eye gaze locations on the screen at certain points. As can be seen, the English learner tended to look up. Reasons for this could include that they did not have enough time to read everything, it took more time for them to understand the information being shown, or they had to read it multiple times. The native speaker, on the other hand, tended to focus at the bottom of a screen most of the time. By comparing these tendencies, we found it clear that we could distinguish English level simply by analyzing the number of vertical saccades. This illustrates the concept of using eye gaze movement and that by making it interactive, we could assist in helping them understand viewers' English levels more efficiently.

3.2 System Configuration

To detect the user's eye movement, we adopted the Tobii Eye Tracker 4C, which we attached to the bottom of an external desktop monitor with the spec of Core i7 CPU and 64GB RAM. The device captures X and Y coordinates with timestamps. From there, we sent the data to another laptop that renders movies. (Figure 2) The Tobii Eye Tracker is capable of capturing coordinates at 90 Hz, however, due to the rendering frame rate, we recorded at 60Hz, a level which is equal or greater to those used in related work. The eye tracker was calibrated once before starting the experiment for each participant. We used OpenFrameWorks to render images [24].



(a) Eye Gaze Movement of User with Good English Level

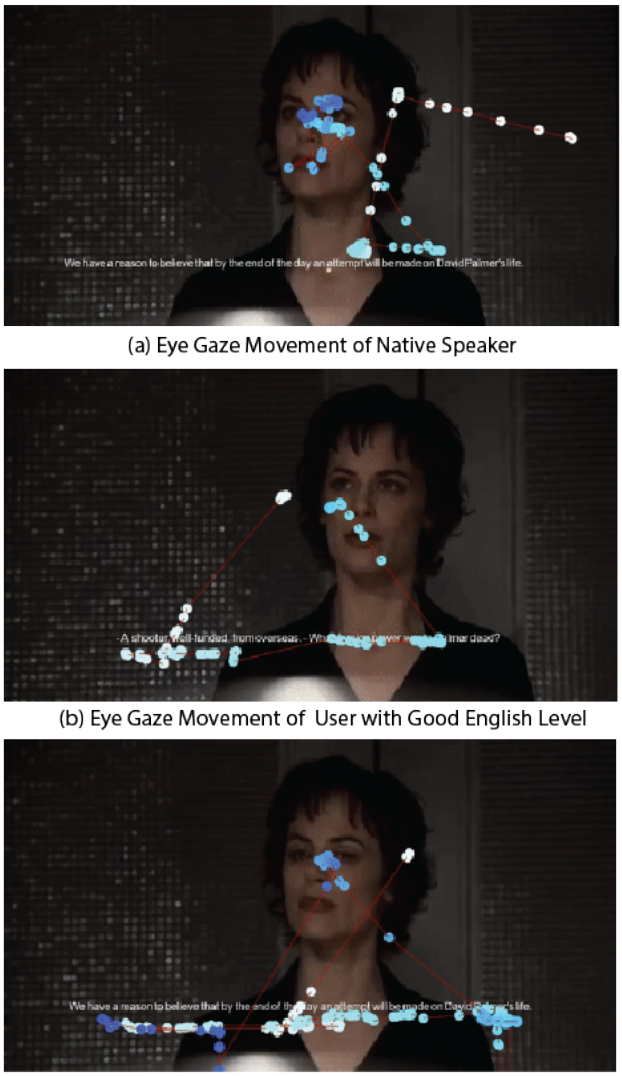
(b) Eye Gaze Movement of User with Low English Level

Figure 3: Visualization of fixations and saccades of users eye gaze movements while watching the introduction of Star Wars. Users with low level of English users tend to have more vertical saccades than the ones with better level.

3.3 Analysis of Eye Movements

After these findings, we asked 15 people to watch a movie with subtitles. Users were asked to touch any key on the keyboard when they did not understand a part of the subtitles after the subtitles had finished or had shifted to the next one. This was because we needed to capture their whole eye gaze movements for that line of subtitles. After carefully explaining about the procedure and giving them a few trials, the users started the experiments. As the movie contents were rather long, we also told them that they could stop the experiments at anytime if they became bored or too tired. For the movie, we picked episodes of the TV series "24" and "Friends", as these two are some the most popular series used for helping people learn English.

Figure 4 shows a part of the movie with the visualization of the fixations to saccades. For written text, the time of fixations can help us to find a word that users cannot understand, but in movies or series, as the screen and lines of subtitles keep moving, we found out that people do not stop at each word that they do not understand. No matter what their English level may be, they do not stop. There are a number of factors as to why English learners would not direct their visual attention at a word they do not know. One reason is of course that the scene continues and they just want to watch the video. Another is that if they stop, they may not understand the overall meaning and so stopping would be more detrimental to



(a) Eye Gaze Movement of Native Speaker

(b) Eye Gaze Movement of User with Good English Level

(c) Eye Gaze Movement of User with Low English Level

Figure 4: Visualization of fixations and saccades of users eye gaze movements while watching an American series 24. The way users look at subtitles differs depending on their English levels. The lower their English level is, the longer they spend on subtitles

their study. One more possibility is that the audio may have lead them to understand the word or phrase they were having trouble with, regardless of the subtitles. Therefore, only using fixation data would not be sufficient.

The bar graphs in figure 5 show the number of "total fixations" and "total fixation on subtitles" per line of subtitles. The first bar indicates how much in total the system detected users fixations, whereas the second bar indicates how long the users spent looking at the line of subtitles. The graph on the top represents subtitles that a user understood while watching the video, and the graph at the bottom is for subtitles that the same user could not understand. As you can see, the user tended to spend more time looking at subtitles when they could not understand them. As mentioned above, we can not use a number of fixations per word while watching a video to detect a user's English level, but we can take advantage of the number of fixations on subtitles in relation to the number of total fixations.

4 RANDOM FOREST

4.1 Parameters

Using the results of this experiment, we used RANDOM FOREST[26] in regards to the parameters that we explained above as input parameters: "total fixations" and "total fixation on subtitles". We also took into account some additional factors.

(1) Total Saccades
Saccades are eye gaze movement shifts from one fixation to another. In the experiment, we found out that users tend to look at people's faces in the movie. When several people appear in one scene, their eye gaze movements go back and forth between each person's face and the subtitles. We could potentially take advantage of saccade parameters assuming the lower the saccades are, the more time they spend time on only reading the subtitles.

(2) Subtitle Length
Even if every word in a line subtitle is relatively easy, there are cases in which users would not understand because the sentence is too long. Unlike reading written texts, subtitles change at a certain speed, and users can not go back after the screen has transitioned. Therefore, this should also be considered as one of the input parameters.

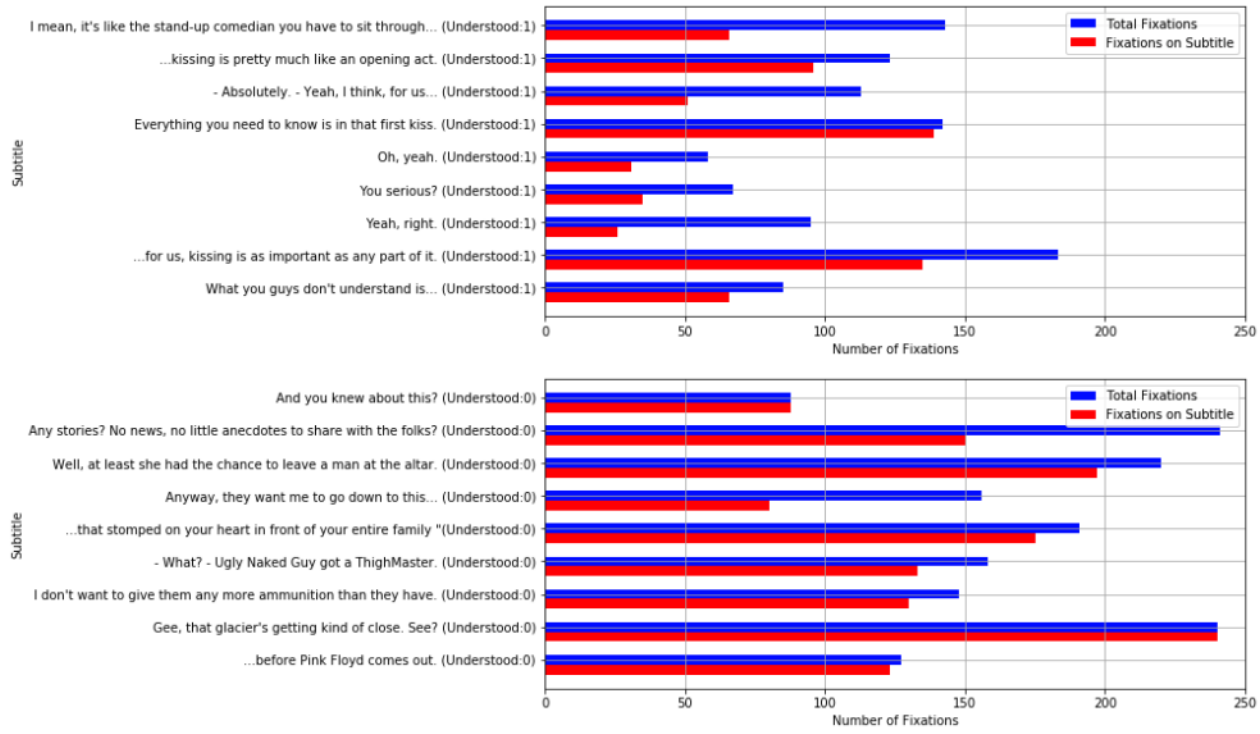
(3) Window Size
A window size is a time in milliseconds while a line of subtitles is displayed. This corresponds to the speaking speed of an actor/actress. When a window size is small, but the length of the line of subtitles is long, users most likely find it difficult to understand or catch.

(4) Subtitle Level
It goes without saying that, the more difficult the word is, the harder it becomes to understand it. The formula for calculating the difficulty of words is explained in the application section below.

Teacher Data is a binary number which represents *understood/not-understood*. Table 1 shows the list of the whole parameters.

4.2 Learning Process

We first split the dataset into 80% samples for training data and 20% samples for test data. We then conducted under-sampling to fill the gap of unbalanced data. (The number of *understood* was more than that of *not_understood*). For tuning hyperparameters, we used the Randomized Search Algorithm to investigate the best performances under this condition. Figure 2 shows the classification report of the model. We put emphasis on "finding" *not_understand* rather than "missing" *not_understand*. Therefore, we focused more on the recall rate than the precision rate or accuracy rate. After creating a model, we extracted some feature importance for the algorithm




```
User4 has 168 understood /9 not understood
Percentage_of_fixations_on_subtile  61.0% understood /85.0% not understood
```

Figure 5: The number of total fixations and total fixations on subtitles. The graph on the top represents subtitles that a user understood while watching the video, and the graph at the bottom is for subtitles that the same user could not understand. The user tended to spend more time looking at subtitles when they could not understand them.

Table 1: Classification Report. The recall rate was focused more because we put emphasis on finding not understand rather than missing not understand.

	Precision	Recall	F1-Score
Weighted Avg	0.82	0.79	0.80

and found out the algorithm clarifies data focusing more on the ratio of the fixations The algorithm seems to classify data focusing on the ratio of the fixations. The model created was able to classify data with the accuracy rate of 80.0%. This model can be improved by obtaining more data, changing the weights and parameters, or by using a different machine learning algorithm.

5 APPLICATION

We have an algorithm to briefly detect a user's English level. We created a system called SubMe, an Interactive subtitle system that detects English level and, depending on their english levels, it helps users to comprehend the content better by using interactive help.

5.1 Word Suggestion

As a proof of concept prototype, we used a pop-up style user interface. The pop-up words become optimized based on the English level of the user. After detecting their English level, we made personalized word lists for their next video. The users were to watch this in following experiments. We extracted words used in the subtitles beforehand and sorted them into difficulty levels.

The word suggestion was conducted in the following way.

(1) Pre-Processing for Each Word
We first calculated a difficulty level for each word. In order to do so, we used the frequency of the words provided by The Corpus of Contemporary American English (COCA) [25]. We assumed that the the less frequently seen words are, the more difficult they are. The frequency rates are mapped in a linear conversion to score points. Table 2 is a part of the score list for words used in the episode of "Friends"we used.
(2) Calculating User English Level
After our Random Forest outputs subtitles that a user possibly did not understand, the system splits the words for each line of subtitles and then takes an average score for all the subtitles possibly not understood by the user.
(3) Extracting New Words

subtitle	subtitle_level	understood	window_size	ratio_on_fixations_subtitle	subtitle_length	fixations_on_subtitle	saccades	fixations
You know, as someone who's recently been dumped.	446.000000	0	3086.0	100.000000	9	185.0	1179.124381	185.0
Well, you may want to steer clear of the word ...	441.000000	0	3461.0	86.956522	12	180.0	2881.764899	207.0
Chances are he's going to be this broken shell...	192.000000	0	4211.0	100.000000	15	252.0	4617.671671	252.0
So you should try not to look too terrific. I ...	226.000000	1	4045.0	100.000000	15	243.0	2123.839550	243.0
Or, you know, hey, I'll go down there, and I'l...	112.000000	1	4379.0	86.311787	16	227.0	3404.110143	263.0
And you can go with Carol and Susan to the OB-...	1226.000000	1	4337.0	77.011494	12	201.0	2354.394696	261.0
You've got Carol tomorrow.	1226.000000	1	2460.0	88.435374	5	130.0	1600.508460	147.0
When did it get so complicated?	703.000000	1	1292.0	100.000000	7	118.0	813.853840	118.0

Figure 6: The list of input parameters to put for Random Forest Machine Learning Algorithm

Table 2: The list of Words with the frequency of the words provided by The Corpus of Contemporary American English (COCA) and converted scores

Word	Frequency	Score
mastodon	99	2314
clang	156	1547
stomp	240	1200
erect	401	778
nausea	1949	541
paranoid	2173	79

The higher the value of 2), the more words a user knows. Thus, we used the value as a threshold to filter new words from the list of words. We show these words with dictionary definitions while watching a movie. In this way users can theoretically learn new words through a dynamic method.

5.2 User Interface

As a proof of concept prototype, we developed two types of user interface. (Figure 7) At first, we adopted a pop-up style user interface, however, some of the users found it hard to absorb the information because of the increase in eye gaze movements shift. They need to watch the scene and subtitles, and adding the pop-ups was seemingly too much information got them to deal with at one time. Given this feedback, we implemented a dual line style user interface with a combination of movie pausing. We showed the definition of the words on top of the subtitle to decrease the additional eye gaze movements shift, and the movie pauses for a certain number of seconds so that the users have enough time to process the information. At any moment, the users can tap the keyboard if they do not want to pause the movie.

(a) Pop-Up Style User Interface

(b) Dual-Line Style User Interface

Figure 7: User Interface of Application. a) a pop-up style user interface b) a dual line style user interface with the combination of movie pausing. We adopted b) because some users found a) hard to catch up with the information because of the increase of eye movements shift.

6 APPLICATION EXPERIMENT

6.1 Procedure

To evaluate the effectiveness of SubMe, we instructed our five participants to watch the first seven minutes of two more different

episodes of "Friends". In random order, one episode was viewed with our system, and the other was viewed with traditional subtitles Beforehand, we explained the guidelines of our experiment to these five participants. The participants consist of 3 male students, 1 male and 1 female adults, and their mother tongue is Japanese. Their TOEIC scores were 830, 830, 830, 850, 850 respectively. A small test, based on the words shown in the subtitles would be conducted afterwards. Participant were told that they should take their time and try to learn as many words as they could. They were then instructed on how to rewind and pause the video so as they may repeat any difficult parts as much as necessary.

As a condition for the small test, we ask the participants to inform us as to whether they had already known the words that they came across in the test before they had watched the movie. This self-reporting mechanism is commonly used in vocabulary-learning evaluations for foreign-language learning [27]. After it was completed, the participants were given a survey based on how effective they consider this activity to be. Particularly, we wanted to know at what level of difficulty they found studying with this tool was, if the video was understandable, if it was entertaining, and a key factor in whether this method can be successful or not. The questionnaire was rated on a 5-point Likert scale. The following list is actual questions.

(1) Q1: Did you find using this tool efficient to learn English?
(2) Q2: Was it easier to comprehend the contents of the movie using this tool?
(3) Q3: Could you enjoy watching the movie more when using this tool?
(4) Q4: Was the level of the word suggestion appropriate for your English level?

In addition to these questions, we also inquired as to if they had any additional comments or not.

6.2 RESULTS

Figure 8 shows the results of the experiment. The participants could define the words significantly more when using our system. (t=3.7,p=0.02) The system also helped them learn new words more. (t=8.5, p=0.001). When SubMe was not used, the participants showed either of two types of behavior when looking up words: they wrote down all unknown words and looked them all up after the movie was finished, or they paused the movie and looked them up one by one. In both ways, they seemed to be looking up words out of context of the movie and picked up irrelevant definitions for some words. For example, in the movie, *senior* was used as in the 4th grade of college, but in the test some defined it as equivalent to *older*. Whereas with SubMe, the users could learn the word with the context , thus it resulted in better test scores.

Table 3 shows the results of the surveys. All the participants thought it helpful and they also indicated that they could enjoy the movie without being bothered to check the words in a dictionary. One participant, however, mentioned the level of the word suggestions was not appropriate. The improvement of English level estimation might solve this problem, but we also need to take into account other factors. During the experiment, this participant seemed not understand the contents of the movie because of a lack of vocabulary, expressions, or grammar. Only assisting with words might

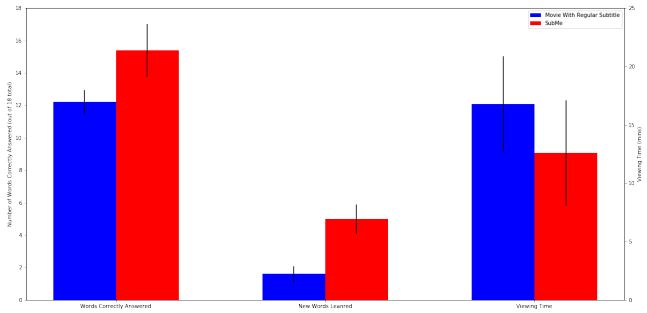


Figure 8: In the experiment we conducted, SubMe not only encouraged users to learn new vocabulary, it also helped them enjoy the movie itself. The users spent less time to remember words when using SubMe.

Table 3: Questionnaire. Q1: Did you find using this tool efficient to learn English? Q2: Was it easier to comprehend the contents of the movie using this tool ? Q3: Could you enjoy watching the movie more when using this tool? Q4:Was the level of the word suggestions appropriate for your English level? (5: Strongly Agree - 1: Strongly Disagree)

	User1	User2	User3	User4	User5
Q1	5	4	4	4	4
Q2	5	4	5	5	4
Q3	5	4	3	3	3
Q4	4	1	4	4	3

not be enough for users with a certain English level. Further research should be explored in order to discover what kind of aid is appropriate for users with their various English levels taken into consideration.

The followings are the summary of the additional comments.

- It was a bit frustrating that the movie paused when showing the words that I already knew
- User Interface can be improved. Speed Control might be ideal.
- First, I was worried that pausing a movie could be annoying, but considering that it is an English learning tool, I think this behavior is helpful and can allow us to enjoy the movie more by assisting with unknown words.
  We would like to take into account these opinions and improve the system better for the further exploration.

7 DISCUSSION AND FUTURE WORK

7.1 UI/UX Design

This time, we adopted a dual line style user interface, but we also came up with several other aids for English learning. One of these was speed-control. Sugai and others conducted research to clarify the effect of speaking rate and pause duration on listening comprehension [31]. They found that a pause (450ms) inserted in a passage provides listeners with additional information processing

time and enhances the comprehensibility of the aural input. As a movie contains audio and visual sources, we need to explore a feasible way to utilize this technique for movies. This time, we designed it so that if the system detects that a user is not able to understand a piece of content, it can slow down the video. This is less invasive and may be able to give them time to make a guess as to what unknown words may mean. Also it would be helpful for improving their listening skills.

7.2 Words vs Expressions

In this research, we suggested unknown words based on the frequency rate of the Corpus vocabulary list. The suggestion accuracy can be explored since the difficulty of a word might differ depending on its context and also there are several existing indexes which define a word's difficulty level [29, 30]. Additionally, it goes without saying that English has a large variety of expressions. Even combinations of simple words can shift the meaning of a phrase to something completely different. Variations of English dialects, such as British versus American, as well as the use of proper nouns and pop-culture references can also greatly influence the results. With written text, many times, there are anecdotes which explain content which may not be so standard or common knowledge. As such, the misunderstanding of pop-culture references for unknown words or phrases should not be indicative of a person's lack of English skills. We need to implement a more sophisticated tool that covers words and expressions by taking into account context.

7.3 Scalability

This time we used the desktop-attachable eye tracker, which detects near-infrared spectrum light, and also benefits from active illumination in that spectrum. The same company also produces a wearable version of it, but there are those who claim that it is not affordable. Also, with the improvement of internet speed, many service companies have begun offering online streaming movie platforms available for smartphones or tablets. Given this background, researchers have been exploring a method to detect eye movement using a webcam [28]. If we can find a way to detect eye movement by using a smartphone's built-in camera, we can collect more data sustainably and provide a more personalized interactive system for English learning.

8 CONCLUSION

This experiment confirmed many assumptions we had already made, while also proving that this popular trend of using movies to study English can in fact be harnessed as an efficient language learning tool. Implementing SubMe, an Interactive subtitle system that detects English levels through tracking eye gaze movement, we found that this tool can help users to comprehend content better by utilizing interactive help. With this, we hope to have further opened up this field of study, which began with monitoring the eye gaze movements of readers, to a much broader range. Due to the advent of life-logging, the accuracy of data collection, and the inevitable lowering of costs, customized English study applications tailored to the user's specific needs will be the future of education. In the world of machine learning, where the prospect of language barriers seems to all but be disappearing, there needs to be tool that can help maintain the motivation of those who wish to actually study and become proficient in other languages. Using machine learning and our current trends, SubMe can assist in promoting this. In the experiment we conducted, SubMe not only encouraged users to learn new vocabulary, it also helped them enjoy the movie itself. As for now, we are focused on English, but there is definitely the possibility that it could be used for multiple other languages, including, for instance, Japanese.

REFERENCES

[1] d'Ydewalle, Gery*Foreign-Language Acquisition by Watching Subtitled Television Programs*Journal of Foreign Language Education and Research. 12.

[2] Secules, T., Herron, C., and Tomasello, M*The effect of video context on foreign language learning*. The Modern Language Journal 76, 4 (1992), 480-490.

[3] MARIE-JOSEE BISSON, WALTER J. B. VAN HEUVEN, KATHY CONKLIN, RICHARD J. TUNNEY *Processing of native and foreign language subtitles in films: An eye tracking study* Applied Psycholinguistics 35 (2014), 399/418 doi:10.1017/S0142716412000434

[4] Fukunaga, N*Those anime students: Foreign language literacy development through Japanese popular culture.* Journal of Adolescent & Adult Literacy 50, 3 (2006), 206-222.

[5] Conklin, K., Pellicer-Sanchez, A. (2016). *Using eye-tracking in applied linguistics and second language research.* Second Language Research, 32(3), 453-467.

[6] Montero Perez, M., Peters, E., Desmet, P. (2015). *Enhancing Vocabulary Learning Through Captioned Video: An Eye?Tracking Study.* The Modern Language Journal, 99(2), 308-328.

[7] Rayner, Keith*Eye movements in reading and information processing: 20 years of research.*Psychological Bulletin, Vol 124(3), Nov 1998, 372-422

[8] Georg Buscher and Andreas Dengel*Gaze-based filtering of relevant document segments. In International World Wide Web Conference (WWW).* 2019, 20-24.

[9] Bulling, A., Ward, J. A., Gellersen, H., and Troster,G*Robust recognition of reading activity in transitusing wearable electrooculography.* In Proc. ofPervasive 08,19-37

[10] Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A.,and Kliegl, R. *Coregistration of eye movements andeeg in natural reading: analyses and review.* Journalof Experimental Psychology: General 140, 4 (2011),552

[11] Kazuyo Yoshimura, Koichi Kise, Kai Kunze *The eye as the window of the language ability: Estimation of English skills by analyzing eye movement while reading documents* 2015 13th International Conference on Document Analysis and Recognition (ICDAR),23-26 Aug. 2015, Electronic ISBN: 978-1-4799-1805-8

[12] Yevgeni Berzak, Boris Katz, Roger Levy *Assessing Language Proficiency from Eye Movements in Reading* Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1,Pages 1986-1996. 10.18653/v1/N18-1180.

[13] Olivier Augereau, Kai Kunze, Hiroki Fujiyoshi, Koichi Kise *Estimation of english skill with a mobile eye tracker* UbiComp '16 Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, Pages 1777-1781,Heidelberg, Germany September 12 - 16, 2016, ISBN: 978-1-4503-4462-3

[14] Kunze, Kai and Kawaichi, Hitoshi and Yoshimura, Kazuyo and Kise, Koichi *Towards inferring language expertise using eye tracking* CHI'13 Extended Abstracts on Human Factors in Computing Systems, Pages 217-222. 10.1145/2468356.2468396.

[15] Kiyomi Chujo and Kathryn Oghigian.*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.

[16] Pascual Mart nez-Gomez 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. ACM, 95-104.

[17] Copeland, Leana and Gedeon, Tom and Mendis, Balapuwaduge *Predicting reading comprehension scores from eye movements using artificial neural networks and fuzzy output error* Artificial Intelligence Research. 3, 2014, 10.5430/air.v3n3p35. 8

[18] S. Wang and L. Wang *Movie dictionary: A multimedia lexicon tool for language learning*IEEE International Conference on Computer Science and Information Technology, 2010, pp.484-487.

[19] Y. Zhu et al. *ViVo: Video-Augmented Dictionary for Vocabulary Learning* CHI Conference on Human Factors in Computing Systems, 2017, pp.5568-5579.

[20] Geza Kovacs *Smart subtitles for language learning* CHI EA '13 CHI '13 Extended Abstracts on Human Factors in Computing Systems, Paris, France April 27 - May 02, 2013, ISBN: 978-1-4503-1952-2

[21] Geza Kovacs, Robert C. Miller *Smart subtitles for vocabulary learning* CHI '14 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Pages 853-862, Toronto, Ontario, Canada April 26 - May 01, 2014, ISBN: 978-1-4503-2473-1

[22] Kuno Kurzhals, Emine Cetinkaya, Yongtao Hu, Wenping Wang, Daniel Weiskopf *Close to the Action: Eye-Tracking Evaluation of Speaker-Following Subtitles* CHI

'17 Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Pages 6559-6568, Denver, Colorado, USA May 06 - 11, 2017, ISBN: 978-1-4503-4655-9

[23] Qikun Ma, Shiyang Wang, Jie Liu, Nianlong Li *InteractiveSubtitle: Subtitle Interaction for Language Learning* ChineseCHI '18 Proceedings of the Sixth International Symposium of Chinese CHI, Pages 116-119 ,Montreal, QC, Canada April 21 - 22, 2018, ISBN: 978-1-4503-6508-6

[24] openframeworks. http://www.openframeworks.cc/.

[25] The Corpus of Contemporary American English (COCA). https://corpus.byu.edu/coca/

[26] Prinzie, A., Van den Poel, D. *Random Forests for multiclass classification: Random MultiNomial Logit* Expert Systems with Applications 2008, 34(3), 1721-1732.

[27] Wesche, M., and Paribakht, T. S.*Assessing Second Language Vocabulary Knowledge: Depth Versus Breadth*Canadian Modern Language Review 53, 1 (1996), 13-40.

[28] Alexandra Papoutsaki*Scalable Webcam Eye Tracking by Learning from User Interactions*CHI EA '15 Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, Pages 219-222, Seoul, Republic of Korea April 18 - 23, 2015, ISBN: 978-1-4503-3146-3

[29] Flesch R *A new readability yardstick* Journal of Applied Psychology. 32: 221-233. doi:10.1037/h0057532

[30] Dale E, Chall J*A Formula for Predicting Readability*Educational Research Bulletin. 27: 11-20+28.

[31] Sugai, K., Yamane, S., & Kanzaki, K. *The Time Domain Factors Affecting EFL Learners Listening Comprehension: a study on Japanese EFL Learners* Annual Review of English Language Education in Japan, 27,2016, 97-108