Utilizing Gaze Detection to Simulate the Affordances of Paper in the Rapid Serial Visual Presentation Format

Gustav Öquist [1], Staffan Björk[2] and Mikael Goldstein[3]

[1] Uppsala University, Department of Linguistics, PO Box 527
751 20 Uppsala, Sweden
gustav@stp.ling.uu.se
[2] Interactive Institute, PLAY Studio, PO Box 620
405 30 Göteborg, Sweden
staffan.bjork@interactiveinstitute.se
[3] Ericsson Research, Interaction & Usability Lab, Torshamnsgatan 23,
164 80 Kista, Sweden
mikael.goldstein@era.ericsson.se

Abstract. We present how gaze detection can be used to enhance the Rapid Serial Visual Presentation (RSVP) format, a dynamic text presentation technique suitable for mobile devices. A camera mounted on the device is used to monitor the reader's gaze and control the onset of the text presentation accordingly. The underlying assumptions for the technique are presented together with a description of a prototype, Smart Bailando, as well as our directions for further work.

1 Introduction

On paper the average reading speed for English text is between 220-340 words per minute (wpm) [9]. Reading speed on large screens is today likely to be more or less the same due to improved resolution [12]. Increased resolution will surely improve legibility on small screens as well but readability will remain low due to the limited screen space available [3]. In many cases, a paper copy can simply solve the problem of having to read from a screen at all, but users of handheld devices do not always have access to printing facilities. Developers therefore try to make reading as easy as possible by improving display quality and user interfaces but handheld devices still have an inherent problem in their limited screen space.

This dilemma does however presuppose that the text is presented in the traditional page format. One approach to overcome the size constraint may be to make use of the possibilities actually offered by mobile devices and trade space for time [1]. Dynamic text presentation via Leading or Rapid Serial Visual Presentation (RSVP) requires much smaller screen space compared to traditional text presentation with maintained reading efficiency [1, 6, 8, 11, 12]. Leading, or the Times Square Format, scrolls the text on one line horizontally across the screen whereas RSVP presents the text as chunks of words or characters in rapid succession at a single visual location [10]. From a physiological perspective RSVP appears to suit the natural reading process better since the text then moves successively rather than continuously [17].

In a repeated-measurement experiment, Goldstein et al. [4] found that neither reading speed nor comprehension differed from paper text reading for longer texts. However, the NASA-TLX (Task Load Index) revealed significantly higher task load for RSVP conditions compared to paper reading for most factors. One explanation to the high cognitive load may have been that each text chunk was exposed for the same fixed duration of time. Just and Carpenter [7, p. 330] have found that "there is a large variation in the duration of individual fixations as well as the total gaze duration on individual words" when reading text from paper. Adaptive RSVP [5, 17] attempts to match the reader's cognitive text processing pace more adequately by adjusting the exposure time of each chunk with respect to the characteristics of the text being shown. In a usability evaluation Öquist and Goldstein [17] found that adaptation could indeed decrease task load for most factors. In an experiment with a similar approach Castelhano and Muter [2] found that the introduction of punctuation pauses, interruption pauses and pauses at clause boundaries made RSVP significantly more liked. Although these evaluations are not fully comparable they all seem to indicate that the RSVP format has some potential but also some flaws that yet remains to be resolved.

We believe that dynamic text presentation can improve reading efficiency on small screens, moreover we recognize that also relatively new formats like RSVP must adhere to the fundamental principles of reading that has evolved over time in order to be usable. This notion has led us to explore ways of simulating ordinary reading via RSVP using sensors.

2 Paper and Screen Affordances of Traditionally Presented Text

The trade-off between time and space that RSVP offers comes with the additional mental cost that the act of reading for the user is changed. The natural eye movements when reading traditionally presented text involve performing fixation-saccade-fixation patterns including regressions and return sweeps [10]. One inherent difference of RSVP is that it demands the reader to continuously fixate his gaze at one single location in the text presentation window.

It is very common that thought and gaze is frequently diverted from the text during traditional paper reading due to external distractions or periods of reflection. Paper-presented text supports this activity as the text stays in the same place and it is easy to resumes reading. By using Donald Norman's [13] term *affordances,* one could argue that this is an affordance that traditionally presented text on paper or on screen offers. This kind of affordances does not apply to the RSVP format due to its dynamic nature. Thus, readers are forced to continuously monitor themselves when using the RSVP format. If the gaze strays away, the RSVP presentation has to be stopped manually. This may be one reason for the high cognitive demand score obtained in earlier experiments [1, 6, 17] as readers can have felt an urge to fixate on the text continuously since looking away can lead to missing information.

If one enhanced the RSVP application with sensors that register the reader's gaze, *gaze detection*, the application could become context-aware [15] and automatically stop/start the text presentation when the reader looked away from the text. A pre-

condition for this would be that the terminal using the RSVP format would have a built-in camera focused on the reader's eyes continuously during RSVP reading. Mobile phones are currently being released on the market with such a camera integrated into their design (e.g. the Sony Ericsson P800 and the Nokia 7650) and cameras can be bought as add-on modules for PDAs (e.g. the HP Pocket Camera) soon making this requirement very easy to fulfill. Based on the observations presented above, we believe that adding gaze detection functionality to RSVP reading on handheld PDAs and cellular phones is one feasible route to making reading on small devices as convenient as ordinary screen or paper reading.

3 Smart Bailando

In one of our threads to explore the possibilities of RSVP with gaze detection we are currently supervising a Master's thesis [16] where a RSVP application, *Smart Bailando*, is being developed in which the stop/start of the text presentation is controlled by eye movement (Fig 1).

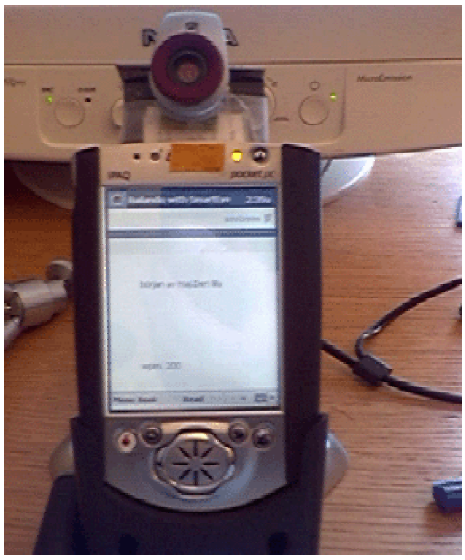

Fig. 1. The Smart Bailando prototype with the gaze detection sensor attached

Gaze detection is provided by a software platform for real-time measurements of eye movement developed by Smart Eye AB (www.smarteye.se). The platform allows gaze tracking using a standard PC equipped with one or several digital video cameras including web cameras, making it a quick and easy prototyping platform. The platform is written in C++, and as it runs on Windows OS, a Pocket PC version of the system is feasible as soon as PDAs have the required computational powers (which with the current technological development will be within 2 years). As the current gaze detection can only run on a PC, Smart Bailando is built as the client of a client-

server application where the Smart Eye platform functions as the server running on the PC. The camera used must be connected to the PC, but by dismantling the camera and mounting it on the PDA a good approximation can be achieved even though it enforced a restriction of how the PDA can be moved.

The prototype application currently runs both on a Pocket PC PDA (a Compaq iPAQ 3630 with a wireless LAN network) and in emulation mode on a PC. The system requires a calibration session and is currently in a refinement phase where the overall performance of the system is being improved by modifying individual parts of the whole system including but not limited to the making the calibration process easier, providing user feedback of the tracking process, improving network performance (all by the students), and modifications of the eye tracking algorithms (by Smart Eye).

Smart Bailando builds upon the Bailando application [5, 17], which was developed at Ericsson Research's Usability & Interaction Lab in Kista, Sweden, to explore Sonified and Adaptive RSVP. Bailando was developed for PDAs running the Pocket PC operating system, which offers the application sufficient processing power for experimenting with resource demanding tasks such as sound playback on a mobile device [5, 6]. The actual text presentation in Smart Bailando is exactly that of Bailando, i.e. it uses the Adaptive RSVP techniques first presented in that prototype.

4 Conclusions and Future Work

We have presented Smart Bailando, a proof-of-concept prototype that uses gaze detection to enhance the RSVP format. We have motivated the use of gaze detection by the desire to mimic the affordances provided by traditional text presentation. Although the current prototype shows that it is technologically possible to use eye movement to control RSVP, we mostly see it as a first step in exploring the feasibility of using real-time eye tracking techniques in combination with RSVP. Based on the prototype, we have identified the following research issues that are currently being pursued: formally evaluating the prototype, building a fully mobile prototype, and evaluating the prototype in a mobile setting.

Acknowledgements

The authors would like to thank Per Åkervall and Roger Granath for all their work with the Smart RSVP prototype. We would also like to thank the people at Smart Eye AB for providing and customizing the gaze detection software to our needs.

References

1. Bruijn, O. and Spence, R. (2000). Rapid Serial Visual Presentation: A space-time trade-off in information presentation. *Proceedings of Advanced Visual Interfaces, AVI'2000*.

2. Castelhano, M.S. and Muter, P. (2001). Optimizing the reading of electronic text using rapid serial visual presentation. *Behavior & Information Technology,* 2001, 20(4), 237-247.
3. Duchnicky, R.L. and Kolers, P.A. (1983). Readability of text scrolled on visual display terminals as a function of window size. *Human Factors*, 25, 683-692.
4. Goldstein, M., Sicheritz, K. and Anneroth, M. (2001). Reading from a small display using the RSVP technique*, Nordic Radio Symposium, NRS01,* Poster session, Nynäshamn, 4-6 April 2001 (full paper available on CD-Rom only).
5. Goldstein, M., Öquist, G., Bayat-M., M., Björk, S. and Ljungberg, P. (2001). Enhancing the reading experience: Using Adaptive and Sonified RSVP for reading on small displays, *Proceedings of 3rd International Mobile HCI Workshop.*
6. Goldstein, M., Öquist, G. and Björk, S. (2002). Immersion does not guarantee Excitement: Evaluating Sonified RSVP *(Full paper submitted to NordiCHI'2002).*
7. Just, M.A. och Carpenter, P.A. (1980). A theory of reading: From eye fixations to comprehension, *Psychological Review,* 87(4), 329-352.
8. Joula, J.F., Ward, N.J. and MacNamara, T. (1982). Visual search and reading of rapid serial presentations of letter strings, words and text. *J. Exper. Psychol.: General*, 111*,* 208-227.
9. Kump, P. (1999). *Break-trough rapid reading.* New Jersey: Prentice-Hall Press.
10. Mills, C.B. och Weldon, L.J. (1987). Reading text from computer screens. *ACM Computing Surveys,* Vol. 19, No. 4, ACM Press.
11. Muter, P. (1996). Interface design and optimization of reading of continuous text. In van Oostendorp, H. & de Mul, S. (Eds.), *Cognitive aspects of electronic text processing,* 161-180. Norwood, N.J.: Ablex.
12. Muter, P. and Maurutto, P. (1991). Reading and skimming from computer screens and books: The paperless office revisited? *Behavior & Information Technology*, 10, 257–266.
13. Norman, D.A. (1988). *The Psychology of Everyday Things.* Doubleday, NY, USA.
14. Rahman, T. and Muter, P. (1999). Designing an interface to optimize reading with small display windows. *Human Factors,* 1(1), 106-117, Human Factors and Ergonomics Society.
15. Schilit, B.N., Adams, N. and Want, R (1994). Context-aware computing applications. *Proceedings of the IEEE Workshop on Mobile Computing Systems and Applications.*
16. Åkervall, P. and Granath, R. (2002). *Eye Controlled RSVP on Handhelds.* Forthcoming Master's Thesis.
17. Öquist, G. and Goldstein, M. (2002). Towards an improved readability on mobile devices: Evaluating Adaptive Rapid Serial Visual Presentation. *Proceedings of Mobile HCI'2002.*