Understanding Mobile Reading via Camera Based Gaze Tracking and Kinematic Touch Modeling

Wei Guo
University of Pittsburgh
Pittsburgh, PA, USA
weg21@cs.pitt.edu

Jingtao Wang*
Google Cloud AI
Beijing, China
jingtaow@google.com

ABSTRACT

Despite the ubiquity and rapid growth of mobile reading activities, researchers and practitioners today either rely on coarse-grained metrics such as click-through-rate (CTR) and dwell time, or expensive equipment such as gaze trackers to understand users' reading behavior on mobile devices. We present Lepton, an intelligent mobile reading system and a set of dual-channel sensing algorithms to achieve scalable and fine-grained understanding of users' reading behaviors, comprehension, and engagement on unmodified smartphones. Lepton tracks the *periodic* lateral patterns, i.e. saccade, of users' eye gaze via the front camera, and infers their muscle stiffness during text scrolling via a Mass-Spring-Damper (MSD) based kinematic model from touch events. Through a 25-participant study, we found that both the periodic saccade patterns and muscle stiffness signals captured by Lepton can be used as expressive features to infer users' comprehension and engagement in mobile reading. Overall, our new signals lead to significantly higher performances in predicting users' comprehension (correlation: 0.36 vs. 0.29), concentration (0.36 vs. 0.16), confidence (0.5 vs. 0.47), and engagement (0.34 vs. 0.16) than using traditional dwell-time based features via a user-independent model.

CCS CONCEPTS

• Human-centered computing→Ubiquitous and mobile computing→Ubiquitous and mobile computing systems and tools

KEYWORDS

Text reading; mobile computing; gaze tracking; Mass-Spring-Damper; machine learning; intelligent user interfaces

ACM Reference format:

Wei Guo and Jingtao Wang. 2018. Understanding Mobile Reading via Camera Based Gaze Tracking and Kinematic Touch Modeling. In *Proceedings of the 2018 International Conference on Multimodal Interaction (ICMI'18),* Oct. 16-20, 2018, Boulder, CO, USA. ACM, NY, NY. 10 pages. https://doi.org/10.1145/1234567890

ICMI '18, October 16–20, 2018, Boulder, CO, USA

ACM ISBN 978-1-4503-5692-3/18/10…$15.00
https://doi.org/10.1145/3242969.3243011

1 INTRODUCTION

Mobile reading is experiencing rapid growth in the era of smartphones [59]. According to a recent survey, the time people spend on the mobile reading activities, such as reading articles on a social media app, reading email messages, or chatting with friends via instant messaging, is around 2 hours per day in the United States in 2016 [46], accounting for 15% of waking activity time [46]. Despite the enormous progress, reading non-pleasure contents on mobile devices for work or learning is still challenging. Recently, Neilson discovered that comprehension drops from 39.18% to 18.93% after switching from desktop screens to mobile-sized screens [40]. Indeed, compared to media consumption channels such as watching videos [54], the passive nature of mobile reading and the distracting environment often lead to declined attention and increased non-linear reading patterns [33].

Understanding users' read behaviors is a crucial first step towards improving mobile reading. However, most practitioners today still rely heavily on coarse-grained metrics such as click-through-rate (CTR) [21][56] and dwell time to investigate reading behaviors on mobile devices. Such approaches have been proven to be inadequate [17][21][56] due to the sparsity and ambiguity of the click and dwell signals. For example, extended dwell time may be caused by desirable content, increased difficulty, or external distractions.

We present Lepton (Figure 1), an intelligent mobile reading system and a set of dual-channel sensing algorithms, to achieve scalable and fine-grained understanding of users' reading behaviors on unmodified smartphones. Lepton tracks the *periodic lateral patterns*, i.e. saccade, of users' eye gaze via the front camera and infers users' *muscle stiffness* during text scrolling via a Mass-Spring-Damper (MSD) based kinematic model from touch events. Overall, Lepton combines a robust periodic saccade tracking channel via the front camera and a muscle stiffness tracking channel from text scrolling events to monitor and understand mobile reading.

This paper offers three major contributions:

- We propose a set of robust features on top of periodic saccade patterns of eye gaze from noisy gaze estimations of the front-facing camera in a smartphone.
- We use a kinematic model of hand-arm dynamics (MSD model) to quantify users' muscle stiffness during scrolling operations and then infer users' attention in reading.
- By combining rich features from both the periodic saccade tracking channel and muscle stiffness tracking channel, the

* This research was in-part conducted when JW was at the University of Pittsburgh. JW is currently with Google AI China Center.

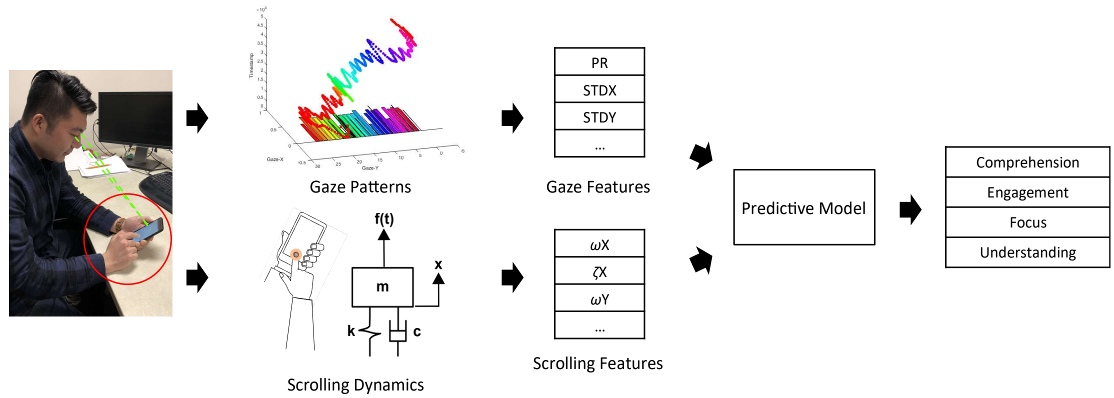


Figure 1. The architecture of Lepton: the visual channel (top) tracks the periodical patterns of users' eye gaze via the embedded front-facing camera of a smartphone. The kinematic touch channel (bottom) analyzes users' scrolling touch behavior via a Mass-Spring-Damper (MSD) model.

new features can significantly improve the accuracies in predicting users' comprehension and engagement in reading.

2 RELATED WORK

Reading is a process of translating signs and symbols into meanings and incorporating the new information into existing cognitive and affective structures [49]. Reading in digital formats was first prototyped by Alan Kay in 1968 [28], and received an increasing attention in recent decades. With the rise of mobile devices, mobile reading (the digital reading behavior based on handhold reading devices) has expanded in popularity [59].

Mobile reading has both pros and cons. It provides flexibility and interactivity when compared with reading in print. Besides reducing the size and weight of books and having scalable font size, mobile reading also supports active searching [41], and recording reading process [52]. Although digital reading was reported to be efficient in the lab environment [52], the more diversified usage environments and smaller screens can induce new challenges, such as declined attention and increased non-linear reading patterns [33][40]. We believe the rich sensors on smart phones and tablets provide an exciting opportunity to better understand and improve mobile reading.

Comprehension and engagement are critical indicators of successful reading. Comprehension is a process where a reader builds mental representations of text information [36][45] and reading engagement is a multidimensional construct consisting of users' cognitive, affective, behavioral characteristics during reading.

Many efforts have been explored to understand reading, including theoretical analysis [16][32][44][45][47], self-report questionnaires, clickstream analysis (e.g. CTR), and dwell time [1][56][21]. Unfortunately, all existing approaches have inherent limitations. For example, although theoretical models were essential for researchers to understand reading activities for desktop computers, but their impacts on mobile reading are limited due to the lack of proper theory and enabling technologies for scalable and fine-grained analysis. Offline measurements, such as self-report questionnaires, lack the ability to understand the moment-by-moment decisions of a reader during reading [19]. The sparsity and ambiguity of dwell time and CTR tracking also make them inadequate to understand reading. Verbal self-report (a.k.a. think aloud) is capable of tracking continuous and direct signals during reading, however, its reliability and validity are still under debate [1][5].

With the improvements in sensing technology, activities during reading, such as eye gaze movements [48], scrolling motions [12], screen-touching motions [42], and physiological signals [31] can be detected easier and more accurately, implying better opportunities to understand reading.

Eye-gaze Tracking

With the help of eye tracking technology, researchers have discovered a strong relationship between eye movements and cognitive processes.

Eye movements during reading are usually interpreted by low-level visual information comprising the basic characteristics such as saccades, fixations, return sweep, and blinking. A saccade is a rapid movement of eyes and a fixation is the 200-300ms relatively still of eyes between two saccades [11]. Such eye movement features were used to interpret the cognitive process [30], comprehension [48][7][35], proficiency [4][14][27][58], and engagement [41] in reading. Please refer to the survey by Rayner [48] on low-level visual information in reading. Unfortunately, robust gaze tracking requires either on-body electrooculography (EOG) sensors [31][10][11] or remote IR-based eye trackers [6][8][35][52]. Such equipment is both expensive and difficult to carry around in mobile environments.

In comparison, gaze tracking via a webcam is convenient but the accuracy is much lower when compared with dedicated eye trackers [34][55]. Inspired by SwitchBack [34], which tracked the periodic return sweep of gaze via a front camera to estimate the reading position, we propose a set of robust features on top of periodic saccade patterns of eye gaze. Such features can be extracted from noisy gaze estimations via front cameras of smartphones and can be used to analyze both low level reading behaviors and higher-level comprehension and engagement in reading.

Pointing & Scrolling

Although the computer mouse has a smaller throughput than eye gaze [43], some researchers showed that the mouse cursor could still serve as a good proxy for low-level visual information such as gaze [21][25][13][50]. For example, Chen [13] showed that the staying of a mouse cursor within an area of interest indicated locations of gazes with more than 75% accuracy; Huang [25] visualized the Euclidean distance between mouse cursors and gaze coordinates on a search engine page and showed a strong correlation between mouse cursors and gaze coordinates.

Mouse motion can also be used to infer users' affections [29][53] and subjective preferences [39]. SenticMouse [29] leveraged a pressure sensor on a mouse to predict users' affection (correlation>0.75) during image browsing. Claypool [15] found a strong correlation between mouse motions and reading engagement. Moustress [53] predicted users' stress with around 70% accuracy from common mouse activities, such as clicking, dragging, and steering. Inspired by Sun et al. [53] and Hill [24], which used mass-spring-damper (MSD) system to understand the dynamics of human arm motion when doing two-dimensional tasks, we apply this model to understand finger scrolling motions on smartphones. Other than the 2D mouse movements analyzed in Moustress, scrolling events on smartphones provide us more informative features such as scrolling pressure and touch size. We used both simple features such as the number of scrolls as in [15] as well as MSD features to predict users' reading comprehension and engagement.

Touch-screens are ubiquitous on digital reading devices nowadays [8]. Many studies have been conducted to understand or interpret scrolling activities [3][12][20][57]. Among them, Grusky [20] used scrolling to reveal the online viewport locations so as to understand reading. Campbell [12] classified different habits of scrolling among users. With compared to existing research on text scrolling, Lepton takes into account a new set of muscle stiffness features from an MSD model. Lepton also revealed a strong correlation between muscle stiffness features, reading comprehension, and engagement.

3 DESIGN OF LEPTON

Lepton uses a *periodic saccade tracking channel* and a *kinematic/muscle stiffness tracking channel* to understanding mobile reading, especially for metrics such as comprehension and engagement.

3.1 Periodic Saccade Tracking

Traditional eye-gaze features are derived from the first-order statistics of gaze fixations and saccades, e.g. the mean and standard deviation of the durations of the fixations and the lengths of saccades [48]. Gaze fixations and saccades are usually collected from commercial eye trackers. In Lepton, we intentionally chose a low-cost, low-resolution alternative – i.e. using the front cameras of smartphones as the sensing channel of eye gaze during reading. Although webcam-based gaze tracking has been explored by the research community for more than one decade, this approach is still inferior to dedicated gaze trackers in both accuracy and signal-noise ratio (SNR). As a result, most tried and true features from first-order statistics of gaze do not work well for camera-based gaze tracking. In comparison, the robust features of periodic saccade in Lepton were inspired by two observations in mobile reading: 1) Although the estimations of gaze fixations are noisy and inaccurate from webcam-based techniques, there is a strong periodic lateral movement pattern caused by line-by-line reading; 2) the small screen of mobile devices can afford to display fewer words per line, leading to more return sweeps of eye gaze. Such return sweeps can be clearly discovered even from noisy gaze estimations (Figure 2). In summary, although camera-based gaze tracking generates less accurate estimations of gaze fixations, the strong periodic lateral movement pattern and the frequent return sweeps of eye gaze allow us to have robust and accurate estimations on which line the reader is looking at during mobile reading. Inspired by the observations above, we propose the periodic saccade pattern-based eye gaze features in Table 1.

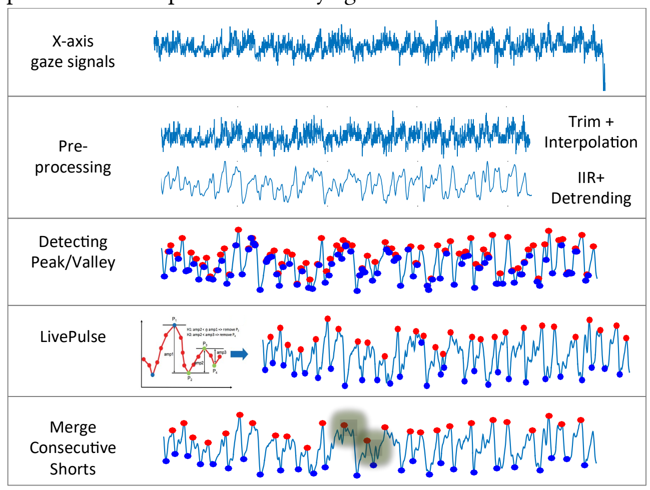


Figure 2. The framework for getting periodic saccade patterns. Top to Bottom: the process begins with horizontal (x-axis) gaze signals; through preprocessing, LivePulse algorithm and merging consecutive shorts, the final periodic saccade patterns were achieved.

Feature	Definition
PR	Predicted periodic lateral patterns divided by number of lines in reading material
STDX	Standard deviation of x-axis of gazes
STDY	Standard deviation of y-axis of gazes
rMSLL	The square root of the mean squared adjacent predicted line lengths' differences
rMSLD	The square root of the mean squared adjacent predicted line durations' differences
M1ADLL	Mean of absolute deviation of predicted line lengths
M1ADLD	Mean of absolute deviation of predicted line durations
M1ADLY	Mean of absolute deviation of line mean Y-axis of gazes
MADLL	Median of absolute deviation of predicted line lengths
MADLD	Median of absolute deviation of predicted line durations
MADLY	Median of absolute deviation of line mean Y-axis of gazes
STDLL	Standard deviation of predicted line lengths
STDLD	Standard deviation of predicted line durations
STDLY	Standard deviation of line mean Y-axis of gazes

Table 1. Periodic saccade pattern based eye gaze features.

We use the front camera of a Google Nexus 5X smartphone to extract gaze coordinates. The front camera captures each image frame during reading. The frames are then passed through

Qualcomm Snapdragon SDK [22] to accelerate the speed of gaze estimation. Our algorithm can achieve a rate of 20 frames per second on a Google Nexus 5X. The battery can last around 3 hours with gaze tracking on.

In our research, we run our text reading interface in portrait mode. To illustrate gaze estimations, we define the top-left corner as the origin. The x-axis increases from left to right and the y-axis increases from top to bottom. When a user is reading, the horizontal axis (x-axis) of her gazes will appear in a zig-zag periodic saccade pattern as she finishing reading a line and her gaze sweeping back to read another line [34].

Figure 2 illustrates the workflow of extracting the periodic saccade patterns from a user' gaze movements. Our algorithm takes a stream of a reader's horizontal gaze estimations (x-axis values) as input and goes through four steps, i.e. 1) preprocessing; 2) detecting peaks/valleys; 3) removing false peaks/valleys; and 4) merging consecutive short patterns.

During the preprocessing stage, Lepton will interpolate, scale and detrend the gaze signals. An Infinite Impulse Response (IIR) low-pass filter (2.5 Hz cutoff frequency) is used to remove gaze jittering. During the peak/valley detection stage, all local maximums and minimums of eye gaze are labeled as potential peaks and valleys. Then we use the LivePulse [23] algorithm, which is proven to be efficient in separating noises from stronger signals [23], to remove most of the false peaks and valleys. Lastly, the consecutive short return sweeps are merged since our reading materials are paragraph-based – it is unlikely to have multiple short return sweeps in a paragraph. The gaze pattern between two selected valley points is marked as a periodic saccade pattern, representing a user starting a line (the first valley), reaching its end (the middle peak), and sweeping the gaze back to the beginning of the next line (the second valley).

The rationale behind our periodic saccade pattern is – we are replacing the noisy gaze fixations with more robust "line fixations". Furthermore, gaze fixations require accurate coordinate estimations and calibrations, while the periodic saccade patterns only rely on the "zig-zag" regularity of gaze trajectories and are calibration-free.

We propose 14 robust features on the top of the periodic saccade patterns (Table 1). These features include descriptive statistics and temporal characteristics of the patterns.

Since the number of periodic saccade patterns equals to the number of lines, we defined two metrics to evaluate the accuracy of detecting periodic saccade patterns: A) the number of lines read by a reader, and B) the existence of non-linear reading actions such as reread and skip.

A. The number of lines

We can use the number of valley-to-valley periodic saccade patterns to estimate the number of lines read by a user.

B. Non-linear actions (reread & skip)

We propose two methods, 1) *X-line-counting* method, and 2) *Y-only* method, to detect the existence and location of the non-linear reading actions (skip/reread).

The *X-line-counting* method first estimates the number of periodic saccade patterns, and then compares it with the actual number of lines in the current page. If the ratio between the estimated line number and the actual line number is larger than a given threshold, it has a reread action in this page. Similarly, there exists a skip action if the ratio is smaller than a given threshold. We used grid search on training data to determine the optimal cut-off threshold for reread and skip action detection.

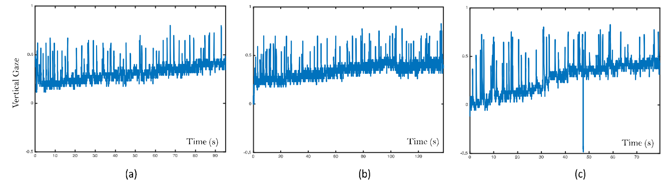


Figure 3. A reader's vertical gaze shape (y values) when (a) reading consecutively from top to bottom, (b) reading from top to bottom but rereading a set of paragraphs, (c) reading from top to bottom but skipping a set of paragraphs.

Another approach to detect reread/skip activities is to use vertical (y-axis) gaze signals based on the following observations: 1) when a reader reads line by line, the vertical axis (y-axis) of gazes follows an increasing step shape within a fixed viewport of the reading material (Figure 3.a); 2) When the reader rereads, her vertical gaze location plunges, and then follows the increasing step shape (Figure 3.b); and 3) When she skips, the vertical gaze data shoots up, and then follows the increasing step shape (Figure 3.c) . For the *Y-only* method, the vertical gaze data is passed through an outlier removal algorithm, a Finite Impulse Response (FIR) filter, and a sliding window action classifier to predict the reread and skip actions. The outlier removal algorithm aims to remove the occasional peaks of y-gaze signals caused by eye blinking (Figure 3). Then, a simple FIR average filter is used to remove the signal noise. Lastly, we use a sliding window action classifier (Figure 4) to predict the existence of non-linear reading actions. In this classifier, we define a y value at time t as $\mathbf{f(t)}$, and we classify a sample as reread/skip if there is at least one window having the slope direction change.

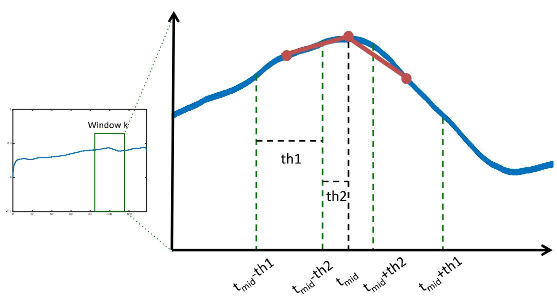


Figure 4. Sliding window reread action classifier. We classify a sample as reread if there is at least one window which has: $\frac{\mathbf{f(t_{mid})-f(t_i)}}{\mathbf{t_{mid}-t_i}} < \mathbf{0} < \frac{\mathbf{f(t_j)-f(t_{mid})}}{\mathbf{t_j-t_{mid}}}$ where $\mathbf{t_{mid} - th1 < t_i < t_{mid} - th2}$ and $\mathbf{t_{mid} + th2 < t_j < t_{mid} + th1}$.

3.2 Muscle Stiffness Tracking

Besides eye movements, we also extract users' muscle stiffness during text scrolling to infer comprehension and engagement in reading. Muscle activity/tension can be affected by cognitive and emotional states. Researchers have discovered that such muscle changes can be detected by mass-spring-damper (MSD) system via two-dimensional mouse steering and target acquisition tasks

[53][24]. However, it is still unclear whether an MSD model is applicable to reading activities on mobile devices. Taking the advantage of the rich sensors in smartphones, we propose to track, understand and use the muscle stiffness of readers via an MSD model and then infer their comprehension and engagement in reading.

In mobile reading, an MSD system consists of a mass (m) representing the reader's arm and finger(s), attached to a spring component (spring constant k), and a viscous damper (damping coefficient c) representing the muscle elements of the arm and finger(s). During reading, the mass oscillates at a rate related to the tension of the spring, and the oscillation decays exponentially based on the friction of the damper. Therefore, the damping frequency (ω) and damping ratio (ζ) of each MSD dimension can describe the scrolling motions of such dimension. We adopt the correlation between the parameters and muscle stiffness in [53]: $\boldsymbol{\omega} \propto \sqrt{\mathbf{k}}$ and $\boldsymbol{\zeta} \propto \frac{\mathbf{c}}{\sqrt{\mathbf{k}}}$. The MSD model takes the force from the finger(s) and arm as input, and then outputs the scrolling characters such as trajectory.

Since we aimed to predict muscle stiffness by the observed scrolling characteristics, we use linear predictive coding (LPC) to invert the input (muscle stiffness) and output (scrolling characters) of the MSD model. LPC model predicts future signals based on the linear combination of the observed signals in the past:

$$\hat{\mathbf{x}}_{\mathbf{n}} = \sum_{\mathbf{i=1}}^{\mathbf{p}} \mathbf{a_i x_{n-i}}$$

where $\hat{\boldsymbol{x}}_{\boldsymbol{n}}$ is the predicted signal value, $\mathbf{x_{n-i}}$ is the previous observed values, $\mathbf{a_i}$ is the predictor coefficient, and $\mathbf{p}$ is the order of the predictors [26] (p=4 in our design). We leverage the Least Square Fitting to estimate $\mathbf{a_i}$.

LPC takes the input of the observed scrolling change along each dimension, e.g. the list of the displacements on the x-axis, and produces a sequence of coefficients that defines the characteristic polynomial of the MSD system. We then take the complex root $(\mathbf{r})$ of the predicted polynomials, which reveals the damping characteristics of the MSD model in this case: damping frequency $\boldsymbol{\omega} = |\mathfrak{I}(\mathbf{r})|$, damping ratio $\boldsymbol{\zeta} = \frac{|\mathfrak{R}(\mathbf{r})|}{\|\mathbf{r}\|}$ [53]. Figure 5 shows the higher-level workflow from touch events to the estimation of the damping frequency and ratio.

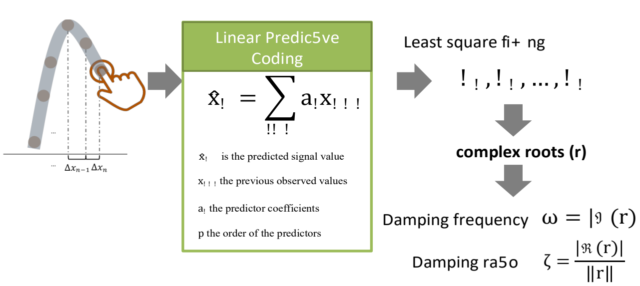


Figure 5. The illustration figure of extracting damping frequency and damping ratio features of a scrolling's horizontal dimension.

Besides displacements on horizontal (X) and vertical (Y) scrolling dimensions as in [53], displacements on three more dimensions were included: touch-size (S), touch-pressure (P) and touch-orientation ratio (R). Therefore, each scrolling extracted 2 MSD features ×5 dimensions. We then aggregate all scrolling features within an article/page reading into a feature vector via the descriptive statistics, such as mean and max.

4 EXPERIMENT

Our experiment consisted of two tasks. In the first task, we quantified the detection accuracy of the periodic saccade patterns. In the second task, we studied the performance of 1) saccade pattern-based gaze features, 2) MSD-based kinematic features, and 3) traditional dwell-time features on predicting comprehension and engagement in reading.

Figure 6. Sample participants our experiment.

4.1 Participants and Apparatus

25 subjects (9 females) ranging from 19 to 35 years old (μ = 26.32, σ = 3.96) were participated in this study. All participants had experiences with news reading on smartphones. None of the participants had dyslexia or emotional disorder.

We used a Google Nexus 5X smartphone in our study. The reading materials were displayed in portrait mode with a 15px display font size. Each screen can show 25 lines of text and each line has 11 words on average.

4.1.1 Task 1

In this task, we quantify the performance of: 1) estimating the number of lines read by a reader, and 2) predicting the existence of non-linear reading actions, i.e. skip and reread, based on the techniques descripted in section 3.

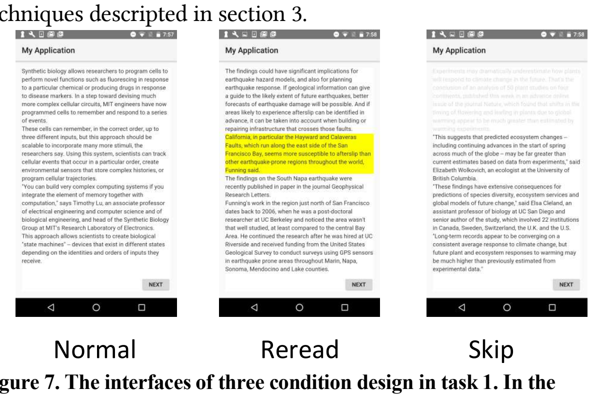


Figure 7. The interfaces of three condition design in task 1. In the normal condition, the participants read each line once, and line by line in a sequence. In the reread condition, participants read a randomly highlighted section twice. In the skip condition, participants skip a randomly grayed-out section when reached.

We investigated a total of three reading conditions, i.e. normal, reread, and skip in our study. In the normal condition, the participants were required to read the given article line by line in sequential order, only once for each text line (Figure 7 left). In the reread condition, participants were required to read a randomly highlighted section twice (Figure 7 middle) once they reach that paragraph. In the skip condition, participants were required to skip a randomly grayed-out section (Figure 7 right).

Each participant read one article per condition. The articles were chosen from New York Times, with comparable lengths (μ=588.5 words, σ=35.43) and difficulties (average Flesch-Kincaid reading ease = 29.03, σ = 8.10).

In this task, since *Y-only* method was applied in detecting non-linear reading actions, we chose a flipping-page design to avoid the possible confounding changes of vertical gaze by scrolling. Each article was divided into 3 pages (23~25 lines in each page). A flipping-page button was placed at the bottom right of each page. Participants clicked the flipping page button to continue to the next page (or stop after finishing with the article).

4.1.2 Task 2

In this task, our goal was to investigate whether lateral saccade pattern-based features and muscle stiffness features can be used to predict the comprehension and engagement in reading.

We used a scrolling-page design in this task. Each participant read 3 news articles in three different topics, including gaming, astronomy, and fitness. The articles in use had comparable lengths (μ=500.33, σ=58.29) and difficulty (average Flesch-Kincaid reading ease = 33.3, σ = 6.59). Each article had around 59 lines on a single scrollable page.

4.2 Procedure

The whole study lasted around one hour for each participant. After completing a background survey, participants went through two reading tasks in sequential order. There was a 10 min training and warm-up session before each task. After finishing each article, participants answered three short-answer questions to measure their comprehension. The questions included both literal and inferential contents. Then they reported the reading engagement including concentration level, confidence of understanding, and engagement after reading each article. We used a within-subject design in which both reading tasks and the orders of conditions were randomized.

4.3 Design and Analysis

4.3.1 Task 1

The data from one participant was discarded due to corruption. In total, we had 24 subjects × 3 conditions × 1 article × 3 flipping pages = 216 page level samples. We had two evaluation metrics 1) predicting number of lines read in a page; and 2) predicting the nonlinear reading actions.

We used the mean absolute error, mean absolute percentage error, root mean squared error and correlation to evaluate the number of lines. Precision-recall curve was used to analyze the prediction algorithm.

4.3.2 Task 2

The participants were instructed to read each article according to their reading habit. They may scroll, skip, or reread the article whenever they desired (Figure 6).

We used the accuracies in comprehensive questions to measure the comprehension in reading. We used self-reported concentration, engagement, and confidence of understanding to understand users' reading engagement on a 7-point Likert scale.

We used the forward-stepwise feature selection method to investigate the features from different channels and their effects on reading comprehension and engagement. Based on the selected features, the gains of reading comprehension and engagement were evaluated by 1) the root mean square error and R^2 value of linear fitting, and 2) the correlation coefficients of leave-one-subject-out user-independent validation.

5 RESULTS

5.1 Task 1 – Predicting the number of lines

We compared two baselines, i.e. SwitchBack [34] and ReadAllLines, with our periodic saccade detection algorithm in Lepton. We reproduced the SwitchBack algorithm based on descriptions in [34]. We did a parameter sweep to derive best thresholds for SwitchBack. The other baseline (ReadAllLines) assumed that each user would read each and every line once on each page.

	SwitchBack	ReadAll Lines	Periodical Pattern Detection
Mean absolute error	4.91	5.68	**3.56**
Mean absolute percentage error	0.2	0.31	**0.16**
Root mean squared error	6.66	7.66	**4.65**
Correlation	0.7	0.15	**0.83**

Table 2. Line detection results via SwitchBack (baseline 1), read all lines once (baseline 2) and periodical pattern detection methods.

As shown in Table 2, our proposed method achieved the highest correlations (0.83) and the lowest errors (e.g. mean absolute percentage error = 0.16) when compared with SwitchBack and ReadAllLines.

5.2 Task 1 – Detecting non-linear actions

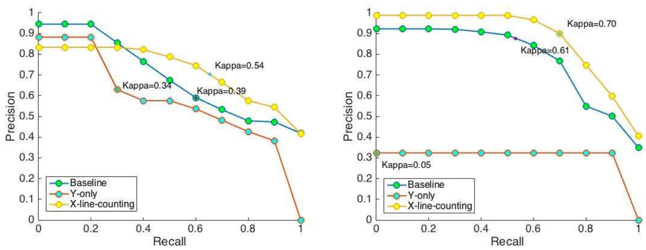
We calculated the precision and recall of each action (reread, skip, normal) versus all other actions. We compared our proposed methods (*X-line-counting* and *Y-only*) with a baseline. The baseline compares the actual dwell time with an expected dwell time threshold derived from a cutoff speed: a page reading action was treated as a skipping action if the dwell time is shorter than a cutoff threshold for skip actions. Similarly, an action was treated as a reread action if the dwell time is longer than a cutoff threshold for reread. By trying cut-off speeds ranging from 100 to 500 words per minute (wpm), we found that the optimal cutoff threshold for reread to be 150wpm and the optimal cutoff threshold for skip to be 250wpm. The precision-recall curves were shown in Figure 8. According to the area under curve (AUC) for reread and skip (Table 3), we found that *X-line counting* method outperforms both baseline and the *Y-only method* in our study.

In summary, both results in tasks 1 confirmed the accuracy and robustness of the features from periodic saccade patterns in mobile text reading.

5.3 Task 2 – Comprehension and engagement

We investigated three channels of features: traditional, periodic saccade patterns (gaze), and muscle stiffness (kinematic) (Table 4) on predicting the comprehension and engagement in mobile reading.

We used the forward step-wise feature selection method to select significant features from the three channels. According to Table 5, the gaze channel and the kinematic channel had significant features on different aspects of reading. We found that PR, MADLL, and MADLD were the significant features from the gaze channel. In the kinematic channel, the features related to vertical scrolling movement and scrolling pressures were significant. When a reader focuses or engages with reading, the increased muscle stiffness of the reader lead to an increased MSD damping ratio. The damping frequency increases when users are focusing. The finger scrolling pressure increases with decreased variance when the reader reads interesting articles.



(a) Reread (b) Skip

Figure 8. Reread and Skip condition precision-recall curve.

Method	Reread	Skip
Baseline	0.69	0.79
Y-only	0.58	0.31
X-line-counting	**0.73**	**0.88**

Table 3. Area under curve (AUC) for reread and skip conditions.

We updated the three sets of selected features (each included 3 features) to better predict users' comprehension and engagement: 1) all three features from the traditional channel; 2) The meanCY, meanCP, and stdCP features from the kinematic channel; and 3) the PR, MADLL, and MADLD features from the gaze channel.

Sources	Feature Group (Count)	Examples
Traditional (TF)	Dwell time (1)	Page duration.
	Speed Related (2)	Reading speed on word and character levels.
Kinematic (SF)	MSD related (40)	Statistical features related to five dimensions of MSD parameters.
	Scroll trail related (5)	Number of scrolls, mean duration of scrolls, etc.
Gaze (GF)	Periodic lateral pattern based (14)	Details in Table 1.

Table 4. Three sources of features: traditional, gaze and kinematic, for predicting reading comprehension and engagement.

With the selected features, we evaluated the root mean square error and R^2 value of from a liner model. In Table 6, we found that the combination of three sets worked the best for predicting reading engagements and comprehension.

We also evaluated the correlation coefficients in user-independent linear regression models via leave-one-subject-out validation. In Table 7, we found that the features in kinematic and gaze channel helped improve the correlation coefficients than TF on predicting concentration (0.36 vs. 0.16), confidence (0.5 vs. 0.47), engagement (0.34 vs. 0.16) and comprehension (0.36 vs. 0.29).

Features	Concentra-tion	Confidence	Engageme nt	Comprehen-sion
mean CY	1.62**	-	2.10**	-
min WY	2.07**	-	-	-
min WX	2.15**	-	-	-
mean CP	-	-	1.20*	0.78*
std CP	-2.44**	-3.05*	-	-
mean WR	-	-	-	6.94***
max WR	2.90*	-	-	-
MAD lineDur	-	-1E-03**	-	-
MAD lineLen	-	-	-3E-04*	-

Table 5. The correlation and corresponding p-values of the features selected for reading engagement and comprehension via forward step-wise method, where *: p-value <0.05; **: p-value <0.01; ***: p-value <0.001.

Features	Concentra-tion	Confidence	Interesting-ness	Comprehe-nsion
TF (3)	1.00(0.11)	1.07(0.33)	1.46(0.09)	0.86(0.17)
SF (3)	0.96(0.19)	1.22(0.14)	1.41(0.14)	0.90(0.09)
GF (3)	1.03(0.05)	1.22(0.13)	1.47(0.07)	0.91(0.07)
TF+SF	0.92(0.28)	1.04(0.39)	1.36(0.23)	0.85(0.22)
TF+GF	1.00(0.14)	1.04(0.39)	1.44(0.14)	0.86(0.20)
SF+GF	0.96(0.21)	1.17(0.23)	1.39(0.20)	0.89(0.15)
TF+SF+GF	**0.92(0.30)**	**1.02(0.44)**	**1.36(0.26)**	**0.84(0.26)**

Table 6. The root mean square errors (the smaller the better) and the corresponding R^2 value (the larger the better) for reading concentration, confidence, engagement and comprehension via tradition features (TF), scrolling features (SF) and gaze features (GF).

6 DISCUSSIONS

When designing Lepton, our major goal is to achieve scalable understanding of mobile reading activities. Such a goal has at least two implications in design: 1) we choose *support* rather than *change* existing reading behaviors among mobile users. For example, we assume that users will read an article in portrait mode; 2) We choose not to include additional sensors (e.g. gaze trackers, and EEG headbands) or hardware modifications to existing smartphones. Such changes will prevent us from deploying Lepton in large scale; 3) We choose to complete all the sensing and inference algorithms on device. Otherwise intermittent Internet connections may break Lepton. Even so, turning on the front camera during reading may still raise concerns from privacy-sensitive users.

6.1 Periodic Saccade Tracking

There are two advantages for the periodic saccade tracking channel in Lepton. First, it achieves a good balance in both accuracy and robustness when compared with alternative approaches such as dwell time and camera-based gaze fixation tracking; Second, this periodic saccade tracking channel is calibration free. It relies on the periodic changes of lateral gaze movement rather than absolute locations of gaze fixations. Essentially speaking, our approach replaces *word-level* fixation tracking to *line-level* periodic saccade tracking. Robust line-level reading process tracking can help us to have a deeper understanding of mobile reading activities in large scale.

Features	Concentr-ation	Confidence	Engageme nt	Comprehen-sion
TF (3)	0.16	0.47	0.16	0.29
SF (3)	0.29	0.16	0.24	0.19
GF (3)	-0.17	0.21	0.13	0.05
TF+SF	**0.36**	0.49	0.33	**0.36**
TF+GF	0	0.49	0.18	0.25
SF+GF	0.17	0.28	0.30	0.20
TF+SF+GF	0.25	**0.5**	**0.34**	0.31

Table 7. Correlation coefficients by leave-one-subject-out validation on linear regression models via different feature sources and different combinations of feature sources

The error rate of our reproduced SwitchBack algorithm was higher than that in the original literature [34] (mean absolute percentage error increased from 3.9% to 20%). We suspect the difference was caused by two reasons: First, Lepton runs in portrait mode rather than the landscape mode of SwitchBack [34]. The lateral gaze movement distance in landscape mode is at least 1.5 times longer than the distance in portrait mode. As such, a global threshold in SwitchBack [34] could not detect the line break accurately. The landscape mode also leads to fewer number of lines per screen, hence reducing the space of possible line numbers; Second, SwitchBack highlights the next line to read if a reader switches visual attention. As such, SwitchBack won't be able to generate a line number larger than the total number of lines. Meanwhile, Lepton allows rereading and a user can read more lines per screen than the number of lines displayed.

The *Y-only action detection* also had a much lower accuracy when compared with *X-line-counting action detection* in our study. After taking a closer look at the failure cases together with experimental videos recorded, we noticed that most of the failures were triggered by large body movements. We noticed that posture adjustments in reading have a much stronger impact on gaze estimations in the y-axis than the x-axis. We suspect accelerometer signals may give us hints when a user is adjusting body posture in reading. Such information can help us improve *Y-only action detection* in the future.

6.3 Modality Comparison

As shown in section 5.3, the combination of the periodic saccade channel and the kinematic channel in Lepton can significantly improve the prediction accuracy of comprehension and engagement when compared with mainstream signals such as dwell time. According to Table 5, periodic saccade features worked better in predicting reading confidence, while scrolling signals alone worked better in predicting reading comprehension, concentration, and engagement. One possible explanation could be - confident users have smooth paces in reading, i.e., all lines are read at a steady speed, except for the short lines.

The periodic saccade channel and the kinematic channel can complement each other in signal frequency and usage environments. The periodic saccade channel can give us continual observations on line-by-line reading processes. Meanwhile there are fewer scrolling operations per page. For example, in task 2, there were 4 to 78 scrolls per article (μ=18.87, σ=13.48), accounting for around one fourth of the total reading time (μ=23.58%, σ=0.18). In comparison, there were around 24 periodic saccade patterns per page. There are also advantages in the kinematic channel. The kinematic channel in Lepton is not sensitive to posture changes and illumination changes, while the periodic saccade channel is sensitive to major posture changes and will not work in dark environments.

7 CONCLUSIONS AND FUTURE WORK

We presented Lepton, an intelligent mobile reading system and a set of dual-channel sensing algorithms to achieve scalable and fine-grained understanding of users' reading behaviors, comprehension, and engagement on unmodified smartphones. Lepton tracks the periodic lateral patterns, i.e. saccade, of users' eye gaze via the front camera, and infers their muscle stiffness during text scrolling via a Mass-Spring-Damper (MSD) based kinematic model from touch events. Lepton leverages signals from these two channels to infer users' comprehension and engagement during reading. Through a 25-subject study, we found that both the periodic saccade patterns and muscle stiffness signals captured by Lepton can be used as expressive features to infer users' comprehension and engagement in mobile reading. Overall, our new signals lead to significantly higher performances in predicting users' comprehension (+53% in R^2), concentration (+173%), and confidence (+33%) than using traditional dwell-time based features.

We plan to explore the following directions in the near future. First, Lepton primarily focuses on understanding line-level reading progress and page-level comprehension and engagement, can we use Lepton, together with supplemental information such as application logs, to understand high-level reading strategies on mobile devices? For example, how could a user search, compile, and read a set of articles to understand a controversial topic, such as "*mountaintop coal mining removal*"; Second, we plan to explore interactive technologies, such as personalized recommendation, smart highlighting, or in-situ quizzes when low engagement is detected; Third, we are interested in exploring privacy-preserving techniques to minimize users' concerns on camera-based gaze tracking during reading; Fourth, we are interested in exploring supplemental sensing channels, such as motion and location, in mobile reading. For example, Bronzaft and McCarthy [9] discovered that the environmental noises had a significant impact on comprehension. We believe that understanding users' mobile context will be important towards facilitation their reading experiences as well.

We thank Xiang Xiao, Xiangmin Fan, Phuong Pham, Shumin Zhai, Zhenyuan Yang, and anonymous reviewers for the constructive feedback. This research was in-part supported by a gift from Byte Dance Telecommunications to the University of Pittsburgh.

8 REFERENCES

[1] Peter Afflerbach, and Byeong-Young Cho. 2009. "Identifying and describing constructively responsive comprehension strategies in new and traditional forms of reading." *Handbook of research on reading comprehension*: 69-90.

[2] Deepak Agarwal, Bee-Chung Chen, and Xuanhui Wang. "Multi-faceted ranking of news articles using post-read actions." *Proceedings of the 21st ACM international conference on Information and knowledge management.* ACM, 2012.

[3] Pär-Anders Albinsson, and Shumin Zhai. "High precision touch screen interaction." *Proceedings of the SIGCHI conference on Human factors in computing systems.* ACM, 2003.

[4] Olivier Augereau et al. "Towards an automated estimation of English skill via TOEIC score based on reading analysis." *In Pattern Recognition (ICPR)*, 2016 23rd International Conference on, pp. 1285-1290. IEEE, 2016

[5] Tom Baranowski. 1988. "Validity and reliability of self report measures of physical activity: an information-processing perspective." *Research Quarterly for Exercise and Sport* 59.4: 314-327.

[6] David Beymer, and Daniel M. Russell. "WebGazeAnalyzer: a system for capturing and analyzing web reading behavior using eye gaze." *CHI'05 extended abstracts on Human factors in computing systems.* ACM, 2005.

[7] Ralf Biedert, Georg Buscher, and Andreas Dengel. 2010. "The eyebook–using eye tracking to enhance the reading experience." *Informatik-Spektrum*, 33(3), 272–281.

[8] Ralf Biedert, Andreas Dengel, Georg Buscher, and Arman Vartan. 2012. "Reading and estimating gaze on smart phones." *Proceedings of the symposium on eye tracking research and applications.* ACM.

[9] Arline L. Bronzaft and Dennis P. McCarthy. 1975. "The effect of elevated train noise on reading ability." *Environment and behavior* 7.4: 517-528.

[10] Andreas Bulling, and Daniel Roggen. "Recognition of visual memory recall processes using eye movement analysis." *Proceedings of the 13th international conference on Ubiquitous computing*. ACM, 2011.

[11] Andreas Bulling, Jamie A. Ward, Hans Gellersen, and Gerhard Tröster. 2008. "Robust recognition of reading activity in transit using wearable electrooculography." *International Conference on Pervasive Computing*. Springer Berlin Heidelberg.

[12] Christopher S. Campbell, and Paul P. Maglio. 2001. "A robust algorithm for reading detection." *Proceedings of the 2001 workshop on Perceptive user interfaces.* ACM.

[13] Mon Chu Chen, John R. Anderson, and Myeong Ho Sohn. "What can a mouse cursor tell us more?: correlation of eye/mouse movements on web browsing." *CHI'01 extended abstracts on Human factors in computing systems.* ACM, 2001.

[14] Shiwei Cheng, Zhiqiang Sun, Lingyun Sun, Kirsten Yee, and Anind K. Dey. 2015. "Gaze-Based Annotations for Reading Comprehension." *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* 1569–1572.

[15] Mark Claypool, Phong Le, Makoto Wased, and David Brown. 2001. "Implicit interest indicators." *Proceedings of the 6th International Conference on Intelligent User Interfaces.* 33–40.

[16] Edmund Burke Huey. 1908. "The psychology and pedagogy of reading." *The Macmillan Company.*

[17] Georges Dupret, and Ciya Liao. 2010. "A model to estimate intrinsic document relevance from the clickthrough logs of a web search engine." *In Proceedings of the third ACM international conference on Web search and data mining.* ACM. 181–190.

[18] Lyndsey Franklin, Kristina Lerman, and Nathan Hodas. 2017. "Will Break for Productivity: Generalized Symptoms of Cognitive Depletion." *arXiv preprint arXiv*:1706.01521.

[19] Simon Garrod, and Meredyth Daneman. 2003. "Reading, Psychology of." *Encyclopedia of Cognitive Science*, 848–854.

[20] Max Grusky, Jeiran Jahani, Josh Schwartz, Dan Valente, Yoav Artzi, and Mor Naaman. 2017. "Modeling Sub-Document Attention Using Viewport Time." *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.* 6475–6480.

[21] Qi Guo, and Eugene Agichtein. 2012. "Beyond dwell time: estimating document relevance from cursor movements and other post-click searcher behavior." *Proceedings of the 21st international conference on World Wide Web.* ACM.

[22] Wei Guo, and Jingtao Wang. 2017. "SmartRSVP: Facilitating Attentive Speed Reading on Small Screen Wearable Devices." *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems.* ACM.

[23] Teng Han, Xiang Xiao, Lanfei Shi, John Canny, and Jingtao Wang. 2015. "Balancing Accuracy and Fun: Designing Camera Based Mobile Games for Implicit Heart Rate Monitoring." *In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* 847-856.

[24] AV Hill. 1938. "The heat of shortening and the dynamic constants of muscle." *Proceedings of the Royal Society of London. Series B, Biological Sciences* 126, 843, 136–195.

[25] Jeff Huang, Ryen W. White, and Susan Dumais. "No clicks, no problem: using cursor movements to understand and improve search." *Proceedings of the SIGCHI conference on human factors in computing systems.* ACM, 2011.

[26] Leland B. Jackson. 2013. "Digital Filters and Signal Processing: With MATLAB® Exercises." *Springer Science & Business Media.*

[27] Jakob Karolus, et al. "Robust Gaze Features for Enabling Language Proficiency Awareness." *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.* ACM, 2017.

[28] Alan C. Kay, FLEX-a flexible extendable language. No. TR-4-7. UTAH UNIV SALT LAKE CITY SCHOOL OF COMPUTING, 1968.

[29] Dana Kirsch. "The sentic mouse: Developing a tool for measuring emotional valence." *Bachelor of Science in Brain and Cognitive Sciences*, MIT (1997).

[30] Reinhold Kliegl, Antje Nuthmann, and Ralf Engbert. "Tracking the mind during reading: The influence of past, present, and future words on fixation durations." *Journal of experimental psychology*: General 135.1 (2006): 12.

[31] Kai Kunze, Susana Sanchez, Tilman Dingler, Olivier Augereau, Koichi Kise, Masahiko Inami, and Terada Tsutomu. 2015. "The Augmented Narrative: Toward Estimating Reader Engagement." *Proceedings of the 6th Augmented Human International Conference*, 163–164.

[32] David Laberge, S. Jay Samuels. 1974. "Toward a theory of automatic information processing in reading." *Cognitive Psychology*, 6, 293–323

[33] Ziming Liu. 2005. "Reading behavior in the digital environment: Changes in reading behavior over the past ten years." *Journal of documentation* 61.6 700-712.

[34] Alexander Mariakakis, Mayank Goel, Md Tanvir Islam Aumi, Shwetak N. Patel, and Jacob O. Wobbrock. 2015. "SwitchBack: Using Focus and Saccade Tracking to Guide Users' Attention for Mobile Task Resumption." *In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 2953-2962). ACM.

[35] Pascual Martínez-Gómez, and Akiko Aizawa. 2014. "Recognition of understanding level and language skill using measurements of reading behavior." *Proceedings of the 19th international conference on Intelligent User Interfaces.* ACM.

[36] Danielle S. McNamara, and Joe Magliano. "Toward a comprehensive model of comprehension." *Psychology of learning and motivation* 51 (2009): 297-384.

[37] Claire Cain Miller, and Julie Bosman. 2011. "E-Books Outsell Print Books at Amazon." New York Times 19.

[38] Emiliano Miluzzo, Tianyu Wang, and Andrew T. Campbell. 2010. "EyePhone: activating mobile phones with your eyes." *Proceedings of the second ACM SIGCOMM workshop on Networking, systems, and applications on mobile handhelds.* ACM.

[39] Vidhya Navalpakkam, and Elizabeth Churchill. "Mouse tracking: measuring and predicting users' experience of web-based content." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 2012.

[40] Jakob Nielsen. 2011. "Mobile Content Is Twice as Difficult." Nielsen Norman Group. https://www.nngroup.com/articles/mobile-content-is-twice-as-difficult-2011.

[41] Heather L. O'Brien, Luanne Freund, and Richard Kopak. "Investigating the role of user engagement in digital reading environments." Proceed*ings of the 2016 ACM on Conference on Human Information Interaction and Retrieval.* ACM, 2016.

[42] JongHwan Oh, SungJin Nam, and Joonhwan Lee. 2014. "Generating Highlights Automatically from Text-Reading Behaviors on Mobile Devices." *In CHI'14 Extended Abstracts on Human Factors in Computing Systems* (pp. 2317-2322). ACM.

[43] Oyewole Oyekoya, and F. W. M. Stentiford. "A performance comparison of eye tracking and mouse interfaces in a target image identification task." (2005): 139-144.

[44] Charles Perfetti, M. G. McKeown, and L. Kucan. 2010. "Decoding, vocabulary and comprehension." *Bringing Reading Research to Life*, 291-303.

[45] Charles A. Perfetti, Nicole Landi, and Jane Oakhill. 2005. "The Acquisition of Reading Comprehension Skill." *The Science of Reading: A Handbook*, 227–247.

[46] Sarah Perez. 2017. "U.S. Consumers Now Spend 5 Hours per Day on Mobile Devices." *TechCrunch*, 3 Mar. 2017, techcrunch.com/2017/03/03/u-s-consumers-now-spend-5-hours-per-day-on-mobile-devices

[47] Josephine A. Piekarz. "Individual Differences in Interpretative Responses in Reading." PhD dissertation. University of Chicago, Department of Education, 1954.

[48] Keith Rayner. 1998. "Eye movements in Reading and Information Processing: 20 Years of Research." *Psychological Bulletin*, 124(3), 372–422.

[49] Mildred Coen Robeck, and Randall Reed Wallace. 1990. "The Psychology of Reading: an Interdisciplinary Approach." Lawrence Erlbaum.

[50] Kerry Rodden, et al. "Eye-mouse coordination patterns on web search results pages.*" CHI'08 extended abstracts on Human factors in computing systems.* ACM, 2008.

[51] Ben Shneiderman. "Touch screens now offer compelling uses." *IEEE software* 8.2 (1991): 93-94.

[52] Eva Siegenthaler, et al. "Comparing reading processes on e-ink displays and print." Displays 32.5 (2011): 268-273.

[53] David Sun, Pablo Paredes, and John Canny. 2014. "MouStress: detecting stress from mouse motion." *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems* 61–70.

[54] Jim Trelease. 2013. The read-aloud handbook. Penguin.

[55] Erroll Wood and Andreas Bulling. 2014. EyeTab: model-based gaze estimation on unmodified tablet computers. *In Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '14).* ACM, New York, NY, USA, 207-210. DOI: https://doi.org/10.1145/2578153.2578185

[56] Xing Yi, Liangjie Hong, Erheng Zhong, Nanthan Nan Liu, and Suju Rajan. 2014. "Beyond clicks: dwell time for personalization." *Proceedings of the 8th ACM Conference on Recommender systems.* ACM.

[57] Jibin Yin, and Xiangshi Ren. "ZWPS and pressure scroll: Two pressure-based techniques in pen-based interfaces." *IPSJ Digital Courier* 3 (2007): 767-778.

[58] Kazuyo Yoshimura et al. "The eye as the window of the language ability: Estimation of English skills by analyzing eye movement while reading documents." *Document Analysis and Recognition (ICDAR),* 2015 13th International Conference on. IEEE, 2015

[59] Liyi Zhang, and Wei Ma. 2011. "Correlation analysis between users' educational level and mobile reading behavior." *Library Hi Tech* 29(3) 424-435.