Understanding and Improving Mobile Reading via Scalable and Low Cost Sensing

by

Wei Guo

Bachelor, University of Pittsburgh, 2014

Submitted to the Graduate Faculty of

the School of Computing and Information in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2019

UNIVERSITY OF PITTSBURGH

SCHOOL OF COMPUTING AND INFORMATION

This thesis was presented

by

Wei Guo

Advisor:

Jingtao Wang, Ph.D., Department of Computer Science

Adriana Kovashka, Ph.D., Department of Computer Science

Committee:

Diane Litman, Ph.D., Department of Computer Science

Byeong-Young Cho, Ph.D., Department of Instruction and Learning

Understanding and Improving Mobile Reading via Scalable and Low Cost Sensing

Wei Guo

University of Pittsburgh, 2019

In recent years, due to the increasing ubiquity of Internet and mobile devices, mobile reading on smart watches and smartphones is experiencing rapid growth. Despite the great potential, new challenges are brought. Compared to traditional reading, mobile reading faces major challenges such as encountering more frequent distractions and lacking portable and efficient technique to deeperly understand and improve it.

Fortunately, the development of the hardware and software of mobile devices provide an opportunity to track users' behavior and physiological signals accurately in a low-cost and portable manner. In this thesis, I explored the usage of low-cost mobile sensors to solve the measurement challenges of reading.

I used the *low-cost mobile sensing techniques* on mobile devices to understand and improve the degree and quality of reading. In this thesis, I first present SmartRSVP, a reading interface on smart watches that leverages eye-gaze contact tracking technique and heart rate sensing technique to facilitate reading under distractions. I then present Lepton, an intelligent reading system on smart phones that tracks eye-gaze periodical patterns and sensing the screen touching behavior to monitor readers' cognitions and emotions during reading. Lastly, I present StrategicReading, which uses the implicitly captured eye gaze patterns, scrolling motions, and log histories to monitor users' reading strategies and performance during multiple-sources online reading.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS **9**

1.0 INTRODUCTION **11**

1.1 READING ON WATCH-SIZE DEVICES **12**

1.2 READING ON POCKET-SIZE DEVICES **14**

1.3 DEVELOPED SYSTEMS **15**

1.4 THESIS STATEMENT AND HYPOTHESIS **19**

1.5 THESIS CONTRIBUTION **20**

1.6 THESIS OUTLINE **20**

2.0 SMARTRSVP: TOWARDS ATTENTIVE SPEED READING ON SMALL SCREEN WEARABLE DEVICES **21**

2.1 INTRODUCTION **21**

2.2 DESIGN **23**

2.2.1 RSVP **24**

2.2.2 Visual Attention Tracking **24**

2.2.3 Cognitive State Inference **26**

2.2.4 Speed regulation **26**

2.3 EXPERIMENTS & RESULTS **27**

2.3.1 User Study 1 ..... **28**

2.3.1.1 Participants **28**

2.3.1.2 Apparatus **28**

2.3.1.3 Procedure **30**

2.3.1.4 Design & Analysis **30**

2.3.1.5 Results **31**

2.3.2 User Study 2 ..... **33**

2.3.2.1 Participants **33**

2.3.2.2 Apparatus **34**

2.3.2.3 Procedure **34**

2.3.2.4 Design & Analysis **35**

2.3.2.5 Results **35**

2.3.3 User Study 3 ..... **37**

2.3.3.1 Participants **37**

2.3.3.2 Apparatus **38**

2.3.3.3 Procedure **38**

2.3.3.4 Design & Analysis **38**

2.3.3.5 Results **38**

2.3.4 User Study 4 ..... **40**

2.3.4.1 Participants **40**

2.3.4.2 Apparatus **41**

2.3.4.3 Procedure **41**

2.3.4.4 Design & Analysis **41**

2.3.4.5 Results **42**

2.4 DISCUSSIONS **43**

2.5 LIMITATIONS AND FUTURE WORK **44**

2.6 CONCLUSION **45**

3.0 LEPTON: UNDERSTAND MOBILE READING VIA CAMERA BASED GAZE TRACKING AND KINEMATIC TOUCH MODELING 46

3.1 INTRODUCTION 46

3.2 DESIGN 48

3.2.1 Visual Tracking Channel 48

3.2.1.1 Line detection 51

3.2.1.2 Action (reread & skip) detection 51

3.2.2 Kinematic channel 54

3.3 EXPERIMENTS 55

3.3.1 Participants 55

3.3.2 Apparatus 56

3.3.2.1 Phase 1 56

3.3.2.2 Phase 2 57

3.3.3 Procedure 57

3.3.4 Design and Analysis 58

3.3.4.1 Phase 1 58

3.3.4.2 Phase 2 59

3.4 RESULTS 59

3.4.1 Phase 1 – Line detection 59

3.4.2 Phase 1 – Action (reread & skip) detection 60

3.4.3 Phase 2 – Reading comprehension and engagements 62

3.5 DISCUSSIONS 66

3.5.1 Periodic Saccade Tracking 66

3.5.2 Modality Comparison **68**

3.5 LIMITATIONS AND FUTURE WORK **69**

3.6 CONCLUSION **70**

4.0 STRATEGICREADING: MEASURING STRATEGIC READING ON SMARTPHONE AUTOMATICALLY **71**

4.1 INTRODUCTION **71**

4.2 RELATED WORK **73**

4.3 DESIGN **76**

4.3.1 Gaze Movements **76**

4.3.2 Scrolling Actions **77**

4.3.3 Logging Paths **78**

4.4 EXPERIMENT **79**

4.4.1 Participants and Apparatus **79**

4.4.2 Experiment Procedure **80**

4.4.3 Metrics Design **81**

4.4.3.1 Reading Performances **81**

4.4.3.2 Reading Strategic Processes **82**

4.5 RESULTS **84**

4.5.1 Rationality of Adopting Reading Strategic Processes from Desktop to Smartphone **85**

4.5.2 Auto-extract Reading Strategies **86**

4.5.2.1 Gaze Channel **86**

4.5.2.2 Scrolling Channel **87**

4.5.2.3 Logging Channel 88

4.5.2.4 Overall Comparison 89

4.5.3 Predict Reading Performance 90

4.5.4 Potential patterns 92

4.5.4.1 Searching vs. Reading 92

4.5.4.2 Domain Reliability Levels 93

4.5.4.3 Scrolling Habits 94

4.6 DISCUSSIONS 97

4.6.1 Channels Comparison 97

4.6.2 Sensing Technique 98

4.6 LIMITATIONS AND FUTURE WORK 98

4.7 CONCLUSIONS 100

5.0 CONLUCSION 101

5.1 CONTRIBUTION 101

5.2 FUTURE WORK 101

ACKNOWLEDGEMENTS

It is an incredible journey from a greenhand of research to getting my PHD diploma. I would call it fortunate.

I mainly own my fortune to my advisor, Dr. Jingtao Wang. At the beginning of my PHD career, I didn't even know what should a research paper contain. My advisor used a few years fed from a baby PHD to a qualified one. He is a very smart person and has a good vision in research. In a gradual process, he started by giving specific guidance and then he slowly released the hands. The guidances could be as specific as bolding the labels in a graph and they could be as deep as planning of life. His vision helped me avoid a lot of detours in my research study. Even if he left Pitt the last year of my PHD study, he was there offering guidance and help whenever I need. If there is one phrase to describe what he meant to my Ph.D. study and even my life, I would like to use "mother".

At the same time, I would like to express my sincere gratitude for Dr. Adriana Kovashka, my co-advisor. She is generous and virtuous. At my most difficult time, she lent me a hand and saved my from miserable state. I appreciate her supportive suggestions and help on my research.

I want to express my gratefulness to my committee member Dr. Boyong-Young Cho. He is also my project mentor and collaborator. He selflessly provided numouse guidances to my research project in his field of expertise. I also want to express my thanks to my committee member, Dr. Diane Litman. She provided a lot of constructive insights and suggestions on my research projects and my thesis.

I am grade that my Ph.D. period was spent in University of Pittsburgh and I feel grateful that I have the best committee members.

I am also very grateful to Dr. Taieb Znati, the dean of my department. He became my solid support and offered to be my plan B when I was frustrating. I really appreciate it.

I would like to thank my lab mates, Xiang Xiao, Xiangmin Fan, and Phuong Pham. As my senior fellow apprentice, they provided me valuable and practical suggestions. I would like to thank my classmates and friends during my Ph.D. period. Thank you all for letting me have a wonderful time in my last years of study. I am sure that the memory will last forever and shine forever.

Many thanks to my parents for giving me my life and guiding me to understand the beauty of life.

Last but not least, I would like to thank my husband. Yes! He evolved from boyfriend to husband! I am grateful to have him with me and growing up with me.

1.0 INTRODUCTION

In recent years, the increasing ubiquity of the Internet and mobile devices enormously affects reading and brings new challenges. Coming with portability, capacity, free sources, and gain of space [76], mobile reading is experiencing rapid growth. Among all kinds of mobile devices, smartphones and pads are commonly used for reading. According to a recent survey, reading activities on smartphones and pads, such as reading articles on a browser or dedicated social media apps and reading messages in email clients or instant messaging (IM) apps, took around 2 hours per day in the United States in 2016, accounting for 15% of the waking activity time [65]. Besides smartphones and pads, reading on small-screen wearable devices is flourishing as well. Taking advantage of staying on users' wrists 24/7, smart watches and wristbands can assist the access of frequent reading tasks amd important notifications. Despite the great potential, both watch-size and phone-size mobile reading faces new challenges.

There are at least three major challenges when a user reads text messages on a smart watch. First, most of today's smart watches use a small display, approximately 1.5-inch diagonal, which only affords showing three or four words per line. Therefore, a user has to rely on more frequent lateral eye gaze movements, i.e. saccade, during reading. Second, more scrolling actions are needed due to the limited number of words per screen. Such scrolling operations not only exacerbate the notorious "fat finger problem" but also occupy both of a user's hands. Third, text reading on smart watches increases the probability of divided attention and interruptions [52].

Although the pocket-size screen reading avoids the problem of watch-size screen reading, it has new challenges due to the complexity of its reading contents and purposes.

When reading on smart watches, readers tend to quickly acquare key information from small messages or emails. Comparatively, the reading on smart phones and pads has various contents and is for diversified purposes. Readers read newspapers, magazines, books, and etc. on smart phones and pads. The challenges are no longer how to display information to the users as on smart watches. Instead, it is important to let the user better consume the contents. The first step is to understand reading on such pocket-size devices.

First, compared to media consumption channels, such as watching videos, the passive nature of mobile reading and the complicated reading context might cause attention decline, non-linear reading patterns and context distractions. Despite the highly diversified contexts, practitioners today still rely heavily on coarse-grained metrics such as click-through-rate (CTR) and dwell time to understand and adapt to users' reading behavior on mobile devices [87][32][25]. These approaches have been proven to be inadequate because of the sparsity and indirectness of the signals. Second, little attention is paid to higher-level mobile reading skills. In a complicated mobile reading environment, the skill set for processing a single print text is insufficient for performing the complicated reading tasks for learning and working purposes [17]. Measuring, understanding and improving higher-level mobile reading skills are important and critical for both research and practice, but they still remain unexplored.

Many new techniques have been proposed to address the challenges faced by watch-size and phone-size mobile reading. Existing techniques have pros and cons.

1.1 Related Work: Reading on Watch-Size Devices

Researchers propose to address the display affordance problem, tedious scrolling problem, distractive environment problem when reading on watch-size screens.

To solve the text display and scrolling problem on watch-size screens, some researchers have used Rapid Serial Visual Presentation (RSVP)[1] as a speed-reading technique on small-screen devices [13][62][77]. RSVP has both spatial and temporal efficacy when compared with traditional reading [51] and it is proved to have the potential of increasing the reading speed without sacrificing the comprehension level [77][85][62][63][75]. The adaptations of words' exposure time in RSVP [62][63][6][85] further reduce the negative effects of RSVP on attentional blink [70], repetition blindness [44], higher visual attention demands [13], higher recovery cost [70], and higher cognitive workload [62][63][77]. However, the content-based speed adaptations are not environmentally adaptive, especially in distractive environments.

To facilitate such reading in a distractive environment, gaze-aware interfaces are usually used to process visual distractions during RSVP reading. For example, Hansen et al [37] demonstrated the feasibility of using a commercial gaze tracker to control RSVP playback. Dingler and colleagues [22] demonstrated gaze controlled RSVP with a head-mounted gaze tracker and visual markers. Although gaze-aware interfaces are attentive to the environments and intuitive to use, they are restricted by the sensing techniques. Current gaze detection techniques rely on the detection of the muscle movement around the eyes or the relative position of the pupil to the rest of eyes. Dedicated eye trackers can accurately detect gaze movements, but they face challenges of costly deployment, uncomfortable usage, and calibration requirement problems. Besides dedicated eye

---

[1] Rapid Serial Visual Presentation (RSVP) is a visualization technique that displays textual information one word at a time1 in sequential order.

trackers, smartphones are also able to estimate eye gaze by processing the image frames captured from the front-cameras. For example, EyePhone [57] monitors the positions of gazes from the front camera of a smartphone and uses the gaze to perform hand-free operations on the smartphone. Leveraging smartphones to track gaze is low-cost and portable but lacks stable accuracy in regards to the phone-to-face distance [55].

Besides gaze tracking, a variety of behaviors and physiological signals, such as heart rates [36][74][43], galvanic skin responses (GSR) [40][86], facial expressions and Electroencephalography (EEG) [73] have been explored to infer learner cognitive and affective states in different interaction tasks, such as learning [43][21], operating user interfaces [74], and gaming [36]. However, existing intelligent interfaces also require extra-dedicated sensors to track behavior or physiological signals.

1.2 Related Work: Reading on Pocket-size Devices

To improve current naiive approaches (i.e. CTR and dwell time) to understand reading and to have a higher level understanding of complex reading strategic processes, researchers proposed and designed the new methods.

Researchers are attempting to replace the CTR and dwell-time based approaches to understand reading behaviors. Afflerbach [1] has determined that theoretical models [65][66][49], self-reports [19][23][17], different sensors especially gaze [11][12][8] are representative and widely used to monitor and understand internet-based reading behaviors, comprehension and emotions. The theoretical models gave a great push to screen reading analysis while making a slow progress on mobile reading due to limited scale analysis. In comparison, self-report continuously plays a centrol role in research works focused on understanding online reading, although its accuracy and convenience

are criticized. With the development of new techniques, understanding reading by observing the reading behaviors, such as gaze, are flourishing rapidly. Eye movement features extracted from eye trackers are used to interpret reading cognitive processes [47], comprehension [7][56], reading proficiency [16], reading engagement, etc. Rayner's survey [71] provided a detailed review on how the visual information related to reading. Besides eye movement, other behaviors, such as mouse control motions [46][81][58][18] are also good indicators of users' affections and cognitions. Scrolling actions on touch-screen reading devices are detected [4][88] and analyzed to understand reading progresses [12][31]. Existing approaches either require dedicated sensors or cannot directly predict reading cognition and affections. The portability, cost, and availability prevented the wide adoption of such methods on smartphones beyond lab settings.

Compared to the flourishing works on understanding single-page mobile readings, the research on understanding online reading strategic processes focuses on desktops [19][17][23][20][5]. Such methods highly rely on verbal reports to understand reading strategy; these methods have at least three major challenges. First, think-aloud reading might cause a negative impact on readers' cognitive engagement and affect the strategic processes [17]. Second, accuracies of readers' self reports on cognitive tasks are considerably various [28]. Some readers lack the ability to accurately express their thoughts. Third, verbal reports are time and labor consuming to generate and grade. These challenges prevent the large-scale understanding of reading strategies.

1.3 Research Overview

In summary, expressive and fine-grained sensing are crucial to understand/improve mobile reading on watch-size and pocket-size mobile devices.

Therefore, this thesis explores the usage of low-cost mobile sensing techniques to understand/improve mobile reading on small-screen wearable devices and smartphones. Three types of low-cost mobile sensing techniques are investigated, including the photoplethysmography (PPG) sensing of users' heart rate signals, camera-based gaze tracking of users' eye movements, and resistive sensing of users' screen touching and scrolling behaviors.

The challenges of reading on watch-size devices are caused by limited screen affordance, therefore, we developed SmartRSVP (Chapter 2), an attentive speed-reading system to facilitate text reading on small-screen wearable devices (Figure 1). SmartRSVP uses camera-based facial alignment and eye gaze tracking techniques to determine whether a user is paying visual attention to Rapid Serial Visual Presentation (RSVP) reading or not, and then leverages the visual attention information to play/pause the presentation of dynamic texts. At the same time, SmartRSVP uses heart-rate variability (HRV) features to infer the user's cognitive workload, which is further used to regulate the speed of RSVP in real time. Overall, SmartRSVP leverages real-time visual attention tracking and implicit physiological signal sensing to make text reading via Rapid Serial Visual Presentation (RSVP) more enjoyable and practical on smart watches.

The challenges of reading on pocket-size devices are caused by complex reading contents and the lack of continuous and deeper understanding of users' cognitive states, therefore, we developed two Systems, Lepton (Chapter 3) and StrategicReading (Chapter 4) to track users' cognitive states during single-page and multi-page reading tasks.

To better understand single-page reading cognitions and affections on unmodified smartphones, we developed Lepton (Chapter 3), an intelligent mobile reading system

with a set of dual-channel sensing algorithms to achieve rich and fine-grained understanding of users' reading behaviors, comprehension, and engagement. 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 for touch events. Overall, Lepton combines a visual tracking channel via the front camera and a kinematic channel by tracking scrolling operations on a touch screen to monitor and improve mobile reading activities.

We then developed StrategicReading (Chapter 4), a system that inherits the behavior sensing techniques from Lepton, to further automate the understanding of higher level cognitive states such as reading strategic processes and reading performance during an multi-page online reading task. In addition, StrategicReading system also tracks the evolvement of cross-page behavior that are detected by logging the searching and clicking history. Overall, StrategicReading reliabaly predicted users strategic processes and their reading performance via unmodified mobile phone sensors during an online reading task.

The three types of sensing techniques can be classified into two groups: physiological signal sensing and users' behavior sensing (Figure 1). *The heart rate sensing of PPG signals* is physiological signal sensing. With the ability to stay on a user's wrist 24/7 and collect the user's physiological signals implicitly, smart watches can be a promising test bed for the next generation of intelligent user interfaces. PPG signals are proven to be efficient on tracking mind wandering while watching online videos [67]. SmartRSVP project further explores it in the context of small-screen reading. *The camera-based gaze tracking technique* is a type of behavior sensing technique. Compared

to gaze tracking with dedicated sensors, the accuracy of camera-based gaze tracking is much lower. To solve this problem, our work avoids using the absolute gaze locations; instead, we use the relative period patterns of gaze movements. In SmartRSVP project, users' gaze can be still during reading because we adopted the optimal recognition point (ORP) of rapid serial visual presentation (RSVP) that displays textual information one word at a time in a sequential order and all words' gaze fixation points stay at a fixed location. Without gaze movements, binary attention detection along with a low-pass filter was used to increase the accuracy. Lepton and StrategicReading systems used periodic lateral pattern based gaze features because the normal text reading enables the line-by-line gaze period movements. *The scrolling sensing* is another type of behavior sensing technique. Compared to existing research on finger scrolling motions, our work leveraged the new scrolling features and provided a direct indication of reading comprehension and engagement. All signals in these projects were tracked via built-in sensors in a regular smartphone, thus eliminating the requirement of dedicated sensors.

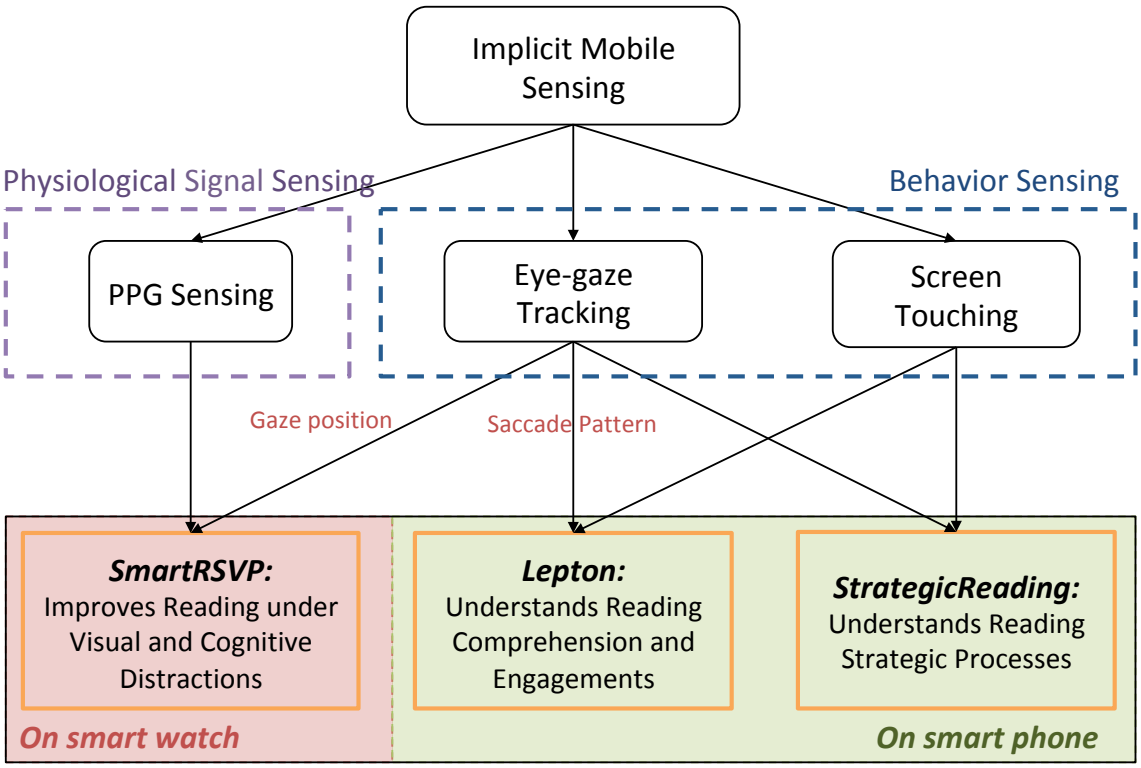

Figure 1. Major Components of The Thesis: Implicitly sensing behavior and physiological data to better understand mobile reading.

1.4 THESIS STATEMENT AND HYPOTHESIS

This thesis systematically explores the usage of behavioral signals and physiological signals implicitly collected via a "sensorless" approach at low-cost to understand, model and improve mobile reading.

We have the following hypothesis:

Hypothesis 1: A physiological and behavior signal-based perceptual and affect-aware interface will effectively control text displaying on watch-size devices.

Hypothesis 2: Users' reading comprehension and engagements will be accurately predicted by camera-based gaze signals and emotion-dependent scrolling changes.

Hypothesis 3: Using users' behavior sensing signals, the authentic online reading strategy and reading performance on smartphones will be accurately predicted.

1.5 THESIS CONTRIBUTION

This thesis is the first attempt of building a perceptual and affect-aware interface to control text displaying on watch-size devices; This thesis creatively investigated the processing of the noisy camera-based gaze signals and emotion-dependent scrolling changes during reading on pocket-size devices and proved the probability to predict reading comprehension and engagements using such signals; This thesis is the first attempt to automate the detection of complex reading strategies via multi-channel sensing techniques.

1.6 THESIS OUTLINE

This chapter has illustrated the background of our research. In the following chapters, I will present the detailed projects. Chapter 2 presents the design and evaluatation of SmartRSVP to facilitate text reading on small-screen wearable devices by leveraging real-time visual attention tracking and implicit physiological signal sensing. Chapter 3 presents the Lepton that facilitates understanding single-page mobile reading by leveraging a visual tracking channel via the front camera and a kinematic channel by tracking scrolling operations on a touch screen. Chapter 4 presents the StrategicReading that automates the tracking of mobile reading strategic processes and reading performance. Chapter 5 summarizes my thesis work.

2.0 SMARTRSVP: TOWARDS ATTENTIVE SPEED READING ON SMALL SCREEN WEARABLE DEVICES

To support reading on watch-size screens, we desgined and implemented SmartRSVP, an attentive speed reading system, on smart watch. SmartRSVP detects users' divided visual attention via eye gaze tracking and internal cognitive workload via physiological signal sensing during mobile reading on smart watch, and provides interventions to make reading more efficient.

2.1 Introduction

Small-screen wearable devices are flourishing nowadays. By staying on users' wrists 24/7, smart watches can assist users' access of frequent tasks and important notifications. Smart watches are also ideal for tracking users' activities and physiological signals for personal wellbeing. Although many new interaction techniques [14] and input modalities [38][45] have been proposed for smart watches during recent years, it remains a major challenge to read textual information on smart watches.

There are at least three major challenges when a user reads text messages on a smart watch. First, most of today's smart watches use a small display approximately 1.5-inch diagonal, which only affords showing three or four words per line. Therefore, a user must rely on more frequent lateral eye gaze movements, i.e. saccade, during reading. Second, more scrolling actions are needed due to the limited number of words per screen. Such scrolling operations not only exacerbate the notorious "fat finger problem" but also occupy a user's both hands. Third, text reading on smart watches increases the

probability of divided attention and interruptions. Paradoxically, the growing amount and type of information accessible via smart watches increases our exposure to such reading interfaces.

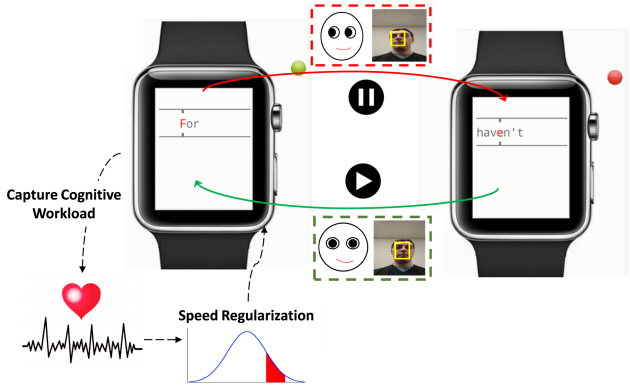

Figure 2. SmartRSVP continuously monitors the visual attention of a user via real-time image processing and infers the user's cognitive workload via implicit physiological signal sensing. SmartRSVP uses the visual attention to play/pause dynamic text presentation and adjusts text displaying speed via the inferred cognitive workload.

To address these challenges, we present SmartRSVP (Figure 2), a novel speed reading system to facilitate text reading on small screen wearable devices. SmartRSVP leverages real-time visual attention tracking and implicit physiological signal sensing to make text reading via Rapid Serial Visual Presentation (RSVP) more enjoyable and practical on smart watches. SmartRSVP uses camera-based facial alignment and eye gaze tracking techniques to determine whether a user is paying visual attention to reading or not and then leverages the visual attention information to play/pause the presentation of dynamic texts. At the same time, SmartRSVP uses heart-rate variability (HRV) features to infer the user's cognitive workload, which is further used to regulate the speed of

RSVP in real time. Overall, SmartRSVP exploits both the spatial and temporal efficacy of the RSVP technique, and reduces its workloads in both visual attention and cognitive. This study offers three major contributions.

1. We present SmartRSVP, a perceptual and affect-aware intelligent interface to facilitate text reading on wearable devices via visual attention tracking and implicit cognitive state sensing.
2. We propose novel algorithms and interaction designs to make text reading via RSVP more enjoyable and practical for text reading on small-screen wearable devices.
3. We show the feasibility, accuracy, robustness, and usability of SmartRSVP via four user studies involving 60 participants in total.

2.2 Design

Figure 2 shows SmartRSVP in action. SmartRSVP displays text via RSVP, and continuously monitors the visual attention and cognitive states of the user. SmartRSVP will pause the text display if 1) there is no human face in the camera viewport, or 2) the user's eye gaze is not in direct contact with the watch screen. SmartRSVP also infers the cognitive workload of the user via implicit PPG sensing through a dedicated PPG sensor or a back camera. The word speed of RSVP will be adjusted based o the cognitive workload.

SmartRSVP includes four major components: 1) The RSVP module; 2) Algorithms for tracking and using the owner's visual attention; 3) A statistical model to predict the internal cognitive states of a user; and 4) The speed regulation module.

2.2.1 RSVP

We use a 20dp monospace font (average height = 8.1mm) to render words in our RSVP module. This font size provides good legibility on a 1.5-inch watch screen, and can display a 12-character-word without line breaking or resizing. SmartRSVP also aligns each word in the Optimal Recognition Point (ORP) [10] and visualizes the ORP in red color (Figure 2). ORP intends to make the gaze fixation point of a word stay at a fixed location to avoid unintended saccades when the gaze fixes on words of different lengths [10]. We use a monospace font to ensure all ORPs having the same width and adjacent words with the same length being aligned at the same location. The display speed of our RSVP module can vary from 200 wpm to 500 wpm. Users can manually tap the watch interface to play or pause the text display on SmartRSVP.

2.2.2 Visual Attention Tracking

Due to the limited availability of front facing cameras on smart watches, we used a Google Nexus 5x smart phone running Android 6.0 to simulate a 42.0mm by 5.9mm smart watch screen. This choice follows practices of existing research on smart watches [14][15][61], we allocate the same physical region on Nexus 5x for display and touch input.

Each image frame captured by the front camera goes through the following three steps to derive a binary outcome on visual attention. 1) Face detection: A Viola-Jones face detector [84] is used to detect the existence and location of a human face; 2) Face Alignment: We use Cascaded Pose Regression [24] to estimate the facial orientation and landmark points on a face; 3) Eye contact estimation: Similar to [55][78], we relied on

the location of the pupil relative to the rest of the eye to estimate the direction of eye gaze. We used the Qualcomm Snapdragon SDK to accelerate the tracking process. The per-frame image processing time was 17ms, and we can achieve 21 frames per second on the hexacore Snapdragon 808 CPU in the Nexus 5x.

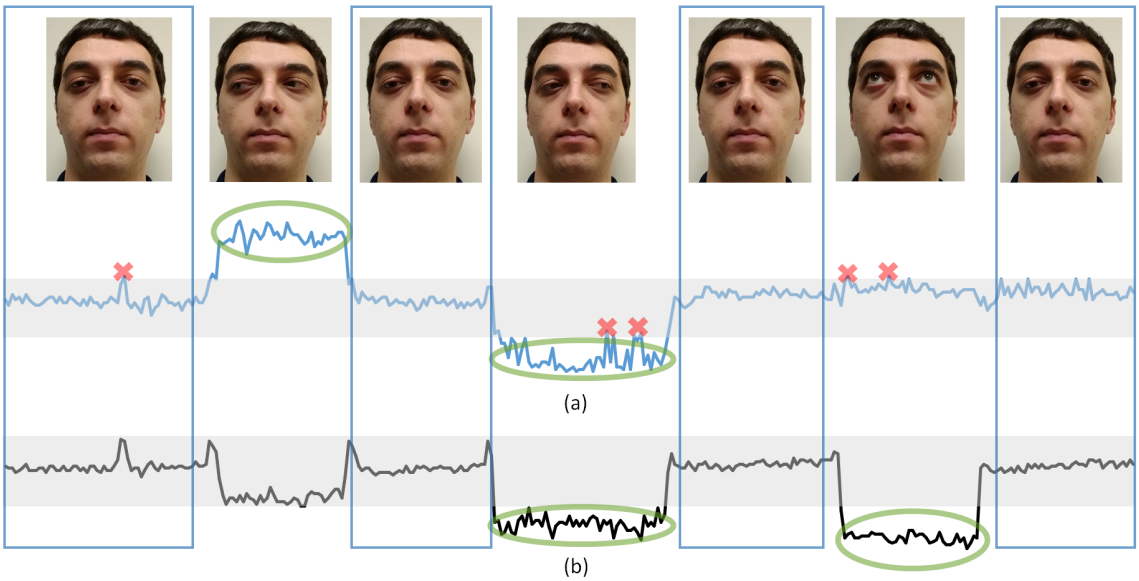

Figure 3. Visual Attention Tracking via face detection and eye contact estimation. The x-axis is time (~17 sec). Row (a) is the predicted horizontal eye gaze locations: Row (b) is the predicted vertical eye gaze locations. Green circles highlight moments when a user is not having gaze contact with his smart watch.

Given the small display size of smart watches, it's not necessary to estimate the absolute location of eye gaze on the watch. Instead, we trained a binary eye contact classifier from five volunteers: taking the union of all volunteers visual attention range, gazes within the union range will be treated as paying visual attention. We also used a 0.5sec low-pass filter to reduce false positives and false negatives from per-frame

estimations. Figure 3 shows the continual output of the eye gaze prediction algorithm and the binary eye contact estimations.

2.2.3 Cognitive State Inference

SmartRSVP infers users' cognitive states via commodity camera-based PPG sensing through the back camera of smart phones. We chose to use the back camera instead of the optical heart rate sensor on a smart watch for PPG sensing because current SensorManager API in Android Wear only reports average heart rates within a fixed time window, rather than raw waveforms of PPG signals. We expect cleaner PPG signals and higher prediction accuracies if we can access raw waveforms from dedicated optical heart rate sensors on smart watches in the future. SmartRSVP uses the LivePulse algorithm [36] to extract the raw PPG waveforms.

We use a fixed-size sliding window to extract features from the temporal PPG signals. We extract 9 dimensions of heart rate and HRV features (Table 1) from each window. After normalization, these 9 dimensions of features are used to train a statistical classifier to predict user cognitive workload within this sliding window.

2.2.4 Speed regulation

We used a one-way binary adaptation strategy similar to BACh [90] to adjust the text display speed dynamically. This adaptation strategy has been proven to be effective by existing research [90] because it avoids the confounders in number, durations and scale of adaptations. SmartRSVP tracks users' PPG signals during reading and decreases the speed of RSVP by 100wpm if a multitasking activity is detected from users' PPG signals.

The reduction speed was chosen by a 4-user pilot study, where 100wpm was the minimum reduction that could be noticed by all users.

Feature	Definition
MHR	Average heart rate
SDHR	Standard deviation of heart rates
rMSSD	The square root of the mean squared adjacent RR intervals' difference
pNN12	Percentage of more than 12ms difference between adjacent RR-intervals
pNN20	Percentage of more than 20ms difference between adjacent RR-intervals
pNN50	Percentage of more than 50ms difference between adjacent RR-intervals
MAD	Median of absolute deviation of RR-interval
AVNN	Average RR-interval
SDNN	Standard deviation of the RR-intervals

Table 1. Heart Rate and HRV Features extracted from raw PPG waveforms.

2.3 Experiments & Results

We ran a total of four user studies to investigate different aspects of SmartRSVP. In the first study, we directly compared SmartRSVP with today's standard reading interface on smart watches and traditional RSVP in a sitting condition. We further investigated the robustness and efficacy of SmartRSVP in standing and walking conditions in the second study. In the third study, we benchmarked accuracies of the

cognitive-state inference module in SmartRSVP. In the fourth study, we evaluated usability and efficacy of the whole SmartRSVP system in action.

2.3.1 User Study 1

This study evaluated the efficacy of the visual attention tracking channel of SmartRSVP and directly compared it with traditional RSVP and normal watch reading interface in a sitting posture.

2.3.1.1 Participants

18 participants (3 females) between 19 and 46 years of age (μ=26) participated in the study. All participants were undergraduate or graduate students from a local university.

None of the participants had previous experience with RSVP.

2.3.1.2 Apparatus

Our experiments were completed on a Google Nexus 5x Smartphone with a 5.2 inch 1920 x 1080 pixels display, 1.8 GHz hexacore SnapDragon 808 CPU running Android 6.0.

There were three conditions: normal watch reading interface (NWR), traditional RSVP (T-RSVP), and SmartRSVP (Figure 4) in this study. NWR used a 20dp sans serif (Droid Sans) font for text display to replicate today's reading interfaces on smart watches, which shows about three or four words per line and eight lines per screen.

Figure 4. Three reading interfaces in the study. From left to right: Normal Watch Reading Interface, Traditional RSVP interface, and SmartRSVP.

Thirty unique email pieces were chosen from Enron email database for this phase, as emails were short, frequent tasks on today's smart watch. They have comparable length (μ = 47 words or 3.5 sentences) and difficulty (average Flesch-Kincaid score = 68.65, σ = 14.97).

Figure 5. Distracters (random 3-digit numbers) appear on a 15-inch laptop screen on the left-hand side of a participant. Left: reading an email message via SmartRSVP; Right: turning left to read the distracter.

To simulate reading activities under divided visual attention, we put a 15-inch laptop on the left-hand side of the participant (Figure 5). When the participant was reading, once every 4 to 6 seconds, the laptop generated a beep sound, and a 3-digit random number was shown on the laptop screen for 2 seconds. Once the participant heard

the beep sound, she was required to look at the 3-digit number (Figure 5, right) and read the number out loud. The participant could resume the reading task after reading the number out loud. We placed 3 distracters for each email message.

2.3.1.3 Procedure

This user study included a single session for 30 minutes. The session started with the introduction of the three interfaces and distractions. Once completed, participants practiced reading on the three interfaces with distractions for 10 min to get familiar with user study system, distractions, the genre of reading materials and comprehensive questions. Participants read a set of 10 emails on each interfaces, 3 sets in total. Both sets and interfaces were randomly selected for each participant. Participants were asked to stop the RSVP whenever they desired. We placed 3 distracters for each email message. After finishing each email, the participant answered one question to test the text comprehension. At the end, participants were asked to complete a questionnaire to provide the subjective feeling of the three interfaces.

2.3.1.4 Design & Analysis

The study used within-subjects design with three interfaces: NWR, T-RSVP, SmartRSVP (Figure 4).

After reading each email, the participant answered one question to test of reading comprehension. There are three levels of text comprehension, i.e. literal, inferential, and evaluative [27]. We only used literal questions, i.e. recalling key information that was explicitly stated in the email, in our study because we focused on evaluating and

comparing reading interfaces rather than testing the language and logical skills of participants. We investigated the following metrics across three interfaces tested:

- False positive and false negative rates of the visual attention tracking module.
- Comprehension rate, equaling to the percentage of correctly answered questions.
- Reading efficiency (E), defined as: $E = \frac{W}{D} \times c$, where D denoted reading duration (including distractions), W was the number of words, and c was comprehension rate [77][42].
- Subjective ratings on a 5-point Likert scale.

2.3.1.5 Results

The false positive rate of visual attention tracking in SmartRSVP was 24.02%, and false negative rate was 3.7%.

Figure 6 shows the average comprehension rates of the three reading interfaces. The comprehension rates and corresponding standard deviations were 52.2% (σ=0.16), 23.9% (σ=0.11), and 57.5% (σ=0.20) respectively.

Repeated measures of analysis of variance showed a significant main effect ($f=23.16$, $p<0.0001$) incomprehension rates among the three reading interfaces. Pairwise mean comparison (t-tests) with Bonferroni correction showed that the comprehension rate of NWR was significantly higher than T-RSVP ($t(17)=-6.27$, $p<0.0001$). The comprehension rate of SmartRSVP was also significantly higher than T-RSVP ($t(17)=-6.32$, $p<0.0001$). However, Difference in the comprehension rates was not significant between NWR and SmartRSVP ($t(17)=0.88$, $p=0.39$). Similar results were discovered on reading efficiency. For NWR, T-RSVP, and SmartRSVP, the reading efficiencies and the

corresponding standard deviations were 65.16 wpm (σ=19.49), 43.93 wpm (σ=21.42), and 67.16 wpm (σ=18.65). There were significant differences in reading efficiency between NWR vs. T-RSVP ($t(17)=-3.02$, $p<0.005$), and between SmartRSVP vs. T-RSVP ($t(17)=-3.37$, $p<0.005$).

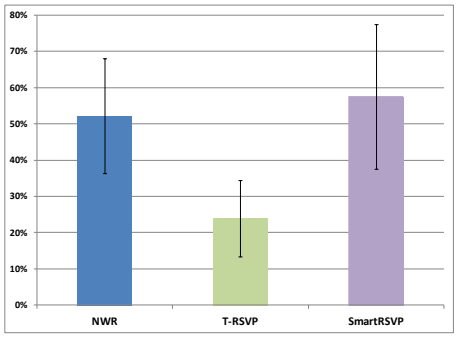

Figure 6. Comprehension rates by reading interfaces

Figure 7 shows the subjective ratings of perceived comfort across three interfaces. The length of each bar represents the average perceived comfort of each platform. The color grids represent the portions of each rating score within the bar. The subjective ratings of perceived comfort were 3.78 (σ = 0.73), 2.06 (σ = 0.96), and 3.28 (σ = 0.94) for SmartRSVP, T-RSVP and NWR respectively. There were significant differences in subjective rating between NWR and T-RSVP ($t(17)=-3.87$, $p<0.0005$), as well as between SmartRSVP and T-RSVP ($t(17)=-6.14$, $p<0.0001$). Although SmartRSVP received higher subjective ratings in comfort when compared with NWR, the difference was not significant ($t(17)=1.75$, $p=0.08$).

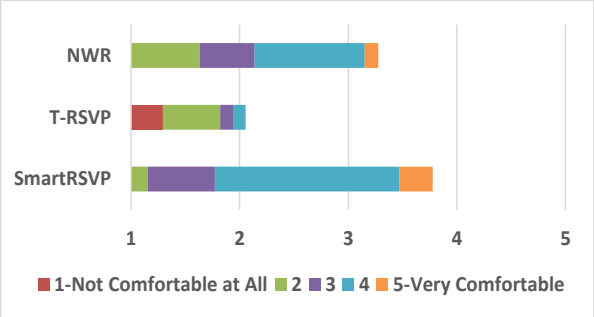

Figure 7. Subjective ratings on perceived comfort on a 5-point Likert scale (1 = not comfortable at all, 5 = very comfortable).

All 18 participants provided positive feedback on the use of eye-gaze as an implicit control channel for RSVP. Among them, 15 participants rated the visual attention tracking and control channel in SmartRSVP "responsive".

2.3.2 User Study 2

This follow-up study further evaluated the usability, efficacy, and robustness of SmartRSVP's visual attention tracking channel during standing and walking conditions.

2.3.2.1 Participants

12 participants (5 females) between 18 and 34 years of age (μ=23) participated in the study (Figure 8). All participants were undergraduate students or graduate students from a local university.

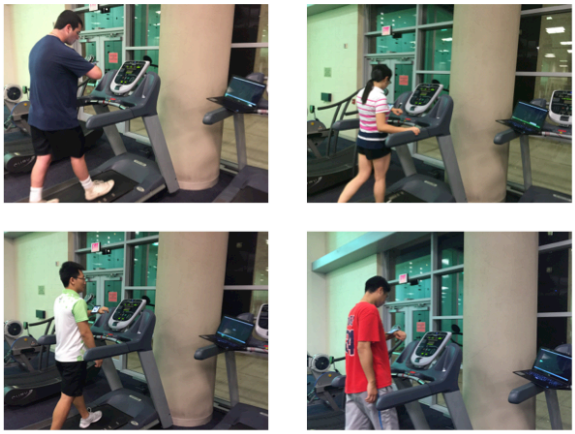

Figure 8. Sample participants in study 2.

2.3.2.2 Apparatus

Apparatus were the same as the user study 1, expect:

1) We excluded the T-RSVP interface to simplify the experimental design. When compared with the sitting posture in study 1, the standing and walking conditions do not bring additional benefit to T-RSVP over SmartRSVP.

2) The participants completed all the tasks on a treadmill in a local gym (Figure 8). The speed of the treadmill was set to 1.5mph for the walking posture.

2.3.2.3 Procedure

User study 2 was a single 30 min session. The session started with the introduction of the two interfaces, i.e. SmartRSVP and NWR. Once completed, participants practiced reading on the two interfaces with distractions for 10 min to get familiar with user study system, treadmill, distractions, the genre of reading materials and comprehensive questions. After finishing each article, the participant answered two literal questions to test the text comprehension.

2.3.2.4 Design & Analysis

The study was a within-subject 2*2 factors design. The factors and levels were:

- Posture: Standing vs. Walking
- Interface: SmartRSVP vs. NWR

Posture and interfaces were randomly ordered for each participant. Each participant completed an article under a unique combination of conditions, leading to 2*2=4 articles for the study.

Following user study 1, we used two literal questions after each article to test users' comprehensions.

2.3.2.5 Results

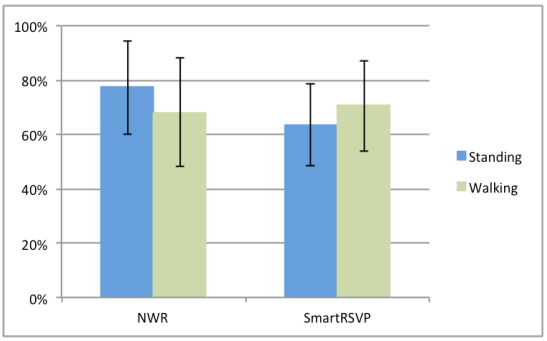

Figure 9. Average comprehension rates by postures.

Figure 9 shows the average comprehension rates of the NWR and SmartRSVP by reading posture. The comprehension accuracies were 77.3% (NWR + Standing, σ=0.69), 68.2% (NWR + Walking, σ=0.81), 63.6% (SmartRSVP + Standing, σ=0.61) and 70.8% (SmartRSVP + Walking, σ=0.66). No significant difference was found on comprehension accuracies in either reading platforms ($t(11)=-0.55$, $p=0.58$) or postures ($t(11)=-0.11$, $p=0.91$). Besides comprehension rates, we also measured reading efficiencies, which

were 177.3 wpm (NWR + Standing, σ=93.98), 163.5 wpm (NWR + Walking, σ=110.17), 129.1 wpm (SmartRSVP + Standing, σ=59.61), and 170.2 wpm (SmartRSVP + Walking, σ=117.67). Again, neither reading platforms (t(11)=-0.71, p=0.48) nor reading postures (t(11)=0.46, p=0.64) had significant impact on reading efficiency.

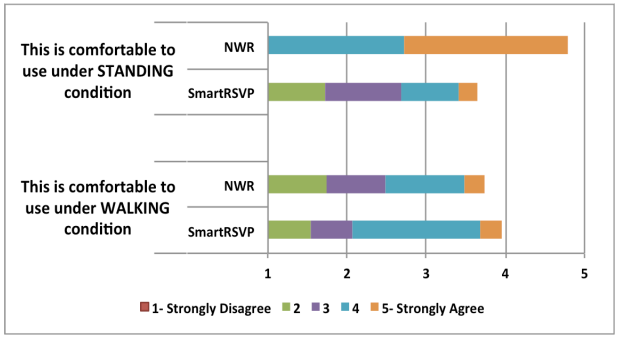

Figure 10. Subjective ratings on perceived comfort for NWR and SmartRSVP (1 = Strongly Disagree, 5 = Strongly Agree).

The subjective ratings of perceived comfort in general were 4.55 (σ = 0.52) for NWR+Standing, 3.18 (σ = 0.98) for SmartRSVP+Standing, 3.27 (σ = 1.01) for NWR+Walking, 3.55 (σ = 0.93) for SmartRSVP+Walking (Figure 10).

Participants preferred NWR over SmartRSVP in standing posture. At the same time, participants preferred SmartRSVP over NWR in walking posture.

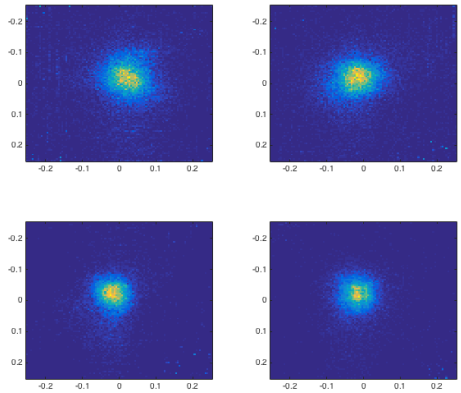

Figure 11. HeatMaps of aggregated eye gaze by postures. Top row: NWR+Standing and NWR+Walking; Bottom Row: SmartRSVP+Standing and SmartRSVP+Walking.

We also collected users' raw gaze data (the estimated gaze point from our algorithm) to quantify the impact of reading techniques and body movements on gaze patterns, and to investigate the robustness of camera-based gaze tracking technique in SmartRSVP (Figure 11). From corresponding Heatmaps we can see that the raw eye gaze locations were less scattered for SmartRSVP (bottom row) than NWR. At the same time, walking (right column) can cause slightly more distributed gaze distributions than standing (left column).

2.3.3 User Study 3

This study evaluates the cognitive workload detection module of SmartRSVP.

2.3.3.1 Participants

The same participants were recruited as user study 1.

2.3.3.2 Apparatus

The same device and the same environment were used as user study 1.

We adopted the color counting task [73] to induce the internal cognitive workload changes. We included two conditions: the focused condition and the multitasking condition. In focus condition, users were asked to concentrate on reading without external distractions. In multitasking condition, a computer placed on the side spoke the names of nine different colors randomly at the speed of one second per word. Participants were told to read as well as count the number of times the two target colors, i.e. “yellow” and “white”, were spoken. Conditions and articles were all assigned to users in random orders.

2.3.3.3 Procedure

This user study includes a single 30min session. Participants read two news articles, one under the focused condition, the other under the multitasking condition. Both articles and conditions were randomly assigned. After reading each article, participants rated their focus level, and answered 5 questions to test the comprehension.

2.3.3.4 Design & Analysis

The study used a within-subjects design with two cognitive workload levels on SmartRSVP interface.

2.3.3.5 Results

The average comprehension rates for focus and multitasking were 56.67% and 26.67% respectively.

We trained both user-independent and user-dependent models on the collected PPG waveforms to predict a participant's focus/multitasking status when using SmartRSVP. We used the leave-one-participant-out technique to train the user-independent models and used two-fold cross-validation to train the user-dependent models.

For each PPG sequence collected, we went through the following three steps to extract the features for our statistical models. 1) Signal segmentation: we used a moving window to split each PPG sequence into fixed-size chunks; 2) Feature extraction: we extracted 9 dimensions of heart rate and HRV features (Table 1) from each chunk via the LivePulse algorithm [36]; 3) Feature Normalization: we normalized features via MeanSTD. We tested different combinations of window number, window size (10s, 15s… 40s), window overlaps (0%-50%), and different initial padding to find the optimal parameters for the classification models.

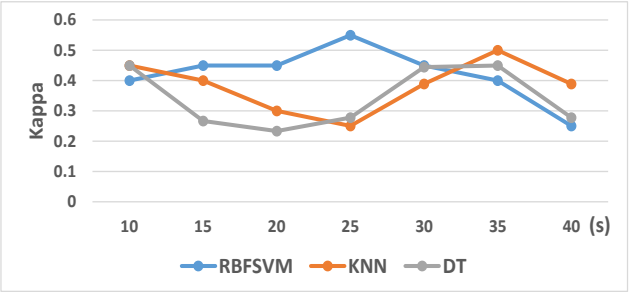

Figure 12. The classifiers' optimal kappa by local window size in user independent models.

We explored the use of support vector machine with radial basis function kernel (RBFSVM), K-nearest neighbor (KNN) and Decision Tree (DT) classifiers to train and predict users' cognitive workloads.

We found the RBFSVM model using two 25s windows (0% overlap, a 6s gap between windows) and ignoring the first 20 seconds of signals led to the best overall performance (accuracy = 77.5%, precision = 78.3%, recall = 85.0%, kappa=0.55) (Figure 12) for user-independent classification. Figure 13 shows the results for user dependent models by window size and classifier. The highest Kappa was 0.64 when using a 40 sec local window and a RBFSVM classifier.

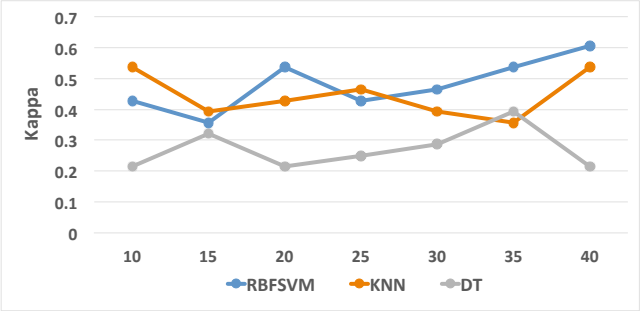

Figure 13. The classifier's optimal Kappa by local window size in 2-fold user dependent models.

2.3.4 User Study 4

This study investigated the usability and efficacy of the speed regulation module of SmartRSVP in action. Our goals were two-fold: 1) determine whether SmartRSVP was able to identify users' focus/multitasking internal cognitive state in everyday tasks, 2) whether the dynamic speed adjustments by SmartRSVP were effective.

2.3.4.1 Participants

14 participants (6 females) between 25 and 33 years of age (μ=29) participated in the study. All participants were undergraduate or graduate students from a local university. Only one participant had previous experiences in RSVP.

2.3.4.2 Apparatus

We adopted the same device, internal distractors and reading environment as in user study 3.

2.3.4.3 Procedure

This user study included a 20min training session and a 40min testing session. For each participant, the training session was conducted approximately at the same time of the day but one week before the testing session.

The training session followed the same procedures as study 3. At the beginning of the testing session, participants were informed that the real-time speed regulation function was enabled. Participants then read in a different and random ordered article for each trail. At the end of each trail, participants first reported counted color if the distraction was enabled, and then completed five short answer questions about the article.

2.3.4.4 Design & Analysis

We used a within-subjects design in this study. The testing session has a 2 (focused vs. multitasking) * 2 (simple threshold classifier vs. RBFSVM classifier) factorial design. Therefore, each participant completed 2+2*2=6 sessions in the study.

We adopted a simple threshold based classifier (TH classifier) as the baseline. The TH classifier calculates the one-dimensional MHR in HRV features to determine a

user's cognitive state. We used this condition to simulate traditional practices [74] that predict cognitive workload from heart rate variability signals in the HCI community. A multitasking event was trigger by the TH classifer if the user's MHR of testing window aligned outside $\mu_{focus} \pm 2\sigma_{focus}$ range of the focus distribution, and the MHR aligned to the side of μ_{focus} as $\mu_{multi-tasking}$ did. RBFSVM classifier was chosen since it performed the best among three classifiers tested in user study 3.

2.3.4.5 Results

We discarded two users due to corrupted data. In the training phase, we used the same three-step procedures to process users' raw PPG signals and got training instances. To avoid carry over effects, we used 20s initial padding in both training and testing phase.

	Baseline (TH)	**RBFSVM**
Accuracy	50%	70%
Precision	50%	83.33%
Recall	40%	50%

Table 2. The live performances SmartRSVP in study 4.

As shown in Table 2, our RBFSVM based classifier consistently achieved better prediction accuracies than the TH classifier (baseline). The relative improvement in accuracy was 40% (Table 2).

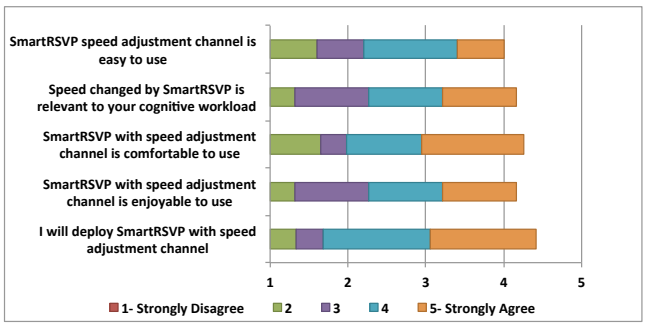

Figure 14. Subjective ratings on a 5-Point Likert scale.

Overall, our users reported positive experiences with speed adjustment module in SmartRSVP (Figure 1). Users considered SmartRSVP's speed adaptations were relevant to their cognitive workload (μ=3.8, σ=1.03) and would like to use SmartRSVP in the future (μ=4.1, σ=0.99).

2.4 Discussions

This research is our initial attempt to design an attention and cognitive state-aware speed reading interfaces on wearable devices. We have intentionally make the following trade-offs to achieve a good balance among robustness to environmental changes, ease of use, and minimal calibration efforts.

First, we focused on detecting the type of cognitive workload (i.e. focused vs. multitasking) rather than detecting the continual levels of each type. We found such coarse grained detection results are sufficient to regulate the speed of RSVP dynamically with good accuracies and robustness.

Second, we chose a one-way, fixed speed adaptation strategy [90] because we wanted to ensure that our adaptive algorithm will at least do no harm to the reading process. Multi-way detections will reduce the detection accuracy and more importantly,

incorrectly increasing the reading speed may lead to disruptive experiences from users. Both experimental results and qualitative feedback from users confirmed this design decision.

2.5 Limitations and Future Work

Despite promising results, we have only scratched the surface of the design space of SmartRSVP. There are several limitations of this research that to be explored in the future. First, our studies were conducted in indoor and consistent lighting conditions. It is harder to track users' visual attention reliably outdoors with inconsistent lighting conditions, e.g. the camera may be overexposure under direct sunshine. In addition to designing more robust algorithms, it would be important to leverage built-in motion sensors such as the GPS, accelerometers and gyroscope in the watch to both infer the context of the users (i.e. indoor, outdoor, moving, not moving) and estimate the orientation and dynamic posture of the smart watch for more accurate predictions; Second, as discovered in our study, there were both challenges and opportunities to provide feedback for the current text presentation when a user was not paying visual attention to the display. In this project, we didn't include extra feedback other than speed reduction. We believe tactile feedback could play an important role here. It will be interesting to explore the feasibility, type, and level of tactile feedback in no visual contact state of SmartRSVP in the future; Third, although we confirmed the feasibility of speed adaptation according to users' cognitive workloads in SmartRSVP, principled research is necessary to further investigate the design space of dynamic speed adaptation (e.g. optimal latency and scale of speed change) which is lack in this project; Fourth, some participants reported unwanted speed reductions which could be eliminated by

inventing a mixed- initiative approach for the fine-grained control of display speed, where both users and the intelligent interface can change or confirm the reading speed in a complementary manner. The use of wrist gestures could be served as a mixed-initiative control channel in SmartRSVP; To evaluate the visual and cognitive control on RSVP display, we exclude the existing content-based adaptation of RSVP, such as adjusting the word- level display durations based on the predicted importance. Such adaptations are orthogonal to SmartRSVP adaptations, therefore could be explored in context of SmartRSVP. Other than that, enabling regressions via gesture-based interactions [53][54], and reminding users (via tactile feedback, sound, or visualizations) about important upcoming messages, could also be explored in the context of SmartRSVP.

2.6 Conclusion

We proposed SmartRSVP, a novel speed reading system to facilitate text reading on watch-size screens. SmartRSVP leverages camera-based visual attention tracking and implicit physiological signal sensing to make text reading via Rapid Serial Visual Presentation (RSVP) more enjoyable and practical on smart watches. In a set of four user studies, we found that SmartRSVP lead to significantly higher comprehension rate (57.5% vs. 23.9%) when compared with traditional RSVP. The current implementation of SmartRSVP was capable of supporting more realistic conditions such as walking in a gym with satisfactory performance and subjective preference. SmartRSVP can predict users' cognitive workloads with an average accuracy of 77.5% (kappa=0.55) in a user-independent model. Finally, SmartRSVP can adjust the speed of RSVP based on users' cognitive workload with 83.33% precision.

3.0 LEPTON: UNDERSTAND MOBILE READING VIA CAMERA BASED GAZE TRACKING AND KINEMATIC TOUCH MODELING

To explore the opportunities to achieve rich and fine-grained understanding of users' mobile reading behaviors, we propose Lepton, an intelligent mobile reading system that captures readers' reading comprehension and engagements through a set of dual-channel sensing algorithms. 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 for touch events.

3.1 Introduction

Mobile reading is experiencing rapid growth in the era of smartphones [91]. According to a recent survey, mobile reading activities, such as reading articles on a browser or dedicated social medias apps, and reading messages in email clients or instant messaging (IM) apps, is around 2 hours per day in the United States in 2016 [65], accounting for 40% of the daily time with mobile devices [65]. Despite such promising progressions, reading non-pleasure contents on mobile devices for working 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 [59]. Indeed, compared to media consumption channels such as watching videos [83], the passive nature of mobile reading and the more complicated reading context lead to declined attention and increased non-linear reading pattern [52].

Despite the highly diversified contexts, practitioners today still rely heavily on coarse-grained metrics such as click-through-rate (CTR) [87][32] and dwell time to understand and adapt to users' reading behavior on mobile devices. Such approaches have been proven to be inadequate [25] because of the sparsity and indirectness of such signals. For example, extended dwell time may imply desirable content, increased difficulty, or even distractions.

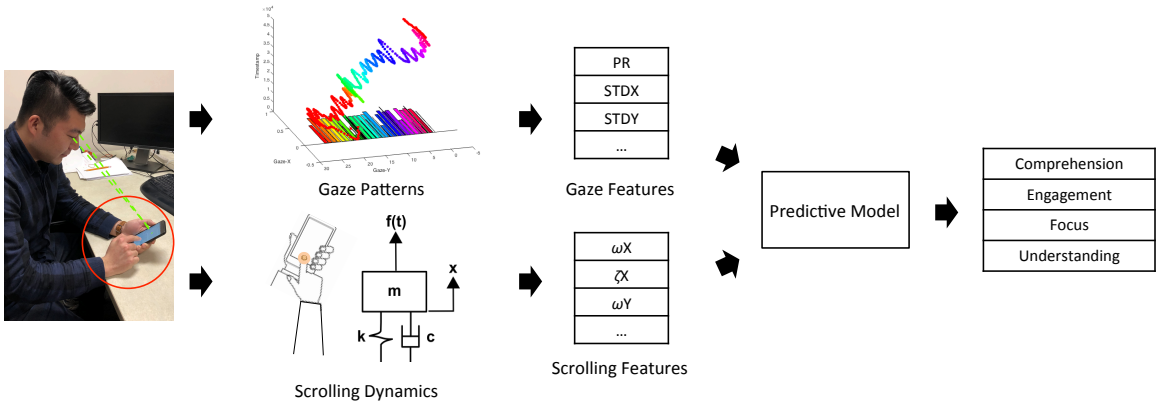

Figure 15. The architecture of Lepton: the visual channel (top) tracks the periodical patterns of users' eye gaze via the smart phone embedded front-facing camera. The kinematic touch channel (bottom) analyzes users' scrolling touch behavior via a Mass-Spring-Damper (MSD) model.

We present Lepton, an intelligent mobile reading system and a set of dual-channel sensing algorithms, to achieve rich 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 their muscle stiffness during text scrolling via a Mass-Spring-Damper (MSD) based kinematic model for touch events. Overall, Lepton combines a visual tracking channel via the front camera and a kinematic

channel by tracking scrolling operations on a touch screen to monitor and improve mobile reading activities.

Lepton system offers three major contributions:

- We propose a novel set of periodic lateral pattern-based gaze features that can be tracked via a widely-used but low-accuracy embedded front-facing camera of a smart phone.
- We use a physiological model of hand-arm dynamics (MSD model) to quantify users' muscle stiffness during scrolling operations and then infer their attention in reading on smart phone.
- By combining rich features from both the visual tracking channel and the kinematic touch channel, we show that we can significantly improve the accuracies to predict users' comprehension and engagements in reading.

3.2 Design

Lepton uses a visual tracking channel and a kinematic channel to facilitate understanding reading comprehension and engagements.

3.2.1 Visual Tracking Channel

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' difference
rMSLD	The square root of the mean squared adjacent predicted line durations' difference

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 3. Periodic lateral pattern based eye gaze features

Traditional eye features are related to gaze fixations and saccades, e.g. mean and standard deviation of fixations' durations and saccades length [71]. Since we trade the accuracy from embedded sensors to improve feasibility, traditional eye features are no longer suitable. We proposed the periodic lateral pattern-based eye gaze features (Table 3). As our proposed gaze features are built upon periodic lateral pattern detection, we defined two metrics to evaluate the periodic lateral pattern detection accuracy: A) Lines read by a reader, and B) the existence of reread/skip action.

In this paper, we aimed to understand English text reading on portrait smartphone, which reads left-to-right in the horizontal direction, and top-to-bottom in the vertical direction. To represent gaze points, a traditional computer graph coordinate system is used to represent the smart phone screen. Such system uses the top-left corner as the origin and the y-axis goes downward increasing in value,, and expresses how a reader reads down the screen. If a reader reads line by line from the beginning to the end of a text page, the horizontal axis (x-axis) of his/her gazes should appear in a zig-zag periodic

lateral pattern as he/she consumes a line and sweeps back to consume another line [55]. Therefore, the x values of gaze were used to predict the periodic lateral patterns.

The front-facing camera of a Google Nexus 5x smart phone captures each image frame during reading. The frames are then passed through Qualcomm Snapdragon SDK to extract gaze coordinates [33]. The per-frame image processing time is 17ms, which can achieve around 20 frames per second on the hexacore Snapdragon 808 CPU in the Nexus 5x.

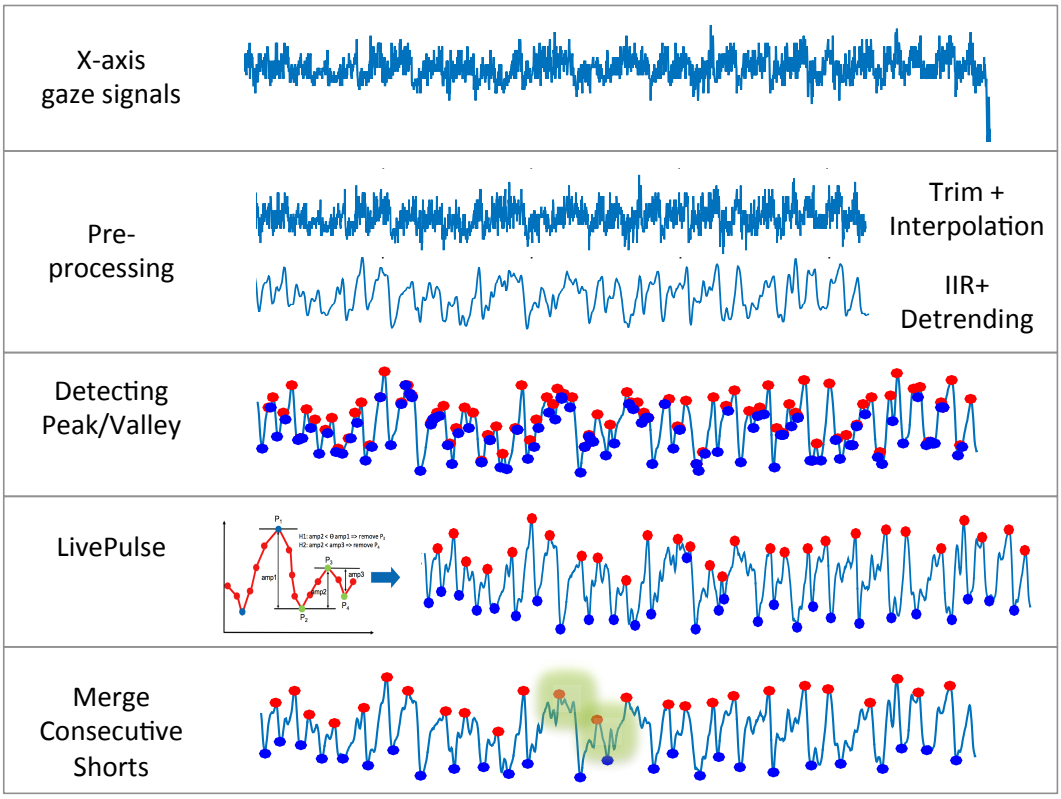

Figure 16. The framework for getting periodic lateral patterns. Top to Bottom: the process begins with horizontal (x-axis) gaze signals; through preprocessing, LivePulse algorithm and merging consecutive shorts, the final periodic lateral patterns were achieved.

3.2.1.1 Line detection

Our goal is to detect the lines a user read on a portrait phone screen with normal display text size (around 25 lines per screen). For each reading sample (Figure 16), the list of a reader's horizontal gaze data (x values) went through preprocessing, detecting peak/valley of periodic lateral pattern, and merging consecutive short patterns. For the preprocessing, we interpolated, scaled and detrended the gaze signals after the removal of noisy data at the beginning and the end. Then an IIR filter (around 2.5 Hz cutoff frequency) was used to process the gaze signals. For the peak/valley detection, all local maximums and minimums were first labeled as potential peaks and valleys. Then we used the LivePulse algorithm [36] to shrink the number of potential peaks and valleys. Lastly, the consecutive short line predictions were merged since our reading materials are paragraph-based. A set of gaze points between two selected valley points was considered as a periodic lateral pattern.

3.2.1.2 Action (reread & skip) detection

We used a X-line-counting action detection method to detect reread and skip actions, which calculated the number of lines as in A., and then compared this number with the number of lines on the page. If the detection number was larger than a certain ratio of the ground truth number, it was considered as reread. Similarly, we classified a sample as skip condition if the ratio was smaller than a certain ratio. We iterated through different values of ratios in experiment data, and found the optimized ratios for reread and skip condition.

Another option of action (reread/skip) detection was using vertical (y-axis) gaze data 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 (Figure 17. a) within a fixed viewport of the reading material; 2) When the reader rereads, his/her vertical gaze location plunges to a certain point, and then follows the increasing step shape (Figure 17. b); and 3) When the reader skips, the vertical gaze data shoots up to a certain point and then follows the increasing step shape (Figure 17. c).

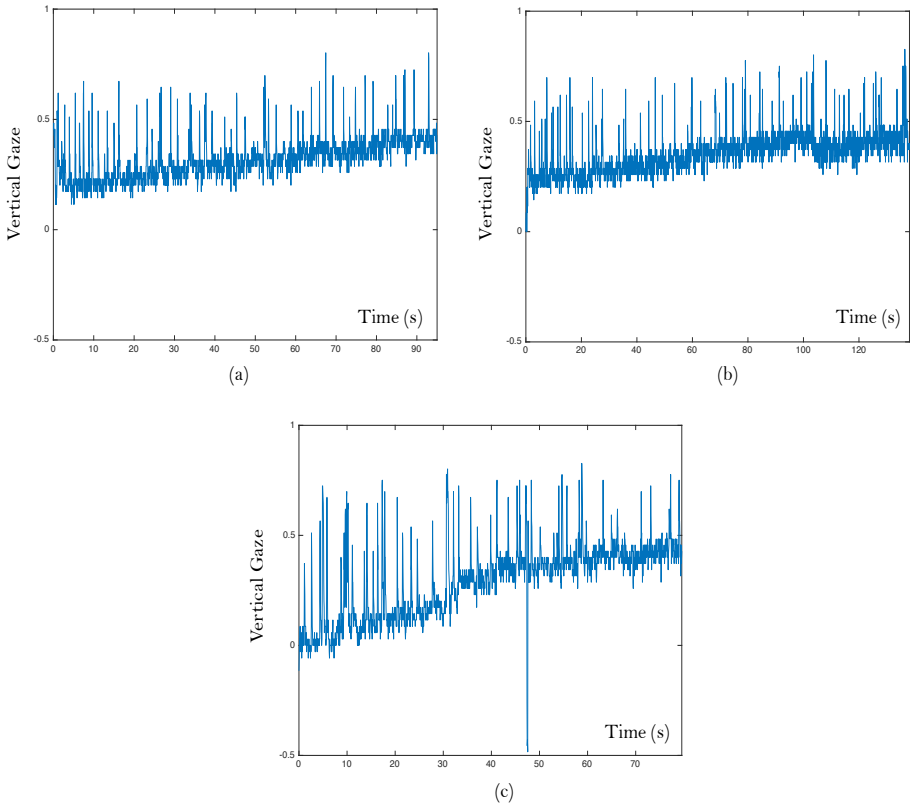

Figure 17. A reader's vertical gaze shape (y values) when (a) reading consecutively from top to bottom, (b) reading from top to bottom but rereading a set of paragraphs, (c) reading from top to bottom but skiping a set of paragraphs

In Y-only action detection method, we passed the vertical gaze data through outlier removal, an FIR filter, and sliding window action classifier to detect the actions.

Outlier removal was aimed to remove the occasional peaks on y values (Figure 17) caused by eye blinking. A simple FIR average filter was used to remove signal noises. In sliding window action classifier, we defined a y value at time t as f(t), and we classified a sample as reread if there is at least one window that satisfies:

$$\frac{f(t_{mid}) - f(t_i)}{t_{mid} - t_i} < 0 < \frac{f(t_j) - f(t_{mid})}{t_j - t_{mid}}$$

$$\text{where } t_{mid} - th1 < t_i < t_{mid} - th2$$

$$\text{and } t_{mid} + th2 < t_j < t_{mid} + th1$$

Similarly, we classify a sample as skip if:

$$0 < \frac{f(t_{mid}) - f(t_i)}{t_{mid} - t_i} < \frac{f(t_j) - f(t_{mid})}{t_j - t_{mid}}$$

$$\text{where } t_{mid} - th1 < t_i < t_{mid} - th2$$

$$\text{and } t_{mid} + th2 < t_j < t_{mid} + th1$$

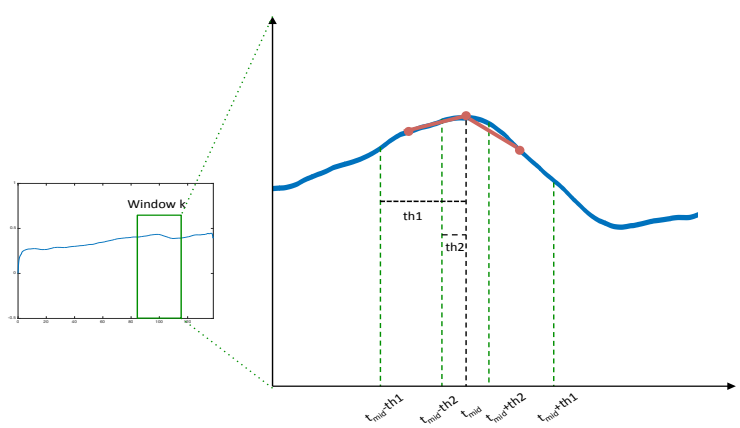

Figure 18. Sliding window reread action classifier. We classify a sample as reread if there is at least one window which has: $\frac{f(t_{mid})-f(t_i)}{t_{mid}-t_i} < 0 < \frac{f(t_j)-f(t_{mid})}{t_j-t_{mid}}$ where $t_{mid} - th1 < t_i < t_{mid} - th2$ and $t_{mid} + th2 < t_j < t_{mid} + th1$.

3.2.2 Kinematic channel

Besides eye movements, scrolling motion is one of the most important information sources during reading. Muscle activity/tension can be affected by emotions such as engagement. Researchers proved that the emotions causing muscle changes could be detected by mass-spring-damper (MSD) system when doing two-dimensional tasks [81][39]. However, this theory has seldom been used in reading analysis. Taking advantage of the rich sensors in smart phones, we proposed to track, understand and use the effects of reading emotions on scrolling action via a single MSD system.

The MSD model takes the input force from the finger(s) and arm, and outputs the scrolling characters such as trajectory. The 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. 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 damped frequency (ω) and damping ratio (ζ) of each MSD dimension can describe the scrolling movements of such dimension. We adopted the correlation between the parameters and muscle stiffness in [81]: $\omega \propto \sqrt{k}$ and $\zeta \propto \frac{c}{\sqrt{k}}$.

The input of the MSD model is the force from the finger(s) and arm and the output is the scrolling characters. However, only the scrolling characters could be observed in our study. Therefore, we used linear predictive coding (LPC) to invert the input and output of the MSD model. LPC signal model predicts future signals based on the linear combination of the observed signals in the past:

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

where $\hat{x}(n)$ is the predicted signal value, x_{n-i} the previous observed values, a_i the predictor coefficients, and p the order of the predictors [41]. To optimize a_i, we used least square error.

LPC takes the input of the observed scrolling change along each dimension, e.g. dx and the corresponding time t on the x-axis scrolling dimension, and produces a sequence of coefficients that defines the characteristic polynomial of the MSD system. We take the complex root (r) of the predicted polynomials, which reveals the damping characteristics of the MSD model in this case: damping frequency $\omega = |\Im(r)|$, damping ratio $\zeta = \frac{|\Re(r)|}{\|r\|}$ [81].

Besides x and y scrolling dimension as in [81], three more dimensions were included, i.e. touch size, touch pressure and touch orientation ratio during scrolling.

3.3 Experiments

Our experiment consisted of two phases. In phase 1, we qualified the gaze periodic lateral pattern detection accuracy to prove the validity of periodic lateral pattern based gaze features. In phase 2, we used two proposed channel features along with the traditional dwell time related features to understand reading comprehension and reading emotions.

3.3.1 Participants

25 participants (9 females) ranging from 19 to 35 years old (μ =26.32, σ =3.96) were recruited for the study. All participants have experiences with reading news on smart phone. None of the participants have dyslexia or emotional disorder.

3.3.2 Apparatus

Our experiment was completed on a Google Nexus 5x Smartphone with a 5.2 inch 1920 x 1080 pixels display, 1.8 GHz hexacore SnapDragon 808 CPU (64-bit ARMv8-A), 5MP front-facing camera and running Android 6.0. We used 15px display text size and portrait screen mode for both phases.

3.3.2.1 Phase 1

For each participant, the three reading articles in this task were randomly chosen from four New York Times articles, ranging 559 to 640 words (μ=588.50, σ=35.43) with comparable difficulties (average Flesch-Kincaid reading ease = 29.03, σ = 8.10). In phase 1, we aimed to test the performance of X-line-counting action detection method and Y-only action detection method. Therefore, a flipping page design was used to avoid the possible confounding changes of vertical gaze data caused by the changing of viewport by scrolling. Each article was divided into 3 pages (around 25 lines each flipping page, more than 90% of full screen). A flipping page button was designed at the bottom right of each page.

The study of this phase had three conditions, i.e. normal, reread and skip. Under the normal condition, we asked the participants to read each line once, and line by line in a sequence. After finishing each page, participants clicked the flipping page button to continue to the next page (or stop if finished with the article). Under the reread condition,

participants read in a similar manner but were required to read a randomly highlighted section twice. Under the skip condition, participants were asked to skip a randomly grayed-out section when reached.

3.3.2.2 Phase 2

In phase 2, each participant read 3 articles. The articles in this task were chosen from the New York Times, ranging 459 to 567 words (μ=500.33, σ=58.29) with comparable difficulty (average Flesch-Kincaid reading ease = 33.3, σ = 6.59). In order to include both the visual tracking channel and the kinematic channel in this phase, we used scrolling design (X-line-counting action detection method was used for action detection) in order to enable scrolling tracking. Each article had around 59 lines on a single page.

3.3.3 Procedure

The entire procedure was around an hour for each participant. Each procedure began with an entrance survey questionnaire about the participant's personal background and reading experiences. Once finished, two reading phases were conducted in a sequence. Before each phase, a 10 min warm-up session was conducted to introduce and let the participant get familiar with our setup. In each phase, the participant was asked to answer three short-answer comprehension questions as well as rate his/her reading emotions including focus level, confidence of understanding and engagement after each article reading.

Figure 19. Sample participants in experiment (phase 2).

3.3.4 Design and Analysis

3.3.4.1 Phase 1

We used a within-subject design with three conditions in phase 1, i.e. normal, reread and skip as described in the apparatus. Under each condition, a participant read one article with 3 flipping pages. We randomized both the conditions and the articles. One participant's data was discarded due to the corrupted data. In total, we have 24 subjects * 3 conditions * 1 article * 3 flipping pages = 216 page level samples. Since we were aimed to evaluate the periodic lateral pattern detection accuracy, we evaluated the two metrics: line detection and condition (normal/reread/skip) classification in each sample.

We used mean absolute error, mean absolute percentage error, root mean squared error and correlation to evaluate the line detection. Precision-recall curve was used to analyze the condition predictions.

3.3.4.2 Phase 2

In phase 2, we conducted a within-subject study to evaluate the feasibility of detecting reading comprehension and emotions via visual and kinematic channel features. Each of the participants read 3 news articles about different topics including gaming, astronomy, and fitness. We randomized the sequence of the articles for each participant. The participant was asked to read each article according to their reading habit, and scroll, skip or reread whenever they desired (Figure 19). In this paper, we used comprehensive questions answer rate to measure reading comprehension, and users' self-rated concentration (cognitive), interestingness (affective), confidence of understanding (affective) to understand users' reading engagement. After reading each article, the participant was asked to answer three comprehension questions as well as rate reading engagements including concentration level, confidence of understanding, and interestingness on a 7-point Likert scale. Forward-stepwise features selection method was used to investigate the visual and kinematic channel features and their effects on reading comprehension and emotions. Based on the selected features, the gains of understanding reading comprehension and emotions were evaluated by the root mean square error and R2 value of linear analysis.

3.4 Results

3.4.1 Phase 1 – Line detection

We compared three methods, i.e. SwitchBack, Read All Lines, and Periodical Patterns Detection in line detection, where our proposed Periodical Pattern Detection outperformed the other methods in all aspects (Table 4).

	SwitchBack	Read All Lines	Periodical Pattern Detection
Mean absolute error	4.91	5.68	3.56
Mean absolute percentage error	0.20	0.31	0.16
Root mean squared error	6.66	7.66	4.65
Correlation	0.70	0.15	0.83

Table 4. Line detection via SwitchBack (baseline 1), read all lines once (baseline 2) and periodical pattern detection methods.

Two baseline approaches were used to evaluate line prediction performance. SwitchBack method was implemented as a baseline with the same procedure and parameter as described in [55]. SwitchBack iterated different values to get the best threshold of gaze-periodic length for detecting the change of lines. The other baseline was made based on the assumption that each user will read every line once within a flipping page.

As shown in Table 4, our proposed periodic lateral pattern detection method achieved 0.83 correlations, and outperformed both baselines in all aspects.

3.4.2 Phase 1 – Action (reread & skip) detection

For each action (condition), we calculated the precision and recall of this condition versus all other conditions. The cutoff speed was used as our baseline: the reading is considered under skipped condition when a reader reads a page faster than a skip cutoff speed. Similarly, reread detection is defined when the reading is slower than a reread cutoff speed. Through the iteration between 100wpm~500wpm, we found that the optimized reread cutoff speed is 150wpm and skip cutoff speed is 250wpm for our participants.

The precision-recall curve was shown in Figure 20 and Figure 21 corresponding to reread and skip condition. Through the calculation of the area under curve (AUC) for reread condition and skip condition (Table 5), we found that periodic lateral pattern counting outperforms both baseline and the y-only method. This result also testifies the sufficiency of periodic lateral pattern detection accuracy.

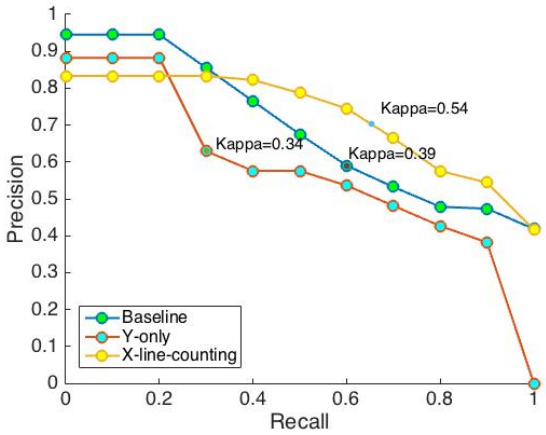

Figure 20. Reread condition precision-recall curve

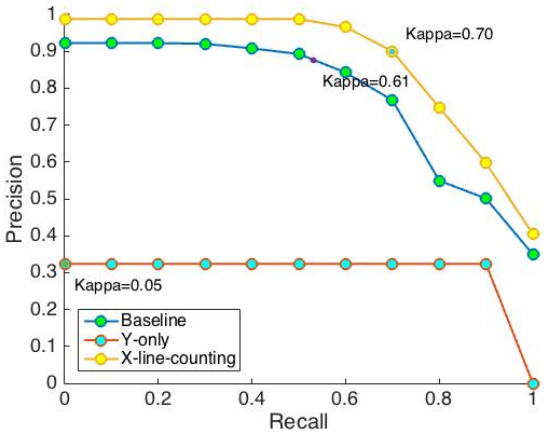

Figure 21. Skip condition precision-recall curve

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

Table 5. Area under curve (AUC) for reread and skip conditions.

3.4.3 Phase 2 – Reading comprehension and engagements

To predict reading comprehension and emotion, we proposed three sources of features: traditional, gaze and kinematic (Table 6).

Sources	**Feature Group (Count)**	**Examples**
Traditional (TF)	Dwell time (1)	Page duration, etc.
	Speed Related (2)	Speed on word level, etc.
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	Details in **Table 3**.

	pattern based (14)	

Table 6. Three sources of features: traditional, gaze and kinematic, for predicting reading comprehension and engagements.

Forward step-wise features selection method was used to select the significant features from the three sources of features (Table 7).

We found that PR, MADLL, and MADLD were the most influential ones in gaze source.

Features	Concentration	Confidence	Interesting-ness	Comprehension
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 7. The correlation and corresponding p-value of the features selected for reading engagements and comprehensions via forward step-wise feature selection method from all features bundle, where *: p-value <0.05; **: p-value <0.01; ***: p-value <0.001.

In Kinematic source, the features related to vertical scrolling movement and scrolling pressures were the most influential ones in the kinematic source. When a reader scrolls during reading, the vertical direction movements are more obvious than horizontal direction. As the reader focuses or engages with reading, the tightly loaded muscles of the readers' arm and fingers cause the increase of MSD model damping ratio. The damping frequency increases when the muscles stretched during focusing and decreases when the muscle thickened during engaging. When the reader has high engagement, the finger pressure increases with low variance.

Features	Concentration	Confidence	Interestingness	Comprehension
TF (3)	1.00(0.11)	1.07(0.33)	1.46(0.09)	0.86(0.17)
SF (16)	0.96(0.32)	1.20(0.29)	1.42(0.27)	0.92(0.21)
GF (3)	1.03(0.05)	1.22(0.13)	1.47(0.07)	0.91(0.07)

TF+SF	0.92(0.40)	1.07(0.47)	1.38(0.34)	0.87(0.32)
TF+GF	1.00(0.14)	1.04(0.39)	1.44(0.14)	0.86(0.20)
TF+SF+GF	0.89(0.46)	1.04(0.51)	1.37(0.38)	0.84(0.39)

Table 8. The root mean square error and the corresponding R^2 value when linearly modeling reading concentration, confidence, interestingness and comprehension via tradition features (TF), scrolling features (SF) and gaze features (GF).

Therefore we updated the three sets of selected features to better predict users' comprehension and emotions: 1) In kinematic source, MSD related features were reduced from 5 dimensions to only vertical scrolling and pressure dimensions; and 2) In gaze source, only PR, MADLL and MADLD were included. As shown in Table 8, we found that the combination of three sets worked the best for predicting reading emotion and comprehension. Scrolling features, among three single sets, worked the best.

To validate the model with selected features, we used 10-fold cross validation method to observe the correlation change. After 100 times iteration, the result in Table 9 proved that the understand of reading comprehension and emotions can be improved by taking pattern-based gaze features and kinematic features into consideration.

Features	Concentration	Confidence	Interestingness	Comprehension
TF (3)	0.26	0.50	0.25	0.30
SF (16)	0.22	0.21	0.21	0.11
GF (3)	0.12	0.29	0.15	0.18

TF+SF	0.27	0.38	0.26	0.20
TF+GF	0.23	**0.55**	0.19	**0.31**
TF+SF+GF	**0.36**	0.47	**0.28**	0.26

Table 9. The correlation coefficient on 10-fold cross validation via linearly modeling reading concentration, confidence, interestingness and comprehension via tradition features (TF), scrolling features (SF) and gaze features (GF).

3.5 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.

3.5.1 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.

The error rate of our reproduced SwitchBack algorithm was higher than that in the original literature [55] (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 [55]. 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 [55] 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.

3.5.2 Modality Comparison

As shown in section 3.4.3, 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 7, 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. A theoretic analysis on it is lack in this research and could be explored in the further.

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.

3.5 Limitations and Future Work

We have limitations in this research and 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. Last, we evaluated periodic lateral pattern tracking as well as actions detection on flipping page design and applied it in scrolling based page design. As claimed in previous sections, flipping page design was chosen to test the validity of horizontal and vertical gaze signals, which helped avoid possible confounding changes of vertical gaze signals caused by the viewport changes during scrolling. However, the validity of inheritating flipping-page-based gaze features for analyzing scrolling-page-based reading decrease the explanation power of our result.

3.6 Conclusion

We presented Lepton, a dual-channel mobile reading system and algorithms, to understand readers' comprehension and emotions during mobile reading. Lepton tracks the periodic lateral patterns of readers' eye gaze in the visual tracking channel and models readers' scrolling behavior in the kinematic channel. Lepton leverages signals from these two channels to infer readers' comprehension and emotions during reading. We found Lepton was able to 1) detect readers' periodic lateral pattern with 0.83 correlation on line detection and 0.73 (reread) and 0.88 (skip) AUC sizes on condition detection; 2) predict users' comprehension (318.18% increment in R-square), focus (54.55%), confidence (322.22%), and engagement (129.41%) more accurately compared to using traditional features.

4.0 STRATEGICREADING: MEASURING STRATEGIC READING ON SMARTPHONE AUTOMATICALLY

Our previous studies have shown the validity of gaze- and touch- tracking techniques on unmodified smartphones to help us better understand comprehension and emotions during single page text readings. In this chapter, I leverage such techniques to understand online reading strategies and performance, and the difficulties of negotiating and learning from complex digital resources.

4.1 Introduction

The pervasive acceptance, mobile convenience and powerful functionality of smartphones makes them an alternative choice to reading devices that replace offline papers and PCs for many users, especially adolescents and college students [89]. The smartphone is also becoming an influential platform for educational purposes, offering diverse domains of reading and learning assistance such as providing course work reading tasks [29] and offering self-learning reading materials [30][80].

To successfully perform the complicated online smartphone reading tasks, users' skill set for processing a single page of text is insufficient. For example, in such tasks, it is important before comprehending a text to first understand how to locate it and how to validate if it is worth reading. Therefore, understanding how users strategically perform in such online reading tasks and having a portable and intelligent way to measure their performance are important for both research and practical purposes [3][23][5][17].

A widely accepted model of online reading strategic processes has four components: information location, meaning making, source evaluation and self-monitoring. The quantity and quality of each model component affect the reading performance. Existing literacy studies [19][17] mainly rely on readers' verbal reports during think-aloud reading to interpret and measure the model of the strategic processes. However, using verbal reports to understand reading strategies has at least three major challenges. First, think-aloud reading might affect readers' strategic processing and consequently might negatively impact researchers' abilities to accurately observe and interpret users' cognitive engagement [17]. Second, the accuracy of readers' verbal reports on cognitive tasks are considerably varied, depending on individual and contextual factors [28]. For example, some readers lack accurate expression of their thoughts. Third, verbal reports are time-consuming and labor-intensive to analyze, evaluate, and interpret. These challenges prevent an efficiently valid and large-scale data-informed understanding of strategies employed by readers.

To automatically extract users' strategic processes of online reading tasks without verbal reports and users' extra efforts, we took advantage of the mobility and functionality of smartphones: sensing users' unnoticeable reading behaviors including users' gaze movements, scrolling actions and logging paths during reading. The gaze channel captures the perceptual activities in reading. The scrolling channel represents the kinematic activities and the logging information records the sessional activities of users. We posit that these three channels are complementary and could help us to get a richer and deeper understanding towards users' reading strategies. We created a novel system called StrategicReading--a multimodal interface, which automates the detection and

evaluation of reading strategies to predict online reading performance on unmodified smartphones. First, StrategicReading tracks the periodic lateral patterns of users' gaze movements via the front camera. Second, it infers their muscle stiffness via the Mass-Spring-Damper (MSD) based kinematic model and determines scrolling types during text scrolling. Last, it monitors the evolution of cross-page behavior by logging the searching and clicking history. Overall, StrategicReading combines the gaze channel, scrolling channel, and logging channel to improve the detection, evaluation, and understanding of reading strategies and reading performance.

To the best of our knowledge, the novelty of this project includes:

1) This study is the first attempt to understand authentic online reading strategies used on smartphones,

2) This study is the first attempt to automate the detection of complex reading strategies via multi-channel sensing techniques.

4.2 Related Work

A reader's reading comprehension strategies are developmental in nature, and they are learned and practiced until fluency of strategy use is achieved [1]. Research on reading strategies has been developed for more than 40 years in order to better understand the intricate workings of the mind [64]. Reading strategies can be formulated into systematic types of processes, including prior knowledge use, inferential reasoning, self-regulation, and affective variables related to efficacy and motivation [19].

Research on online reading strategy has started in the early 90s, where an emerging group of research works have examined the differences of reading strategies between print text reading and online text reading [50][72][68][9]. Coiro provided a good

summary of such differences [19], including the differences caused by visualization of contents, quality of contents, and supported functions of online reading, etc. Currently, one of the well-accepted conceptual frameworks of online reading strategy includes four core online reading processes including (a) information location, (b) meaning making, (c) source evaluation, and (d) self-monitoring [1][17].

Information location involves generating web-search words, strategically surveying the search-results pages, and selecting useful hyperlinks from numerous conjoined ones. This allows opportunities for readers to learn from texts. Source evaluation helps readers to locate reliable and relevant information. Since online information varies in accuracy, reliability and usefulness, readers need to be skeptical, tentative, and critical in examining information sources and detecting consistencies or conflicts among the sources. Meaning making is a process of constructing meaning from the located information. The meaning making process allows readers to identify important ideas, build specific intertextual linkage, and elaborate a cross-textual understanding. The last process is self-monitoring, according to which a reader knows and adjusts her own thinking at a metacognitive level. Readers' self-monitoring guide how they select, apply, and evaluate strategic actions in response to texts, tasks, and situations of reading.

To examine reading strategies, a variety of methods have been used, including protocol analysis of verbal reports, theoretical task analyses, eye tracking, protocol logs, observations of readers, and self-reports. Afflerbach provided a detailed review of these methods [1], among which, verbal protocol analysis, the examination of spoken records of readers' thinking and behavior, is the most frequently picked method to understand

online reading strategies [19][5][17][23][20]. The popularity of verbal protocol analysis derives from its possibility of getting inside the minds [69], capability of uncovering potential cognitive processes [79], and the maturity of its long developed methodology [26].

However, using verbal reports to understand reading strategy has at least three major challenges. First, think-aloud reading might cause negative impact on readers' cognitive engagement and affect their strategic processes [17]. Second, accuracies of readers' verbal reporting on cognitive tasks are considerably various [28]. The consistency of doing and saying and the ability of readers' verbal expressions are not always guaranteed. Third, verbal reports are time and labor consuming to generate and grade. This limitation prevents the large-scale understanding of reading strategies.

With the development of technologies, leveraging sensors to track readers' behavior and thoughts is a potential approach to solving such challenges. For example, Rayner gave a clear and detailed summary regarding how eye movements relate to information processing during reading [71]. Researchers also attempted to interpret cognitive processes in reading [47], comprehension strategies [71][7][56], reading proficiency [16], and reading engagement via gaze tracking. However, the portability and quantitative nature of measurement limits the ability to reveal higher-level strategic processes via eye tracking, especially for the task of reading, which involves locating and comprehending multiple pages on a smartphone.

To increase the portability, StrategicReading tracks the gaze movement via the front camera of a smartphone. The required accuracy level can be decreased by tracking the periodic return sweep of gaze to estimate the reading position and to predict reading

behavior. StrategicReading also infers the users' muscle stiffness during typing, clicking and text scrolling and monitors the reading log data during online reading tasks in order to undertake a fine-grained analysis of reading strategies. Overall, StrategicReading combines (a) visual tracking channel, (b) kinematic channel, and (c) log channel to improve the detection and understanding of reading strategies.

4.3 Design

To automatically extract users' strategic processes of online reading tasks on a smartphone without verbal reports and users' extra efforts, we designed the StrategicReading system which leveraged users' gaze movements, scrolling actions and logging paths during reading.

4.3.1 Gaze Movements

A strong relationship has been discovered between eye movements and cognitive processes with the help of eye tracking technologies. Most of the traditional gaze features are derived from micro and concise eye movements, such as saccades and fixations, which require dedicated eye trackers to accurately observe.

To prevent disturbing users' reading with extra devices such as the mounted eye tracker, we chose to use the front-facing camera of the smartphone to track users' gazes. Such a low-cost and low-resolution eye tracking method cannot accurately track the exact fixations', saccades' locations, durations and lengths. Therefore, the traditional gaze features (e.g. the statistics of fixations and saccades) are no longer applied. Our previous project Lepton [35] used periodic pattern based eye gaze features to predict users' comprehension and engagement in mobile reading. The Lepton project proved the

accuracy of tracking the periodic patterns of saccades and proved the feasibility of using such features to predict reading comprehensions. We adopted the features related to period lateral gaze patterns to investigate the users' reading strategies and reading performance.

Feature	Definition
PR (f1)	Predicted periodic lateral patterns divided by number of lines in reading material
STDX (f2)	Standard deviation of x-axis of gazes
STDY (f3)	Standard deviation of y-axis of gazes
rMSLL (f4)	The square root of the mean squared adjacent predicted line lengths' difference
rMSLD (f5)	The square root of the mean squared adjacent predicted line durations' difference
M1ADLL (f6)	Mean of absolute deviation of predicted line length
M1ADLD (f7)	Mean of absolute deviation of predicted line durations
M1ADLY (f8)	Mean of absolute deviation of line mean Y-axis of gazes
MADLL (f9)	Median of absolute deviation of predicted line lengths
MADLD (f10)	Median of absolute deviation of predicted line durations
MADLY (f11)	Median of absolute deviation of line mean Y-axis of gazes
STDLL (f12)	Standard deviation of predicted line lengths
STDLD (f13)	Standard deviation of predicted line durations
STDLY (f14)	Standard deviation of line mean Y-axis of gazes

Table 10. The periodic saccade pattern based eye gaze features.

4.3.2 Scrolling Actions

In addition to gaze behaviors, the touch-screen of the smartphone offers another available behavior channel--the scroll channel. Our previous project, Lepton, which proved that the Mass-Spring-Damper (MSD) modeled scrolling behaviors, can reliably predict users' comprehension and reading engagement [35]. Therefore, we adopted its interpretation method: considering users' arm and finger(s) as a mass (m) attached to users' muscles represented as a spring component with spring constant k and a damper with damping coefficient c. When the users' are scrolling, the users' scrolling forces are determined by their muscle stiffness: The mass oscillates at a rate related to the tension of the spring with the damping frequency ω, where $\omega \propto \sqrt{k}$. And the oscillation decays

exponentially based on the friction of the damper with the damping ratio ζ, where $\zeta \propto \frac{c}{\sqrt{k}}$. Our MSD model uses the forces of arm and finger(s) as the input to predict the changes of scrolling dimensions, such as scrolling movement on the horizontal direction. To observe the scrolling characters ω and ζ, we used linear predictive coding (LPC) to invert MSD model's input (muscle stiffness) and output (scrolling characters).

LPC linearly combines the observed signals to predict the future signals: $\hat{x}_n = \sum_{i=1}^{p} a_i x_{n-i}$, where $\hat{x}_n$ is the predicted signal value, x_{n-i} is a previously observed value at the i-th order, a_i is the predictor coefficient at i, and p is the order of the predictors. In this process, we used the observed xn and x_{n-i} to calculate the coefficient a_i. The calculated coefficient a reveals the MSD model damping frequency and damping ratio $\omega = |\Im(r)|$, $\zeta = \frac{|\Re(r)|}{\|r\|}$, where r is the complex root of a.

For each scroll, we extracted 5 dimensions of scrolling and each with 2 MSD features (ω and ζ). The 5 dimensions included the scrolling movements on horizontal (X) and vertical (Y) axes, scrolling touch-pressure (P), touch-size (S) and touch orientation ratio (R).

When applying scrolling features, we aggregated the features of each individual scroll into a feature vector using descriptive statistics, such as mean and max.

4.3.3 Logging Paths

For the logging path channel, we used two straightforward features. We used the number of pages visited by a user as well as the number of search terms used by a user.

4.4 Experiment

In this project, our goal was to answer the following three questions:

G1: Can we automatically extract users' behavior information on a smartphone to predict their reading strategies?

G2: Can we then accurately predict users' phone-based online reading performance?

G3: How is users' reading performance affected by other reading behavior information?

To investigate the answers for these questions, we conducted a study that adopted the design of the online question-generation reading tasks in previous research [17]. In the question-generation reading task, users were asked to generate a critical question on a controversial topic based on the information and knowledge gained from the online sources. In this project, we also adopted the controversial topic "mountaintop removal coal mining" [17], because this topic had been proven to be significant when evaluating desktop-based online reading strategic processes for adolescent students.

4.4.1 Participants and Apparatus

Forty freshmen and sophomores (11 males) ranging from 18 to 25 years old (=18.68, =1.23) were recruited from a local college. Their majors included computer science, engineering, biology, political science, psychology, business, linguistics, nursing and pharmacy. The limitation to students at these specific academic levels ensured a plentiful source of participants and achieved maximum possible significance for the chosen topic.

Among 40 participants, 39 participants preferred reading in the portrait mode of the smartphone. Therefore, our study was conducted on the portrait model. One participant had dyslexia. Among all 40 participants, 37 participants' data were valid.

To avoid the users' reading performance changes due to the capabilities of their own phone, our study used a Google Nexus 5X lab phone for all participants to perform this task. Our lab phone had StrategicReading, which was a system utilized the Google search engine. The users could search and view all the sources from the Internet.

4.4.2 Experiment Procedure

The whole study lasted around 2 hours for each participant in two separate days a week apart. The first day took around 0.5 hour and the second day around 1.5 hours.

The study consisted of 6 sessions:

Session 1 (Day 1): Participants were asked to respond to a researcher-developed background knowledge assessment related to the controversial topic. The rest of the study was continued at least one week after this session to minimize the carry-over influence of the assessment material on the online reading task.

Session 2 (Day 2): Participants were then asked to finish a training session, in which participants familiarized themselves with our system and question generation task. The training session also aimed to help participants understand the characteristics of high-quality questions through examples and discussions.

Session 3 (Day 2): Participants then conducted the 20-minute online reading task on the StrategicReading app in a Nexus 5X smartphone connected to the Internet. Participants were encouraged to use up the entire 20 min for online reading. During this

session, the phone screen was recorded simultaneously and participants were not allowed to write notes.

Session 4 (Day 2): After the completion of the online reading task, participants repeated the knowledge assessment in session 1.

Session 5 (Day 2): After finishing the post-reading knowledge measurement, participants then generated and wrote up a critical question regarding the topic. Participants were not allowed to look back to the phone during this session.

Session 6 (Day 2): Finally, participants reviewed the in-session recorded videos and gave retrospective verbal reports of their moment-to-moment thinking processes.

4.4.3 Metrics Design

Our study was designed to understand and prove the possibility of automating the prediction of *reading performance* and *reading strategic processes* via users' behavior information.

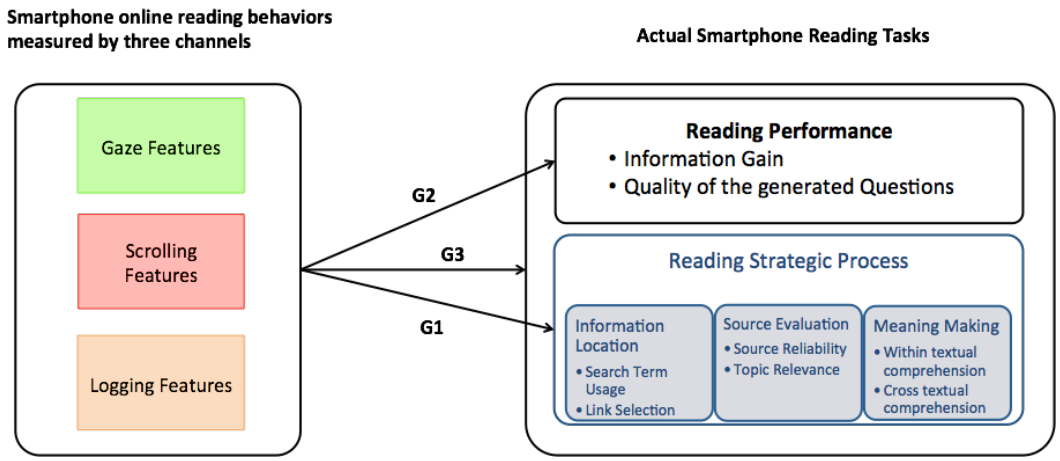

Figure 22. Study Overview

4.4.3.1 Reading Performances

Two types of reading performance were targeted in this project.

The first type was the knowledge gain, which was the growth of users' knowledge and understanding of the topic after performing this online reading task. To collect users' knowledge gain, we implemented knowledge assessment sessions before and after users' online reading task (study session 1 and session 4). We adopted the same knowledge assessment questionnaire of the topic "mountaintop removal coal mining" from previous research [17]. The questionnaire had 10 multiple-choice questions, 10 true-or-false questions and 5 short-answer questions. These questions were targeting general information, in-depth knowledge, and critical thinking on the topic. The information gain was the improvement of scores from prior to post knowledge assessments.

The second type of reading performance was the quality of the user's generated question. Following the grading rubrics in previous research [17], the generated critical question was graded from three perspectives: validity, relevance and significance. The grading of validity was based on the soundness of the supported details users' provided in the generated questions to support the topic. Relevance was graded by the closeness between the generated questions and the topic. Significance addressed how well such generated critical question would facilitate critical and multi-perspective thinking. Each of the three perspectives had four quality levels when grading: lacking-0, partial-1, adequate-2 and complete-3. The score range of a generated critical question was from 0 to 9.

4.4.3.2 Reading Strategic Processes

The online reading task required users to process, filter, comprehend and digest multiple sources of contents. The reading strategies for complex online reading were

summarized into four categories: information location, meaning making, source evaluation and self-monitoring [17]. These strategies were investigated and proved by desktop online reading tasks [17]. Given the similarity of desktop and smartphone online reading, our project inherited such strategic modeling for our project.

Information Location

Information location is the core strategy for users to navigate to the information. During this process, users apply meaningful search terms to conduct the information searches. Among the search results, users select relevant links and reject irrelevant ones.

Meaning Making

Meaning making is the information absorbing process. Given the contents, users understand important information within each link or page, as well as building intertextual relationships among the absorbed information and elaborating a meta-level understanding of the overall concept.

Source Evaluation

Source evaluation is the strategic process that helps users to filter out the unrelated or unreliable sources in order to increase the efficiency of online reading. During the source evaluation process, users determine the reliability of a source and the source's significance to investigate the issue. In addition, users also determine the relevant sources and they progress and link the sources to their reading needs.

Self-Monitoring

Self-monitoring is a strategic process that helps users to adjust and refine their own understanding of the concept by interacting with the online contents. This process includes the gradual changes within each of the other three strategic processes.

According to previous research, the self-monitoring process is at a metacognitive level and hard to materialize, and can be primarily represented by the other three strategic processes. Therefore, in this project, we only include the first three. Table 11 represents the strategies that we would like to automatically predict.

Information Location	s1	Search terms (Applying search terms & conducting information searches)
	s2	Links (Selecting relevant links & rejecting irrelevant ones)
Meaning Making	s3	Within textual comprehension (understanding important information)
	s4	Cross textual comprehension (building intertextual relationships and elaborating a metalevel understanding)
Source Evaluation	s5	Topic relevance (determining relevant sources and linking it to topic)
	s6	Source reliability (Author & discerning reliable sources & assessing each source's significance to read)

Table 11. The reading strategic processes

The reading strategies were to be predicted via our three channel features. Therefore, it was important to have the ground truth of their scores to check the performance of our predictions. We counted the numbers of the appearances of a user's strategic reading processes (s1-s6) from his/her retrospective verbal report as the labels of the ground truth. Domain experts authorized the quality of the grading.

4.5 Results

Following our three goals, we analyzed the results in four sections: 1) Before other analyses, we first looked into the rationality of adopting strategic reading processes from desktops to smartphones, we then 2) auto-extracted reading strategies, 3) predicted reading performance and 4) investigated other patterns.

4.5.1 Rationality of Adopting Reading Strategic Processes from Desktop to Smartphone

Although the reading strategic processes that we used in this project were proven to be valid on a desktop online reading environment [17], when adapting from the desktop to the smartphone, the validity of the strategic processes was unexplored. Before we used the strategic processes in the result analysis, we first proved their validity on the smartphone online reading environment.

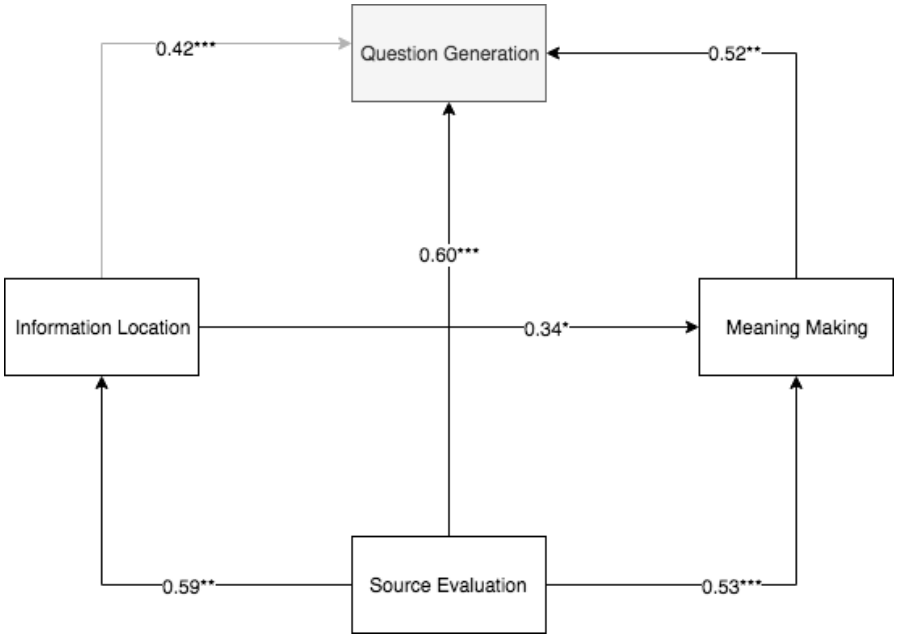

Figure 23. Path analysis of the three strategies (information location, meaning making, and source evaluation) and the critical questioning quality: The correlation coefficient between the frequencies of strategies and the quality of critical questions ($p<0.001$ ***, $p<0.01$ **, $p<0.05$ *).

A path analysis, similar to a previous project [17], was generated using the correlations among these strategies' appearance frequencies as well as the correlations between strategic processes and the quality of the generated questions (Figure 23).

Significant correlations were found among strategic processes and between strategic processes and the quality of the generated questions. Such results were consistent with the analysis of the desktop environment. These results proved that the strategies could reliably predict reading performance during online reading tasks on smart phones.

4.5.2 Auto-extract Reading Strategies

We processed and analyzed the three types of readers' behavior data (gaze, scrolling and logging data) and used them to predict users' reading strategies.

4.5.2.1 Gaze Channel

First, we observed the linear correlation coefficients between gaze features and the appearance frequencies of strategic reading processes across the participants (Table 12).

Among all the gaze features, the predicted number of lines (PR) had positive correlation with the within-textual comprehension of the meaning making process (s3), where the correlation coefficient was 0.59 ($p<0.001$). In addition, features related to line durations differences had a negative correlation with within-textual comprehension, for example rMSLD had the correlation coefficient -0.54 ($p<0.001$), M1ADLD with correlation coefficient -0.52 ($p<0.01$) and STDLY with correlation coefficient -0.54 ($p<0.001$).

		Information Location		Meaning Making		Source Evaluation	
		s1	s2	s3	s4	s5	s6
Gaze Features	PR	0.18	0.15	**0.59*** **	0.3	**0.37***	0.07
	STDY	0.01	0.31	**-0.34***	-0.24	-0.03	-0.02
	rMSLD	-0.21	-0.1	**-0.54*** **	-0.3	**-0.35***	-0.01
	M1ADLD	-0.23	-0.09	**-0.52** **	-0.29	**-0.35***	0.01
	M1ADLY	0	0.27	**-0.35***	-0.23	-0.03	-0.05
	MADLL	-0.13	-0.15	**0.35***	0.18	0.28	0.01
	MADLD	-0.17	0.02	**-0.38***	-0.23	-0.28	0.06
	STDLD	-0.21	-0.1	**-0.54*** **	-0.3	**-0.35***	-0.01
	STDLY	-0.01	0.28	**-0.39***	-0.25	-0.05	-0.04
Scrolling Features	meanWX	-0.05	0.16	0.09	0.08	**0.33***	-0.01
	meanCY	0.09	-0.16	**-0.34***	-0.11	0.04	-0.17
	meanWP	-0.2	-0.05	**-0.43** **	-0.25	-0.11	-0.25
	meanWS	0.06	0.2	**-0.42** **	0.05	0.1	0.12
	meanWR	-0.01	0.09	**-0.48** **	-0.14	0.03	-0.1
	stdWX	-0.1	-0.14	-0.11	**-0.34***	-0.14	-0.07
	stdWS	0.14	0.16	**0.39***	-0.04	-0.03	0.09
	stdCS	0.04	-0.17	**0.47** **	-0.03	-0.13	0.01
	varWX	-0.1	-0.13	-0.11	**-0.34***	-0.14	-0.07
	varWS	0.13	0.17	**0.39***	-0.04	-0.03	0.09
	varCS	0.03	-0.18	**0.47** **	-0.03	-0.13	0.01
	medianWP	-0.12	-0.06	**-0.51** **	-0.18	-0.11	-0.13
	medianWS	0.03	0.24	**-0.33***	0.03	0.15	0.09
	medianWR	-0.01	0.09	**-0.47** **	-0.16	0.05	-0.14
	rangeWR	**-0.48** **	-0.08	-0.25	-0.15	-0.26	-0.06
	# scrolls	0.22	**0.33***	-0.19	0.24	0.08	**0.39***
Logging Features	# domains visited	**0.54*** **	**0.43** **	0.01	0.23	0	**0.50** **
	# search terms used	**0.79*** **	0.21	0.04	**0.37***	0.06	0.15

Table 12. Correlation coefficients between three channel behavior features and reading strategies' frequencies (p<0.001 ***, p<0.01 **, p<0.05 *). Features without significant correlation with strategic processes were not listed in this table.

4.5.2.2 Scrolling Channel

The linear correlations between scrolling features and strategic reading processes were also explored. The representative scrolling features, which had significant correlations with some strategic processes, were shown in Table 12.

The scrolling pressures and touch-size could be used to predict users' within-textual comprehension of meaning making (s3): the more a user used within-textual comprehension strategy, the more relaxed the user was when compared with performing the other strategies because we observed decreased muscle damping frequencies of pressure (WP), touch-size (WS) and touch-ratio (WR) caused by muscle relaxation. For example the mean and median of the damping frequency of pressure (WP) had the correlation coefficients -0.43 ($p<0.01$) and -0.51 ($p<0.01$). The results were presented in Table 12.

Besides the within-textual comprehension strategy, some strategic processes were also significantly correlated with scrolling feature. Stable scrolling muscle frequencies on horizontal directions indicated more use of cross-textual comprehension (s4). The standard deviation and variance of WX both had the correlation coefficient -0.34 ($p<0.05$) with cross-textual comprehension. We also observed that the increased number of scrollings led to more use linking selection (s2) and source reliability checking (s6), where the correlation coefficients were 0.33 ($p<0.05$) and 0.39 ($p<0.05$).

4.5.2.3 Logging Channel

For the features in the logging channel, we observed that the information location strategic processes (search terms used strategic process s1 and the linking selection s2) could be highly significantly indicated by the logging features (Table 12). The number of domain visited by a user had 0.54 ($p<0.001$) and 0.43 ($p<0.01$) correlations with s1 and

s2. And the number of search terms used accurately revealed s1 with correlation 0.79 ($p<0.001$).

The user's cross-textual comprehension (s4) was also reflected by using more search terms. And users' source reliability checking (s6) was positively related to the number of domains the users visited.

4.5.2.4 Overall Comparison

With the selected features in Table 12, we evaluated the root mean square error and R^2 value of the linear model (Table 13), and found that the combination of three sets had the largest R^2 value for all six strategies.

	s1	s2	s3	s4	s5	s6
GF	2.3(0.23)	**1.85**(0.41)	4.58(0.64)	3.9(0.17)	2.17(0.19)	3.18(0.15)
SF	2.53(0.32)	1.98(0.5)	6.42(0.48)	3.48(0.51)	2.38(0.28)	2.77(0.52)
LF	1.44(0.62)	1.93(0.19)	6.79(0)	3.54(0.14)	**2.15**(0.01)	**2.58**(0.29)
GF+SF	2.36(0.58)	2.03(0.63)	**3.69**(0.88)	3.41(0.67)	2.27(0.54)	2.83(0.65)
GF+LF	1.28(0.78)	1.87(0.44)	4.37(0.7)	3.51(0.38)	2.25(0.2)	2.83(0.38)
SF+LF	1.54(0.77)	1.98(0.55)	6.58(0.5)	3.11(0.65)	2.27(0.41)	2.72(0.58)
GF+SF+LF	**1.2(0.91)**	2.18(**0.64**)	3.79(**0.89**)	**2.5(0.85)**	2.21(**0.63**)	2.95(**0.67**)

Table 13. The root mean square errors (the smaller the better) and the corresponding R^2 value (the larger the better) for predicting reading strategies via gaze (GF), scroll (SF) and logging channel features (LF).

We also evaluated the correlation coefficients in user-independent linear regression models via a leave-one-subject-out validation (Table 14) and found that the features in gaze and logging channels helped improve the correlation coefficients on predicting strategies.

	s1	s2	s3	s4	s5	s6
GF	-0.07	0.15	**0.92*** **	0.89***	0.89***	0.87***
SF	-0.34*	-0.12	0.78***	0.75***	0.75***	0.73***
LF	**0.75*** **	**0.58*** **	0.9***	0.88***	0.88***	**0.88*** **
GF+SF	-0.05	0.02	0.86***	0.81***	0.8***	0.78***
GF+LF	0.73***	0.55***	**0.92*** **	**0.9*** **	**0.9*** **	**0.88*** **
SF+LF	0.43***	0.3*	0.77***	0.75***	0.75***	0.74***
GF+SF+LF	0.57***	0.29*	0.85***	0.84***	0.82***	0.79***

Table 14. Correlation coefficients by leave-one-subject-out validation on linear regression models via different feature channels and different combinations of feature channels.

From all the above results, every strategic process can be predicted via some channels of users' behaviors (Table 15).

Strategic Processes	Channels of Features
s1	S+L
s2	G+S+L
s3	G+S
s4	S+L
s5	S
s6	G+S+L

Table 15. The strategic processes and their corresponding significantly correlated channels of features. G-Gaze, S-Scrolling, and L-Logging channel.

4.5.3 Predict Reading Performance

The primary reason for understanding users' strategic processes was to better predict and improve the users' reading performance. Therefore, whether we could accurately predict the users' reading performance was critical and important. As mentioned above, we had two important criteria to measure the reading performance: information gain (IG) and the quality of the users' generated questions (CQ).

In this section, we aimed to compare the predictions qualities of reading performance via strategic processes and users' behavior features.

There were almost 100 features of the three behavior sensing channels. To avoid over-fitting, a stepwise feed-forward method was used to select the most significant features. For each iteration, the stepwise feed-forward method found the most significant feature that achieved the lowest p value in predicting the outcome when combined with the existing selected features (if any), and added it into the selections. It continued this process until no additional features combining with selected features could achieve $p<0.05$.

To compare the performance between the strategies and our three channel features, we used R-square and root mean square error (RMSE). Lower RMSE is better, higher R-square is better.

	IG	CQ	CQ-Relevance	CQ-Validity	CQ-Significance
Strategies	NaN	0.33 (0.36)	**0.33 (0.42)**	0.36 (0.30)	0.39 (0.30)
3ChannelFeatures	**0.20 (0.28)**	**0.29 (0.54)**	NaN	**0.30 (0.56)**	**0.33 (0.51)**

Table 16. Linear fitting model of the selected features from each feature-group to predict reading performances. The results were presented in RMSE(R-square). NaN means no significant features found.

	Strategies	**3ChannelFeatures**
IG	/	maxWX+; varWS+; rangeCS+
CQ	S5, S6	meanWS+;varWR+;medianCY-;#domains visited+
CQ-Relevance	S5, S6	/
CQ-Validity	S4, S6	meanCY+;varWR+;medianCY-;# domains visited+
CQ-Significance	S5, S6	stdCY+;stdWR+;medianCY-;# domains visited+

Table 17. The features of each domain selected via the stepwise selection method in predicting reading performance.

The selected features that contributed to predicting reading performances were listed in Table 17. Among all the behaviour features from the three channels, we found that the scrolling features (maxWX, varWS and rangeCS) were good at predicting information gain. The scrolling features (meanWS, varWR, medianCY) and logging features (number of domains visited) were reliable for predicting the quality of the critical questions.

From our results, the three channels' features were performing better than strategic processes in predicting user information gain and the quality of the critical questions. The behavior features were able to assist strategic reading processes in understanding users' reading performance.

In addition to goal 1 and 2, we also explored other potential patterns.

4.5.4 Potential patterns

4.5.4.1 Searching vs. Reading

During the online reading tasks, there were two major behaviours the users performed: searching and reading. We investigated how the users' allocated their time and energies in searching and reading, and how such allocation affected their reading performance.

We used the ratio of pages visited for searching purposes to pages visited for reading purposes as the indicator for energy allocation. The time allocation was the duration ratios between them.

By observing the correlation coefficients between users' resource allocation and the strategic processes and the reading performance (Table 18), we found that the

appliance of search terms (s1) linearly revealed such distribution. A user's energy allocation on searching was significantly reflected by the search terms the user used (correlation coefficient r=0.37, p<0.05), and his/her time spent on searching was highly significantly reflected by the application of search terms (correlation coefficient r=0.65, p<0.001).

Other than the linear correlation, we also tried the Logistic regression and SVM classification to explore the non-linear effects. A very similar result had been observed for logistic regression: there were significant relationships only between the search/reading distribution and the search term usage (s1). For SVM classification, we used a 10-fold cross validation with 50 iterations to test the results (37 users, 33 for training, 4 for testing). And we observed no noticeable results.

	s1	s2	s3	s4	s5	s6	IG	CQ
search#/read#	**0.37***	0.02	-0.09	0.1	-0.08	-0.11	-0.01	-0.07
searchDuration/readDuration	**0.65*****	0.29	-0.08	0.23	0.01	0.07	-0.2	0.14

Table 18. The correlation coefficients of search read ratio vs. strategies and reading performance.

4.5.4.2 Domain Reliability Levels

The source domain of the contents directly determined the quality and reliability of the contents. To better understand the source evaluation strategic processes (s5 and s6), we separated users' visited domains into 5 different reliability levels: .edu, .gov, .org, .com, and others.

We used the proportion of pages visited of each level to predict the users' strategic processes and their reading outcomes. We grouped .edu and .gov into a reliable group, and .org, .com and others into an unreliable group. The following table (Table 19) shows

the correlation coefficient of different reliability groups versus strategic processes and reading performance.

	Reliable Sources	Unreliable Sources
s1	-0.23	-0.07
s2	0.26	-0.18
s3	-0.05	-0.23
s4	-0.13	0
s5	-0.07	-0.12
s6	**0.34***	0.09
IG	-0.16	**-0.34***
CQ	-0.01	-0.11
CQ-relevance	0.01	-0.02
CQ-validity	-0.05	-0.27
CQ-significance	0.01	-0.04

Table 19. The correlation coefficients of reliability-groups count and duration distributions vs. strategies and reading performances. (.edu, .gov and into reliable group; .com and others as unreliable group)

We found that the users who had more source reliability checkings (s6) clicked and read more pages from reliable sources. The more people clicked and read on unreliable sources, the less information the users gained. We also investigated the durations within different reliability levels, and found they were not as informative as the visited counting.

4.5.4.3 Scrolling Habits

To investigate how users' habits of scrolling affected / were affected by their reading abilities, we observed the users' scrolling lengths on different directions as well as the time they spent on each scroll. We sought to find some scrolling patterns of "good readers".

First of all, we extracted the statistical features of the users' scrolling lengths and durations. We did a t-test on each scrolling feature regarding the reading performance. We found that all features rejected the null hypothesis when predicting users' information gain and the quality of the generated critical question. We then observed the linear correlation between such features and their reading performances (Table 20). We found that the vertical length of a scroll was the most informative channel.

	IG	CQ	CQ-Relevance	CQ-Validity	CQ-Significance
mean HorzLength	-0.19	0.03	-0.06	0.16	-0.01
mean VertLength	-0.11	**0.41***	**0.33***	**0.45****	**0.37***
mean Length	-0.11	**0.36***	0.28	**0.41***	0.32
mean Dur	0.05	-0.1	-0.12	-0.05	-0.12
std Horz Length	**-0.34***	-0.04	-0.09	0.05	-0.07
std VertLength	-0.24	0.27	0.24	0.31	0.22
std Length	-0.28	0.3	0.25	**0.34***	0.24
std Dur	0.02	-0.14	-0.16	-0.11	-0.13

Table 20. Linear correlation coefficient between scrolling pattern features and reading performance.

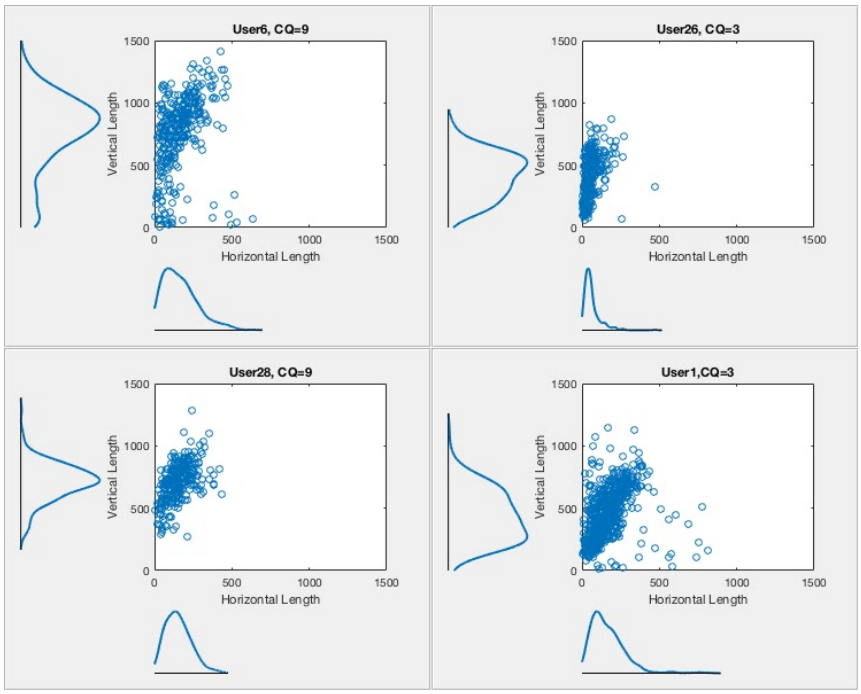

Figure 24. Examples of users' scrolling lengths Comparison

Therefore, we investigated into the vertical lengths of scrolls. By separating vertical scrolling lengths into line-wise (<10% screen length), section-wise (10%~33% screen length) and page-wise (>33% screen length) categories, we observed that the more information a user gained and the higher quality of his/her generated critical question, the less line-wise scrolls he/she would perform. The more section-wise scrolls a user used, the better he/she gained the knowledge. The more page-wise scrolls a user used, the better he/she created the critical question (Table 21).

	IG	CQ	CQ-Relevance	CQ-Validity	CQ-Significance
Line-wise scrolling	-0.1	**-0.34***	-0.24	**-0.33***	**-0.37***
Section-wise Scrolling	**0.33***	-0.31	-0.3	-0.32	-0.24
Page-wise Scrolling	-0.25	**0.44****	**0.39***	**0.44****	**0.39***

Table 21. Correlation coefficient of the proportion ratio of different scrolling types and reading performance.

4.6 Discussions

4.6.1 Channels Comparison

We found that the features of the three behavior channels (gaze, scrolling and logging) worked differently in predicting users' strategic processes and predicting reading performances.

In predicting users' reading strategic processes, the gaze channel was good at predicting information location strategic processes as well as the source reliability checking strategic processes. In comparison, users' meaning making strategic processes were well captured by the scrolling channel features. The logging channel features were helpful in predicting strategic processes except the comprehension within one page.

The raw features in three channels were too coarse to make a reliable and accurate prediction of users' reading performance. For gaze channel, we extracted the gaze features from search pages and reading pages separately. We found that, the gaze features of searching pages could accurately predict the quality of users' generated questions (Table 22). For the scrolling channel, we found that the deeper a user got into the topic, the longer scrolls the users would use. For logging channel, we separated users' visited pages into different reliability level and found that the more time users spent on unreliable sources, the less efficient the users got the information.

	Accuracy	Precision	Recall	Kappa
IG	0.6	0.5	0.4	0.2
CQ	**0.8**	**0.8**	**0.9**	**0.6**
CQ-relevance	**0.8**	**0.8**	**1**	**0.7**
CQ-validity	0.6	0.7	0.8	0.4
CQ-significance	**0.8**	**0.8**	**1**	**0.7**

Table 22. Gaze features in predicting users reading performances (SVM 10-fold, 50 iterations)

For strategic processes and performance predicting purpose, each channel had its own advantages and they could complement each other.

4.6.2 Sensing Technique

From the behavior signal observing perspective, the gaze channel provided a continuous observation, whereas, the scrolling channel and logging channel had scattered observations. During the 20 minutes reading, users' have 164 to 752 scrolls (=363.14, =111.49), accounting for 43.23% of their total reading time. The gaze channel, although had 100% coverage of reading time, was very sensitive to users' posture changes and the environment illumination changes. This limited the usability of such channel in a complex environment. However, scrolling channel and logging channel were stable in different environments.

4.6 Limitations and Future Work

Other than gaze tracking and screen sensing techniques, other signals are explored in existing research, such as brainwave [71]. We choose not to include such signals so that no additional sensors are required during the online reading task. Without extra sensors, we preserved the users' reading behaviors as they read everyday on smart phones.

We adopted the design of the online question-generation reading tasks in a previous research [17]. In the task, users were asked to generate a critical question on a controversial topic based on the information and knowledge gained from the online

sources. We also adopted the controversial topic "mountaintop removal coal mining" [17], as this topic had been proven to be significant when evaluating desktop-based online reading strategic processes for adolescent students. We used the retrospective verbal reports to interpret and measure the model of the strategic processes. However some limitations should be noted when interpreting the results. First, the complicated strategies require a general and systematic analysis, while the single topic chosen in this project might affect the explanatory power of our analysis. Second, the participants' source (freshmen and sophomore) and population (40 participants) might also affect the explanatory power of our analysis. Follow-up studies on topics of different perspectives with large scale of participants could assist this work. Third, the accuracy of the grading of verbal reports might introduce errors. The verbal report itself might introduce errors due to the readers' expression skills and memory bias. We attempted to minimize the grading error by letting experts train the grader and letting experts review the grading samples.

The StrategicReading system requires turning on the front camera and tracking users' logging histories and gazing and scrolling behaviors, which might also raise concerns from the users.

Despite these limitations, this study creatively uses technologies to understand higher level reading strategies. Previous research attempted to use behavior sensing to understand low level and local activities such as area of Interests, attention, and reading speed. The use of technologies to understand higher level reading strategies is very limited and the efforts in this dissertation is the first attempt to bridge this gap. We also plan to explore the following directions in the future. First, considering all previous

projects observing desktop online reading behaviors, a comparison among platforms (e.g. smartphones, desktops and pads). Second, this project's success of using behavior-sensing technique on an educational problem provides the possibility of the appliance of such technique to other psychology and educational problems. Third, we are interested in applying NLP techniques to understand users' reading contents together with analyzing users' behavior. Lastly, we plan to explore interactive technologies to give the users their personal training of how to read well.

4.7 Conclusions

We created a novel system called StrategicReading--a multimodal interface, which automates the detection and evaluation of reading strategies to predict online reading performance on unmodified smartphones. First, StrategicReading tracks the periodic lateral patterns of users' gaze movements via the front camera. Second, it infers their muscle stiffness via the Mass-Spring-Damper (MSD) based kinematic model and determines scrolling types during text scrolling. Last, it monitors the evolution of cross-page behavior by logging the searching and clicking history. Overall, StrategicReading combines the gaze channel, scrolling channel, and logging channel to improve the detection, evaluation, and understanding of reading strategies and reading performance.

5.0 CONLUCSION

This thesis systematically explores the usage of implicit behavioral signals and implicit physiological signals collected via a "sensorless" and low-cost approach to understand, model and improve mobile reading.

We explored the usage mobile eye gaze sensing and heart rate signal sensing and screen touch sensing to facilitate reading under external and internal distractions (SmartRSVP) and monitor readers' cognitions and emotions during reading (Lepton). We also creatively applied such behavior sensing techniques to psychology-based educational problems to automatically measuring users' higher-level cognitive process (StrategicReading).

5.1 Contribution

This thesis is the first attempt of building a perceptual and affect-aware interface to control text displaying on watch-size devices; This thesis creatively investigated the processing of the noisy camera-based gaze signals and emotion-dependent scrolling changes during reading on pocket-size devices and proved the probability to predict reading comprehension and engagements using such signals; This thesis is the first attempt to automate the detection of complex reading strategies via multi-channel sensing techniques.

5.2 Future Work

We developed SmartRSVP (Chapter 2), an attentive speed-reading system to facilitate text reading on small-screen wearable devices, to solved the challenges of

reading on watch-size devices that are caused by limited screen affordance. Despite promising results, we have only scratched the surface of the design space of SmartRSVP. There are *technology* and *design* of this research that could be explored in the future. Regards *technology*, the gaze tracking of SmartRSVP is sensative to lighting conditions (camera may be overexposure under direct sunshine). There are two potential ways to solve this problem: 1) designing more robust algorithms, and 2) leveraging built-in motion sensors such as the GPS, accelerometers and gyroscope in the watch to both infer the context of the users (i.e. indoor, outdoor, moving, not moving) and estimate the orientation and dynamic posture of the smart watch for more accurate predictions. Regards the *design*, providing feedback for text presentation and exploring the type and level of adaptation could be two future direction. In current SmartRSVP design, we choose to not include feedback other than binary, one-way speed reduction to avoid confounding results. We believe that providing feedbacks, such as tactile feedback, to the users when having speed reduction and having a more intelligent type and level of speed reduction could even benefit users seeing the contents on watch-size screens. Other than that, enabling regressions via gesture-based interactions [53][54], and reminding users (via tactile feedback, sound, or visualizations) about important upcoming messages, could also be explored in the future.

The challenges of reading on pocket-size devices are caused by complex reading contents and the lack of continuous and deeper understanding of users' cognitive states, therefore, we developed two Systems, Lepton (Chapter 3) and StrategicReading (Chapter 4) to track users' cognitive states during single-page and multi-page reading tasks. The future work of reading on pocket-size devices should be in breath and depth. We

investigated using users' eye movements and scrolling motions in Lepton and StrategicReading. In the future, exploring supplemental sensing channels, such as gyroscope and accelerometers for users' motions and GPS sensors for users' locations, in mobile reading could be the breath direction. We believe these sensors could benefit in understanding users' mobile context that will be important towards facilitation their reading experiences. In depth, the next step after understanding reading is to improve reading. We plan to explore interactive technologies, such as personalized recommendation, smart highlighting, or in-situ quizzes when low engagement is detected.

Overall, this thesis is the first attempt to bridge the gap between technologies and higher-level reading strategies. Its success provides the possibility of the appliance of such technique to other psychology and educational problems in the future.

Bibliography

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

[2] Abraham, Jeffrey C., et al. "Media content location awareness and decision making." U.S. Patent No. 8,725,680. 13 May 2014.

[3] Alexander, James RM, and Frances Martin. "Norming tests of basic reading skills." *Australian Journal of Psychology* 52.3 (2000): 139-148.

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

[5] Anmarkrud, Øistein, Ivar Bråten, and Helge I. Strømsø. "Multiple-documents literacy: Strategic processing, source awareness, and argumentation when reading multiple conflicting documents." *Learning and Individual Differences* 30 (2014): 64-76.

[6] Benedetto, S., Carbone, A., Pedrotti, M., Le Fevre, K., Bey, L.A.Y. and Baccino, T., 2015. Rapid serial visual presentation in reading: The case of Spritz. *Computers in Human Behavior*, *45*, pp.352-358.

[7] Biedert, Ralf, Andreas Dengel, Georg Buscher, and Arman Vartan. "Reading and estimating gaze on smart phones." In *Proceedings of the symposium on eye tracking research and applications*, pp. 385-388. ACM, 2012.

[8] Biedert, Ralf, Georg Buscher, and Andreas Dengel. "The eyebook–using eye tracking to enhance the reading experience." *Informatik-Spektrum* 33.3 (2010): 272-281.

[9] Bolter, Jay David. *Writing space: The computer, hypertext, and the history of writing*. Lawrence Erlbaum Associates, Inc, 1991.

[10] Brysbaert, M. and Nazir, T., 2005. Visual constraints in written word recognition: evidence from the optimal viewing-position effect. *Journal of Research in Reading*, *28*(3), pp.216-228.

[11] Bulling, Andreas, et al. "Robust recognition of reading activity in transit using wearable electrooculography." *International Conference on Pervasive Computing*. Springer, Berlin, Heidelberg, 2008.

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

[13] Castelhano, M.S. and Muter, P., 2001. Optimizing the reading of electronic text using rapid serial visual presentation. *Behaviour & Information Technology*, *20*(4), pp.237-247.

[14] Chen, X.A., Grossman, T. and Fitzmaurice, G., 2014, October. Swipeboard: a text entry technique for ultra-small interfaces that supports novice to expert transitions. In *Proceedings of the 27th annual ACM symposium on User interface software and technology* (pp. 615-620). ACM.

[15] Chen, X.A., Grossman, T., Wigdor, D.J. and Fitzmaurice, G., 2014, April. Duet: exploring joint interactions on a smart phone and a smart watch.

In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 159-168). ACM.

[16] Cheng, S., Sun, Z., Sun, L., Yee, K. and Dey, A.K., 2015, April. Gaze-based annotations for reading comprehension. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 1569-1572). ACM.

[17] Cho, Byeong-Young, et al. "Examining Adolescents' Strategic Processing During Online Reading With a Question-Generating Task." *American Educational Research Journal*54.4 (2017): 691-724

[18] Claypool, Mark, et al. "Implicit interest indicators." *Proceedings of the 6th international conference on Intelligent user interfaces*. ACM, 2001.

[19] Coiro, Julie, and Elizabeth Dobler. "Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for and locate information on the Internet." *Reading research quarterly* 42.2 (2007): 214-257.

[20] DeSchryver, Michael. "Web-Mediated Knowledge Synthesis for Educators." *Journal of Adolescent & Adult Literacy* 58.5 (2015): 388-396.

[21] D'Mello, S., Picard, R.W. and Graesser, A., 2007. Toward an affect-sensitive AutoTutor. *IEEE Intelligent Systems*, *22*(4).

[22] Dingler, T., Rzayev, R., Schwind, V. and Henze, N., 2016, September. RSVP on the go: implicit reading support on smart watches through eye tracking. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers* (pp. 116-119).

[23] Goldman, Susan R., et al. "Comprehending and learning from Internet sources: Processing patterns of better and poorer learners." *Reading Research Quarterly* 47.4 (2012): 356-381.

[24] Dollár, P., Welinder, P. and Perona, P., 2010, June. Cascaded pose regression. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on* (pp. 1078-1085). IEEE.

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

[26] Ericsson, Karl Anders, and Herbert Alexander Simon. *Protocol analysis: Verbal reports as data, Rev*. the MIT Press, 1993.

[27] Fialding, L. G., and P. David Pearson. "Synthesis of research reading comprehension: What works." *Educational Leadership*51 (1994): 62-62.

[28] Garner, Ruth. "Verbal-report data on cognitive and metacognitive strategies." *Learning and study strategies*. 1988. 63-76.

[29] Gikas, Joanne, and Michael M. Grant. "Mobile computing devices in higher education: Student perspectives on learning with cellphones, smartphones & social media." *The Internet and Higher Education* 19 (2013): 18-26.

[30] Godwin-Jones, Robert. "Mobile apps for language learning." (2011).

[31] Grusky, Max, et al. "Modeling Sub-Document Attention Using Viewport Time." *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2017.

[32] Guo, Qi, and Eugene Agichtein. "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, 2012.

[33] Guo, Wei, and Jingtao Wang. "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, 2017.

[34] Guo, Wei, and Jingtao Wang. "Towards Attentive Speed Reading on Small Screen Wearable Devices." Proceedings of the 2018 on International Conference on Multimodal Interaction. ACM, 2018.

[35] Guo, Wei, and Jingtao Wang. "Understanding Mobile Reading via Camera Based Gaze Tracking and Kinematic Touch Modeling." Proceedings of the 2018 on International Conference on Multimodel Interaction. ACM, 2018.

[36] Han, Teng, et al. "Balancing accuracy and fun: Designing camera based mobile games for implicit heart rate monitoring." *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2015.

[37] Hansen, J.P., Biermann, F., Madsen, J.A., Jonassen, M., Lund, H., Agustin, J.S. and Sztuk, S., 2015, September. A gaze interactive textual smartwatch interface. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers* (pp. 839-847). ACM.

[38] Harrison, C. and Hudson, S.E., 2009, October. Abracadabra: wireless, high-precision, and unpowered finger input for very small mobile devices. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology* (pp. 121-124). ACM.

[39] Hill, A. Vo. "The heat of shortening and the dynamic constants of muscle." *Proc. R. Soc. Lond. B* 126.843 (1938): 136-195.

[40] Hussain, M.S., AlZoubi, O., Calvo, R.A. and D'Mello, S.K., 2011, June. Affect detection from multichannel physiology during learning sessions with AutoTutor. In *International Conference on Artificial Intelligence in Education* (pp. 131-138). Springer Berlin Heidelberg.

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

[42] Jackson, M. D., McClelland, J. L., 1979. Processing determinants of reading speed. *Journal of Experimental Phychology: General* 108(2): 151

[43] Jraidi, I., Chaouachi, M. and Frasson, C., 2013, December. A dynamic multimodal approach for assessing learners' interaction experience. In *Proceedings of the 15th ACM on International conference on multimodal interaction* (pp. 271-278).

[44] Kanwisher, N.G., 1987. Repetition blindness: Type recognition without token individuation. *Cognition*, *27*(2), pp.117-143.

[45] Kim, J., He, J., Lyons, K. and Starner, T., 2007, October. The gesture watch: A wireless contact-free gesture based wrist interface. In *Wearable Computers, 2007 11th IEEE International Symposium on* (pp. 15-22). IEEE.

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

[47] Kliegl, Reinhold, 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.

[48] Kunze, Kai, et al. "The augmented narrative: toward estimating reader engagement." *Proceedings of the 6th augmented human international conference*. ACM, 2015.

[49] LaBerge, David, and S. Jay Samuels. "Toward a theory of automatic information processing in reading." *Cognitive psychology* 6.2 (1974): 293-323.

[50] Landow, George P. *Hyper/text/theory*. Johns Hopkins University Press, 1994.

[51] Lee, B., Savisaari, O. and Oulasvirta, A., 2016, May. Spotlights: Attention-Optimized Highlights for Skim Reading. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 5203-5214). ACM.

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

[53] Luzhnica, G., Veas, E., & Pammer, V. (2016, September). Skin reading: Encoding text in a 6-channel haptiv display. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers* (pp. 148-155). ACM.

[54] Luzhnica, G., & Veas, E. (2018, March). Investigating Interactions for Text Recognition using a Vibrotactile Wearable Display. In *23rd International Conference on Intelligent User Interfaces* (pp. 453-465)

[55] Mariakakis, A., Goel, M., Aumi, M.T.I., Patel, S.N. and Wobbrock, J.O., 2015, April. 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.

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

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

[58] Navalpakkam, Vidhya, 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.

[59] Nielsen, Jakob. "Mobile content is twice as difficult." *Alerbox, available: http://www. useit. com/alertbox/mobile-contentcomprehension. html* (2011).

[60] Oh, JongHwan, SungJin Nam, and Joonhwan Lee. "Generating highlights automatically from text-reading behaviors on mobile devices." *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2014.

[61] Oney, S., Harrison, C., Ogan, A. and Wiese, J., 2013, April. ZoomBoard: a diminutive qwerty soft keyboard using iterative zooming for ultra-small devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2799-2802). ACM.

[62] Öquist, G. and Goldstein, M., 2001. Adaptive rapid serial visual presentation. *Unpublished Thesis, Uppsala University, Uppsala, Sweden*.

[63] Öquist, G. and Lundin, K., 2007, December. Eye movement study of reading text on a mobile phone using paging, scrolling, leading, and RSVP. In *Proceedings of the 6th international conference on Mobile and ubiquitous multimedia* (pp. 176-183).

[64] Owen, John. "National Institute of Child Health and Human Development, Maternal-Fetal Medicine Units Network: Midtrimester endovaginal sonography in women at high risk for spontaneous preterm birth." *JAMA* 19.286 (2001): 1340-1348.

[65] Perez, Sarah. 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

[66] Perfetti, Charles A., Nicole Landi, and Jane Oakhill. "The acquisition of reading comprehension skill." *The science of reading: A handbook* (2005): 227-247.

[67] Pham, Phuong, and Jingtao Wang. "Understanding emotional responses to mobile video advertisements via physiological signal sensing and facial expression analysis." *Proceedings of the 22nd International Conference on Intelligent User Interfaces*. ACM, 2017.

[68] Ploutz-Snyder, L. L., et al. "Cardiorespiratory and metabolic adaptations to hyperoxic training." *European journal of applied physiology and occupational physiology* 73.1-2 (1996): 38-48.

[69] Pressley, Michael, and Christine McCormick. *Advanced educational psychology for educators, researchers, and policymakers*. Harpercollins College Division, 1995.

[70] Raymond, J.E., Shapiro, K.L. and Arnell, K.M., 1992. Temporary suppression of visual processing in an RSVP task: An attentional blink?. *Journal of experimental psychology: Human perception and performance*, *18*(3), p.849.

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

[72] Reinking, David. "Me and my hypertext: A multiple digression analysis of technology and literacy (sic)." *The Reading Teacher* 50.8 (1997): 626-643.

[73] Rodrigue, M., Son, J., Giesbrecht, B., Turk, M. and Höllerer, T., 2015, March. Spatio-temporal detection of divided attention in reading applications using EEG and eye tracking. In *Proceedings of the 20th International Conference on Intelligent User Interfaces* (pp. 121-125). ACM.

[74] Rowe, D.W., Sibert, J. and Irwin, D., 1998, January. Heart rate variability: Indicator of user state as an aid to human-computer interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 480-487). ACM Press/Addison-Wesley Publishing Co.

[75] Rubin, G.S. and Turano, K., 1992. Reading without saccadic eye movements. *Vision research*, *32*(5), pp.895-902.

[76] Shimray, Somipam R., Chennupati Keerti, and Chennupati K. Ramaiah. "An overview of mobile reading habits." *DESIDOC Journal of Library & Information Technology* 35.5 (2015).

[77] Sicheritz, K2000. Applying the rapid serial visual presentation technique to small screens. *master's thesis, Dept. of Linguistics, Uppsala Univ*.

[78] Smith, B.A., Yin, Q., Feiner, S.K. and Nayar, S.K., 2013, October. Gaze locking: passive eye contact detection for human-object interaction. In *Proceedings of the 26th annual ACM symposium on User interface software and technology* (pp. 271-280). ACM.

[79] Spires, HILLER A., and THOMAS H. Estes. "Reading in web-based learning environments." *Comprehension instruction: Research-based best practices* (2002): 115-125.

[80] Stockwell, Glenn, and Philip Hubbard. "Some emerging principles for mobile-assisted language learning." *The International Research Foundation for English Language Education* (2013): 1-15.

[81] Sun, David, Pablo Paredes, and John Canny. "MouStress: detecting stress from mouse motion." *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2014.

[82] Thornton, Patricia, and Chris Houser. "Using mobile phones in education." *Wireless and Mobile Technologies in Education, 2004. Proceedings. The 2nd IEEE International Workshop on*. IEEE, 2004.

[83] Trelease, Jim. *The read-aloud handbook*. Penguin, 2013.

[84] Viola, P. and Jones, M., 2001. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on* (Vol. 1, pp. I-I). IEEE.

[85] Wobbrock, J.O., Forlizzi, J., Hudson, S.E. and Myers, B.A., 2002, October. WebThumb: interaction techniques for small-screen browsers. In *Proceedings of the 15th annual ACM symposium on User interface software and technology* (pp. 205-208). ACM.

[86] Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D. and Picard, R., 2009. Affect-aware tutors: recognising and responding to student affect. *International Journal of Learning Technology*, *4*(3-4), pp.129-164.

[87] Yi, Xing, et al. "Beyond clicks: dwell time for personalization." *Proceedings of the 8th ACM Conference on Recommender systems*. ACM, 2014.

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

[89] Yu, Fuxin, and A. R. Conway. "Mobile/smart phone use in higher education." *Proceedings of the 2012 Southwest Decision Sciences Institute* (2012): 831-839.

[90] Yuksel, B.F., Oleson, K.B., Harrison, L., Peck, E.M., Afergan, D., Chang, R. and Jacob, R.J., 2016, May. Learn piano with BACh: An adaptive learning

interface that adjusts task difficulty based on brain state. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 5372-5384). ACM.

[91] Zhang, Liyi, and Wei Ma. "Correlation analysis between users' educational level and mobile reading behavior." *Library Hi Tech* 29.3 (2011): 424-435.