Your Eye Tells How Well You Comprehend

Jiajia Li Grace Ngai Hong Va Leong Stephen Chan
Department of Computing
The Hong Kong Polytechnic University
Hong Kong

Abstract—**Systems that adapt to changes in human needs automatically are useful, built upon advancements in human-computer interaction research. In this paper, we investigate the problem of how well the eye movement of a user when reading an article can predict the level of reading comprehension, which could be exploited in intelligent adaptive e-learning systems. We characterize the eye movement pattern in the form of eye gaze signal. We invite human subjects in reading articles of different difficulty levels being induced to different comprehension levels. Machine-learning techniques are applied to identify useful features to recognize when readers are experiencing difficulties in understanding their reading material. Finally, a detection model that can identify different levels of user comprehension is built. We achieve a performance improvement of over 30% above the baseline, translating over 50% reduction in detection error.**

I. Introduction

In an intelligent e-learning system, we would expect it to behave much like a human teacher to "observe" its student to gauge his/her level of understanding, in order to "adapt" the way on how and what teaching materials are organized or presented to enhance student learning effectiveness. Based on human-centered computing research, human affects or feelings could be detected via the processing of various signals acquired when a user interacts with the computer.

Reading is one of the most common human-computer interaction activities and also one of the most fundamental means of knowledge acquisition. However, reading is a complex cognitive task that depends on human attention level, comprehension ability, visual interest, oculomotor processing constraints, etc. There have been works on understanding the relationship between human cognitive states such as attention and comprehension and human behaviors during reading [7].

Human behaviors are often reflected by human-oriented signals. Studies have shown that eye movement behavior during reading is closely related to human comprehension and attention [9][10]. Eye movements during reading can be categorized into saccades and fixations, which alternately occur during reading [9]. A saccade is a fast movement of the eye, which is usually in a direction parallel to that of the text. It is from left to right in English. A fixation is the maintaining of the visual eye gaze on a single location. The purpose of a saccade is to locate a point of interest on which to focus, while processing of visual information takes place during fixations. Humans typically alternate saccades and fixations in daily life. Readers experiencing difficulty in processing the visual information tend to make more fixations and the fixation duration becomes longer [4][8]. Moreover, under these circumstances, readers often make backwards, or regressive, saccades, to re-read the materials and get a better comprehension of text [5]. The mouse is a common movement tracking and event selection device. The mouse log can often provide useful information in e-learning [17]. However, using the mouse log in reading tasks would be challenging, since mouse use is oftentimes limited.

We make use of a commodity eye tracker to track the eye movement, and commonly available English articles without analyzing their content in building up the model. Finally, we build user-independent models to cope with new unseen users. Human subjects are invited to carry out experiments in reading a variety of articles while recording their eye movements with a commodity eye tracker. The captured eye movements are analyzed to identify specific eye behaviors. We then apply machine-learning techniques to identify a subset of useful features capable of assisting in the determination of human comprehension level in order to produce a resilient user-independent model, capable of serving new unseen users.

Our contributions can be summarized as: (1) investigate comprehension level detection based on a common reading task; (2) identify effective features in describing specific eye movement behaviors at different comprehension levels; (3) apply machine-learning to build user-independent models to recognize the reading comprehension level; and (4) verify our approach empirically through experiments with human subjects. We believe that our work opens up interesting techniques for building adaptive systems.

The rest of this paper is organized as follows. In Section II, we describe our recognition framework based on eye movement behaviors and the extraction of useful features for these behaviors. Section III describes the experimental setups and the experimentation with human subjects carrying out reading tasks, as well as the associated machine-learning algorithms. We then evaluate the effectiveness and accuracy of our models in the next section under different situations. Finally, Section V gives a brief conclusion.

II. System Framework

We analyze the eye movement patterns captured by a commercial eye tracker to detect the level of comprehension of a user when reading an article. We will describe the actual feature extraction mechanisms based on the stream of eye movement signals, followed by the mechanism to select the set of useful features and finally the means by which we classify the level of human comprehension when reading an article.

A. Identifying Eye Movement Behaviors

The eye gaze data analyzed in this work is captured by the Tobii X1 Light Eye Tracker. It represents the position of each eye as a timestamped sequence in screen coordinates, normalized to [0, 1] if the eye gaze detection is effective. Occasionally, the eye tracker may lose track of the eye, when the person moves the head rather rapidly sideway, or turns his/her head around, or blinks the eye. The output value for any unrecognized eye position is given a special value of (-1, -1).

There are two challenges in working with the eye gaze data. First, the eye tracker occasionally fails to detect one eye. The eye tracker reports a normal pair of screen coordinates for the detected eye, while that of the other eye is (-1,-1). It happens when the infrared beam from the eye tracker does not quite get reflected from the bottom of the eye. This constitutes less than 5% of the data, and are discarded. Non-detection of both eyes may be due to eye blinks or when the reader moves his/her head too violently, to be considered when we detect eye blinks.

Second, the sampling rate of the eye tracker is not constant [16]. In our pilot study, we observed that the sampling rate varies between 20 to 30Hz. To facilitate data analysis and pattern recognition, the eye gaze data is down-sampled to 15 Hz. A median filter, shown to be effective in preserving the characteristics of the original signal without introducing signal artifacts [2], is applied to the down-sampled eye gaze signal to further remove noisy artifacts. The window size of the median filter is set to be 120ms, small enough to retain short pulses indicating eye gaze movements.

After preprocessing, the resultant denoised, subsampled eye gaze data can be represented as a sequence E of *eye position vectors*, $E_\text{i} = [t_i, e_{l_x}, e_{l_y}, e_{r_x}, e_{r_y}]$, e_{l_x}, e_{l_y} are the normalized *x* and *y* coordinates of the gaze location of the left eye, captured by the eye tracker, and e_{r_x}, e_{r_y} are the corresponding values for the right eye. We simplify the representation by computing the mean value of the coordinates of the two eyes. The eye gaze feature vectors can be compressed into a sequence of *eye gaze vectors*, $EG_i = [t_i, EG_h, EG_v]$, where EG_h is the horizontal component of the eye gaze location, average of e_{l_x} and e_{r_x} for the same timestamp t_i, and EG_v is the corresponding vertical component, average of e_{l_y} and e_{r_y}.

1) *Detecting Eye Blinks*

Given the eye gaze sequence, EG, eye blinks can be easily detected by identifying the moments in which EG_h and EG_v are both equal to -1. The duration of the blink is defined as the length of the sequence of consecutive eye blink points. In previous work, eye blinks is defined as eyelid closure for a duration of 50 to 500ms [12]. We follow same convention to discard eye closures beyond 50ms and 500ms.

2) *Detecting Eye Fixations*

Eye fixations are defined as periods in which the eye gaze remains stationary on a specific location. The inherent error present in the eye tracker makes detecting fixations from the eye gaze signal more than simply looking for periods during which the eye coordinates do not change. To determine the extent of the inherent sensor error, a pilot experiment was carried out with 3 subjects, who were asked to sit at a normal distance away from the screen and fix their gaze on a single word displayed in the center of the screen for 10 seconds. The recorded eye gaze signal gives us an estimated error of the eye tracker, which was measured to be around 1% of the potential range of the horizontal component of the eye gaze signal. This is then used as the *fixation amplitude threshold* th_{FA}. Defining D_i as the Euclidean distance between successive gaze locations EG_i and $EG_{\text{i}+1}$, a binary vector F can be constructed from EG, where F_i is set to 1 whenever D_i is less than the threshold th_{FA}, i.e., the moments when the eye gaze could be considered to be stationary. Since it has been found that fixations are rarely shorter than 100ms and usually in the range of 200 to 400ms [11], we label fixations as continuous sequences that last longer than 100ms but shorter than 500ms. Too long a potential fixation may indicate abnormal or sometimes erroneous situation. We filter out those relatively uncommon scenarios to clean up the data stream, by removing outliers that could affect the statistical measures that we use as potential features.

3) *Detecting Eye Saccades*

Once the eye blinks and fixations have been identified, the sequences in between the fixations are considered for saccadic eye gaze movements. We classify eye saccades into three categories: line change saccades, forward saccades and regressive saccades. Consider Fig. 1. Line change saccades involve large, fast eye movements that happen when the reader finishes reading one line and the eye jumps to the beginning of the next line (grey sharp-drop columns). Forward saccades follow the direction of the text (left to right in English text) during normal reading. Regressive saccades go against the flow of the text when the eye position moves back to re-read previously-read content (green columns).

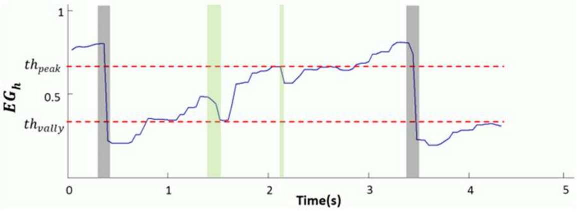

Fig. 1. Horizontal component of eye gaze

Our saccade detection algorithm makes use of the horizontal component EG_h of the eye gaze position to identify the different saccadic types. We use a two-level saccade detection algorithm that first identifies line-change saccades, followed by forward and regressive saccades. Fig. 1 illustrates the horizontal component of the eye gaze signal EG_h. It can be seen that the signal experiences periodic fall-off “cliffs” when the eye gaze switches from the far right of the page (closer to 1) to the left edge (closer to 0). It gives us periodic “peaks” and “valleys” which correspond to line-change saccades. To identify line-change saccades, we define two thresholds: th_{peak} and th_{valley}, corresponding to the maximum and minimum values of the eye gaze signal before and right after the fall-off “cliff”. Line change saccades are defined as movements that start at positions to the right of th_{peak} and end with a position to the left of th_{valley}, determined empirically through an experiment with 4 subjects, who are asked to read articles presented on a LCD. The gaze data from both eyes is then visualized onto images of the screen display, allowing us to observe the trajectory of the eye movements through manual inspection. The best performance (100% recall and 100% precision) is achieved with $th_{peak} = 0.6$ and $th_{valley} = 0.3$.

The second level saccade detection distinguishes between forward and regressive saccades. Previous research [14] has shown that saccadic eye movements during reading usually span a distance of about 7 to 9 characters, equivalent to about 2 degrees of visual angle, with duration between 10 to 100ms. We define the *saccade amplitude* s_{amp} as the total change in distance along the *x*-dimension over a saccade, i.e., $s_{amp} = \sum_{EG \in s} |EG_{i_h} - EG_{i-1_h}|$, where s is a saccade, EG_{i_h} and EG_{i-1_h} are the horizontal components of successive gazes. Given the saccade amplitude, and knowing that normal reading activities generate saccades that span approximately 2 degrees of the visual angle, we can calculate the required amplitude that would be expected. We define the *saccade amplitude threshold* as $th_{SA} = \frac{\pi d EG_{hrange}}{90W}$, where d is the distance from the eyes to the screen, $EG_{h_{range}}$ is the range of EG_h, W is the width of the displayed article on the screen. In our experiment setup, $W = 31.5$ cm, $d = 60$ cm and $EG_{h_{range}} = 0.68$. This leads to $th_{SA} = 0.045$. Since a full line of text has on average 73 characters in our setup, a saccade that spans 7 to 9 characters would give us saccades ranging between 0.065 to 0.084. Combining this with the calculated th_{SA} gives us the parameter settings of $th_{SA_{min}} = 0.045,\ th_{SA_{max}} = 0.084$.

Given a candidate saccade S with n eye gaze points tracked, we identify the first and last eye gaze points EG_1 and EG_n. The difference $EG_1 - EG_n$ allows us to classify the saccade as forward or regressive saccade. If the value is between $th_{SA_{min}}$ and $th_{SA_{max}}$, it is considered a forward saccade, and it is between $-th_{SA_{min}}$ and $-th_{SA_{max}}$, it is considered a regressive saccade. We also cleanse the saccadic stream for outliers. Saccadic segments beyond 20ms and 200ms are discarded, assuming angular saccade speed of 20 degree/s [13].

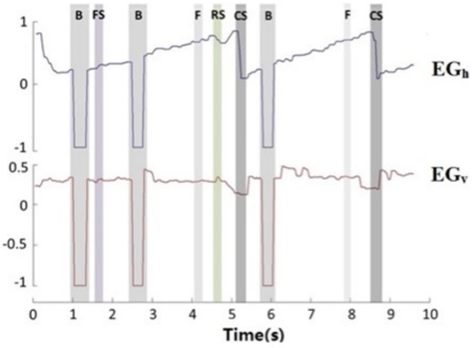

Fig. 2. Identifying eye movement patterns

Fig. 2 depicts an example of eye movement behaviors given the eye gaze positions. The eye blinks, fixations, as well as forward, regressive and line change saccades are illustrated. Here **F** indicates a fixation, **B** an eye blink, **FS** a forward saccade, **RS** a regressive saccade, and **CS** a line-change saccade. A blink is identified as a period of non-eye-detection. A line-change saccade is reflected by backward changes in both x and y in a sequence of eye gaze points, whereas a regressive saccade is reflected by backward changes in x.

B. Features Describing Eye Movement Behaviors

Once the different eye movement behaviors have been identified from the eye gaze points, we can construct the features that will be used to describe these behaviors. Our objective is to automatically detect when the user is having difficulty with a text, or when there is a reduced level of reading comprehension. We consider the eye movement behaviors to define the useful features.

For each saccade type, we define 6 different useful features. We measure the mean and standard deviation (*stdev*) of the metrics of interest for each saccade type, namely, the amplitude (screen distance covered in the saccade), the duration and the speed. With three types of saccades, three metrics and two statistical measures, this gives 18 potentially useful features. Finally, we are interested in the global manifestation of each type of saccadic behavior throughout the period of the article reading task. We measure the number of saccades of each type over the window W_{EG} that spans over the duration of each individual article reading task. This gives 21 saccadic features, depicted in Table I, in describing the saccadic eye behaviors adopted in building our comprehension level recognition model.

TABLE I. FEATURES DESCRIBING SACCADIC EYE BEHAVIORS

Feature	Meaning	Formulation
$s_{1_{FS}}, s_{1_{RS}}, s_{1_{CS}}$ $s_{2_{FS}}, s_{2_{RS}}, s_{2_{CS}}$	Saccade distance	*Mean* (s_1) / *stdev* (s_2) of screen distance covered by forward, regressive, line-change saccades
$s_{3_{FS}}, s_{3_{RS}}, s_{3_{CS}}$ $s_{4_{FS}}, s_{4_{RS}}, s_{4_{CS}}$	Saccade duration	*Mean* (s_3) / *stdev* (s_4) of forward, regressive, line-change saccade duration
$s_{5_{FS}}, s_{5_{RS}}, s_{5_{CS}}$ $s_{6_{FS}}, s_{6_{RS}}, s_{6_{CS}}$	Saccade speed	*Mean* (s_5) / *stdev* (s_6) of forward, regressive, line-change saccade speed
$s_{7_{FS}}, s_{7_{RS}}, s_{7_{CS}}$	Saccade rate	Number of forward, regressive, line-change saccades in window W_{EG}

We extract the features to describe fixations in a similar manner. These include the rate of fixation normalized by W_{EG}, i.e. the time spent in reading the article**.** We also calculate the mean and standard deviation of the duration of the fixations, as well as the interval between successive fixations. Table II details the 5 fixation-related features used in our work.

TABLE II. FEATURES DESCRIBING FIXATIONS

Feature	Meaning	Formulation
f_1, f_2	Fixation duration	*Mean* (f_1) / *stdev* (f_2) of fixation duration
f_3, f_4	Fixation interval	*Mean* (f_3) / *stdev* (f_4) of elapsed time between successive fixations
f_5	Fixation rate	Number of fixations in window W_{EG}

We calculate 5 eye blink features, extracted by calculating the rate of blink normalized by W_{EG}, the mean and standard deviation of the duration of the eye blinks, and the interval between successive blinks, as depicted in Table III.

TABLE III. FEATURES DESCRIBING EYE BLINKS

Feature	Meaning	Formulation
b_1, b_2	Eye blink duration	*Mean* (b_1) / *stdev* (b_2) of duration of eye blinks
b_3, b_4	Eye blink interval	*Mean* (b_3) / *stdev* (b_4) of elapsed time between successive eye blinks
b_5	Eye blink rate	Number of eye blinks in window W_{EG}

In addition to behavior-based features, we use gaze-based features to capture the general characteristics of the unfiltered eye gaze positions. These include the kurtosis and skewness of the horizontal component of the EG_h signal, in Table IV.

TABLE IV. FEATURES DESCRIBING EYE MOVEMENTS

Feature	Meaning	Formulation
e_1, e_2	Eye movement variation	*Kurtosis* (e_1) and *skewess* (e_2) of horizontal component of gaze locations

Finally, we incorporate some features that capture the context of the reading task through basic information on the read article and the overall task. These are the per-line reading speed and the repetition rate of reading, as shown in Table V.

TABLE V. CONTEXTUAL FEATURES

Feature	Meaning	Formulation
c_1	Reading speed	Normalized no of lines in article segment
c_2	Repetition rate	Normalized no of line-change saccades

III. EXPERIMENTS

We evaluate our model for detecting the level of reading comprehension in a real-world setup. The experiment subject reads articles of varying difficulty in a full-screen mode. Articles of varying difficulty are used to induce different levels of comprehension in the subjects. Though the level of difficulty of article is roughly related to the reader level of comprehension, it varies across different persons. A hard article could be hard to comprehend for someone, but just medium for another. An easy article may also take some weak readers much effort to understand. So we rely on subjects to report their level of comprehension when reading a specific article. To enhance the reliability of the self-reported information, we brief all the subjects about self-reporting and precautions to take before the experiments. The eye gaze tracking logs for individual subjects are recorded in real-time during the experiments, and pre- and post-surveys are used for further data collection about the subject. We also check the pre- / post-surveys for consistency, and follow up with an interview whenever a discrepancy is discovered.

A. Participants and Experiment Setup

We recruited 10 subjects (age 20-33 years, *M* = 24.6, *SD* = 4.2), all undergraduate or graduate students and non-native English speakers, four of them are female. A pre-experiment survey shows that they are all comfortable using the computer, and are able to read and write in English, though there is some variation in their grasp of the language.

B. Experiment Design

The English articles used in our experiment are drawn from standardized sources that are used to evaluate English reading comprehension ability: GRE (Graduate Record Examination), TOEFL (Test of English as a Foreign Language), and CET-4 (College English Test Band 4). TOEFL is widely used for admission of non-native speakers to English-speaking colleges and universities. GRE is required by many graduate schools in North America as an admission criterion. CET-4 is used for calibrating the level of English ability for university students in China. Two articles were chosen from each of the three tests, and the length of each article was constrained to be around 500 words. To make sure that the subjects were focused on the reading task, a post-reading guarding procedure was imposed. Subjects were told that they should give a detailed explanation of each paragraph of the article to the experiment instructor after finishing each article. After reading all the articles, they were asked to write down some feedback about the setting of the experiment and provide further suggestion. Considering the background and English level of our subjects, we expect them to have a reasonable understanding of the TOEFL articles, while the GRE would be considered to be difficult, and the CET-4 to pose no difficulty. Post-experiment surveys confirmed that the subjects indeed did find that the articles were of different levels of difficulty and they did experience different levels of comprehension. In general, our experimental subjects also found the setup of the experiment to be comfortable and non-intrusive.

The subjects were informed in advance about the eye gaze tracker and that their gaze movements were being recorded during the experiment. To minimize the impact of the ordering effects, the order of the articles was randomized so each subject would be presented with the articles at different levels of difficulty following one of the possible permutations. The subjects were not constrained for reading time. Immediately after reading each article, the subjects were asked to label their perceived level of comprehension as “low”, “medium” or “high” while their memory was still fresh.

C. The Dataset

After the experiment, the collected eye gaze data is visually inspected to ensure that it is usable. In a few cases, large degree of body movements and askew sitting postures from the subject results in eye tracker failure for prolonged periods of time. These segments are removed from the dataset.

The final dataset contains 41 instances, corresponding to 41 article-reading activities. The total length of the dataset is 228 minutes. According to the subjects’ self-reported labeling, 9 (22.0%) instances belong to high-level-of-comprehension, 17 (41.4%) medium-level-of-comprehension, and 15 (36.6%) low-level-of-comprehension. We still consider this to be a reasonably well-mixed dataset since the deviation from the completely even distribution (33.3%) is small enough. The standard baseline of the dataset is 41.4%, by outputting the label of the largest class to achieve the “best” result.

IV. EVALUATION

We evaluate our approach with a user-independent model that combines the dataset from all subjects.

A. Feature Evaluation and Selection

Based on Sections II-A and II-B, 35 features are extracted from the reading activity eye gaze data. These include 21 saccade features, 5 fixation features, 5 eye blink features, 2 eye movement features, and 2 contextual features. The duration of the window W_{EG} is set to the duration of reading one article.

Our initial feature set contains 35 features, too many to be effective for practical real-time recognition. We apply feature selection to remove non-indicative features to improve performance. We adopt the wrapper method for feature selection, which is reported to outperform the filter method [15]. It considers the selection of a set of features by comparing different feature combinations to identify a subset that is best for the chosen classifier. We use Linear Support Vector Machines (SVM) for classification and its error rate to indicate performance of a feature subset. We adopt the best first searching approach which searches the space of attribute subsets by greedy hill climbing augmented with a backtracking facility, with a comparison window of 5 consecutive non-improving search nodes. This strikes a balance between computational efficiency and effectiveness of features selected. We analyze how the feature evaluator ranks the features by performing a 10-fold-cross-validation. The number of training sets that each feature is selected in during cross-validation is used to measure the degree of importance of the feature.

Table VI shows the top 10 most indicative features, together with the number of partitions (in 10-fold cross-validation) in which they are found to be of potential contribution in recognition. Unsurprisingly, the most indicative features are the reading speed and the rate of the forward saccades. This makes sense, as people who are having difficulties in understanding the reading material will tend to slow down and have longer fixations [9]. They also tend to make more regressive saccades and re-read sections of the text, which is also evidenced by the fact that the repetition rate is selected as one of the most indicative features, and regressive saccades feature heavily dominates among the top 10 most-useful features as well. We note that some of the top ten ranked features are actually indicative only for a small number of training partitions instead of being useful across the board and they may not always work synergistically together. Some may still be useful when combined with other features.

TABLE VI. TOP 10 MOST INDICATIVE FEATURES

Feature	Description	Indicative sets
c_1	Per-line reading speed	10
$s_{7_{FS}}$	Rate of forward saccades	8
e_2	Skewness of eye movements	4
$s_{3_{RS}}$	Mean duration of regressive saccades	3
e_1	Kurtosis of eye movements	3
c_2	Repetition rate	2
f_5	Rate of fixations	2
$s_{3_{FS}}$	Mean duration of forward saccades	2
$s_{6_{RS}}$	Stdev of duration of regressive saccades	2
f_4	Stdev of elapsed time between fixations	2

To select the best feature subset, we execute the feature selection algorithm with the same feature evaluator to arrive at the most effective subset. Having trimmed the set of potentially useful features from 35 to 10, we systematically explore different combinations of subsets of features drawn from the list of 10 potentially useful features. Again, we adopt SVM for classification with 10-fold cross-validation. Finally, there are 6 important features that together give us a best performance. This final list of features selected in building the user-independent model is illustrated in Table VII.

TABLE VII. FINAL SET OF SELECTED FEATURES

Feature	Description
c_1	Reading speed
$s_{7_{FS}}$	Rate of forward saccades
e_2	Skewness of eye movements
c_2	Repetition rate
f_5	Rate of fixations
$s_{3_{FS}}$	Mean duration of forward saccades

Since we are building user-independent models, the gold standard in evaluation is the leave-one-subject-out cross-validation test. From the set of n = 10 subjects, we train the user-independent model using the data from n − 1 = 9 subjects and test the model on the left-out subject. We repeat the experiment n times, each time leaving out a different subject, and the average performance for the 10 experiments is reported.

Table VIII presents the normalized confusion matrix for our leave-one-subject-out reader comprehension level detection experiment. Table IX illustrates the classification performance of the classifier, averaged over 10 subjects.

TABLE VIII. CONFUSION MATRIX FOR COMPREHENSIION DETECTION

Comprehension level Ground truth	High	Medium	Low
High	0.78	0.22	0.00
Medium	0.18	0.64	0.18
Low	0.00	0.20	0.80

TABLE IX. LEAVE-ONE-SUBJECT-OUT COMPREHENSION DETECTION

Performance Comprehension level	Precision	Recall	F-measure
High	0.70	0.78	0.74
Medium	0.69	0.65	0.67
Low	0.80	0.80	0.80
Overall	0.73	0.73	0.73

From Table VIII, we can compute the correctly classified rate (CCR), defined to be the proportion of the correctly classified instances over all the instances. This CCR is found to be 73.2%, significantly higher than the baseline of 41.4%, with an improvement of 31.8%, achieving 54.2% error reduction. Most of the errors come from misclassifying similar comprehension classes, i.e., conflating *low* with *medium* and *medium* with *high*. The classifier never erroneously classifies an instance into an extreme class, i.e., conflating *low* with *high* and vice versa. Table IX indicates that we are able to perform well with an overall F-measure of 0.73. The performance of each individual class indicates that we are able to achieve a high precision as well as a high recall, without having to sacrifice one metric for the other. In fact, we can achieve a precision and a recall up to 0.80 and even the worst recall stands at 0.65, far better than the baseline performance of 41.4%. We are able to accurately recognize low levels of comprehension, but comparatively not that well for medium ones. This is perhaps because low level of comprehension is often associated with lengthy reading with many regressive saccades, which is easier to detect. Medium levels are somewhere in between and more prone to misclassification.

B. *Performance for Existing Users*

The leave-one-subject-out evaluation simulates the case when the model encounters new unseen users. However, it is common to re-encounter existing users in real-life scenarios. One would expect better performance in those scenarios. We now keep all subjects in the 10-fold cross-validation in the experiment and compare the performance with the leave-one-subject-out setting, depicted in Table X and XI. We observe some improvement across the board for all the metrics. Table XII summarizes the key performance metrics between two sets of experiments under the two application scenarios. The actual message is far more encouraging than the mild improvement observed. It does demonstrate that our approach is very ***robust***, capable of delivering good performance even for new unseen users based on training data from a small population of only 9 subjects. Our work represents a successful attempt to reading comprehension detection based on the eye movement patterns.

TABLE X. ALL-SUBJECTS-INCLUDED COMPREHENSION DETECTION

Performance Comprehension level	Precision	Recall	F-measure
High	0.69	1.00	0.82
Medium	0.82	0.53	0.64
Low	0.76	0.87	0.81
Overall	0.77	0.76	0.74

TABLE XI. CONFUSION MATRIX WITH ALL-SUBJECTS-INCLUDED

Comprehension level Ground truth	High	Medium	Low
High	1.00	0.00	0.00
Medium	0.23	0.53	0.24
Low	0.00	0.13	0.87

TABLE XII. PERFORMANCE WITH UNSEEN USERS

Evaluation method	**CCR**	**F-measure**
Leave-one-subject-out with unseen users	73.2%	0.73
All-subjects-included without unseen users	75.6%	0.74

C. Discussion

The features selected by the attribute evaluator reveals useful features for reading comprehension detection. It is notable that eye blinks are neither top-ranked nor selected, despite the fact that blink frequency and duration have been widely reported to be useful for visual engagement measurement [6], fatigue detection [3], and activity recognition [2]. It seems that there is no obvious relationship between eye blinks and human comprehension of reading material. There has been research demonstrating that blink rate is unreliable as a measure of the difficulty of the reading task when the difficulty is varied by introducing glare conditions or auditory distractions [1]. Our work seems to further confirm this by demonstrating that eye blinks are not particularly helpful when the difficulty is varied by the reading materials.

Of the top 10 ranked features selected, 8 belong to eye gaze features, i.e., 4 are saccade features and 2 are fixation features, with 2 more that describe general eye movement statistics (kurtosis and skewness). This suggests that good classification performance relies on the use of a mixture of features, and corroborates previous work that finds that fixations, regressive saccades and reading speed are strong indicators of reading behavior [1].

V. CONCLUSION

This paper investigates eye movements as a modality with which to recognize when humans are having difficulty with reading material. Our method uses only consumer-grade devices, namely, a consumer eye tracker, with some very basic information gathered from the article and the overall task. We identify eye movement behaviors from the eye gaze locations captured by the tracker, and extract features to describe these behaviors. Machine-learning techniques are then used to model the data and build a user-independent model that is capable of recognizing the comprehension level for new unseen users. We evaluate our approach via reading tasks, in which subjects are induced to different levels of comprehension by articles of varying difficulty. Using feature selection, we identify 10 most indicative features from our pool, and then a subset of 6 that combine to give the best performance in reducing the error rate by 50%.

VI. ACKNOWLEDGEMENTS

We wish to thank the anonymous reviewers and the experiment subjects for their time. This work was supported in part by grant PolyU 5222/13E from the Hong Kong Research Grants Council.

REFERENCES

[1] Bitterman, M.E. Heart rate and frequency of blinking as indices of visual efficiency. *Journal of Experimental Psychology*, **35**(4):279–292, 1945.

[2] Bulling, A., Ward, J.A., Gellersen, H., and Tröster, G. Eye movement analysis for activity recognition using electrooculography. *IEEE Trans on Pattern Analysis and Machine Intelligence*, **33**(4):741–753, 2011.

[3] Divjak, M. and Bischof, H. Eye blink based fatigue detection for prevention of computer vision syndrome. In *Proceedings of IAPR Conference on Machine Vision and Applications*, 350–353, 2009.

[4] Inhoff, A.W. and Rayner, K. Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, **40**(6):431–439, 1986.

[5] Just, M.A. and Carpenter, P.A. A theory of reading: From eye fixations to comprehension. *Psychological Review*, **87**(4):329–354, 1980.

[6] Palomba, D., Sarlo, M., Angrilli, A., Mini, A., and Stegagno, L. Cardiac responses associated with affective processing of unpleasant film stimuli. *International Journal of Psychophysiology*, **36**(1):45–57, 2000.

[7] Poole, A. Gender differences in reading strategy use among ESL college students. *Journal of College Reading and Learning*, **36**(1):7–20, 2005.

[8] Rayner, K., Sereno, S.C., and Raney, G.E. Eye movement control in reading: a comparison of two types of models. *Journal of Experimental Psychology. Human Perception & Performance*, **22**(5):1188–2000, 1996.

[9] Rayner, K. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, **124**(3):372–422, 1998.

[10] Rodrigue, M., Son, J., Giesbrecht, B., Turk, M., and Höllerer, T. Spatio-temporal detection of divided attention in reading applications using EEG and eye tracking. In *Proceedings of International Conference on Intelligent User Interfaces*, ACM, 121–125, 2015.

[11] Salvucci, D.D. and Goldberg, J.H. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of Symposium on Eye Tracking Research & Applications*, ACM, 71–78, 2000.

[12] Schleicher, R., Galley, N., Briest, S., and Galley, L. Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired? *Ergonomics*, **51**(7):982–1010, 2008.

[13] Sen, T. and Megaw, T. The effects of task variables and prolonged performance on saccadic eye movement parameters. In *Theoretical and Applied Aspects of Eye Movement Research*, Elsevier, 103–111, 1984.

[14] Sibert, J.L., Gokturk, M., and Lavine, R.A. The reading assistant. In *Proceedings of ACM Symposium on User Interface Software and Technology*, ACM, 101–107, 2000.

[15] Talavera, L. An evaluation of filter and wrapper methods for feature selection in categorical clustering. In *Advances in Intelligent Data Analysis*, Springer, 440–451, 2005.

[16] Tobii X1 Light Eye Tracker. Specification of Gaze Precision and Gaze Accuracy. http://www.tobiipro.com/siteassets/tobii-pro/technical-specifications/tobii-pro-x2-60-technical-specification.pdf.

[17] Tsoulouhas, G., Georgiou, D., and Karakos, A. Detection of learners' affective state based on mouse movements. *Journal of Computing*, **3**(11):9–18, 2011.