

# Enhancing the Reading Experience: Using Adaptive and Sonified RSVP for Reading on Small Displays

MIKAEL GOLDSTEIN<sup>1</sup>, GUSTAV ÖQVIST<sup>1</sup>, MANDANA BAYAT-M<sup>1</sup>,  
PETER LJUNGSTRAND<sup>2</sup>, STAFFAN BJÖRK<sup>2</sup>

<sup>1</sup>*Ericsson Research, Usability & Interaction Lab, Kista, Sweden*

<sup>2</sup>*PLAY group, Interactive Institute, Gothenburg, Sweden*

{mikael.goldstein, gustav.oqvist, mandana.bayat-m}@era.ericsson.se  
{peter.ljungstrand, staffan.bjork}@interactiveinstitute.se  
Telephone: +46 (8) 757 3679  
Fax: +46 (8) 757 3100

Ongoing work on self-paced reading using two variants of the RSVP (Rapid Serial Visual Presentation) technique on a PDA device with small display screen is reported. The two variants, Adaptive and Sonified RSVP, have been developed with different goals: improving subjective reading convenience and enhancing reading experience. Adaptive RSVP aims at making RSVP reading less mentally demanding and frustrating by taking into account the cognitive processing of a text by adjusting the exposure time of each text chunk to content as well as context. Sonified RSVP aims at enhancing the reading experience by linking nomic auditory icons to the text, creating a multi-modal experience. Adaptive and Sonified RSVP have been implemented in the Bailando text viewer prototype, which is described.

*Keywords: RSVP, adaptive RSVP, sonified RSVP, visualization techniques, auditory icons,*

## Background

Rapid Serial Visual Presentation (RSVP) is a way of presenting a text dynamically on a screen [16]. RSVP presents chunks of text in a rapid succession at a single visual location. By thus minimising the need for saccadic eye movements during reading, a trade-off between space requirements and duration of text presentation is enabled. Juola et al. [13] found that for equal reading rates, comprehension in the RSVP mode was similar to page mode comprehension whereas Masson [15] found that on-line reading *comprehension* for RSVP text was poorer than for page-presented computer text. The difference may be attributable to the insertion of a blank window of between 200-300 ms appearing between sentences in the Juola et al. study. In a repeated-measurement experiment using the RSVP Reader 1.0 prototype [19], Goldstein et al. [11] found that neither reading speed in words per minute ( $M \approx 290$  wpm) nor comprehension ( $M = 82\%$ ) differed from paper text reading of long Swedish texts (3000-6000 words). The RSVP technique has also gained significant interest in recent years as can be noted by several commercial implementations [6, 8].

The ability to trade display space requirements for display time without affecting comprehension makes RSVP a good candidate for displaying long text documents

(e.g., a novel or a text document attached to an e-mail) when display real estate is small and the user is on the move and lacks printing facilities. This context scenario is typical for modern mobile phones and PDAs.

## Adaptive RSVP

Previous implementations utilizing the RSVP technique have typically exposed each chunk of text on the screen for a *fixed* duration of time. The exposure time of each text chunk has usually been related to the reader's selected reading speed level measured in wpm, the number of characters or words to present in each text chunk and the average word length. The evaluation of the RSVP Reader V1.0 prototype, indicated that objective metrics such as reading speed and comprehension were equal to reading on paper. However the NASA-Task Load Index, measuring the subjective factors Mental, Physical and Temporal demands as well as Performance, Effort and Frustration level revealed significantly higher task load for the RSVP conditions compared to paper reading for most factors [11, 19]. One explanation may be that this RSVP application did not take into account the variation how fast different words and different parts of a sentence are processed. Just and Carpenter 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 [14:330]. Adaptive RSVP attempts to mimic the reader's text processing pace more adequately by adjusting each chunk exposure time in respect to content as well as to context [20]. Assuming the *eye-mind hypothesis* [14], i.e. the eye remains fixated on a text chunk as long as it is being processed, the needed exposure time of a text chunk can be assumed to be proportional to the predicted gaze duration of that text chunk when reading a normal (non-RSVP visualized) text. Since very common, known or short words are usually processed faster than infrequent, unknown or long words, the text chunk exposure times can be adjusted accordingly for RSVP. Likewise, a shorter sentence is usually processed faster than a longer one. Thus, processing time differs both between and within sentences. For example, most information tends to be introduced late in sentences; ambiguity and references tend to be resolved at the end of sentences. Thus, text chunks exposing the end of a sentence require more processing time and consequently a longer exposure time [17, 18]. Adaptive RSVP is currently under evaluation, the results and more details remains to be published [20].

## Sonified RSVP

In order to enhance self-paced reading experience when interacting with electronic media, a possible enhancement would be to attach appropriate sounds when certain passages in the text are read. One example of this form of augmentation can be found in the Listen Reader [2], which synchronises audio playback with the (silent) self-paced reading of text and images printed on *traditional paper pages*. Using embedded RFID tagging [1] for page identification and electric field sensing to detect the actions of the person reading the book. Ambient sound is added to each page and played as the reader's fingers touch different parts of the page. However, synchronisation between text and sound may be poor.

In the case of Sonified RSVP [10], appropriate sounds are played when a certain text chunk is displayed. Two different types of appropriate sound augmentation were

available when designing sonified RSVP: *earcons* and *auditory icons*. Earcons [5, 7] are structured sounds that have been introduced as an alternative to graphical icons to improve usability in GUIs. Earcons are abstract, synthetic tones that are constructed from motives using timbre, register, intensity, pitch and rhythm. Auditory icons [9] are sounds from the everyday environment that are used to create an intuitive link to the object or action that they represent. They are based on natural sounds and have been classified into three different types: symbolic, nomic and metaphorical. Symbolic mappings rely on social convention for meaning, such as applause for approval. Nomic mapping is physical such as the sound of a closing metal cabinet for closing a file. Metaphorical mappings represent similarities, such as a falling pitch for a falling object, making auditory icons using them similar to earcons.

Nomic auditory icons were chosen as the elements for Sonified RSVP for several reasons. Compared to earcons, nomic auditory icons can more directly be linked to the content of the text in a one-to-one fashion. Of the three forms of auditory icons identified, nomic auditory icons require the least training to recognize. Finally, the auditory content of nomic auditory icons can be designed to augment the self-paced reading experience in a fashion similar to that of sound effects in film. For example, if the text chunk describes a door being slammed, the sound of a slamming door can be synchronised to the text to create an augmented reading experience. If the plot takes place on a sailing boat at sea, the nomic auditory icon of the wind can be continuously playing to indicate the duration of the action confined to the ship.

The use of recorded or computer-generated speech for augmenting the reading experience offers many of the advantages that nomic auditory icon lacks. However, there are several reasons why speech was not used for Sonified RSVP. Speech is intrusive and attention-demanding but more importantly, it is much slower than self-paced reading. Moreover, listening to a recorded or synthesised voice for a longer time can also become tiresome. However, it has been proposed that the spoken information can be conveyed in a less intrusive way by letting a musical message imitate the sound of a corresponding spoken message [12].

## **The Adaptive and Sonified RSVP Prototype Bailando**

A prototype incorporating both Adaptive and Sonified RSVP features has been developed at the Ericsson Research's Usability & Interaction Lab in Kista, Sweden. The prototype is called Bailando and runs on the Pocket PC (iPAQ) operating system. (Figure 1) The software is written in C++ and is easily portable to other platforms and PDA's.

### **Adaptive RSVP**

When a text document is opened an approximate readability rating, LIX ("Läsbarhetsindex") comparable to the Flesh index [4] is calculated, based on the percentage of long words and the average sentence length. An appropriate reading speed is then suggested depending on the current readability rating and the reader's reading profile. Bailando is currently able to present text in three RSVP modes: fixed RSVP, adaptive RSVP, and content and context adaptive RSVP. Content adaptation adjusts text chunk exposure time to the amount of text information displayed while context adaptation adjusts text chunk exposure time to sentence structure and word frequency. Content adaptation is language independent while context adaptation is language dependent since it uses word frequency data [20].

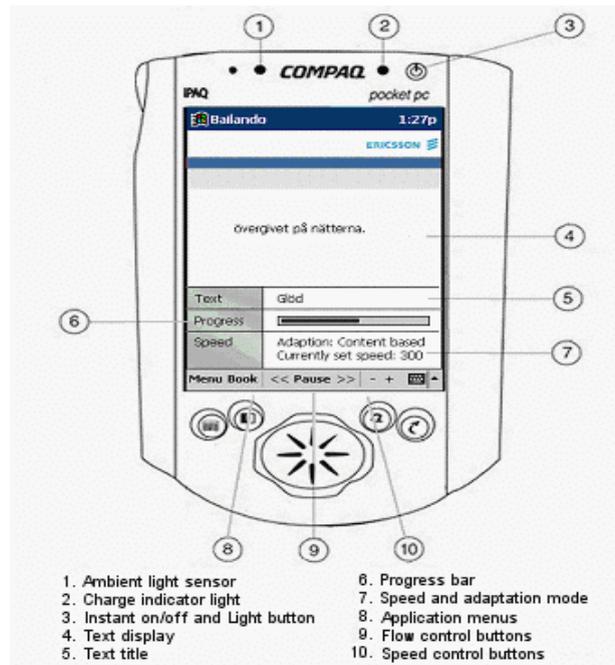


Figure 1. The Bailando prototype running on a Compaq iPAQ.

Bailando features a graphical user interface which gives the reader full control over the reading, including actions to go back, forward and to pause in the text and adjust the reading speed (in wpm). The RSVP text display is 25 characters wide and each text chunk is presented left justified in a sans-serif typeface. In order to support memory of spatial location while reading there is a progress bar which also helps the reader to estimate the text length and how long time it will take to read.

### Sonified RSVP

The Bailando prototype supports sonified RSVP by using tags in HTML and XML. An excerpt from a XML tagged text file in Swedish from the Swedish author August Strindberg's *The Red Room* is provided below as an example of Sonified RSVP tagging:

```

Det var en afton i början av maj. Den lilla trädgården på Mosebacke hade
ännu icke blivit <SOUND SRC="SOUND108.wav">öppnad för allmänheten och
rabatterna voro ej uppgrävda;</SOUND> snödropparne hade <SOUND
SRC="SOUND713.wav"> arbetat sig upp genom fjolårets lövsamlingar och
höllo just på att sluta sin korta verksamhet för att lämna plats åt de
ömtåligare <SOUND SRC="SOUND1.wav" PERPETUAL> saffransblommorna, vilka
tagit skydd under ett ofruktsamt päronträd; syrenerna väntade på sydlig
vind för att få gå i <SOUND SRC="SOUND136.wav">blom, men lindarne bjödo
ännu kärleksfilter i sina obrustna knoppar åt bofinkarne, som börjat
bygga sina lavklädda bon mellan stam och gren;</SOUND>

```

When the program encounters a <SOUND> tag in the text chunk to be displayed, the appropriate audio file is played. Thus, exact synchronisation between word and sound is

easily attainable and works at any selected reading speed. When a `</SOUND>` tag is encountered, all sounds playing are stopped. Since sound files tend to become quite large, the Bailando prototype supports the fetching of audio data from an external source by employing the URL format. Sonified RSVP is currently under evaluation, the results and more details remains to be published [3].

## References

1. Back M, Cohen J. Page detection using embedded tags. *Proceedings of UIST 2000*, ACM Press
2. Back M, Cohen J, Gold R, Harrison S, Minneman S. Listen Reader: An electronically augmented paper-based book. *In Conference Proceedings of CHI 2001*; 3(1):23-29, ACM Press
3. Bayat-M M. *Sonified RSVP*. Forthcoming Master's Thesis, Department of Computer Science, Royal Institute of Technology, Stockholm, 2001, Sweden
4. Björnsson CH. *Läsbarhet*. Liber, Stockholm 1968
5. Blattner MM, Sumikawa DA Greenberg RM. Earcons and Icons: Their Structure and Common Design Principles. *Human-Computer Interaction*, 1989; 4:11-44. Lawrence Erlbaum Associates, Inc
6. Brainspeed Inc. More information at: <http://www.ebrainspeed.com/> (June 2001)
7. Brewster S, Raty VP, Kortekangas A. Earcons as a method of providing navigational cues in a menu hierarchy. *Proceeding of Human-Computer Interaction'96* by MA Sasse, RJ Cunningham and RL Winder (Eds.), People and Computers XI, 1996:169-183, Springer-Verlag, London
8. Fastreader by Brainclone Enterprises. Available at: <http://www.brainclone.com/> (June 2001)
9. Gaver W. Auditory icons: Using sound in computer interfaces. *Human-Computer Interaction*, 1986; 2: 167-177, Lawrence Erlbaum Associates, Inc.
10. Goldstein M, Anneroth M, Sicheritz K, Lundberg T. Enhancing the user experience: On-line speed reading from a PDA screen. *The Fourth Swedish Symposium on Multimodal Communication, SSoMC 2000*. Available at: <http://www.dsv.su.se/~ssomc> (November 2000)
11. Goldstein M, Sicheritz K, Anneroth M. Reading from a small display using the RSVP technique. *Nordic Radio Symposium, NRS01*, Poster Session, Nynäshamn, April 2001, Sweden. Full paper available on CD-ROM only
12. Gärdenfors D. Auditory Interfaces – A Design Platform. Available at: <http://www.designsounds.net/katalog.html> (June 2001)
13. Jouta JF, Ward NJ, MacNamara T. Visual search and reading of rapid serial presentations of letter strings, words and text. *J. Exper. Psychol. General*, 1982; 111: 208-227.
14. Just MA, Carpenter PA. *The psychology of reading and language comprehension*. Boston: Allyn and Bacon, 1987
15. Masson MEJ. Conceptual processing of text during skimming and rapid sequential reading. *Mem. Cognition*, 1983; 11: 262-274
16. Mills CB, Weldon, LJ. Reading text from computer screens. *ACM Computing Surveys*, 1987; 19(4), ACM Press
17. Rayner K, Pollatsek A. *The psychology of reading*. Englewood Cliffs, 1989, NJ: Prentice Hall
18. Reichle E, Rayner K. *Comparing the E-Z Reader Model to Other Models of Eye Movement Control in Reading*. Available at CogPrints: <http://cogprints.soton.ac.uk/> (May 2000)
19. Sicheritz K. *Applying the Rapid Serial Presentation Technique to Personal Digital Assistants*. Master's Thesis, Department of Linguistics, Uppsala University, 2000, Sweden
20. Öqvist G. *Adaptive Rapid Serial Visual Presentation*. Forthcoming Master's Thesis, Department of Linguistics, Uppsala University, 2001, Sweden.