Towards Attentive Speed Reading on Small Screen Wearable Devices

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

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

ABSTRACT

Smart watches can enrich everyday interactions by providing both glanceable information and instant access to frequent tasks. However, reading text messages on a 1.5-inch small screen is inherently challenging, especially when a user's attention is divided. We present SmartRSVP, an attentive speed-reading system to facilitate text reading on small-screen wearable devices. 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. Through a series of three studies involving 40 participants, we found that 1) SmartRSVP can achieve a significantly higher comprehension rate (57.5% vs. 23.9%) and perceived comfort (3.8 vs. 2.1) than traditional RSVP; 2) Users prefer SmartRSVP over traditional reading interfaces when they walk and read; 3) SmartRSVP can predict users' cognitive workloads and adjust the reading speed accordingly in real-time with 83.3% precision.

CCS CONCEPTS

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

KEYWORDS

Smart watch; Text Reading; RSVP; Gaze tracking; PPG; Heart rate variability; Cognitive workload; Visual attention

ACM Reference format:

Wei Guo and Jingtao Wang. 2018. Towards Attentive Speed Reading on Small Screen Wearable Devices. In *Proceedings of 20th ACM International Conf. on Multimodal Interaction (ICMI'18), October 16-20, 2018, Boulder, CO, USA.* ACM, NY, NY, USA, *10 pages.* https://doi.org/10.1145/1234567890

1 INTRODUCTION

Small-screen wearable devices are flourishing nowadays. By staying on users' wrists 24/7, smart watches can assist users' access to 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 [6] and input modalities [25][30] have been proposed for smart watches during recent years, it remains a major challenge to read textual information on smart watches.

There are three major challenges when a user reads text messages on a smart watch. First, most of today's smart watches use a small display size approximately 1.5-inch diagonal, which only affords to show 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 increase our exposure to such reading interfaces.

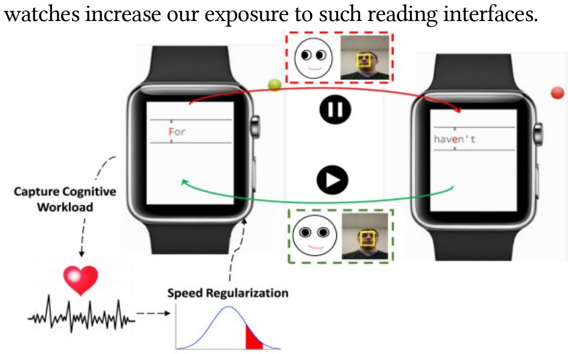

Figure 1. SmartRSVP continuously monitors the visual attention of a user via real-time image processing and infers the user's cognitive workload via implicit physiological signal sensing. SmartRSVP uses the visual attention to play/pause dynamic text presentation and adjusts text display speed via the inferred cognitive workload.[*]

To address these challenges, we present SmartRSVP (Figure 1), 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 text messages and then leverages the 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,

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

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

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

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 users' workloads in both visual and cognitive attentions.

This paper offers three major contributions.

- 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.
- We propose novel algorithms and interaction designs to make text reading via RSVP more enjoyable and practical on small-screen wearable devices.
- We show the feasibility, accuracy, robustness, and usability of SmartRSVP via three user studies involving 40 participants.

2 RELATED WORK

RSVP

Rapid Serial Visual Presentation (RSVP) is a visualization technique that displays textual information one word at a time[1] in a sequential order. Although RSVP was originally invented to examine the temporal characteristics of human attention, Gilbert [21] first used RSVP for text reading in 1959. Forster [19] investigated the comprehension and processing of written language in 1970. With the rise of personal computing and digital media, researchers began to treat RSVP as a speed-reading technique for computing devices during the past forty years [5].

RSVP has both spatial and temporal efficacy when compared with traditional reading. However, RSVP may not always be efficient due to the lack of an easy mechanism for regressions. RSVP achieves spatial efficacy by displaying unlimited text within a limited space. RSVP also achieves temporal efficacy when compared to traditional reading. In traditional reading, a reader processes word information through eye gaze in three major types of movements [52]: 1) fixing her eye gaze on a word; 2) making a saccade to the next fixation; 3) moving to the next line via a return sweep. In RSVP, the user can keep the same eye gaze fixation during reading. As a result, the duration of saccades and return sweeps can be highly reduced or even eliminated, and content is actively presented for readers' attention [32]. On the other hand, the missing of regressions causes the weakness of RSVP. Regressions are letter-level, word-level, line-level, and even section-level backward saccades, which is 10-15% of all saccades [46]. Such regressions are vital for text comprehension [31][46][47][48], but are missing in RSVP reading. Despite lacking regressions, RSVP also faces challenges such as attentional blink [45], repetition blindness [29], demanding higher visual attention [5], higher recovery cost [45], and higher cognitive workload [40][41][53].

The obvious strength and weakness of RSVP lead to the mix results when compared with traditional reading. Some studies show that traditional reading surpasses the performances of RSVP, especially on large computer displays [2][5][12]. There are also some studies that demonstrate the advantage of RSVP for speed-reading over traditional reading. For example, Rubin [52] found that RSVP reading rates were consistently higher than traditional reading rates, with adequate comprehension levels even on large screen devices. Wobbrock et al [57] reported that as many as 720 wpm (12 words per second) are readable via RSVP, whereas the traditional reading speed is around 250 wpm. Most implementations of RSVP on pocket-sized devices achieved comparable reading speed [40][53] and comprehension level [40][41][53] as traditional reading.

Researchers proposed adaptations of RSVP to improve the RSVP reading experience. Chien [10] adapted RSVP by decreasing the speed from 350wpm to 150wpm, and achieved 17.3% increase in comprehension. Masson [36] adapted RSVP by adding a 500ms pause between sentences of reading materials, which enhanced user comprehension by 10%-20%. More complex adaptations such as adapting each word's exposure time based on its length [2][40][41][57], font, the type of the followed punctuation [57], and its frequency and position in the sentence [40] were also explored.

Different from content-based speed adaptation policies adopted by existing research, we explore how unique sensors, such as the front camera[2] and the Photoplethysmography (PPG) sensor on a smart watch can be used to adapt RSVP depending on users visual and cognitive workload in real-time. Therefore, SmartRSVP can exploit both spatial and temporal efficacy of RSVP on small-screen wearable devices without sacrificing practicability. Guo and Wang [20] verified the feasibility of detecting users' cognitive workload in RSVP via pilot studies and offline benchmarking. We further designed, implemented, and evaluated an interactive system via principled research systemically.

Eye-gaze Aware Interfaces

Eye-gaze tracking has been widely used to understand user behavior and attention [42] in human-computer interaction tasks. Eye-gaze has also been used as an expressive input modality to facilitate pointing, typing, and multi-modal switching, even on mobile devices [3][16] and small screen devices [17][24].

In the context of reading, SocialReading [8] visualized teachers' gaze movements on academic readings in order to improve student comprehension. SwitchBack [37] tracked the periodic return sweep of eye gaze via a front camera to estimate and highlight the current sentence in a reading task where the user was interrupted by divided attention.

Our visual attention tracking feature was inspired by the "gaze locking" technique by Smith et al. [54] and the seeTXT technique by Dickie et al. [11]. Gaze locking refers to the robust binary sensing of eye contact in a static image via computer vision algorithms. seeTXT relies on a customized infrared eye-contact sensor (ECS) to augment media consumption on mobile devices. In comparison, our SmartRSVP technique focuses on improving the speed-reading experiences on smart watches. In addition to real-time visual attention tracking, SmartRSVP can also infer cognitive workload from implicit PPG sensing and regulate the speed of reading accordingly. Further, we also directly compare the users'

[1] Or one visual item a time for stimuli such as pictures.

[2] Front cameras are already available on smart watches today, e.g. Z80 3G Android watch, Zeblaze THOR S, and GT08 from Lichip Electronic Co., etc. More device manufacturers may include cameras once compelling usage scenarios for cameras are discovered.

performance and preferences with SmartRSVP against alternative techniques in three studies.

Hansen et al [23] demonstrated the feasibility of using a commercial gaze tracker to control RSVP playback running on a PC. Dinger and colleagues [13] demonstrated gaze-controlled RSVP with a head-mounted gaze tracker and visual markers. In comparison, SmartRSVP does not rely on any external sensors and achieves portability, which is the vital characteristic of smart watch. Besides being portable and enabling gaze-based control, SmartRSVP also offers implicit cognitive state sensing for the dynamic speed regulation of RSVP. This paper goes beyond a feasibility test by running controlled experiments to quantify and compare the performances of alternative techniques.

Affect/Emotion Aware Interfaces

Building computers that can understand and respond to user affect, emotion, and cognitive states [44] has been a compelling vision driving the research on intelligent user interfaces and ubiquitous computing. A variety of physiological signals, such as heart rates [22][51][28], galvanic skin responses (GSR) [26][58], facial expressions, Electroencephalography (EEG) [50], and eye-gaze [15][50] have been explored to infer users' cognitive and affective states in different interaction tasks, such as learning [15][28], operating user interfaces [51], and gaming [22]. In this paper, we propose the sensing and modeling of PPG signals to facilitate speed-reading on smart watches. We believe that with the ability to stay on a user's wrist 24/7 and collect the user's physiological signals implicitly, smart watches will become a promising testbed for the next generation affect/emotion aware interfaces.

Smart Watch Interfaces

The portability, size, and sensing power of smart watches have brought both opportunities and challenges to interaction design in recent years [6][7][9][25][30][39]. Existing research has focused on 1) designing new interaction techniques [17][24], especially cross-device interactions [6] and authoring environments [9]; 2) enabling more expressive and space-efficient input modalities [25][30]; 3) providing efficient text input for ultra-small touch screens [7][39].

While most existing research has focused on the input and interaction techniques with smart watches, SmartRSVP is an attempt to create a streamlined and enjoyable text output paradigm on smart watches. SmartRSVP addresses the inherent limitations of both smart watches and classic RSVP through sensing, inferring, and adapting to user attentional and cognitive states in reading.

3 The DESIGN OF SMARTRSVP

Figure 1 shows SmartRSVP in action. SmartRSVP displays text via RSVP, and continuously monitors the visual attention and cognitive states of a user. SmartRSVP pauses the text display if there is no human face in the camera viewport, or 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 presentation speed of RSVP will be adjusted based on the cognitive workload.

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

3.1 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) [4] and visualizes the ORP in red color (Figure 1). 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 [4]. We use a monospace font to ensure all ORPs[3] 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 tap the watch screen to play or pause the text displayed on SmartRSVP.

3.2 Visual Attention Tracking

Due to the limited availability of front facing cameras on smart watches at the time of our experiment, we used a Google Nexus 5X[4] smartphone running Android 6.0 to simulate a 42.0mm by 35.9mm smart watch screen. This choice follows practices of existing research on smartwatches [7][39], we allocated 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 users' gaze point. 1) Face detection: A Viola-Jones face detector [56] is used to detect the existence and location of a human face; 2) Face Alignment: We use Cascaded Pose Regression [14] to estimate the facial orientation and landmark points on a face; 3) Eye contact estimation: similar to [37][54], 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 process 21 frames per second on the hexacore Snapdragon 808 CPU in the Nexus 5X.

Compared to dedicated eye trackers, camera-based gaze tracking lacks the accuracy to detect the absolute locations of eye gazes [38], however, such accuracy is not necessary given the small display size of smart watches. Therefore, we proposed a calibration-free binary eye contact algorithm optimized for the visual attention tracking requirement in the context of SmartRSVP. We trained a binary eye contact classifier from five volunteers: taking the union of all volunteers' visual attention range of the simulated watch screen. Therefore, the gazes within the union range will be treated as paying visual attention. A 0.5s low-pass filter was used to reduce false positives and false negatives from per-frame estimations. Figure 2 shows the continual output of the eye gaze prediction algorithm and the binary eye contact estimations.

[3] We use the following heuristics to determine ORP: 1) Choose the first character as ORP for words with no more than three characters; 2) Choose the second character for words with length 4 or 5; 3) Choose the third character for words with length 6; 4) Choose the fourth character for the rest.

[4] Google Nexus 5X embeds a 5.0MP front camera, which has the same image quality as the front camera of Zeblaze THOR S 3G smart watch released in late 2017.

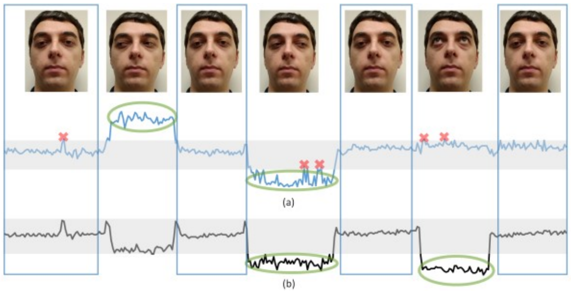

Figure 2. Visual Attention Tracking via face detection and eye contact estimation. The x-axis is time (~17 sec). Row (a) is the predicted horizontal eye gaze locations at each timestamp, and Row (b) is the predicted vertical eye gaze locations. Green circles highlight moments when a user is not having gaze contact with his smart watch.

3.3 Cognitive State Inference

SmartRSVP infers users' cognitive states via commodity fingertip PPG sensing through the back camera of a smartphone. We choose this option rather than relying on built-in heart rate sensors on smartwatches because today's smart watch APIs do not give access to raw heart beat waveforms. We expect cleaner PPG signals and higher prediction accuracies if we can access raw waveforms from dedicated optical heart rate sensors on smart watches.

We used the LivePulse algorithm [22] to preprocess the temporal PPG signals and then used a fixed-size sliding window to extract features from them. We extracted 9 dimensions of heart rate and HRV features (Table 1) from each window. After normalization, these features were used to train a statistical classifier to predict user cognitive workload within each sliding window.

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.

3.4 Speed regulation

A one-way, binary adaptation strategy was used to adjust the RSVP speed dynamically. This adaptation strategy avoids the confounders in number, duration, and scale of speed changes in adaptations. This paradigm has been proven to be effective by existing research [59] and suits our design purpose. SmartRSVP tracks users' PPG signals during reading and decreases RSVP speed if a multitasking activity is detected. The amount of speed reduction was chosen by a 4-user pilot study, and finalized to be 100wpm since it was the minimum reduction that could be noticed by all participants.

4 USER STUDIES

We ran three user studies to investigate different aspects of SmartRSVP. The first two studies aimed to evaluate the visual control module of SmartRSVP and the last one aimed to evaluate its speed regulation module. 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 evaluated usability and efficacy of the whole SmartRSVP system in action.

4.1 User Study 1

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

4.1.1 Participants

18 participants (3 females) between 19 and 46 years of ages (μ=26) participated in the study. None of the participants had experiences with RSVP.

4.1.2 Apparatus

There were three interfaces in this study: normal watch reading interface (NWR), traditional RSVP (T-RSVP), and SmartRSVP (Figure 3). NWR used a 20dp sans serif (Droid Sans) font for text display to replicate today's reading interfaces on smart watches. NWR shows around four words per line and eight lines per screen. T-RSVP had the same appearance as SmartRSVP (details in Design of SmartRSVP session). Similar as in SmartRSVP interface, a user could also tap the screen to play/pause the text display in T-RSVP.

Figure 3. Three reading interfaces in the study. From left to right: Normal Watch Reading (NWR) Interface, Traditional RSVP (T-RSVP) interface, and SmartRSVP.

Thirty unique email pieces were chosen from Enron email database as the reading materials. Reading short email messages is a frequent task on today's smart watches. The selected emails have comparable lengths (μ = 47 words or 3.5 sentences) and difficulties (average Flesch-Kincaid score = 68.65, σ = 14.97).

We also designed and deployed visual distractions in this study. A 15-inch laptop was put on the side of a participant (Figure 4) to generate distracters. When the participant was reading, for every 4 to 6 seconds, the laptop generated a distracter—a 3-digit random number on the screen along with a beep sound. Once the participant heard a beep sound, she was required to look at the laptop screen

(Figure 4, right) and read the number out loud. Then the participant could resume the reading task.

Figure 4. 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.

4.1.3 Procedure
This user study included a single session for 30 minutes. The session started with an introduction of the three interfaces and distractions. Once completed, participants practiced reading on the interfaces with distractions for 10 min to get familiar with the interfaces, distractions, the genre of reading materials and the comprehensive questions. After the practice, participants read a set of 10 emails on each interface, 3 sets (30 emails) in total for all interfaces. Both sets and interfaces were randomly ordered for each participant. We placed 3 distracters for each email message. After reading an email, the participant answered one question to test the reading comprehension. At the end, participants were asked to complete a questionnaire to provide the subjective feedback on the three interfaces.

4.1.4 Design & Analysis
The study used a within-subjects design with three interfaces: NWR, T-RSVP, and SmartRSVP (Figure 3).

We investigated the following metrics across the three interfaces:

- False positive and false negative rates of the divided visual-attention, which were calculated by comparing the recorded play/pause timestamps during reading with the distracters.
- Comprehension rate—the percentage of correctly answered questions. There are three levels of text comprehension, i.e. literal, inferential, and evaluative [18]. 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.
- Reading efficiency (E), defined as: $\mathrm{E} = \frac{\mathrm{W}}{\mathrm{D}} \times \mathrm{c}$, where D denotes reading duration (including distractions), W is the number of words, and c is comprehension rate [27][53].
- Subjective ratings of comfort on a 5-point Likert scale.

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

The average comprehension rates and corresponding standard deviations were 52.2% (σ=0.16), 23.9% (σ=0.11), and 57.5% (σ=0.20) respectively (Figure 5). 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, the difference of the comprehension rates between NWR and SmartRSVP was not significant ($t(17)=0.88$, $p=0.39$).

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 efficiencies between NWR vs. T-RSVP ($t(17)=-3.02$, $p<0.005$), and between SmartRSVP vs. T-RSVP ($t(17)=-3.37$, $p<0.005$).

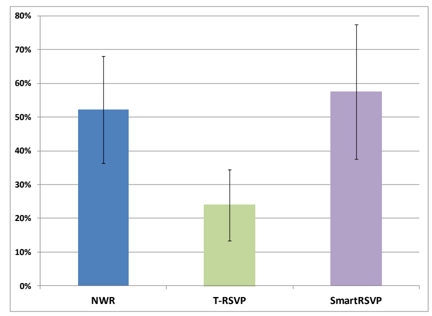

Figure 5. Comprehension rates by reading interfaces

Figure 6 showed 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 are the portions of each rating score within the bar. The subjective ratings of the 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 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 ($t(17)=1.75$, $p=0.08$). All 18 participants provided positive feedback on the use of eye-gaze as an implicit control channel for RSVP. More than 80% participants thought the SmartRSVP's visual attention control channel was "responsive".

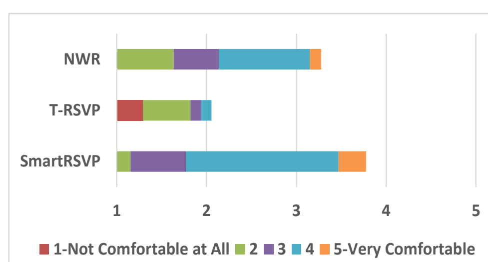

Figure 6. Subjective ratings on perceived comfort on a 5-point Likert scale (1 = not comfortable at all, 5 = very comfortable).

4.1.6 Discussions
By leveraging camera-based visual control channel, SmartRSVP overcame the recovery cost of divided attention on T-RSVP and

achieved significantly higher comprehension, reading efficiency, and perceived comfort than T-RSVP.
Meanwhile, SmartRSVP had a comparable performance as NWR. We believe current results are still promising for three reasons: 1) our participants had more than 10 years of experience in normal text reading interfaces. In comparison, the results in SmartRSVP were immediate pick-up performances with less than an hour of exposure; 2) The sitting posture in this study could benefit NWR by reducing the tremor of the watch screen (easier fixations) and sparing the hands and arms for scrolling purpose (both hands available); 3) SmartRSVP is a hands-free interface, which can facilitate reading when no hands are available to scroll. In this study, SmartRSVP achieved comparable performance as NWR even when users' both hands were available. The results suggest that SmartRSVP can be served as a complementary interface to NWR in sitting posture without sacrificing a user's reading performances, especially when both of the user's hands are occupied.

4.2 User Study 2

This study further evaluated the usability, efficacy, and robustness of SmartRSVP's visual control channel during standing and walking conditions when reading longer articles.

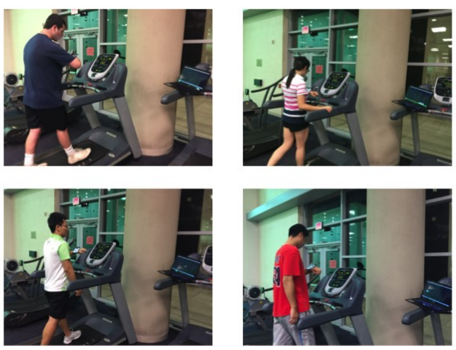

Figure 7. Sample participants in study 2.

4.2.1 Participants & Apparatus & Procedure

12 participants (5 females) between 18 and 34 years of ages (μ=23) participated in the study (Figure 7). Three of them had participated in user study 1. None of the rest had experience with RSVP.
The setting of this study was the same as user study 1, except:
1) We excluded the T-RSVP interface to simplify the experimental design. When compared with the sitting posture in user study 1, the standing and walking conditions do not bring additional benefits to T-RSVP over SmartRSVP.
2) The participants completed all the tasks on a treadmill in a local gym (Figure 7). The speed of the treadmill was set to 1.5mph in the walking posture.
3) The reading materials were changed from short email messages to longer articles to test the range of SmartRSVP's application. We selected four news articles from the New York Times, ranging 300-400 words (μ=369, σ =24) with comparable difficulties (average Flesch-Kincaid score = 49, σ = 8.46).
4) We updated the binary eye-contact classifier by including 5 more volunteers to increase ethnic diversity.
This study had the same procedure as user study 1.

4.2.2 Design & Analysis

The study used within-subjects design with two-by-two factors, including Standing and Walking postures, as well as SmartRSVP and NWR interfaces. Participants completed articles under each unique combination of postures and interfaces, leading to 2×2=4 articles for the study. Posture and interfaces were randomly ordered for each participant.
The evaluation metrics were the same as user study 1, except we used two literal questions after each article to test users' comprehensions.

4.2.3 Results

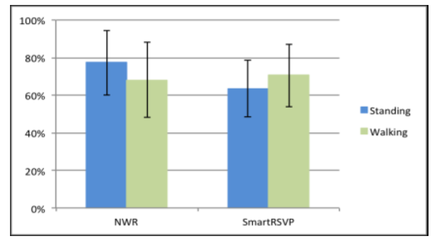

Figure 8. Average comprehension rates by postures.

By applying the updated binary eye-contact classifier in SmartRSVP, the average false positive rates and average false negative rates of divided visual tracking dropped to 5.05%, 4.5% in Standing posture, and 6.06%, 9.09% in Walking posture.
The average comprehension rates of the NWR and SmartRSVP by reading postures were shown in Figure 8. The comprehension accuracies were 77.3% (σ=0.69) on NWR in Standing posture (N+S), 68.2% (σ=0.81) on NWR in Walking posture (N+W), 63.6% (σ=0.61) on SmartRSVP in Standing posture (S+S) and 70.8% (σ=0.66) on SmartRSVP in Walking posture (S+W). 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 (N+S, σ=93.98), 163.5 wpm (N+W, σ=110.17), 129.1 wpm (S+S, σ=59.61), and 170.2 wpm (S+W, 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.

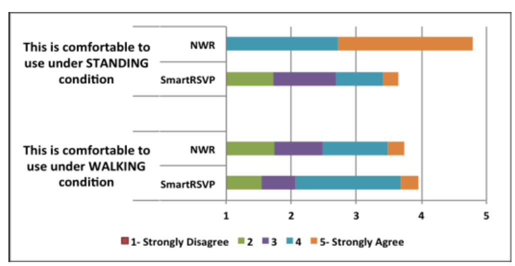

Figure 9. Subjective ratings on perceived comfort for NWR and SmartRSVP (1 = Strongly Disagree, 5 = Strongly Agree).

The subjective ratings of perceived comfort were 4.55 (σ = 0.52) for N+S, 3.18 (σ = 0.98) for S+S, 3.27 (σ = 1.01) for N+W, 3.55 (σ = 0.93) for S+W (Figure 9). Participants preferred NWR in standing posture, while preferred SmartRSVP in walking posture.
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 the camera-based gaze tracking technique in SmartRSVP (Figure 10). 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).

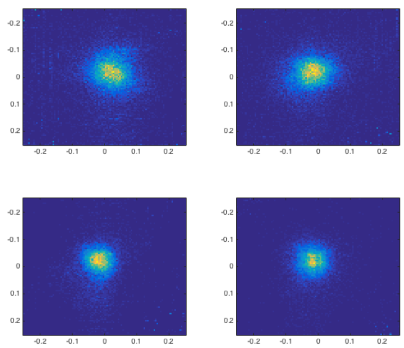

Figure 10. Heatmaps of aggregated eye gaze by postures. Top row: NWR+Standing and NWR+Walking; Bottom Row: SmartRSVP+Standing and SmartRSVP+Walking.

4.2.4 Discussions

These results were still encouraging considering that: 1) participants were able to learn to use SmartRSVP "in the wild" with little training and no calibration; 2) Our visual attention tracking algorithm was able to tolerate the constant but irregular motions during reading in standing and walking postures. These findings also suggest that SmartRSVP could serve as an effective complement to NWR when a user read during walking and both hands are occupied.

4.3 User Study 3

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

4.3.1 Participants & Apparatus

14 participants (6 females) between 25 and 33 years of ages (μ=29) participated in the study. Only one participant had previous experiences in RSVP.

We adopted the color counting task [50] to induce the internal cognitive workload changes. Two workloads were included: the focused and multitasking. To induce different cognitive workloads, a computer was placed on the side that spoke the names of nine different colors randomly at the speed of one second per color. Participants were told to read and ignore the background audio in the focus condition. In the multi-tasking condition, they read as well as counted the number of times the two target colors were spoken, i.e. "yellow" and "white". Conditions and articles were assigned to users in random orders.

The reading articles in this study had an average of 794 (σ = 30.13) words and comparable difficulties (average Flesch-Kincaid Reading Ease score = 38.64, σ = 4.57).

4.3.2 Procedure

This study included a 20min training session and a 40min testing session.

In the training session, the participants read a news article under each cognitive workload condition. Participants rated their focus level, and answered comprehensive questions after each reading.

The testing session was conducted one week after the training sesion, where the real-time speed regulation function was enabled. In the testing session, two cognitive-workload classifiers were used to adjust the RSVP speed. One was the SmartRSVP embedded classifier and the other was a baseline method (details in 4.3.3). A participant read an article under each unique combination of classifiers and cognitive-workload conditions.

4.3.3 Design & Analysis

We used a within-subjects design in this study. The training session had 2 cognitive-workload conditions. The testing session used two-by-two (2 cognitive workload conditions and 2 speed adaptive classifiers) factorial design. Therefore, each participant completed 2+2×2=6 articles in the study.

Two user-dependent cognitive-workload classifiers were evaluated in this study. First, we adopted a simple threshold-based classifier (TH classifier) as the baseline. To simulate traditional practices [51] that predict cognitive workload from heart rate variability signals in the HCI community, the TH classifier calculates the one-dimensional MHR in HRV features to determine a user's cognitive workload. A multitasking state was triggered if the observed MHR had less than 5% probability in training-focus distribution and more than 5% probability in training-multi-tasking distribution. Second, a RBFSVM classifier was used in SmartRSVP due to its promising results in previous research [20]. For each real-time input window (50s HRV signals), the RBFSVM classifier predicted a probability of being multitasking ranging from 0~100%. To increase the precision, we classified one as multitasking only when the probability is >90%,

To assess the performance of SmartRSVP's real-time speed regulation, we compared 1) the accuracy, precision, and recall of speed regulation under different cognitive conditions (reliability), and 2) users' subjective feedbacks on SmartRSVP (effectiveness).

4.3.4 Results

	Baseline (TH)	**RBFSVM**
Accuracy	50.0%	70.0%
Precision	50.0%	83.3%
Recall	40.0%	50.0%

Table 2. The live performance of SmartRSVP in study 4.

In the training phase, the same three-step procedures were used to process users' raw PPG signals and get training instances as in Guo and Wang [20]. We used 5s stripe (the gap between the starting points of two consecutive windows), 20s initial padding, and 50s local window size for both training and testing, which achieved optimal performance in a 2-fold user-dependent RBFSVM classifier model (accuracy = 68%, kappa = 0.35) in the training dataset.

As shown in Table 2, SmartRSVP's RBFSVM classifier achieved better real-time prediction accuracies than the baseline classifier. The relative improvement in accuracy was around 40%.

Overall, our users reported positive experiences with the speed adjustment module in SmartRSVP (Figure 13). Users considered SmartRSVP's speed adaptations were reliable (μ=3.8, σ=1.03) and would like to use SmartRSVP in the future (μ=4.1, σ=0.99).

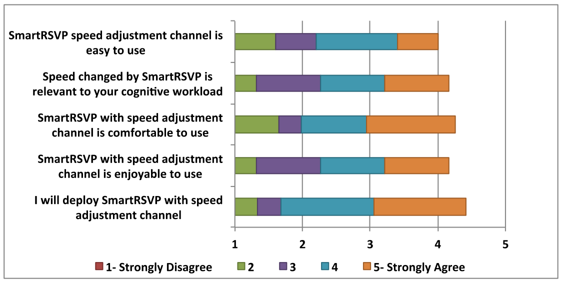

Figure 13. Subjective ratings on a 5-Point Likert scale.

4.3.5 Discussions

In our experiment, the calibration process of the speed adaptation module includes 1) choosing the favorite RSVP speed; and 2) gathering heart rate signals when the participants were reading two articles, one reading in focus and the other one in multitasking. The two-step calibration is for first-time users only.

SmartRSVP could detect users' cognitive status with 70% accuracy in real-time. By leveraging the real-time adaptation algorithm of SmartRSVP, users reported positive experiences.

As expected, RBFSVM classifier worked better than the TH classifier (baseline) because multiple dimensions of HRV features were taken into account. When further investigating users' MHR, we found that even for a single reader, his $\mathrm{MHR}_{\mathrm{multi-tasking}}$ did not always align to the same side of $\mathrm{MHR}_{\mathrm{focus}}$. Therefore, users' cognitive workload during RSVP reading was not directly correlated to users' MHR. Instead, users' perceived difficulties and interestingness of the reading materials had stronger correlations with MHR: harder reading materials were related to higher MHR (Pearson correlation=0.38) and interesting reading material lead to lower MHR (Pearson correlation = -0.35).

We intentionally made two 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 were sufficient to regulate the speed of RSVP dynamically with good accuracies and robustness; Second, we used a one-way and fixed-speed adaptation strategy [59] because it ensured our algorithm would do no harm to the reading process. Multi-way detections will reduce the detection accuracy and incorrectly increasing the reading speed can be disruptive to reading experiences.

5 FUTURE WORK

While SmartRSVP was optimized for smart watches and smart wristbands, it could also be helpful for emerging interaction technologies, such as smart glasses, augmented reality (AR) displays, and heads-up displays (HUDs), where there are limited screen estates, restricted input modalities, or insufficient cognitive bandwidth to display, navigate, or process textual information.

Despite promising results, we have only scratched the surface of the design space of SmartRSVP. There are several important topics 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. 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). Inspired by Yuksel et al [59], we plan to adjust the displaying speed of SmartRSVP in a binary manner from low speed to high speed; Fourth, it will be interesting to invent 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. We plan to explore the use of wrist gestures [30] as a mixed-initiative control channel in SmartRSVP; Last but not least, several alternative presentation techniques, such as showing important words in different colors, content-based RSVP speed adaptations such as adjusting the word-level display durations based on the predicted importance, enabling regressions via gesture-based interactions [34][35], and reminding users (via tactile feedback, sound, or visualizations) about important upcoming messages, could be explored in the context of SmartRSVP.

6 CONCLUSIONS

We proposed SmartRSVP, a novel speed-reading system to facilitate text reading on small screen wearable devices. 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 three 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. Finally, SmartRSVP can adjust the speed of RSVP in real-time based on users' cognitive workload with 83.3% precision.

REFERENCES

[1] Afergan, D., Peck, E.M., Solovey, E.T., Jenkins, A., Hincks, S.W., Brown, E.T., Chang, R. and Jacob, R.J., 2014, April. Dynamic difficulty using brain metrics of workload. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (pp. 3797-3806). ACM.

[2] 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.

[3] Biedert, R., Dengel, A., Buscher, G. and Vartan, A., 2012, March. Reading and estimating gaze on smart phones. In *Proceedings of the symposium on eye tracking research and applications* (pp. 385-388).

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] Chi, P.Y.P. and Li, Y., 2015, April. Weave: Scripting cross-device wearable interaction. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 3923-3932). ACM.

[10] Chien, Y.H. and Chen, C.H., 2007. The use of dynamic display to improve reading comprehension for the small screen of a wrist watch. *Human Interface and the Management of Information. Methods, Techniques and Tools in Information Design*, pp.814-823.

[11] Dickie, C., Vertegaal, R., Sohn, C. and Cheng, D., 2005, October. eyeLook: using attention to facilitate mobile media consumption. In *Proceedings of the 18th annual ACM symposium on User interface software and technology* (pp. 103-106). ACM.

[12] Dillon, A., 1992. Reading from paper versus screens: A critical review of the empirical literature. *Ergonomics*, *35*(10), pp.1297-1326.

[13] 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).

[14] 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.

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

[16] Drewes, H., De Luca, A. and Schmidt, A., 2007, September. Eye-gaze interaction for mobile phones. In Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology (pp. 364-371). ACM.

[17] Esteves, A., Velloso, E., Bulling, A. and Gellersen, H., 2015, November. Orbits: Gaze interaction for smart watches using smooth pursuit eye movements. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (pp. 457-466). ACM.

[18] Fialding, L.G. and Pearson, P.D., 1994. Synthesis of research reading comprehension: What works. *Educational Leadership, 51*, pp.62-62.

[19] Forster, K.I., 1970. Visual perception of rapidly presented word sequences of varying complexity. *Attention, Perception, & Psychophysics*, *8*(4), pp.215-221.

[20] Guo, W., & Wang, J. (2017, May). SmartRSVP: Facilitating Attentive Speed Reading on Small Screen Wearable Devices. *In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 1640-1647). ACM.

[21] Gilbert, L.C., 1959. Speed of processing visual stimuli and its relation to reading. *Journal of Educational Psychology*, *50*(1), p.8.

[22] Han, T., Xiao, X., Shi, L., Canny, J. and Wang, J., 2015, April. Balancing accuracy and fun: designing camera based mobile games for implicit heart rate monitoring. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 847-856). ACM.

[23] 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.

[24] Hansen, J.P., Lund, H., Biermann, F., Møllenbach, E., Sztuk, S. and Agustin, J.S., 2016, March. Wrist-worn pervasive gaze interaction. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications* (pp. 57-64). ACM.

[25] 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.

[26] 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.

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

[28] 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).

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

[30] 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.

[31] Larson, K. (2004). The science of word recognition. *Advanced Reading Technology, Microsoft Corporation*

[32] 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.

[33] Lu, G., Yang, F., Taylor, J.A. and Stein, J.F., 2009. A comparison of photoplethysmography and ECG recording to analyse heart rate variability in healthy subjects. *Journal of medical engineering & technology,* 33(8), pp.634-641.

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

[35] 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). ACM.

[36] Masson, M.E.J., 1983 Conceptual processing of text during skimming and rapid sequential reading. *Memory & Cognition* 11, 3 (pp. 262–274).

[37] 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.

[38] Miluzzo, E., Wang, T., and Campbell, A.T. 2010. EyePhone: activating mobile phones with your eyes. *Proc. SIGCOMM '10*, 15–20.

[39] 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.

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

[41] 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).

[42] Oulasvirta, A., Tamminen, S., Roto, V. and Kuorelahti, J., 2005, April. Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 919-928). ACM.

[43] Parnami, A., Gupta, A., Reyes, G., Sadana, R., Li, Y. and Abowd, G., 2016, September. Mogeste: mobile tool for in-situ motion gesture design. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct* (pp. 345-348).

[44] Picard, R.W. and Picard, R., 1997. *Affective computing* (Vol. 252). Cambridge: MIT press.

[45] 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.

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

[47] Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological science in the public interest*, *2*(2), 31-74.

[48] Rayner, K., Slattery, T. J., & Bélanger, N. N. (2010). Eye movements, the perceptual span, and reading speed. *Psychonomic bulletin & review*, *17*(6), 834-839.

[49] Ren, S., Cao, X., Wei, Y. and Sun, J., 2014. Face alignment at 3000 fps via regressing local binary features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1685-1692).

[50] 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.

[51] 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.

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

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

[54] 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.

[55] Vadas, K., Patel, N., Lyons, K., Starner, T. and Jacko, J., 2006, September. Reading on-the-go: a comparison of audio and hand-held displays. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services* (pp. 219-226). ACM.

[56] 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.

[57] 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.

[58] 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.

[59] 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.