RSVP on the Go - Implicit Reading Support on Smart Watches Through Eye Tracking

Tilman Dingler, Rufat Rzayev, Valentin Schwind, Niels Henze
University of Stuttgart – VIS, Stuttgart, Germany
{firstname.lastname}@vis.uni-stuttgart.de

ABSTRACT
While smartwatches have become common for mobile interaction, one of their main limitation is the limited screen size. To facilitate reading activities despite these limitations, reading with Rapid Serial Visual Presentation (RSVP) has been shown to be feasible. However, when text is presented in rapid sequence, single words are easily missed due to blinking or briefly glancing up from the screen. This gets worse the more the reader is engaged in a secondary task, such as walking. To give implicit control over the reading flow we combined an RSVP reading application on a smartwatch with a head-worn eye tracker. When the reading flow is briefly interrupted, the text presentation automatically pauses or backtracks. In a user study with 15 participants we show that using eye tracking in combination with RSVP increases users' comprehension compared to a touch-based UI to control the text presentation. We argue that eye tracking will be a valuable extension for future smartwatch interaction.

Author Keywords
Reading interfaces; RSVP; eye-tracking; eye gaze interaction; mental load; comprehension

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous; Miscellaneous

INTRODUCTION
Reading with Rapid Serial Visual Presentation (RSVP), a concept originally introduced by Forster [6], has recently gained considerable attention. RSVP for text presents words sequentially. As it requires screens to present only single words, this presentation technique is especially promising when screen real-estate is limited. Smartwatches with particularly small screens are therefore well suited for it. However, reading via RSVP requires high user attention since words are flashed in rapid sequence. This makes reading while performing an additional task, *e.g.* walking, challenging. Even small distractions cause users to briefly look up from the screen and hence lose part of the currently displayed sentence.

ISWC '16, September 12-16, 2016, Heidelberg, Germany
© 2016 ACM. ISBN 978-1-4503-4460-9/16/09...$15.00.
DOI: http://dx.doi.org/10.1145/2971763.2971794

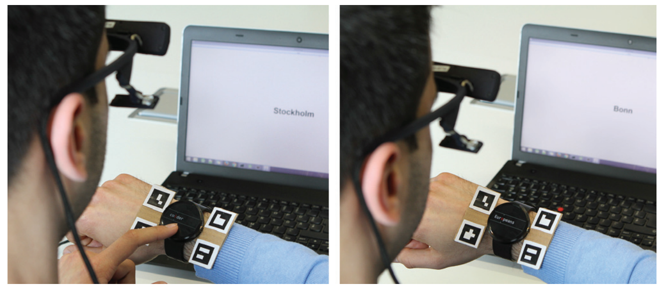

Figure 1. Participant reading a text on a smartwatch while monitoring city or country names on a secondary display. Left: the reading flow is controlled explicitly through touch. Right: the reading flow is controlled implicitly through eye-gaze.

To avoid missing vital text parts, controls are required to pause and resume the flashing of words. RSVP is typically controlled using explicit user input. However, pressing a button or performing a gesture requires time, attention and accuracy. Explicit input can therefore be challenging for short and frequent interruptions. Looking up from the screen while walking to adjust the walking direction, for example, can take less than a second. For such brief interruptions, we created a system, which allows users to implicitly control RSVP text presentation. Therefore, we augmented readers with a mobile eye tracker to detect when the reading flow is interrupted and pause the sequence of words accordingly. In case of longer interruptions, the text backtracks to the beginning of the sentence to restore the reader's context. Hence, we aim to combine the advantages of normal reading, which allows readers to freely switch their focus between text and environment, with the advantage of RSVP requiring very little screen space. In this paper we present a system that uses RSVP to present text on a smartwatch and allows users to explicitly or implicitly control the text presentation. We conducted a study with 15 participants to investigate the advantages of implicit eye gaze control. The contribution of this paper is as follows:

1. We present a system for smartwatches that uses eye tracking to control the reading flow with RSVP.

2. A study that shows the advantage of implicit RSVP control while engaged in a secondary task.

RELATED WORK
Limited screen size poses a challenge for reading User Interfaces (UIs), which has been subject to previous investigations: reading performance on mobile devices generally increases with bigger font size, even more so it affects the

readers' subjective preference and lowers levels of perceived difficulty [1, 2]. A secondary task, such as walking, further decreases reading performance while increasing cognitive load [14]. RSVP has been suggested for usage on mobile devices and wearables in order to trade space for time [3]. As smartwatches with an even more limited screen real estate become more prevalent, this technique has also gained commercial interest[1]. Text can be displayed larger and of higher resolution as only a single word needs to fit on the display at a time. Originally coined by Forster [6], RSVP describes an experimental model for temporal characteristics of attention, in which users focus on visual stimuli sequentially presented in the same place. Thereby subsequent targets are potentially missed especially when they occur in rapid succession (180-450ms), which Raymond *et al.* [13] described as *attentional blink*. When users are engaged in a secondary task, more stimuli are potentially missed. This is especially critical for reading activities, where missing parts of a sentence may severely inhibit readers' text comprehension. In a former study we investigated the feasibility of RSVP to increase reading speed on electronic devices: high reading speed with RSVP lead users to easily miss entire words and thereby lose text context [5]. Hedin and Lindgren [8] investigated RSVP on mobile devices with regard to its effects on reading comprehension and efficiency: by comparing RSVP to reading with scrolling they found that with RSVP speed and comprehension is high, but users were generally uncomfortable with the technique. This may especially be due to a perceived lack of control over the reading flow.

First work on gaze-based interaction focused on the user-computer dialogue in a natural and unobtrusive way [9]. Kern *et al.* [11] investigated the feasibility of using eye-tracking to facilitate the resumption of an interrupted task: they developed a system that provided visual placeholders to highlight the last gaze position which allowed users to efficiently switch between tasks. Hansen *et al.* [7] added gaze-tracking to smartwatches to allow hands-free interaction through gaze gestures. Dickie *et al.* [4] introduced a platform for sensing eye contact on mobile screens based on an infrared camera system. They further discussed a reading application using RSVP and controlling the reading flow through eye-gaze. However, the application was neither implemented nor tested.

Concluding, reading with RSVP seems generally feasible in situations, in which screen real-estate is severely limited, as is the case with watches. But especially with regard to monitoring a second task, users may lose text context by missing words, which compromises text comprehension. In order to give users control over the reading flow, we apply eye-gaze tracking as an implicit interaction technique and hypothesize certain advantages over explicit reading control.

EXPLICIT AND IMPLICIT RSVP CONTROL

We developed a prototype that enables users to explicitly or implicitly control RSVP text presentation. The system consists of a smartwatch with a touchscreen and an eye tracker.

[1]http://spritzinc.com/

Figure 2. RSVP reading interface with four visual markers augmenting the smartwatch to detect when user's eye gaze is directed at the watch.

Therefore, we implemented an Android Wear RSVP application based on an open source framework[2] for the Motorola Moto 360 smartwatch. The Moto 360 runs Android (version 6.0.1) and has a circular 320×290 pixel 1.56″ touchscreen display with a 46 mm diameter. The application displays text, which is shown word by word with a specially colored letter serving as a focal point for the readers eyes to focus on (see Fig. 2). The number of words shown per minute (e.g. 200 words per minute) can be freely selected from the application's setting.

Users explicitly control the text flow through the touchscreen of the smartwatch: reading can be paused and resumed through simple taps (Fig. 1 (left)). To allow readers restore the reading context, the text presentation goes back to the beginning of the current sentence in case of longer pauses (*i.e.* longer than 5 seconds). We use an eye tracker to enable implicit control over the text presentation (Fig. 1 (right)). For tracking the user's eye movements, we use the Pupil Pro eye tracker by Kassner *et al.* [10], which comes with a 3d-printed frame and a software package for calibration, gaze detection, and surface registration. The modular eye tracker consists of a 120Hz head-mounted monocular camera with a resolution of 640×480 pixels and a world camera with 30Hz which delivers the video stream in FullHD. We attached four visual markers, each with a size of $20\,\text{mm}\times20\,\text{mm}$, to the bezels of the smartwatch screen (see Fig. 2), which allow the software to determine whether the eye gaze is directed inside or outside of the marked rectangle, *i.e.* users look at the smartwatch display or not. The gap between markers constitutes 14mm (vertically) and 64 mm (horizontally), so that the diameter of the registered virtual surface corresponds to the diameter of the watch interface. The eye tracker communicates with the smartwatch using the Pupil Capture software[3], which entails the Pupil Server plugin for broadcasting eye gaze data over a network. The Android Wear application on the smartwatch receives the data stream and determines whether the user is currently looking at the watch or not. Hence, the reading flow pauses when the user looks away and automatically resumes when the user looks back at the watch. In case of a longer reading pause (longer than 5 seconds), the reading position is further reset to the beginning of the current sentence.

[2]https://github.com/andrewgiang/SpritzerTextView
[3]https://github.com/pupil-labs/pupil

USER STUDY

To compare explicit with implicit control of RSVP reading flow we conducted a user study where participants read texts with the system described above. We hypothesized that an adaptive reading interface taking into account the user's eye gaze would lead to higher text comprehension and reading confidence.

Study Design

We employed a repeated-measures design with the RSVP control modality being the independent variable resulting in two conditions: 1) manual control through touch interaction (*tap*) and 2) implicit control through eye gaze (*eye gaze*). For each condition we introduced a secondary task: while reading on the smartwatch, participants were asked to monitor words displayed on a desktop monitor in front of them. Shown words were either countries or city names. If a word was a country name, participants had to press a button on the keyboard. City names had to be ignored. Words were shown for 10 to 15 seconds. As dependent variables we measured 1) overall task completion time, 2) comprehension scores, 3) tracked eye movements (i.e. reading pauses), 4) errors on the secondary task, *i.e.* number of countries missed, and 5) measured mental load using a NASA TLX questionnaire after each condition.

Procedure

After welcoming and introducing participants to the purpose of the study we asked them to sign a consent form and recorded demographic data. We then introduced the RSVP reading interface on the smartwatch and set up the mobile eye tracker, which required a brief calibration for each participant. We explicitly asked to keep the arm wearing the watch rather still so that it remained in the camera view of the mobile eye tracker. After they had familiarized themselves with the interface and the available controls, we assigned participants to a starting condition. We then asked them to read while completing the secondary task, whereas we instructed them to treat both tasks as equally important. After each text we administered a 10-item comprehension test. For each condition, participants read 2 texts, after which they completed a NASA TLX questionnaire. We counterbalanced both the sequence of conditions via latin-square as well as the allocation between tasks and reading texts. All texts were taken from a collection used for English as a Second Language (ESL) learners, an adaption from [12]. These texts came with a predefined set of 10 comprehension questions per text and comprised on average 548 words ($SD = 2.87$). Since participants read 2 texts per condition, the maximum comprehension score per condition was 20. The study took about 50 minutes per participant, which was concluded with a final questionnaire.

Participants

We recruited 15 participants (11 males, 4 females) through a university mailing list. With a mean age of 26.5 years ($SD = 3.5$) most had a background in IT or were university students. All reported English to be their second language, 3 were wearing contact lenses (20%), 5 glasses (33.3%). 8 of them indicated to be familiar with the RSVP reading technique (53.3%), 8 were wearing watches on a regular basis (53.3%), none of which were smartwatches. Participants were rewarded 10EUR for taking part in the study.

RESULTS

Each participant read in total 4 texts, 2 in each condition. Table 1 summarizes the descriptive system measurements.

	explicit interaction			implicit interaction		
	Md	M	SD	Md	M	SD
Number of pauses	8	13.5	11.73	29	30.9	6.14
Percentage of missed countries	0	14.02	18.54	6.25	9.84	12.59
Num. of mistakes on secondary task	0	0.6	0.74	1	0.93	1.22
Task completion times (second)	297.2	313.3	46.4	352.4	354.1	45.3
Comprehension Scores	16	15.86	1.88	18	17.33	1.67

Table 1. System measurements for both conditions: explicit (touch) and implicit (eye gaze).

Objective Measures

A Wilcoxon Signed-Ranks Test revealed that the median of the comprehension scores for the eye gaze condition ($Md = 18$) was significantly higher than for the tap condition ($Md = 16$), $Z = 74.5, p = .041$. Thus, participants had a better understanding of the read text when reading with implicit control using the eye tracker. Another Wilcoxon Signed-Ranks Test also revealed that the median of the average number of pauses for the tap condition ($Md = 8$) was significantly lower than for the eye gaze condition ($Md = 29$), $Z = 105.0, p = .001$. In sum, participants made 9 mistakes on the secondary task, *i.e.* cities were selected instead of countries, during the tap condition ($Md = 0$) and 14 mistakes during the eye gaze condition ($Md = 1$). We found no statistically significant differences between the percentage share of missed countries while tapping ($M = 14.019, SD = 18.539$) vs. while using eye gaze ($M = 9.842, SD = 12.599$), $t(14) = 1.549, p = .144, r = .842$. There was a statistically significant difference between task completion times for the tap condition (in seconds: $M = 313.3, SD = 46.4$) and the eye gaze condition ($M = 354.1, SD = 45.3$), $t(14) = -2.383, p = .032, r = .044$, where participants took more time when reading with implicit control over the text presentation.

Subjective Feedback

As for the subjective assessment through the Nasa TLX questionnaire, a student t-test showed no statistically significant difference between the tap condition ($M = 10.278, SD = 2.564$) and eye gaze condition ($M = 10.289, SD = 2.293$), $t(14) = -.014, p = .989, r = .190$. Thus, we found no effect of the control mechanism on the perceived mental load. In the final questionnaire most participants reported the frequent taps to pause the text to be annoying and hence preferred the support through eye-gaze. However, they felt a lack of control, since the implicit pauses entailed no noticeable feedback.

DISCUSSION AND LIMITATIONS

Comprehension was higher for implicit than for touch interaction. Therefore, eye gaze interaction seems to be less distracting than when explicitly having to control the reading flow, which confirms, our hypothesis. While the average number of pauses was also higher in the eye gaze condition, switches between tasks seem to have been done with ease: participants reported in the explicit interaction condition that they sometimes looked up without pausing the text flow in order to take a brief glance at the secondary screen. Hence, they compromised on text comprehension, whereas the eye gaze interaction implied a pause during a quick glance. This is also what contributed to the overall increase in task completion time. The monitoring task we employed as a secondary task was meant as a distraction task. Because of the eye tracker's sensitivity to large head motions, we refrained from having participants perform a walking task. With more robust trackers we would like to test our hypothesis in the context of a navigational and therefore spacial task. Although we designed the eye-gaze interaction with RSVP in mind, a similar approach could be taken to control the flow of automatic scrolling through text displayed on a conventional reading interface. Other reading techniques as proposed by [5] could also benefit from eye-gaze tracking to dynamically adjust reading speed, for example. Further, we envision front-facing cameras soon to be able to detect the user's eye gaze which will render the mobile eye tracker obsolete and allow for more widely applied eye-gaze interactions, also for reading UIs.

CONCLUSION

In this paper we described a working prototype to control reading flow in a RSVP reading UI through the user's implicit eye gaze. When reading on mobile devices screen size is limited and distractions from the environment can hinder the reading flow. In our user study, we showed that gaze interaction is more effective than having users explicitly pause and resume reading through touch. Hence, we showed eye gaze as an implicit interaction modality to support reading on the go while monitoring a secondary task.

ACKNOWLEDGMENTS

We acknowledge the funding through the Future and Emerging Technologies (FET) programme within the 7th Framework Programme for Research of the European Commission, under FET grant number: 612933 (RECALL), as well as through the German Research Foundation (DFG) within the project C04 of SFB/Transregio 161.

REFERENCES

1. Leon Barnard, Ji Soo Yi, Julie A Jacko, and Andrew Sears. 2005. An empirical comparison of use-in-motion evaluation scenarios for mobile computing devices. *International Journal of Human-Computer Studies* 62, 4 (2005), 487–520.
2. Iain Darroch, Joy Goodman, Stephen Brewster, and Phil Gray. 2005. *Proc. of INTERACT '05*. Springer Berlin Heidelberg, Berlin, Heidelberg, Chapter The Effect of Age and Font Size on Reading Text on Handheld Computers, 253–266. DOI: http://dx.doi.org/10.1007/11555261_23
3. Oscar de Bruijn and Robert Spence. 2000. Rapid serial visual presentation: a space-time trade-off in information presentation. In *Advanced visual interfaces*. 189–192.
4. Connor Dickie, Roel Vertegaal, Changuk Sohn, and Daniel Cheng. 2005. eyeLook: Using Attention to Facilitate Mobile Media Consumption. In *Proc. of UIST '05*. ACM, New York, NY, USA, 103–106. DOI: http://dx.doi.org/10.1145/1095034.1095050
5. Tilman Dingler, Alireza Sahami Shirazi, Kai Kunze, and Albrecht Schmidt. 2015. Assessment of Stimuli for Supporting Speed Reading on Electronic Devices. In *Proc. of AH '15*. ACM, New York, NY, USA, 117–124. DOI: http://dx.doi.org/10.1145/2735711.2735796
6. Kenneth I Forster. 1970. Visual perception of rapidly presented word sequences of varying complexity. *Perception & Psychophysics* 8, 4 (1970), 215–221.
7. John Paulin Hansen, Florian Biermann, Emilie Møllenbach, Haakon Lund, Javier San Agustin, and Sebastian Sztuk. 2015. A GazeWatch Prototype. In *Proc. of MobileHCI '15*. ACM, New York, NY, USA, 615–621. DOI: http://dx.doi.org/10.1145/2786567.2792899
8. Björn Hedin and Erik Lindgren. 2007. A Comparison of Presentation Methods for Reading on Mobile Phones. *Distributed Systems Online, IEEE* 8, 6 (2007), 2–2.
9. Robert J. K. Jacob. 1990. What You Look at is What You Get: Eye Movement-based Interaction Techniques. In *Proc. of CHI '90*. ACM, New York, NY, USA, 11–18. DOI: http://dx.doi.org/10.1145/97243.97246
10. Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction. In *Proc. of UbiComp '14 Adjunct*. ACM Press, New York, USA. DOI: http://dx.doi.org/10.1145/2638728.2641695
11. Dagmar Kern, Paul Marshall, and Albrecht Schmidt. 2010. Gazemarks: Gaze-based Visual Placeholders to Ease Attention Switching. In *Proc. of CHI '10*. ACM, New York, NY, USA, 2093–2102. DOI: http://dx.doi.org/10.1145/1753326.1753646
12. Elizabeth Quinn and Ian Stephen Paul Nation. 1974. *Speed reading: A course for learners of English*. Oxford University Press.
13. Jane E Raymond, Kimron L Shapiro, and Karen M Arnell. 1992. Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human perception and performance* 18, 3 (1992), 849.
14. Bastian Schildbach and Enrico Rukzio. 2010. Investigating Selection and Reading Performance on a Mobile Phone While Walking. In *Proc. of MobileHCI '10*. ACM, New York, NY, USA, 93–102. DOI: http://dx.doi.org/10.1145/1851600.1851619