



I am Smartglasses, and I Can Assist Your Reading

Baojie Yuan, Yetong Han, Jialu Dai, Yongpan Zou^(✉), Ye Liu,
and Kaishun Wu

College of Computer Science and Software Engineering, Shenzhen University,
Shenzhen, China

{yuanbaojie2018,hanyetong2017}@email.szu.edu.cn,
{yongpan,ly,wu}@szu.edu.cn

Abstract. Human reading states reflect people’s mental activities and are closely related with learning behaviors. Consequently, it is of great value to identify reading states in many aspects. For example, it can provide timely assistance in learning processes and boost the learning efficiency of learners in a personalized way. Conventionally, researchers usually make use of EEG and fMRI to recognize human reading states. However, these methods have shortcomings of high device cost, low portability and bad user experience. In this paper, we design a real-time reading states detection system named ETist with commodity wearable glasses that can identify four reading states including attention, browsing, mind wandering and drowsiness via tracking eye movement. Through our experiments, we demonstrate that this system can recognize four fine-grained states with an average accuracy of 84.0%, which is applicable in a wide area of applications.

Keywords: Eye tracking · Reading states · Wearable devices

1 Introduction

Reading is a mainstream vision-based information acquisition channel and shows increasing significance due to the rapid update of knowledge. However, reading states show vital impact on the reading efficiency [1]. It is obvious that people needs to spend less time in understanding materials when they are concentrated [2]. On the contrary, the reading efficiency decreases vastly when a reader is distracted by other disturbers [3]. Motivated by such common cases, we ask such a question *What if we can identify a reader’s reading states?* Theoretically, we envision that identifying a reader’s reading states brings about several potential benefits as follows. On one hand, this information is helpful for readers to adjust themselves into the best state timely when they are reminded about being distracted. On the other hand, readers can get timely assistance when they are stuck in difficulties. For instance, for educators, knowing students’ learning states provides guidance of improving teaching methods and optimizing contents organization.

Researchers in psychology and cognitive science have conducted some research work related to recognizing reading states. In general, previous work around this topic can be categorized into three different kinds by supporting technologies, namely, electroencephalogram (EEG)-based [4, 5], functional magnetic resonance imaging (fMRI)-based [6–8] and eye tracking based [9–15]. EEG and fMRI-based approaches detect reading states by monitoring electrophysiological signals varying with brain activities. However, both of them have certain limitations. They require special devices which are too bulky and expensive to carry. The another kind of technique we utilize to recognize cognitive states of users to provide learning strategies and the like is eye-tracking. The rational behind this method is that different reading states induce unique eye movement patterns and allows for discrimination between reading states by analyzing data [16]. Nevertheless, researches before only focus on studying a certain specific state and its impact on reading, instead of identifying different states and providing corresponding assistance. In this paper, we propose a real-time reading cognitive detection system called ETist with commodity wearable glasses that can recognize four fine-grained reading states via tracking eye movements, namely *attention*, *browsing*, *mind wandering* and *drowsiness*. We design ETist system with a sequence of functional blocks that takes eye movements data as input and outputs human reading states. Each part of ETist system is carefully designed for achieving high performance and better user experience. We have also carefully designed experiments to evaluate the performance of ETist under different settings and the final results show that our system can achieve an average accuracy of 84.0% in recognizing four reading states in common reading scenarios. Finally, we also introduce a case study with ETist system to demonstrate that this system can provide appropriate support to readers based on the recognition results of reading states.

Contributions: In summary, our main contributions can be concluded as follows:

- To the best of our knowledge, this is the first attempt to detect fine-grained cognitive reading states and build a reading assistant system which is convenient for users in real time only using eye-tracking glasses.
- In order to identify the user’s reading mental state according to the reading material in real time, we propose line and page segmenting algorithms to segment eye-tracking data according to the reading page.
- By combining with reading states recognition and POIs identification, we build a prototype system and conducts case-study experiments to evaluate its performance. The results show that, compared with traditional methods, our system can effectively assist a user’s reading process.

2 Related Work

2.1 EEG-Based/FMIR-Based Cognitive State Recognition

As EEG signal could reflect the electrical activity of scalp, it can detect the human brain activities. EEG is commonly used to detect the cognitive state.

Berka *et al.* [4] found that the EEG signal contains some cognitive state information such as task engagement level and mental workload. Lan *et al.* [5] propose a classification system by using machine learning approaches to estimate the reading states. However, in order to use EEG to detect the cognitive state, it necessitates to stick the electrode slices on user's head or wear a wireless sensor headset. Compared with the eye tracking method, EEG is not portable and comfortable for user to use. fMRI can also detect the cognitive state by measuring the brain activities. Mather *et al.* [6] explain how fMRI reveals reading states of humans. Wang *et al.* [7] detect cognitive state by using fMRI classifiers. However, this technology can't be widely deployed because of the highly price of the fMRI device. Compared with fMRI, eye tracking prevails by being portable, affordable and comfortable to use.

2.2 Eye Tracking-Based Reading States Recognition

Eye tracking has been intensively used by psychologists to study basic cognitive processes during reading and other types of information processing such as pattern of information processing [10,13], effects of instructional strategies [14], effects of learning strategies [15]. Particularly, some researches use fusing-data for attention detection when reading or other information processing activities. For example, Li *et al.* [11] use eye tracking in conjunction with webcam and mouse dynamics to identify cognitive states. Bixler *et al.* [12] fuse eye gaze data and physiology data to achieve detecting of mind wandering when reading. There are some researches using eye-tracking techniques to detect driver cognitive distraction [17] and drowsiness [18] in real time. However, those works with fusing data need additional hardware, which brings burden to a user and causes other problems. and most researches pay more attention on detecting single state like mind wandering rather than recognizing multiple states. To the best of our knowledge, this is the first work to detect fine-grained cognitive reading states in real time only with low-cost eye-tracking glasses.

3 System Overview and Methodology

3.1 Concepts of Reading States

As aforementioned, we focus on studying four reading cognitive states, namely, attention, browsing, mind wandering and drowsiness. There are several reasons behind this. First, studies have shown that attention and mind wandering are two basic mental states in reading tasks [19]. Mind wandering usually happens when attention involuntarily shifts from reading to other unrelated thoughts, which has been proved to have negative effect on reading comprehension [2]. This suggests that it's possible to improve reading efficiency by attempting to hinder mind wandering in time and reduce its negative effects. Therefore, it is necessary to detect these two states for the purpose of helping people better focus on the task at hand. Second, for a fine-grained identification of reading states, we further divide the two basic states into four fine-grained states. In the following, we shall clarify them based on psychological literatures.

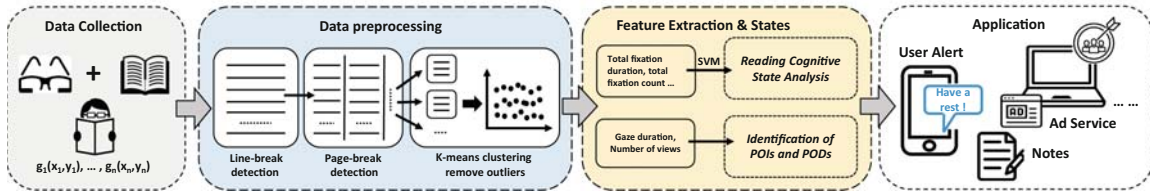


Fig. 1. An overview of the system architecture of ETist

- **Attention:** it is a kind of behavioral and cognitive process during which people concentrate on a discrete aspect of information while ignoring other perceivable information [20]. In our work, it corresponds to a state of focused reading.
- **Browsing:** it is a quick examination of the relevance of a number of objects which may or may not lead to a closer examination or acquisition/selection of (some of) these objects [21]. In this work, it corresponds to the state of fast and fluent reading.
- **Mind wandering:** It's a kind of involuntarily shift from thoughts related about reading to other unrelated thoughts. Mind wandering often causes poor comprehension during reading [2].
- **Drowsiness:** it is a state of strong desire for sleep [22], which can make people stop reading. It can be defined as high-level mind wandering.

3.2 System Overview

Figure 1 shows the architecture of our ETist system which consists of three main parts, namely, data collection, data preprocessing and reading states recognition. When users are doing reading, the smartglasses with two eye cameras and one world camera collect eye movements data containing real-time coordinates of the eye and corresponding timestamps.

Due to the large number of fixation points and the existence of many invalid points, we propose line-break and page-break algorithms to locate the page where each fixation point is, and then use k -means algorithm to remove invalid outliers in preprocessing block. Following that, we perform feature extraction and selection for reading process within different pages, and further recognize reading states with machine learning methods. Based on the above results, our system can be applied in different areas such as advertising and education.

3.3 Methodology

Data Preprocessing. The data processing is composed of three steps. The smartglasses output three-dimensional space coordinates of eye movements at a certain rate. As mentioned above, each eye movement data point can be denoted by $g = (x, y, z)$, of which x , y and z represent three-dimensional space coordinates, respectively. However, it is to be noted that, the raw data output by the

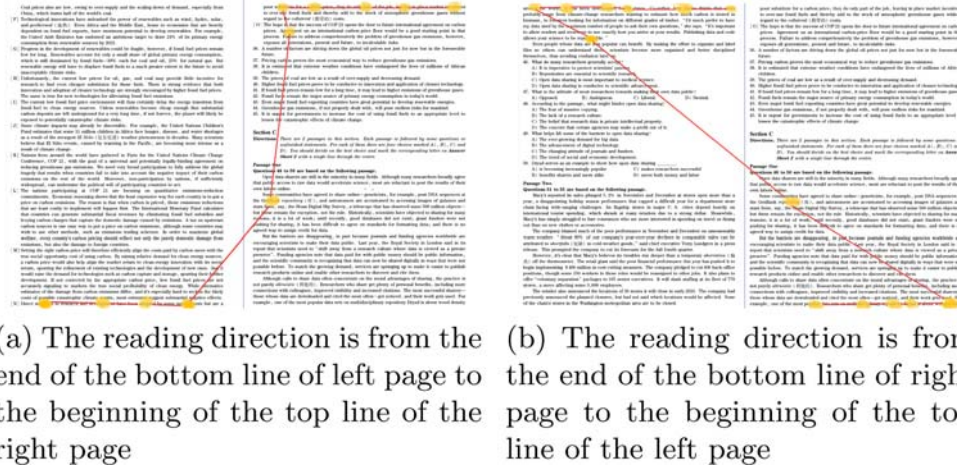


Fig. 2. An example of reading track of page breaks

hardware system do not contain information about which line and page they belong to. So the first step is to segment raw eye-tracking data by the line-break algorithm and store each segment in an array. For a sequence of eye gaze points in the i th line, they can be denoted by $L_i = (g_{i1}, g_{i2}, \dots, g_{il})$ where l is the total number of points in a line. Similarly, we utilize $P_j = (L_{j1}; L_{j2}; \dots; L_{jn})$ to represent the whole data points within j th page. Then, we use a page-break detection algorithm to segment L in order to get lines which are in the same page. This will help reduce the amount of data processed, and can more accurately locate what was being read at the time. After handling pages and lines in reading materials, the third step is to remove outliers from the obtained data. In addition, because the user can not get the information while scanning, we also need to remove the scan points in the data. Through analysis, we adopt the optimized k-means methodolgy to remove outliers.

Detection of Line Breaks. We segment raw eye-tracking data into parts corresponding to reading lines. We use (g_1, g_2, \dots, g_l) to represent a sequence of eye gaze points. Since the reading direction for a line break is from the end of text line to the beginning of a new line, which is against the regular reading direction, an eye gaze g_l must satisfy the first condition. Moreover, because the minimum spacing of two lines is line-spacing, g_l also meets the second condition. In a word, a eye gaze g_l is recognized as a line break if the following two conditions are satisfied.

Condition 1: Considering g_k is the beginning point of new line, the eye gaze direction from the g_i to g_l is opposite to the direction from g_l to g_k ($i < l < k$).

Condition 2: Let y_o , y_n represent the y-coordinate of the beginning point of current line and new line respectively. The absolute value of y_n minus y_o must larger than the line-spacing.

Detection of Page Breaks. In this paper, we mainly focus on the page breaks when a user reads in sequence (shown in Fig. 2). There are mainly two scenarios of page break detection. When a user reads in sequence, the reading direction for a page break is from the bottom line of current page forwards to the top line of a new page. Since the book is divided into left page and right page, there are two kinds of page breaks. One kind is a user reads from left page to the right page (shown in Fig. 2a) and the direction of the eye movement is from the end of the bottom line of left page to the beginning of the top line of the right page. Another kind is the opposite, as shown in Fig. 2b. We use $x(i)$, $y(i)$ to represent the x-coordinate and y-coordinate of the eye-tracking data i respectively. In order to remove the noise point, $x(i)$, $y(i)$ in line j must satisfy:

$$|x(i) - x(i-1)| < P \quad (P > 0) \quad (1)$$

$$|y(i) - y(i-1)| < T \quad (T > 0) \quad (2)$$

The threshold P must smaller than the margin between the two pages and larger than the spacing between two words. Since the reading direction is from the bottom to top and the value of y will decrease until it reaches the first line, the $y(i)$ must satisfy:

$$y(i) - y(i-2) < -C \quad (C > 0) \quad (3)$$

$$y(i) < \min(y) + \frac{\max(y) - \min(y)}{4} \quad (4)$$

where $\min(x)$, $\min(y)$, $\max(x)$, $\max(y)$ represent the minimum, maximum value of x , y lying in reading lines. As for the first kind of page breaks, the value of x will increase because the reading direction is from the left page to the right page. Moreover, $x(i)$ belongs to a new line. Therefore, $x(i)$ must satisfy:

$$x(i) > \max(x) + d \quad (5)$$

$$x(i+1) - x(i) < P \quad (6)$$

where d is half of the margin between two pages. As for the second kind of page breaks, the value of x will decrease because the reading direction is from the right page to left page. In addition, $x(i)$ will lie in a new line. Therefore, $x(i)$ must satisfy:

$$x(i) < \min(x) - (\max(x) - \min(x)) \quad (7)$$

$$x(i+1) - x(i) < P \quad (8)$$

When the system detects the page breaks, it will clear the maximum and minimum value of the x and y and recalculate on the eye-tracking data which lies on the new page.

Feature Extraction. After performing data preprocessing, we extract a set of features for reading cognitive states classification and POIs detection. At the first stage, we extract several base eye movement features, which is fixations, saccades, long fixations and recessions by analyzing the gaze point data. Then we calculate the duration of all reading states using time stamp. Specifically, we embed each of the classification models utilized in the learning stage which include support vector machine (SVM), random forest (RF) and K nearest neighbor (KNN) during feature selection. Then we pick out the feature subset that achieves optimal average accuracy as the final selected features.

Gaze Feature Extraction. Human reading states are closely related with different eye movement states which can be classified into fixation, long fixation, saccade, and revisit according to Lai *et al.* [23]. We identify these movements based on the following metrics.

- *Fixation*: the eyes stay within a certain spatial range for a duration between 100 ms and 300 ms.
- *Long fixation*: the eyes stay within a certain spatial range for a duration more than 300 ms, which indicates a deep thinking state.
- *Saccade*: the eyes move rapidly from one fixation point to another and the duration of eye gaze of less than 100 ms.
- *Revisit*: the order in which the eyes read is the opposite of the order in which the text is arranged.

After identifying different kinds of base eye movements, we further extract eye-movement features covering temporal, spatial and counting aspects. During our experiments, we find that when a user is in a mind wandering state, the depths of his fixation points are greater than the distances from the eye to the reading material plane. when the user is in attention state and drowsiness reading cognitive state, the depth of his fixation points is roughly the same as the distance between the eye to the reading material plane. Figure 3 shows the depth of user’s fixation point in attention state and mind wandering state, respectively. Therefore, the z -axis coordinate value of the gaze point is an important feature. And finally the features that we picked were total fixation duration, gaze duration, total revisited fixation, duration, total saccade duration, total reading time, total fixation count, revisited fixation count, number of blinks and saccade count based on our experiments.

POIs feature extraction. In this section, we will introduce the features that we choose to recognize the POIs of users when they are reading and the reason why we choose these features. As we can see, when a user is interested or has difficulty in reading a piece of material, he/she will spend more time on this part to obtain more detailed information. As a result, his/her gaze duration in the corresponding reading area is longer than that in other areas. Second, repetition is the first principle of all learning [24]. When a user has difficulty in reading a piece of content, the user may tend to repeatedly read the content to deepen the understanding of the content. Therefore, the user’s gaze duration in this area

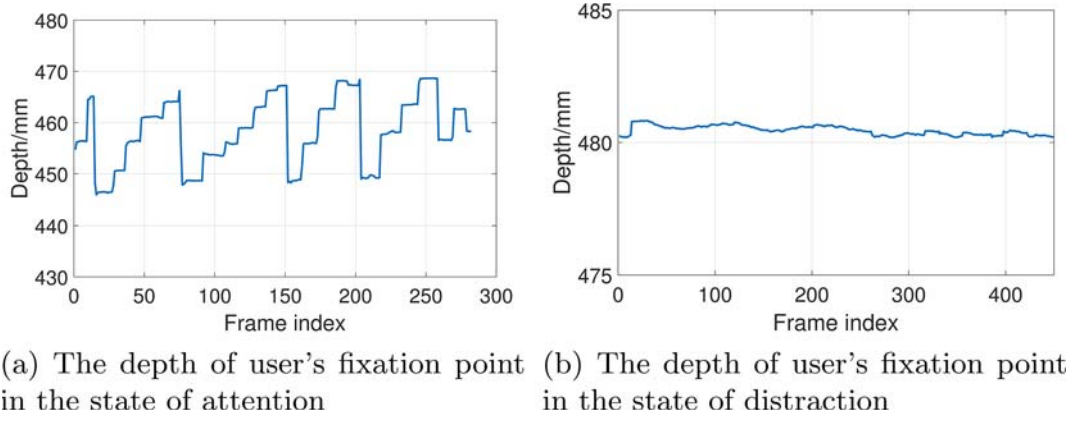


Fig. 3. The depth of user's fixation point in different states

will be longer than in other areas and the number of revisiting times will be longer as well and we decide these two aspects as features to detect POIs. In the process of features extraction of POIs, we perform a linear fit on each line of every fixation point's centroid to obtain a fitting function $f(x)$ first. Since there are generally less than 20 fixation points in each line, calculation will be greatly reduced. When the user revisit certain part of material, the user may read the previous content in the same line, or the content before this line. So the gaze point $g = (x_i, y_i, z_i)$ must satisfy (9) or (10):

$$x_i - x_{i-1} < \alpha, \|y_i - y_{i-1}\| < \beta \quad (9)$$

$$y_{i-1} - y_i > \beta \quad (10)$$

The value of α is three-quarters of the space between two words, and the value of β is the line spacing of reading content. When user end up revisiting the contents, the returned gaze point $g = (x_i, y_i, z_i)$ must satisfy:

$$x_{i+1} - x_i < \alpha, \|y_i - y_{i+1}\| < \beta \quad (11)$$

After obtaining the coordinates of the gaze point that the user normally reads for the first time, we need to know the serial number of lines the user revisit and then increase the revisit number of it. If the gaze point $g = (x_i, y_i, z_i)$ satisfies the (12), it is decided to be on this line.

$$\|f(x_i) - y_{i+1}\| < \beta \quad (12)$$

Then we calculate the number of fixation points n in each line of the reading material. Since the sampling frequency of smart glasses 30 Hz, the duration of gaze t can be calculated by (13).

$$t = n * \frac{1}{30} \quad (13)$$

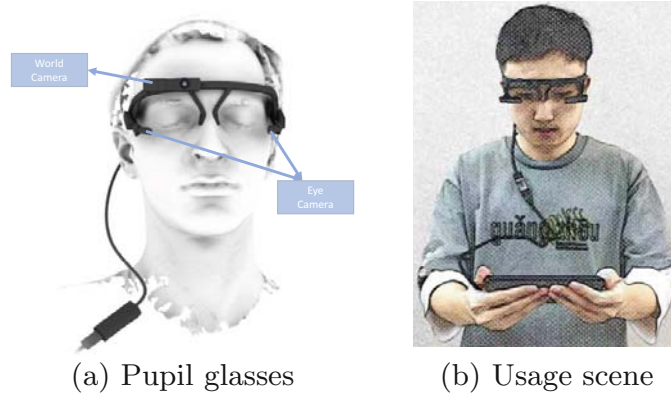


Fig. 4. Hardware of ETist

Classification. We use four kinds of classifiers including SVM, Random Forest, kNN and Ensemble Learning for data training. As a result, we find out that Random Forest classifier performs the best since it provides a high average classification accuracy of 84.0% and is more stable than other machine learning methods in English reading scenario. As for real-time system, we choose Random Forest as our classifier.

4 Implementation and Experiments

4.1 System Implementation

Hardware: The hardware of ETist system mainly include Pupil eye-tracking smartglasses [25], as shown in Fig. 4 and a data processing unit (*e.g.*, a PC). When a user performs reading with smartglasses, the eye camera captures eye movements with a resolution of 640×480 at a sampling rate of 30 fps. Each frame is sent to the connected PC for image processing to extract coordinates of the user's pupils. In our experiments, we utilize a Lenovo desktop with a Intel Core i5-4590 CPU and 8 GB RAM which runs software to capture and process images.

Software: The software of Pupil smartglasses is responsible for two-fold functions. On one hand, it controls the smartglasses to capture images and also extract coordinates of one pupil. On the other hand, as an open-source software, it provides powerful libraries for developers to design customized systems and realize personalized functions. Based on the open-source software, we implement our algorithms for page/line segmenting, outliers removal, and eyes' movements recognition.

4.2 Experiment Design

To assess the performance of our ETist system, we recruit a total number of 50 participants covering different grades of undergraduates and graduates in our campus. Among them, 30 participants are males and the others are females,

with ages between 20 and 30 years old. What is more, 25 participants are short-sighted and have to wear glasses in daily life. To encourage their participation, each one is paid with ¥30 per hour. The system evaluation can be divided into two main parts, namely, reading states recognition and POIs detection. In the following, we shall give details of each part.

Experimental Setting of Recognizing Reading States. In the first part, we make use of different reading materials as stimulus and request participants to read them at different states. Before experiments, we give them an introduction of the experimental process and instruct them to conduct calibration following the procedure provided by the software. Knowing about our experiments, they will do some warm-up practice to help them be more familiar with the whole process. For different states, each participant is assigned with 100 pages of reading materials selected from different sources with different difficulty. In the following, we shall give some details about settings of evaluating different reading states.

- **Attention** The reading materials are selected from reading comprehension articles of CET 6¹. To make them focused, each reading article is attached with several questions and participants are requested to answer them after finishing reading. Additional bonus will be based on their performance of answering questions.
- **Browsing** The reading materials of this setting are primary English compositions which can be easily understood by university students. Different from the above, participants are not required to answer questions after finishing reading.
- **Mind wandering** We refer to the experimental setup of Ylias *et al.* [26]. The reading materials in this setting are same with those in testing attention. The differences lie in two aspects. On one hand, during experiments, popular TV shows are played according to participants' preferences with the purpose of distracting them. On the other hand, the experiments are arranged between 16:00 pm and 17:00 pm during which people are more likely distracted.
- **Drowsiness** We refer to the experimental setup of Hu *et al.* [27]. To obtain the real state of drowsiness, we arrange experiments in midnights during which participants will first sleep at 23:00 pm for about 2.5 h, and then be alarmed to get up for experiments according to our survey of their daily schedules. They need to finish two pages of reading materials which are the same with above setting.

After finishing each page, participants will report their true states during experiments and then have a rest for about 3 mins before beginning a new repetition. If the self-report results differ from experimental setting, we discard corresponding data and restart experiments. The total experiments last for about 4 months.

¹ College English Test Band 6.

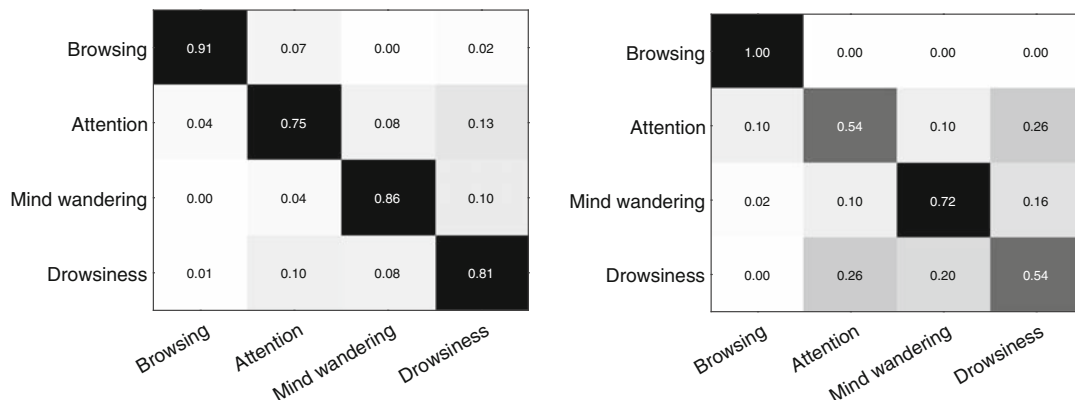
Experimental Setting of Detecting POIs. To evaluate the accuracy of detecting POIs, we request each participant to read 100 page of materials which are selected from comprehension articles of CET 6 and have different contents. During reading each article, participants are requested to mark their POIs on the materials. It is noted that the POIs marked by participants may be some key words, phrases, or short sentences. The ETist syetem detects POIs every 30 s which usually cover about two lines. If the marked POIs belong to the detected lines by smartglasses, it is considered as correct hit.

5 Performance Evaluation

5.1 Performance of Recognizing Reading States

We first evaluate the performance of recognizing reading states with training and testing data collected from the same person, which is referred as the identical-person case. For each participant, we use 5-fold cross-validation method to obtain the corresponding average confusion matrix, in which each row represents the ratios that a certain reading state is classified as others. Obtaining the confusion matrix of each participant, we further average them to get the results of different states. Figure 5a shows the result of identical-person tests. As we can see, the minimum and average accuracy of detecting reading cognitive state is 76.0% and 84.0%, respectively. As for the lower accuracy of recognizing drowsiness, we hypothesize that it is because three participants have experience of working over night. To verify our assumption, we exclude their data from our data set and classify data again. The final accuracy of reading drowsiness state is about 79.8%, with an increase of 3.8% compared with the former one.

Then we evaluate ETist’s performance in cross-person scenario where classification models are trained and tested with samples collected from different persons. We use 200 data samples of one person as training data set to train a



(a) The average confusion matrix of recognizing reading states in identical-person tests (b) The average confusion matrix of recognizing reading states in cross-person tests

Fig. 5. The performance of recognizing different states in tests

model and test it with another person's data. Since different people have different reading abilities and habits, it is impossible to use data of one person to train the data of others. However, we can find people who have similar English reading abilities to achieve this goal. In all of participants, we find two persons have the most similar English reading abilities according to the score of the CET. The result is shown in Fig. 5b in the form of confusion matrix. As you can see, the accuracy of reading cognitive state detection is 70.0%, with a decrease of 14.0%. We hypothesize that although they have similar English reading abilities, there are some different in their reading habits, which leads to different reading tracks. However, we envision that with more training data, the recognition performance of ETist in cross-person scenario can be increased to above 70.0%.

5.2 Performance of Detecting POIs

To evaluate the performance of detecting POIs, we make use of *precision* and *recall* as evaluation metrics. During the experiment, the number of POIs user marked is 56, and the number of POIs system detected is 80, among which the number of marked POIs is 32. As we can see, the total number of missed points is 24, of which 20 points are difficult points. This is because experimenters think that some difficult points of these such as persons' names can be skipped without affecting the reading quality. As for the remaining points, they are users' POIs, but experimenters did not read them repeatedly. Consequently, the system did not detect those points. In addition, the number of false detections of the system is 48, the main reason is that readers understand the difficulties through repeated reading, which leads to false detections of the system. Since points of interest and difficulties in the reading content detected by the system will not interfere with the user's reading, the most important indicator for the detection of reading POIs is the recall rate, which is 57.1%. After excluding the difficulties that do not affect the understanding of reading, the recall rate is 88.9%.

5.3 Case Study

In order to improve user's reading efficiency, we build an app called ETist to provide help for user when they are reading. The app has two mainly function. One is that when user is distracted or the app turn into background more than 2 min in reading, the app will give user a warm tip such as vibrate, determined by user to remind him or her of reading attentively. The other one is translation. When a user focuses on a word more than five seconds, we use OCR technique to extract sentences from user's view captured by eye-tracking glasses. Afterwards, we use google translation to translate from English into Chinese and then send the translation to ETist application. Owing to the limited resolution of eye-tracking glasses, a fraction of words can not be extracted accurately with OCR. However, with high-definition cameras, we believe that this problem can be solved easily. To further evaluate the performance of ETist in boosting up reading efficiency with better user experience, we conduct a real-world case study with all the participants. According to researches before, reading efficiency is reflected by two

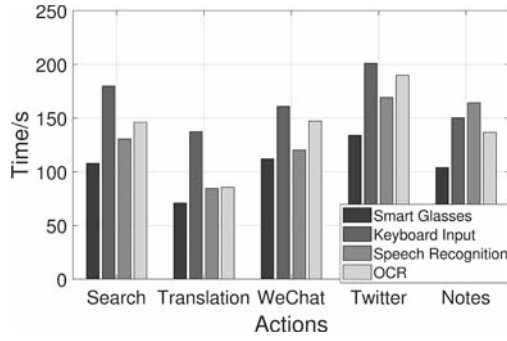


Fig. 6. The comparison of reading efficiency with different methods

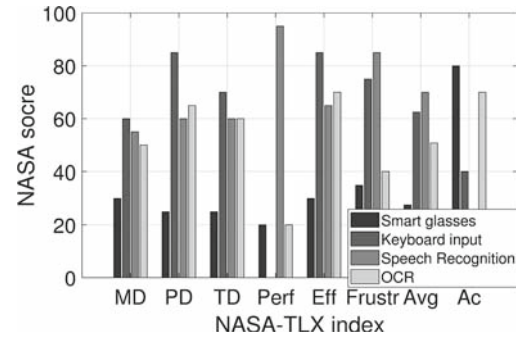


Fig. 7. The NASA scores of our method

main factors, namely, the understanding level and the reading time. Therefore, it can be roughly quantified by the following equation [28].

$$E = R \times C \quad (14)$$

where E , R and C represent reading efficiency, reading speed and understanding level, respectively. In each experiment, the experimenter indicated that he could fully understand the reading content, so we can set the value of C were 100%. Then the main factor that influencing reading efficiency of the experimenter is their reading time. To choose some traditional way for comparison, we conduct a survey by asking all the participants such a question “*what actions will you conduct when you encounter problems in reading?*”. By collecting their answers, we can get a ranking list of different choices according to the frequencies. The top five choices are searching online, translation, turning to friends via instant message applications(*e.g.*, wechat), posting on community websites, and making notes.

In the user-study experiments, we assign each participant with 100 pages of reading materials in each of which an uncommon geographical word is intentionally inserted into a certain position in the first three lines. In each session, participants conduct a certain kind of action mentioned above when they encounter the uncommon word. As comparison, we consider three other alternative ways to perform different actions when reading difficulties, namely, inputting via a keyboard, speech recognition and OCR text recognition. The reading speed is measured by the time consumed for finishing the whole page. Figure 6 shows the corresponding result. As we can see, the experimenter has the shortest reading time when using our ETist eye tracking assistant system in the same operation type. What is more, we also compare the user experience of using different methods. The approach that we used for this evaluation is NASA-TLX [29]. NASA-TLX is a multi-dimensional scale designed to obtain workload estimates from one or more operators while they are performing a task or immediately afterwards. The NASA-TLX score are shown in Fig. 7 which including scores that measure user’s like level. From the average NASA-TLX score, Smart glasses score 35 points less than keyboard input, 42.5 points less than speech recognition, 23.3 points less than OCR text recognition, and users’ reading load has been reduced by 56%,

61%, and 46%, respectively. From the perspective of user's preference, ETist is the highest scoring option of all ways. In summary, the reading assistance system based on smart glasses can improve the user's reading experience.

6 Conclusion

In this paper, we have proposed an eye-tracking based system called ETist to detect fine-grained reading states in real time to provide timely assistant when users are reading. Experiments have been conducted to evaluate the recognition performance of such a system. Results show that ETist can identify human reading states with an average accuracy of 84.0%. Moreover, we have also developed an application based on the output of ETist as a case study which aims to provide assistance for a reader stuck by vocabularies. We expect that more applications related with human mental states can be built upon this system and provide broader applications in the near future.

Acknowledgment. This research was supported in part by the China NSFC Grant (61802264, 61872248), Guangdong NSF 2017A030312008, Shenzhen Science and Technology Foundation (No. JCYJ20180305124807337, No. ZDSYS20190902092853047), Guangdong Science and Technology Foundation (2019B111103001, 2019B020209001), GDUPS (2015), and Natural Science Foundation of SZU (No. 860-000002110537).

References

1. Smallwood, J., et al.: Subjective experience and the attentional lapse: task engagement and disengagement during sustained attention. *Consciousness Cognit.* **13**(4), 657–690 (2004). <https://doi.org/10.1016/j.concog.2004.06.003>
2. Feng, S., D'Mello, S., Graesser, A.C.: Mind wandering while reading easy and difficult texts. *Psychon. Bull. Rev.* **20**(3), 586–592 (2012). <https://doi.org/10.3758/s13423-012-0367-y>
3. Smallwood, J., McSpadden, M., Schooler, J.W.: When attention matters: the curious incident of the wandering mind. *Memory Cogn.* **36**(6), 1144–1150 (2008)
4. Berka, C., et al.: EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviat. Space Environ. Med.* **78**(5), B231–B244 (2007)
5. Lan, T., Adami, A., Erdogmus, D., Pavel, M.: Estimating cognitive state using EEG signals. In: *Signal Processing Conference, 2005 13th European*, pp. 1–4. IEEE (2005)
6. Mather, M., Cacioppo, J.T., Kanwisher, N.: How fMRI can inform cognitive theories. *Perspect. Psychol. Sci.* **8**(1), 108–113 (2013)
7. Wang, X., Hutchinson, R.A., Mitchell, T.M.: Training fMRI classifiers to discriminate cognitive states across multiple subjects. In: *NIPS*, pp. 709–716 (2003)
8. Chen, Y., Fu, S., Iversen, S.D., Smith, S.M., Matthews, P.M.: Testing for dual brain processing routes in reading: a direct contrast of chinese character and pinyin reading using fmri. *J. Cogn. Neurosci.* **14**(7), 1088–1098 (2002)
9. Rayner, K.: Eye movements and attention in reading, scene perception, and visual search. *Quart. J. Exp. Psychol.* **62**(8), 1457–1506 (2009)

10. Rayner, K., Chace, K.H., Slattery, T.J., Ashby, J.: Eye movements as reflections of comprehension processes in reading. *Sci. Stud. Read.* **10**(3), 241–255 (2006)
11. Li, J., Ngai, G., Leong, H.V., Chan, S.C.: Multimodal human attention detection for reading from facial expression, eye gaze, and mouse dynamics. *ACM SIGAPP Appl. Comput. Rev.* **16**(3), 37–49 (2016)
12. Bixler, R., Blanchard, N., Garrison, L., D’Mello, S.: Automatic detection of mind wandering during reading using gaze and physiology. In: *Proceedings of the 2015 ACM on international Conference on Multimodal Interaction*, pp. 299–306 (2015)
13. Brusnighan, S.M., Folk, J.R.: Combining contextual and morphemic cues is beneficial during incidental vocabulary acquisition: semantic transparency in novel compound word processing. *Read. Res. Quart.* **47**(2), 172–190 (2012)
14. Johnson, C.I., Mayer, R.E.: An eye movement analysis of the spatial contiguity effect in multimedia learning. *J. Exp. Psychol. Appl.* **18**(2), 178 (2012)
15. Tsai, M.J., Hou, H.T., Lai, M.L., Liu, W.Y., Yang, F.Y.: Visual attention for solving multiple-choice science problem: an eye-tracking analysis. *Comput. Educ.* **58**(1), 375–385 (2012)
16. Marshall, S.P.: Identifying cognitive state from eye metrics. *Aviat. Space Environ. Med.* **78**(5), B165–B175 (2007)
17. Liang, Y., Reyes, M.L., Lee, J.D.: Real-time detection of driver cognitive distraction using support vector machines. *IEEE Trans. Intell. Transp. Syst.* **8**(2), 340–350 (2007)
18. Devi, M.S., Bajaj, P.R.: Driver fatigue detection based on eye tracking. In: *First International Conference on Emerging Trends in Engineering and Technology. ICETET 2008*, pp. 649–652. IEEE (2008)
19. Schooler, J.W.: Zoning out while reading: evidence for dissociations between experience and metaconsciousness jonathan w. schooler, erik d. reichle, and david v. halpern. *Thinking and seeing: Visual metacognition in adults and children*, 203 (2004)
20. James, W.: *The Principles of Psychology*, vol. 1. Cosimo Inc. (2007)
21. Hjørland, B.: Theoretical clarity is not ‘manicheanism’: a reply to Marcia bates. *J. Inf. Sci.* **37**(5), 546–550 (2011)
22. Shen, J., Barbera, J., Shapiro, C.M.: Distinguishing sleepiness and fatigue: focus on definition and measurement. *Sleep Med. Rev.* **10**(1), 63–76 (2006)
23. Lai, M.-L., et al.: A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educ. Res. Rev.* **10**, 90–115 (2013)
24. Bruner, R.F.: Repetition is the first principle of all learning. Available at SSRN 224340 (2001)
25. Pupil. Pupil eye-tracking glasses. 30 March 2017. <https://pupil-labs.com/>
26. Ylias, G., Heaven, P.C.L.: The influence of distraction on reading comprehension: a big five analysis. *Personality Individ. Differ.* **34**(6), 1069–1079 (2003)
27. Shuyan, H., Zheng, G.: Driver drowsiness detection with eyelid related parameters by support vector machine. *Expert Syst. Appl.* **36**(4), 7651–7658 (2009)
28. Jodai, H.: Reading rate and comprehension. Online Submission 2011
29. Hart, S.G.: Nasa-task load index (nasa-tlx); 20 years later. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, pp. 904–908. Sage Publications Sage CA: Los Angeles (2006)