Evaluating Sonified Rapid Serial Visual Presentation: An Immersive Reading Experience on a Mobile Device

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

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

Abstract. Can the addition of sound enhance the reading experience on small screens when using Rapid Serial Visual Presentation (RSVP) for dynamic text presentation? In this paper we introduce Sonified RSVP and report findings from a usability evaluation where the experience of reading texts enhanced with nomic auditory icons was evaluated. At a comfortable pace 12 subjects read long Swedish texts of equal difficulty with and without the addition of sound on a handheld device. Reading speed (M≈217 wpm) and comprehension (M≈58% correct) did not differ significantly between the two conditions. The evaluation revealed a rather high task load for both conditions but no significant differences. However, the subjective rating of Immersion was rated significantly higher for the Sonified condition. Causes, implications and directions for further work are discussed based on these findings.

1 Introduction

With the advent of the mobile Internet and the ensuing increase in information resources available to mobile devices, such as Personal Digital Assistants (PDAs) and cellular phones, mobile users will eventually have access to the same amount of information as the stationary users. The fact that the mobile devices are much smaller does however constrain the usefulness of this advancement. The limited input capabilities make them primarily suitable for information retrieval, but the limited screen space currently constitutes a bottleneck for such appliances [9]. Since it is the customer's demand for small devices that sets these constraints they are likely to remain in the future as well. This notion, combined with the fact that readability has long been considered important as even small improvements can ease reading for large groups of people [19], has made the issues concerning the reading experience on small screens progressively more important for mobile usability.

Reading from small screens differs in many ways from reading both on paper and on large screens. The average reading speed on paper for English text is around 300 words per minute (wpm) [25], whereas early research on screen reading has revealed that reading speed decreases by 20-30% when reading on large screens [27]. As screen resolution has improved and people have gotten more used to them, readability on large screens has become more or less equal to paper [30]. The evolution in readability on small screens is however not likely to follow the same pattern. Screen resolution will surely improve and thus improve legibility, but decreased readability will most likely remain intrinsic to limited screen space [8].

The quandary with the size-readability trade-off does however presuppose that the text is presented in the traditional page format with a spatial layout. One approach to overcome the size constraints may be to design interfaces that utilize the possibilities offered by mobile devices to trade space for time [5]. Rapid Serial Visual Presentation (RSVP) and Leading are the two major techniques that have been proposed for dynamic text presentation [27]. 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 (Fig. 1). Both formats offer a way of reading texts on a very limited screen space [27, 28]. Comparisons between the formats have so far been inconclusive [20, 23], but at normal reading speeds RSVP appears to be more efficient. From a physiological perspective RSVP also seems more *natural* to use, the reason for this is that the text moves successively rather than continuously [43].

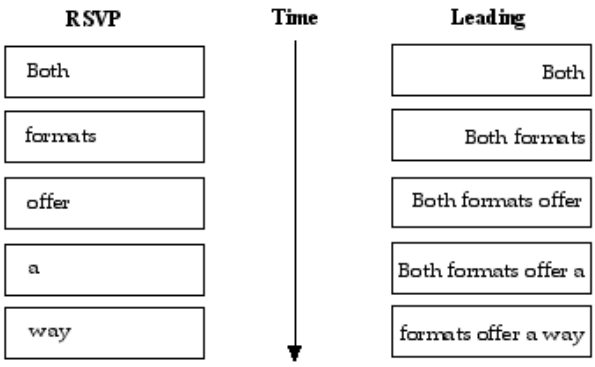

Fig. 1. Snapshots over time of text presentation via RSVP (left) compared to Leading (right)

Since dynamic presentation of text offers a potential improvement in readability on small screens it is important to explore its possibilities. In this paper we focus on the possible enhancement of the RSVP format by adding sounds to the text. First, a background to previous work on RSVP is presented together with motivation for the development of Sonified RSVP. This is followed by a description of a prototype for Sonified RSVP called Bailando. The design of the usability evaluation is then presented along with the experimental results. A discussion based on the findings together with possible implications for future work concludes the paper.

2 RSVP

RSVP originated as a tool for studying reading behavior [10, 13, 35] but has lately received more attention as a presentation technique with a promise of optimizing reading efficiency, especially when screen space is limited [15, 21, 36, 44]. The reason for the interest is that the reading process works a little different when RSVP is used and that it requires much smaller screen space than traditional text presentation. Reading a text with a spatial layout, like this page, consists of three distinct visual tasks: processing information in fixed gazes or *fixations*, performing *saccadic* eye movements to move between fixations, and moving to the next line using *return sweeps*. Whereas the saccadic eye movements and return sweeps are performed very quickly (~40 ms and ~55 ms respectively), fixations take longer time (~230 ms for fast readers and ~330 ms for average readers) [37]. RSVP displays the text as small chunks successively within a small area, which is assumed to minimize the need for saccadic eye movements and return sweeps [35, 38]. Moreover, RSVP is thought to increase reading speed since the format prevents re-reading of adjacent text segments. More important for mobile usability is however that RSVP reduces the need for physical interaction, in the form of paging or scrolling, when reading from small screens [43].

Juola et al. [20] found text comprehension between RSVP and traditional presentation techniques equal whereas Masson [26] found that comprehension of RSVP presented text was poorer than for page-presented text. A possible explanation for the different results may be the insertion of an empty screen (with a duration of 200-300 ms) between the sentences in the Juola et al. study. In a repeated-measurement evaluation, Goldstein et al. [15] compared ordinary reading in a paper book to reading on a PDA via RSVP; the results showed that neither reading speed nor comprehension differed. However, the NASA-TLX (NASA Task Load Index) [17] revealed significantly higher task load ratings for the RSVP conditions. 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 has 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 [22, p. 330].

Adaptive RSVP [16, 43, 44] was developed as an attempt 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 an evaluation of adaptive RSVP, where all conditions were performed on a PDA, Öquist and Goldstein [44] found that adaptation could indeed decrease task load for most factors. The evaluation also showed that RSVP might increase reading speed compared to traditional text presentation on mobile devices. In an experiment with a similar approach, Castelhano and Muter [6] found that the introduction of punctuation pauses, interruption pauses and pauses at clause boundaries made the RSVP format 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. Nevertheless, dynamic presentation formats like RSVP also have some additional advantages over traditional text presentation that can be used to enhance the reading experience.

3 Sonified RSVP

One of the less explored properties of the RSVP format is its capability to present other media simultaneously with exact timing in respect to the text being read. The advantage of adding sound to visual text presentation, compared to using a text-to-speech engine generating audio output, is mainly retrieval speed. Listening to a text read aloud in English is comfortable at a rate of around 150 wpm [42], whereas silent self-paced reading is roughly twice as fast as that [25]. Adding sound to a text or a picture is generally considered not to impose any harmful effects. When watching a movie, sound is even considered to be a mandatory attribute and movies with sound are generally more popular than those without. Does the same line of reasoning apply for adding sound to the text being read?

3.1 Multimodal Aspects

According to Paivio's dual coding theory [33], information is processed through two generally independent channels. One channel processes verbal information (oral or written) while the other channel processes non-verbal communication (e.g. illustrations, pictures, melodies, or ambient sounds). Paivio stated "Human cognition is unique in that it has become specialized for dealing simultaneously with language and with nonverbal objects and events" [34, p. 53]. The use of both channels simultaneously to convey information is called referential processing and has been shown to have an additive effect on recall [34]. Students instructed with redundant multimedia generally performed better when the multimedia supports referential processing [31]. However, adding a redundant modality does not always show the expected beneficial effects. In fact, the addition of an exact audio replica to visually presented text instructions generated negative effects on learning due to excessive working memory load. Kalyuga [24] found that concurrent duplication (auditory explanation appearing simultaneously with the same visually presented text) of information using the different modes of presentation increased the risk of overloading working memory, resulting in a negative effect on learning. Based on these findings, Sonified RSVP [14, 16] has been introduced as a way of augmenting the visual communication channel of RSVP with non-verbal sounds.

3.2 Sounds Suitable for Sonified RSVP

Auditory icons is one form of non-verbal sounds that has been introduced as a way of creating intuitive links to objects or actions occurring within a interface by linking these to natural sounds. Gaver [11] identified three different types of auditory icons: Symbolic, metaphorical and nomic. Symbolic icons are based on social conventions, e.g., using applause for approval, while metaphorical icons use similarities to convey meaning. Nomic auditory icons are straight depictions of the information to be conveyed, e.g. using the sound of a paper being crumbled and being thrown into a trashcan when a file is dragged to the trashcan icon on the desktop.

An alternative to using auditory icons are Earcons, structured sounds introduced by Brewster et al. [3] as a substitute for graphical icons when navigating a hierarchy of nodes in a Graphical User Interface (GUI). Earcons are abstract, synthetic tones constructed from motives using timbre, register, intensity, pitch and rhythm [4]. Although limitations have been identified with all methods of using sound (Learning time for earcons and scalability for auditory icons) [18], nomic auditory icons had several attributes making them feasible for Sonified RSVP.

Nomic auditory icons can be directly linked to the content of the text in a one-to-one fashion and they require little or no training to recognize. Norman [32] stresses the importance of developing a conceptual model that the user understands, and with the one-to-one mapping offered by nomic auditory icons this is easily achieved. Finally, the auditory content of nomic auditory icons can be designed to augment the RSVP reading experience in a fashion similar to that of using sound effect when watching a sound movie and they can be synchronized to appear simultaneously with the RSVP text presentation at any selected reading speed.

3.3 Sonification of a Paper-Based Book

The idea of adding sound to text is not new but has until recently usually taken the form of adding simple sounds to images in children's books. The Listen Reader is a more serious attempt by Back et al. [1] to augment a book with sound. In a museum setting, a children's book was enhanced with ambient sound linked to different pages by electric field sensing, tactile interaction from the reader's hands were detected and used to trigger the sounds. The Listen Reader used a paper-based book in order not to lose the affordances of a real book; the resulting device became a quite large installation where the synchronization between the sounds and the text passage read was put in the user's control. Although many affordances of a real book were kept in the Listen Reader, one important property was lost due to its size: the ability to read a book wherever you want. The authors considered handheld devices to be too intrusive and attention demanding to use and stated that "…thus immersive reading is difficult if not impossible." to accomplish [1, p. 24].

4 BAILANDO: A Mobile Reader with Sonified RSVP

Bailando is a prototype incorporating Sonified RSVP capabilities that has been developed at Ericsson Research's Usability & Interaction Lab in Kista, Sweden [16, 39, 43, 44]. The prototype runs on a Compaq iPAQ 3630 Pocket PC, a small PDA with a touch sensitive high-resolution color display. The primary motivation for using a PDA compared to using a smaller handheld device, e.g. a cellular phone, was that it had the sound and memory capabilities necessary for developing and evaluating Sonified RSVP. Although the iPAQ actually offers far more screen space than is needed for the RSVP format it still shares many properties of smaller handheld devices and the size and weight of the device is similar to a pocket book of approximately 250 pages.

Since a lot of screen space was available all the application controls were implemented in the GUI (Fig. 2). This also makes the Bailando software easier to run on other PDAs, since button assignments differ between devices. The GUI contains flow control buttons to start, pause and to resume text presentation. The text presentation can also be paused and resumed by touching anywhere on the text presentation area of the screen. The user can move backwards in the presentation (<<) if he feels he missed or misunderstood some text and can skip text (>>) if browsing. Reading speed is decreased (-) or increased (+) with the speed control buttons in steps of 10 wpm.

1. Earphone jack
2. Text window
3. Text title
4. Progress bar
5. Speed settings
6. Application menus
7. Flow control buttons
8. Speed control buttons

Fig. 2. The Bailando prototype running in reading mode on a Compaq iPAQ 3630 Pocket PC

In order to support memory of spatial location while reading there is a progress bar that shows the amount of remaining text. Previous studies have shown that the user preference for RSVP increase with the inclusion of a completion meter [36]. The text presentation window has a width of 25 characters and the text was presented left justified in a 10-pt. sans-serif typeface. While presenting text the Bailando prototype tries to fit as many words as possible in each text window. The exposure time for each window was calculated using a content adaptive [16, 43, 44] formula (Eq. 1):

$$time_1 = (nwrd+nchr)/(davg*wpm/60)\ . \qquad (1)$$

The formula uses the number of words (nwrd) and the number of characters (nchr) as a basis for the results. Both arguments are added and divided by the product of the average word length including delimiters (davg) and the currently set speed in words per minute (wpm) divided by 60. The result is a variable exposure time ($time_1$) depending on the content the current text chunk. A delay of 250 ms was added at sentence boundaries and an extra blank window was shown for 250 ms when the end of a sentence completely filled the text presentation window.

The Bailando prototype supports Sonified RSVP by using tags in the XML (eXtensible Markup Language) document format. When a <SOUND…> tag is found in the text chunk to be displayed, the appropriate audio file linked to it is played. Thus, exact synchronization between text and sound is easily attainable and works at any user-selected reading speed:

```
At a high speed a car speeded down the street, as
<SOUND SRC="HONK.WAV"> it honked it's horn all </SOUND>
the children fled to the curb.
```

When a </SOUND> tag is found, all playing sounds are stopped. By using a markup language it is not harder to add sounds to a text than adding images to a web page. Since the file size of sounds tends to become quite large and the memory resources of PDAs tend to be small the Bailando prototype has support for external storage devices. By employing the standard Uniform Source Locator (URL) format it is also possible to retrieve sounds from other locations via wired or wireless connections.

5 The Usability Evaluation

The aim with the evaluation was to see how the addition of sound affected the reading experience when using RSVP on a mobile device. It was important that the same device and software was used for all conditions since the look and feel was likely to bias the assessment. First we will present the method of the experiment and then we will present the findings.

5.1 Method

Sonified RSVP was benchmarked against Unsonified RSVP in a usability evaluation. The following null hypotheses were set for reading with and without sound:

- No difference in Reading speed
- No difference in Comprehension
- No difference in Task load
- No difference in Attitude

The hypotheses were tested in the SPSS V10.0 software using the repeated-measurement General Linear Model (GLM). The significance level was set to 5% and the level of multiple comparisons was Bonferroni adjusted.

Design. A repeated-measurement within-subject design was adopted where each subject participated in both a Sonified and an Unsonified RSVP condition. Two different texts were used and each could either appear as Sonified or Unsonified according to a balanced design, thus creating four different presentation order combinations. All subjects were randomly assigned to one of the four presentation orders in groups of three.

Subjects. Twelve computer literate students varying in age between 20-33 years (M=26 years) from the Department of Computer Science at the University of Stockholm were enrolled with the following selection criteria: Swedish as a native language, experience of using a PDA and a keen interest in reading fiction. Ten were males and two were females, all read an average of one 'paper' book of fiction/month, or more. Two had a slight hearing impairment in one ear and five wore glasses during the experiment. Seven had experience of listening to audio books during childhood; one had also listened to audio books as an adult. None had any prior knowledge of the RSVP paradigm and all but one was familiar with reading longer texts from a large screen.

Apparatus. All texts were presented on a Compaq iPAQ 3630 running the Bailando prototype in content adaptive RSVP mode. In the Sonified condition, earphones were plugged to the audio output of the PDA.

Texts. Two Swedish texts with similar length and readability rating were chosen (Table 1), "Behind rose-red laces" (Text A) and "Markurells in Wadköping" (Text B). The readability ratings were measured with LIX [2], a readability rating developed for Swedish texts comparable to the Flesh index for English texts [40]. One additional text, "The red room" (Text T), was used as a training text.

Table 1. Text material used in the experiment

Text	*Title (in Swedish)*	*Author, chapter*	*Words*	*LIX*	*Sounds*
A	Bakom rosenröda snören	Johan Wahlborg, chapter 1	4040	40	20
B	Markurells i Wadköping	Hjalmar Bergman, chapter 1	3733	39	24
T	Röda rummet	August Strindberg, chapter 1	1702	48	14

Sounds. For the Sonified conditions a number of different nomic auditory icons (14 to 24 sounds) of varying duration were added to the texts (Table 1). Each nomic auditory icon fitted a certain passage in each text (blowing wind, siren, restaurant background sound, birds twittering, church bells ringing, etc.). The duration of the nomic auditory icons were tailored to fit a lowest reading speed of 100 wpm and a highest reading speed of 600 wpm; each subject selected a volume level they felt comfortable with.

Instructions. Each subject was given the following instruction on paper: You will be presented with a new way of reading text on a small PDA screen. When adjusting the reading pace, try to set a pace that is as *comfortable* as possible according to your preferences. You will be given a multiple-choice comprehension, a workload and an attitude inventory after each read text.

Setting. The experiment took place in a dedicated usability lab outfitted with audio and video-recording facilities (Fig. 3, next page). While reading the subject was seated in a comfortable chair in a room separated from the experimenter by a one-way mirror. Before the experiment started each subject had some time to get acquainted with the facilities in order to create a relaxed, and consequently controlled, setting.

Fig. 3. Interior of the usability lab with the observation room (left) and the test room (right)

Training. Each subject participated in two training sessions where they read the same training text twice (Text T, Table 1), first without and then with sound. The reason for reading the same text twice was to give each subject an early success experience and making them more willing to experiment with the user interface. The experimenter first set the reading speed to a low value (app. 100-wpm) and encouraged the subject to read at this pace. Then the reading pace was set to a much higher value (app. 300-wpm) and the subject was encouraged to lower the reading pace until reaching a *comfortable* reading pace. All subjects were instructed that they could alter the reading pace at their own will whenever desired. After the completion of the training sessions, a comprehension inventory of 10 multiple-choice questions with three alternatives was administered. Then each subject had a 20-minute break before the first experimental condition started.

Procedure. Each subject was exposed to one of the texts (Text A or B in Table 1) presented as either Unsonified or Sonified, the subject selected the initial reading speed. After having read the first text, the subject answered a set of inventories. If the first text was read in Unsonified mode, the second text was read in Sonified mode, and vice versa. The same set of inventories was administered after the second text.

Inventories. After each experimental condition three inventories were administered. The first inventory was a comprehension test made up of 10 multiple-choice questions with four-alternatives. The second inventory was the NASA-TLX Task Load Index [17] designed to check Mental, Physical, and Temporal demands, as well as Performance, Effort and Frustration levels. This measure of workload was chosen since the results would then be comparable to previous evaluations where the measure was rewardingly used [15, 44]. Finally, an attitude inventory was administered. It contained seven questions regarding the reading: Experience, Excitement, Comfort, Stimulation, Immersion, Understanding and Speed, using a 10-graded (1 to 10) discrete ordinal scale with two verbal bi-polar anchor points. The questions were formulated in the following form: “How did you perceive the Experience while reading?” (Irritating-Fantastic).

5.2 Results

All subjects completed the experiment and there were few problems with understanding what to do or how to do it. The presentation of the results is divided into four sections: Reading speed, Comprehension, Task load and Attitude. Under each section the null hypotheses regarding no difference is tested.

Reading Speed. Reading speed was calculated as words read per minute based on the *total* time it took for the subjects to read a text including all kind of interruptions like pauses, regressions, speed changes etc. The null hypothesis regarding no difference in mean reading speed between Unsonified (M=215 wpm) and Sonified RSVP (M=218 wpm) was kept. The number of speed changes amounted to an average of 2 per read text.

Comprehension. Comprehension was computed as percent correctly answered multiple-choice questions out of 10. The null hypothesis regarding no difference in mean comprehension between Unsonified (M=57.5 %) and Sonified (M=59.2 %) RSVP was kept.

Task Load. Task load ratings were calculated as percent of millimeters to the left of the tick mark on a 120-mm scale. The null hypothesis regarding no difference in workload between Unsonified and Sonified RSVP was kept since none of the factors turned out to be significant. The Performance ratings reflected the objective metrics of reading speed and comprehension (Fig. 4).

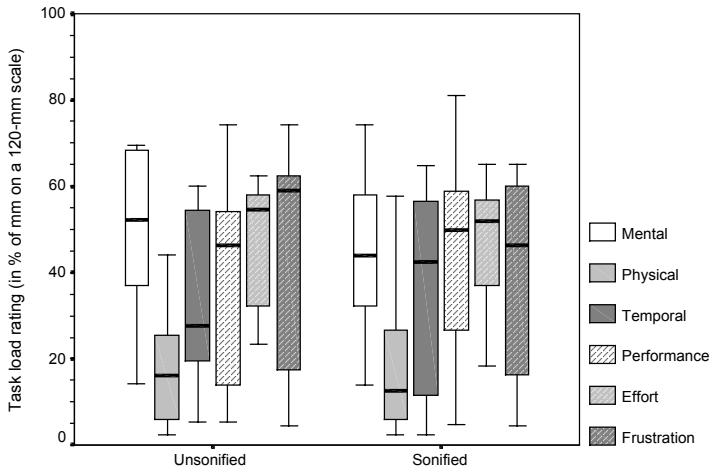

Fig. 4. Box plot of NASA-TLX Task Load Index ratings with median, 25-, 75-percentile and min/max values represented. Unsonfied ratings are clustered to the left and the Sonified ratings are clustered to the right, lower ratings are better

Attitude. Attitude ratings were calculated as mean average rating on the 10-point scale. The null hypothesis regarding no difference in attitude between Unsonified and Sonified RSVP was partly rejected since the rating of perceived Immersion showed significance in favor of Sonified RSVP ($F[1, 11]=11.88$, $p=0.005$). Perceived reading speed and comprehension reflected the objective metrics of reading speed and comprehension (Fig. 5).

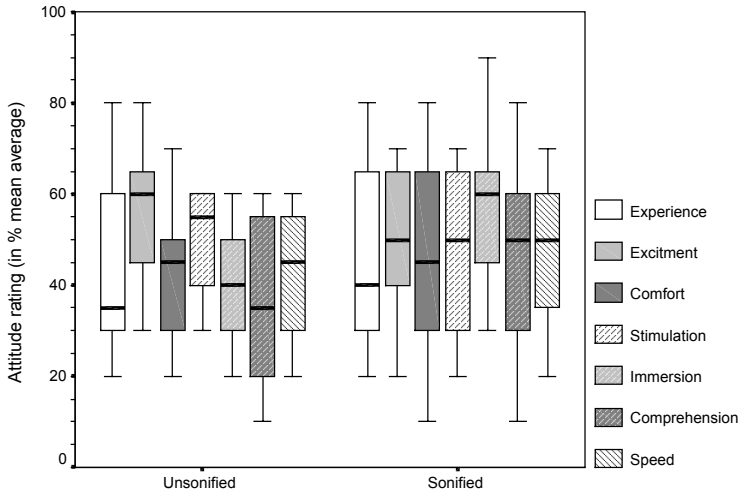

Fig. 5. Box plot of Attitude ratings with median, 25-, 75-percentile and min/max values represented. Unsonfied ratings are clustered to the left and Sonified ratings are clustered to the right, higher ratings are better

6 Discussion

The experimental results indicated no difference in reading efficiency in terms of reading speed, comprehension and task load between Unsonified and Sonified RSVP. However, since the null hypothesis regarding no difference in attitude was partly rejected there was a difference in the reading experience. The discussion will first be based on the obtained results and then we will discuss if the addition of sound is a plausible enhancement of the reading experience on mobile devices.

6.1 Experimental Results

That the addition of non-verbal sound to RSVP in the form of nomic auditory icons did not cause any difference in reading efficiency seems to support the dual coding theory [33]. In any case, the addition of nomic auditory icons did not seem to affect reading efficiency in any negative way.

Reading Speed. The obtained reading speed was close to what Rahman and Muter found [36], but considerably slower than what Goldstein et al. [15] reported. The difference amounted to around 100 wpm (217 vs. 320 wpm). The difference may be attributable to priming; the subjects were instructed to optimize for speed in the Goldstein et al. experiment and to select a comfortable reading speed in this experiment. Further, the Bailando prototype gave the user control over the presentation, which has been shown to yield lower reading speed [7, 29]. The average reading speed on paper for an adult Swedish reader is around 240 wpm [2] and in this light the lower reading speed is not discouraging, as the motivation for the prototype was to facilitate comfortable and user-controlled reading on small displays rather than improving reading speed in general.

Comprehension. Given the simultaneous information conveyance of Sonified RSVP, referential processing [34] implies that comprehension would be greater for the condition adding the non-verbal channel using the aural medium. This was not the case in the experiment and a probable cause might be that the questions that made up the comprehension questionnaire did not relate directly to the sounds, as sound was used to enhance the emotional or social setting rather than act as an informative voice-over.

Task Load. Surprisingly, the NASA-TLX ratings remained high comparable to the Goldstein et al. [15] experiment even though reading speed was substantially lower due to the fact that the subjects were instructed to optimize for comfortable reading. This may imply that the size of the assumed trade-off between reading speed and cognitive task load is negligible for the RSVP format, quite contrary to the size of the well-established speed-accuracy trade-off [41]. An increase in task load is probably inherent to the RSVP format, but more elaborate adaptation models [43] and interaction techniques [45] may decrease task load even further. An increase in task load may even be acceptable if it is compensated with an increase in reading efficiency, especially in the mobile context where time often equals money [44].

Attitude. The only significant difference in the attitude inventory was the rating of Immersion, where Sonified RSVP was rated higher. Although one should always be cautious when relying on subjective measures, this finding seems to suggest that an immersive reading experience is possible to achieve on a handheld device and not impossible as claimed by Back et al. [1]. However, it seems a little surprising that there existed no positive co-dependence between Immersion and the ratings of Stimulation, Excitement and Experience. If Immersion increased significantly these ratings were somehow expected to increase in a similar fashion as well. Therefore, adding nomic auditory icons to text using the RSVP format does not seem to enhance the reading experience in a clear-cut fashion. Since the only significant difference in this evaluation was subjective, it would have been interesting to have more qualitative results to base the discussion on, e.g. user comments, but unfortunately such were not collected in this evaluation.

6.2 Can Sound Enhance the Reading Experience on Mobile Devices?

Although individual and cultural differences in the assessment of what is beneficial for the reading experience are likely to be large, and the experiment conducted here is quite small and limited to RSVP, the answer to the question seems to be yes. It is promising that even simple sonification with nomic auditory icons increased the feeling of immersion while reading using RSVP on a handheld device. The lack of co-dependence with the ratings of Stimulation, Excitement and Experience may be an artifact from the experimental design: the selection of sounds to match the text, the choice of text and target user group, and how the questionnaire was written. However, one can identify at least one other reason for the outcome: the difference between what we chose to call *realistic* and *dramatic* soundscapes.

A realistic soundscape enhances the reading experience by adding information that brings more detail to the description of the physical environment, e.g. playing the sound of a seagull when events in the text take place at a harbor. By giving more detail to the place described in the text, one can assume that users have a greater possibility to feel immersed in the story. In contrast, a dramatic soundscape increases the reading experience by providing additional information about characters internal emotions and reactions or by creating dramatic tension. Gaver [12] mentions the music played in the motion picture Jaws, before the first witnessed shark attack, as an example of such emotion provoking sound.

Thus, adding a dramatic soundscape to a text can evoke emotional responses such as excitement from readers without trying to place the reader within the story. The soundscape used in the experiment was realistic, primarily motivated by the fact that a compelling dramatic soundscape would require a professional sound producer, but also since a simple realistic soundscape was considered less likely to confuse the users conceptual model [32]. Although a dramatic soundscape is usually created by playing continuous music with durations longer than that of certain typical transient nomic auditory icons, one can imagine a dramatic soundscape created only by using auditory icons. One hypothesis is that a dramatic soundscape used for Sonified RSVP would yield higher ratings of Stimulation, Excitement and Experience as well.

Gaver [12] claimed that memory limitations is one reason why sounds are not generally used in interfaces on a larger scale, but adding elaborate soundscapes to texts presented on mobile devices is, at least from a technical perspective, likely to be a viable enterprise quite soon. Earphones are already widely used as accessories to cellular phones (as hands-free attachments) and most mobile devices will eventually have the required capabilities for high-quality sound playback. By defining a standardized markup language for sonification of RSVP it may be possible for publishers of e-books to add sound to texts of various genres. Real-time streaming of sounds over wireless networks may further reduce the size of a reader and cut costs.

On the other hand, although it may be a technically feasible to create applications that use Sonified RSVP, the question remains if this is something that the users want. We believe that, at the end of the day, it is up to the users to decide which presentation formats they want to use and which enhancements they deem appealing. RSVP has been found to have potential for improving readability on small screens, but if it is the most appropriate presentation format to use remains to be answered. Although

the addition of nomic auditory icons did not enhance the reading experience in a clear-cut fashion in this experiment, we have at least shown that Sonified RSVP can increase the feeling of immersion while reading on a mobile device. Based on the findings from this evaluation, we think that further exploration of sound-enhanced text presentation formats may bring forward novel and interesting applications that a wide range of users would find useful and benefit of.

7 Conclusions

We have shown that the addition of sound to text displayed using the RSVP format can improve some aspects of the reading experience on a small screen. More specifically we found that the rating of perceived Immersion was significantly higher when nomic auditory icons were played simultaneously with the text presentation. The addition of a more elaborate soundtrack to strengthen the dramatic presence of the text has been suggested as a possibility to further enhance the reading experience on small screens. The most important finding in this evaluation is however that it actually is possible to create a more immersive reading experience on a small screen by adding sound to the text presentation.

8 Further Work

Bailando is the first prototype for reading with Sonified RSVP. Currently the sound playback capabilities are quite limited and several improvements can be made that may improve the user experience, e.g. the ability to play several sounds simultaneously, to fade in and out, and to repeat sounds perpetually. A follow-up experiment using a dramatic soundscape instead of a realistic one could determine if the findings from the current experiment depend on the design of the soundscape or other factors. It would also be interesting to see if other dynamic text presentation formats could benefit from the addition of sound in a similar way that RSVP did, and last but not least, it would be interesting to collect more qualitative data on how users experience the addition of sound while reading on a small screen.

Acknowledgements

The authors would like to thank Master's Thesis student Mandana Bayat-M for implementing the nomic soundscapes and conducting the experiment. We would also like to thank Peter Ljungstrand at the Interactive Institute for rewarding discussions around the concept of Sonified RSVP.

References

1. Back, M., Cohen, J., Gold, R., Harrison, S. and Minneman, S. (2001). Listen reader: an electronically augmented paper-based book. *Proceedings of the SIGCHI on Human Factors in Computing Systems 2001*, 23-29, Seattle, WA, USA, ACM Press.
2. Björnsson, C.H. (1968). *Läsbarhet*, Stockholm, Liber (in Swedish).
3. Brewster, S., Raty V.P., and Kortekangas, A. (1996). Earcons as a method of providing navigational cues in a menu hierarchy. *Proceeding of Human-Computer Interaction'96* by M.A. Sasse, R.J. Cunningham and R.L. Winder (Eds.), People and Computers.
4. Brewster, S. (1998). Using non speech sounds to provide navigation cues. *ACM Transactions on Computer Human Interaction,* 5(3), 224-259.
5. Bruijn, O. and Spence, R. (2000). Rapid Serial Visual Presentation: A space-time trade-off in information presentation. *In Proceedings of Advanced Visual Interfaces*, AVI'2000.
6. Castelhano, M.S. and Muter, P. (2001). Optimizing the reading of electronic text using rapid serial visual presentation. *Behaviour & Information Technology*, 20(4), 237-247.
7. Chen, H.C. and Chan, K.T. (1990). Reading computer-displayed moving text with and without self-control over the display rate. *Behaviour & Information Technology,* 9, 467-477.
8. 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.
9. Ericsson, T., Chincholle, D. and Goldstein, M. (2001). Both the device and the service influence WAP usability, *IHM-HCI2001,* Volume II, Usability in Practice by J. Vanderdonckt, A. Blandford and A. Derycke (Eds.), Short paper, 10-14 September, Lille, France, 79-85.
10. Forster, K.I. (1970). Visual perception of rapidly presented word sequences of varying complexity. *Percep. Psychophys. 8,* 215-221.
11. Gaver, W. (1986). Auditory icons: Using sound in computer interfaces. *Human-Computer Interaction,* 2, 167-177, Lawrence Erlbaum Associates, Inc.
12. Gaver, W. (1998). Auditory interfaces. In *Handbook of Human-Computer Interaction*, 2nd ed., 1003-1041.
13. Gilbert, L.C. (1959). Speed of processing verbal stimuli and its relation to reading. *J Educ Psychol,* 58, 8-14.
14. Goldstein, M., Anneroth, M., Sicheritz, K. and Lundberg, T. (2000). Enhancing the reading experience: Sonified on-line reading from a small PDA screen using RSVP. *Fourth Swedish Symposium on Multimodal Communication,* SsoMC.
15. Goldstein, M., Sicheritz, K. and Anneroth, M. (2001). Reading from a small display using the RSVP technique. *Nordic Radio Symposium,* NRS01.

16. Goldstein, M., Öquist, G., Bayat-M, M., Ljungstrand, P. and Björk, S. (2001). Enhancing the reading experience: Using adaptive and sonified RSVP for reading on small displays. *Proceedings of Workshop on Mobile Devices* at IHM-HCI 2001, Lille, September 10-14.
17. Hart, S.G. and Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human Mental Workload,* by P.A. Hancock and N. Meshkati (Eds.). Elsevier Science Publishers, B.V. North-Holland.
18. Hearst, M.A. (1997). Dissonance on audio interfaces. *IEEE Expert,* 12, 5, 10-16, IEEE Computer Press.
19. Huey, E.B. (1968). *The Psychology and Pedagogy of Reading.* Massachusetts: MIT Press. (Originally published 1908).
20. 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.
21. Juola, J.F., Tiritoglu, A., and Pleunis, J. (1995). Reading text presented on a small display. *Applied Ergonomics*, 26, 227-229.
22. Just, M. A., and Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review,* 87(4), 329-354.
23. Kang, T.J., and Muter, P. (1989). Reading Dynamically Displayed Text. *Behaviour & Information Technology*, 8(1), 33-42.
24. Kalyuga, S. (2000). When using sound with a text or picture is not beneficial for learning. *Australian Journal of Educational Technology,* 16(2), 161-172.
25. Kump, P. (1999). *Break-trough rapid reading.* New Jersey: Prentice-Hall Press.
26. Masson, M.E.J. (1983). Conceptual processing of text during skimming and rapid sequential reading. *Memory & Cognition,* 11(3), 262-274.
27. Mills, C.B. and Weldon, L.J. (1987). Reading text from computer screens. *ACM Computing Surveys,* 19(4), ACM Press.
28. Muter, P. (1996). Interface Design and Optimization of Reading of Continuous Text. In *Cognitive aspects of electronic text processing.* H. van Oostendorp and S. de Mul (Eds.). Norwood, N.J.:Ablex.
29. Muter, P., Kruk, R. S., Buttigieg, M. A., and Kang, T. J. (1988). Reader controlled computerized presentation of text. *Human Factors*, 30, 473-486.
30. Muter, P. and Maurutto, P. (1991). Reading and skimming from computer screens and books: The paperless office revisited? *Behavior & Information Technology*, 10, 257–266.
31. Najjar, L.J. (1996). Multimedia information and learning. *Journal of Educational Multimedia and Hypermedia,* 5, 129-150.
32. Norman, D.A. (1988). *The Psychology of Everyday Things.* Doubleday, NY, USA.

33. Paivio, A. (1971). *Imagery and Verbal Processes.* New York: Holt, Rinehart & Winston, Inc.

34. Paivio, A. (1986). *Mental Representations*. New York: Oxford University Press.

35. Potter, M.C. (1984). Rapid Serial Visual Presentation (RSVP): A method for studying language processing. In *New Methods in Reading Comprehension Research.* Kieras, D.E. and Just, M.A. (Eds.). Hillsdale, N.J.:Erlbaum.

36. 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.

37. Robeck, M.C. and Wallace, R.R. (1990). *The Psychology of Reading: An Interdisciplinary Approach,* Lawrence Erlbaum Associates, Hillsdale, New Jersey, 2nd edition.

38. Rubin, G. S. and K. Turano. (1992). Reading without saccadic eye movements. *Vision Research*, 32(5), 895-902.

39. The Bailando prototype web page. A tool for research on small screen readability, available at: http://stp.ling.uu.se/~gustav/bailando/ (June 2002).

40. Tekfi, C. (1987). Readability formulas: An overview. *Journal of Documentation,* 43(3), 261-273.

41. Wickens, C. D. (1992). *Engineering psychology and human performance,* 2nd edition, Chapter 8, HarperCollins Publishers Inc., New York.

42. Williams, J. R. (1998). Guidelines for the use of multimedia in instruction. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, 1447-1451.

43. Öquist, G. (2001*). Adaptive Rapid Serial Visual Presentation.* Master's Thesis, Department of Linguistics, Uppsala University, Sweden.

44. Öquist, G. and Goldstein, M. (2002). Towards an improved readability on mobile devices: Evaluating Adaptive Rapid Serial Visual Presentation. *Proceedings of Mobile HCI'2002*, 19-20 September 2002, Pisa, Italy, Springer Verlag.

45. Öquist, G., Björk, S. and Goldstein, M. (2002). Utilizing gaze detection to simulate the affordances of paper in the Rapid Serial Visual Presentation format. *Proceedings of Mobile HCI'2002*, 19-20 September 2002, Pisa, Italy, Springer Verlag.