Master's Thesis

Adaptive Rapid Serial Visual Presentation

Gustav Öquist
gustav@stp.ling.uu.se

Language Engineering Programme
Department of Linguistics
Uppsala University
December 2001

Supervisors: Anna Sågvall Hein, Uppsala University
Mikael Goldstein, Ericsson Research

Abstract

Can readability on small screens be improved by using Rapid Serial Visual Presentation (RSVP) that adapts the presentation speed to characteristics found in the text? In this thesis Adaptive RSVP is introduced together with two algorithms for adaptation. Findings from a usability evaluation, where the ability to read long and short texts on a mobile device in a realistic setting was assessed, are reported. In a balanced repeated-measurement experiment employing 16 subjects two variants of Adaptive RSVP were benchmarked against Fixed RSVP and traditional text presentation. No significant differences were found for reading speed and comprehension for long texts but for short texts all RSVP formats increased reading speed by 33%. For long texts Adaptive RSVP decreased task load compared to Fixed RSVP and repeated use of RSVP, regardless of type, was also found to decrease task load. No differences in task load were however found for short texts. Causes, implications and effects of these findings are discussed together with directions for further work.

Acknowledgements

The author would foremost like to thank Mikael Goldstein for inspiration, supervision and a lot of support during the work with this thesis. Many thanks also to Peter Ljungstrand and Staffan Björk at the Interactive Institute for valuable collaboration. The author would also like to express gratitude to all colleagues at Ericsson Research's Usability & Interaction Lab in Stockholm and Sony Ericsson Mobile Communications in Lund. Much credit is also awarded all of those who participated in the usability evaluation. Last but not least, the author would like to put across many thanks to family and friends for love and support.

Table of Contents

List of Figures

List of Tables

List of Equations

1 Introduction

The mobile Internet has been widely predicted to cause a revolution in communications over the next few years and when the time is ripe it is likely to become an integral part of our everyday lives. However, before the revolution can start there must be devices, services and applications available that people really want to use. Enabling these new technologies, bringing them to market and making desirable content available to them is the challenge facing companies involved in the wireless business today.

Presently the limited screen space on most mobile devices is a bottleneck for efficient and usable retrieval of information (Ericsson et al. 2001). Since it is the customers demand for small devices that sets the limit this constraint is also likely to remain tomorrow. Combined with the expected increase in texts to be read on mobile devices due to the advent of the mobile Internet this quandary has made the issues concerning readability on small screens progressively more important.

Early research on screen reading showed that reading speed decreased by 20-30% when reading on large screens compared to reading on paper (Mills and Weldon 1987). With time, as screen resolution improved and people got more used to them, readability on large screens became more or less equal to paper (Muter and Mauretto 1991). The evolution in readability on small screens is however not likely to follow the same pattern. The resolution will surely get better and thus improve the texts legibility but decreased readability will still be intrinsic to limited screen space (Duchnicky and Kolers 1983). Readability may however still be increased by designing interfaces that display the text in a way more suitable for small screens.

Focus has thus shifted from how screens display static texts to how texts can be dynamically displayed on the screen. Leading and Rapid Serial Visual Presentation (RSVP) are the two major techniques that has been proposed for dynamic text presentation (Mills and Weldon 1987). 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. Both formats offer a way of reading texts on a very limited screen space (Mills and Weldon 1987; Kang and Muter 1989; Joula et al. 1995; Rahman and Muter 1999; de Bruijn and Spence 2000; McCrickard et al. 2001).

In this thesis a potential enhancement of the RSVP format is described. The text output is supposed to become more predictive to the reader by letting the presentation adapt to characteristics found in the text. It is important to explore the possibilities of dynamic text presentation since improved readability on small screens also means improved usability of mobile devices.

> "There's nothing more exciting than what software can do to help users and we see that what we've done so far is really just scratching the surface.“
>
> - Bill Gates, Microsoft, MVP Summit, November 30, 2001.

1.1 The Assignment

The aim with the thesis work was to twofold. Primarily it was to learn whether the RSVP format could be improved by letting the text presentation speed adapt to characteristics found in the text. Secondly, it was to see if the RSVP format could improve readability on mobile devices. In order to reach these objectives a few subsequent tasks had to be performed.

To start with the adaptive RSVP format had to be developed, it already existed as an idea but it had to be put into practice. After that adaptation had to be integrated into an application for RSVP on a handheld device. Finally, the prototype had to be benchmarked against other presentation formats in a usability evaluation. The work with the assignment can thus be roughly divided into the following three tasks:

Design	Find out how adaptation could and should work on basis of previous studies and evaluations.
Implement	Construct and test an application for reading texts on a handheld device that incorporates adaptive RSVP features.
Evaluate	Assess the ability to read using RSVP and other text presentation formats by benchmarking the prototype against other applications in a usability evaluation.

1.2 Outline of this Thesis

Substantial parts of this thesis were originally written to be included in three articles reporting progress and results from the work with this thesis (Goldstein et al. 2001; Goldstein, Öquist and Björk 2001; Öquist and Goldstein 2001). The thesis will keep the inherent structure of the work done during the assignment, with a few additions.

Background	This thesis is all about reading text on small screens and this part will give an overview of the essentials about reading and readability.
Design	The RSVP and adaptive RSVP formats are formally put down in this part together with a novel way of visualizing this form of dynamic text presentations.
Implementation	The implementation of the prototype is presented in this part together with screen shots, design decisions and technical specifications.
Evaluation	In this part the method and design of the usability evaluation is described together with an analysis of the obtained data.
Discussion	On basis of the results from the evaluation the feasibility of using adaptive RSVP for reading on small screens is discussed in this part.
Conclusions	This part summarises the thesis and put forward some insights reached on basis of the experimental results together with some directions for future work.

2 Background

By reading we examine and grasp the meaning of painted, written or printed symbols. As far as we know, pictures were drawn to represent animals and other objects as early as the Stone Age, around 20,000 B.C. The remnants of the oldest systematic reading and writing systems are dated back to North Babylonia, around 8,000 years ago. Alphabet signs, as we now interpret them, were used in Egypt at least 7,000 years ago (Huey 1908; Hill 1999). Reading can essentially be seen as a very intricate form of pattern recognition that has evolved within mankind for centuries. As with pattern recognition in general, the interpretation is highly dependant on convention. The development of language and writing is very much a result of agreeing on what we choose to associate with certain patterns and how we present them.

The traditional page format for presentation of text has been fashioned over thousands of years into its current form. Yet, this thesis puts forward a novel and maybe also unconventional form of text presentation. One question then immediately arises. Why replace something that works perfectly well? The answer to this is that new appliances add new constraints. Today there is an unprecedented need for reading texts on small screens. It is quite obvious that traditional text presentation in the page format requires far more screen space than the most mobile devices have to offer. This is the foremost reason for the development of the dynamic text presentation formats presented in this thesis. If traditional text presentation does not work properly then it might be possible to improve readability by trying some new approaches. However, also new presentation formats must adhere to the principles for reading that has evolved over time. The background part of this thesis begins in section 2.1 with a primer on some of the most important issues regarding the reading process.

The first objective of this thesis was to see if adaptation could improve readability. If we are to say that one text presentation format is better than another, or that one text is easier to read than another, we need to be able to measure readability. In section 2.2 the concept of readability is defined and a few different measures of readability are introduced. The second objective was to learn whether the ability to read on small screens could be improved by using adaptive RSVP. It has already been implied that there is problem with readability on small screens. Section 2.3 focuses on how screen reading and limitations in screen space affects readability. In section 2.4 a few different text presentation formats that can be used for small screens are presented and the background part of this thesis is concluded in section 2.5 by a description of previous RSVP evaluations and implementations.

2.1 The Reading Process

It is important to acknowledge that the key to reading, and language in general, lies deeply embedded within in our mind (Taylor and Taylor 1983; Hill 1999). There has been a lot of research on how we read and researchers have presented several different models of how the reading process works. Some are very detailed whereas others are more generalizing (Reichle et al. 2000). Most agree that we recognize patterns and then mentally process them in some way. Unfortunately, little is known for sure about the processing part and there are many disputes around these issues. However, since we do not known much about the processing, a better starting point for understanding how we read might be in the other end. By observing how the eyes move while reading we can tell how the recognition part works, if we do that we might also learn some about how the processing works as well.

The first evidence of what really happens with the eyes while reading was discovered in the end of 19^{th} century. An overview of the revelation of the reading process follows next. The structure of it is partially adopted from an excellent review over influential eye movement studies by Paulson and Goodman (2000).

2.1.1 The Revelation of Eye Movements

In 1879, the French oculist Emile Javal found that the eye does not sweep smoothly along the line while reading. Instead it proceeds by making short jumps called *saccades*. Between the saccades the eye stays put for a brief time in pauses called *fixations*. Upon the end of a line the eye moves to the beginning of the next line in a single movement called a *return sweep* (reported in Huey 1908 referred to in Paulson and Goodman 2000). The findings of Javal are what triggered the following decades of research in eye movements while reading.

In 1891, Landholt, one of Javal´s contemporaries, discovered that "reading of a foreign language required more pauses, as did also the reading of detached words, numbers and list of proper names" (reported in Huey 1908:19 referred to in Paulson and Goodman 2000). Landholt´s findings are probably the first evidence of that the reading process is not constant, but varying depending on the type of text being read. The Landholt study was also important since it was the first to imply that eye movement studies might tell us something about the cognitive processing that takes place while reading.

In 1900, Dodge reported that the eye does not retrieve any information at all while moving. The experiment had a very simple design, but nonetheless the proof was very convincing. Dodge used a cardboard piece with a slit of 4 mm and behind it he put different coloured cards. The subjects were told to fixate on one side of the slit and then move to the other side in one unbroken eye movement. Dodge found that the subjects could not tell what colour that had been exposed or even if there had been no slit at all (Dodge 1900 referred to in Paulson and Goodman 2000). The experiment showed that it is during the fixations that the reader actually processes information.

In 1908, Edmund Burke Huey provided the first physical records of eye movements while reading. In the records he made a number of interesting findings. He found that the eye sometimes moves backwards to reread words and phrases in movements called *regressions*. He also found that only 20-70% of the words in a line are fixated. Further, the first fixation on a line was not found to be on the first but rather on the second or third word (Huey 1908 referred to in Paulson and Goodman 2000). These findings provided the first evidence of a reading process where the reader chooses where and when to fixate next. The results reached by Huey might also have been the first indication of that reading is not just a simple word identification process, but rather a process where words are processed simultaneously as chunks.

In 1922, Judd and Buswell reported results from the first study where the subject's eye movements had been photographed. The detail level was high and the plates provided accurate records of eye movements and fixation durations. In the data Judd and Buswell found evidence for that readers read differently under different circumstances. They also concluded that reading is not simply a matter of bottom-up word identification but rather a perceptual process that involves interpretations on the reader's part (Judd and Buswell 1922 referred to in Paulson and Goodman 2000).

From the early findings we know that a reader cannot perceive anything while moving the eyes and does not fixate on each word. We also suspect that some kind of processing takes place during the fixations. The *perceptual span* of each fixation is thus decisive to how much information that can be processed at a time. Physiological studies of the eye offer some useful answers on the physical constraints for the reading process.

2.1.2 Physiology of the Eye

The receptive part of the eye, called the *retina*, is essentially a panel full of photosensitive receptors located on the back of the eyeball (Ø ~42 mm) (figure 1). The retina has two types of receptors, *cones* and *rods*. Cones register luminosity and colours whereas rods register light changes. Rods are much more sensitive to light but they cannot detect colors and are also slower to respond. Most of the cones are located in a tiny area at the centre of the retina called the *fovea* (Ø ~0,2 mm). The fovea is surrounded by the *parafovea* (Ø ~3 mm); in this region there are still many cones, but also an increasing amount of rods. Outside the parafovea there are few cones and a decreasing amount of rods, therefore vision becomes progressively less clear in the periphery of the retina (Procter and Procter 1997).

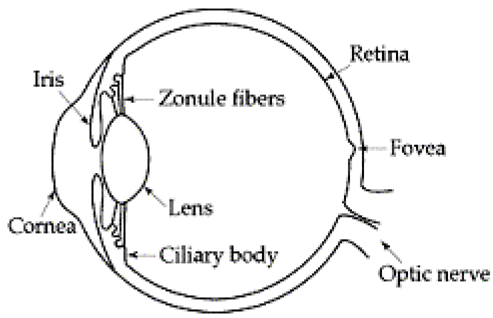


Figure 1. Cross-section of the eye.

When reading a text, the image of the text is inversely reflected upon the retina. The retina has a 240-degree field of vision but the maximum resolution is restricted to the fovea. The fixation target must be located in the fovea since a high concentration of cones is required for accurate recognition. The foveal field of vision is only one or two degrees wide and this means that only six to eight characters can be in focus at a time. The parafoveal region further extends the perceptual span to approximately 20 characters but beyond that acuicity is too low for retrieval. The perceptual span is centered to the right of the fixation point, at least for readers of left-to-right languages (Just and Carpenter 1980). Readers pick up information from approximately eight or nine character spaces to the right of a fixation, and four or so to the left (Rayner and Pollatsek 1989; Robeck and Wallace 1990; Rayner and Serano 1994; Rayner 1998).

After information is processed in a fixation, peripheral vision is used to determine the location of the next fixation. A saccade, or a return sweep is executed to move to the next fixation target, it must not necessarily be a forward movement. Regressions are essentially backward saccades that are used for clarification of incomplete retrieval and they appear about 19% of the time (Just and Carpenter 1980). The length of a saccade is usually between 1-20 characters and they are performed very quickly ~40 ms. Fixations take a little longer time, ~230 ms for fast readers and ~330 ms for average readers (Robeck and Wallace 1990). The duration of the fixations has also been found to vary a lot. In some studies it has ranged between 100-500 ms (Rayner 1998) whereas it in others has been found to vary between as much as 50-1500 ms (Just and Carpenter 1980).

The planning of saccades and the use of regressions for clarification seems to indicate that there is more to reading than meets the eye. The large differences observed in saccade lengths and fixation durations appear to reflect an ongoing process that changes depending on what is being read. The knowledge of what is know about the physiology of the eyes and their movements while reading seems to suggest that perception and recognition is highly dependent on cognitive (i.e. linguistic) processing.

2.1.3 Eye and Mind

Landholt and others were probably quite early to assume a strong bond between the eye and the mind. Unfortunately, research in the field of eye movements and reading came to an abrupt halt around the 1930's. Nothing significantly actually happened until the late 1970's. The reason for the renewed interest was probably that computers then became available. Eye-tracking equipment usually generates huge amounts of data and without computers it is time consuming to analyse the results. Another probable cause for the renewed research efforts might have been the appealing prospect of creating computational models of the reading process. In any case, a few models of the reading process have at least evolved since then. They can be roughly divided into the following two categories: the *oculumotor* and the *processing* driven models (Reichle et al. 2000).

The oculumotor models mostly look at the visual properties of the text (i.e. word lengths) and the physiological limits of the eye (i.e. perceptual span and saccade lengths) in order to determine the location and duration of fixations (Reichle et al. 2000). Oculumotor modelling has been successfully used to predict eye movements, but the models can never explain the whole reading process since they ignore the fact that language evidently has an impact on reading. The processing models on the other hand assign linguistic processing a very central role. The general assumption of these models is that the fixation duration is directly related to the cognitive processing whereas the fixation targets are determined by a combination of linguistic, orthographic and oculumotor factors (Reichle et al. 2000). Since this thesis is mostly concerned with fixation durations the processing models had has most impact on the work, and particularly the work of Just and Carpenter.

In 1980, Just and Carpenter suggested that, "a reader can take in information at a pace that matches the internal comprehension process" (Just and Carpenter 1980:329). From this starting point they developed the most widely known processing model of reading. They began by observing actual *gaze durations*, the sum of all fixations on a word before moving to the next, made by college students reading scientific passages of text. Just and Carpenter found large variations in the duration of individual fixations as well as the duration of fixations on individual words. They also found that almost each content word was fixated and that fixation times were longer on words that were infrequent, thematically important or clarifying the interpretation of previous words. The gaze durations were also found to be longer at the end of a sentence thus indicating integrative processing.

From these findings Just and Carpenter developed a reading model based on two assumptions. The first was the *immediacy hypothesi*s, which state that each word is immediately processed when it is fixated. The second assumption is the *eye-mind hypothesis*, which state that the eyes remain fixated on a word as long as it is processed (Just and Carpenter 1980). Both assumptions have later been criticized mainly because they don't account for context and preview effects, i.e. that other words than the ones fixated can be predicted out of the context, or perceived in the parafovea, and therefore also affect the processing times (Paulson and Goodman 2000; Reichle et al. 2000). The resulting reading model presented by Just and Carpenter is very comprehensible but unfortunately it tries to explain the *entire* reading process, from fixation to long term-memory (Just and Carpenter 1980; Just et al. 1982). Although this made the model quite complex it is still disputed as it is assumed to simplify matters too much (Reichle et al. 2000). However, although the model might have tried to cover too much of the reading process it still has merits, simplifying a complex problem is not necessarily negative, and many approaches used by Just an Carpenter has also been used in this thesis.

There are also other processing models available but these have had a minor impact on the work presented here. Rayner's *E-Z reader* is a model that is similar to the one presented by Just and Carpenter, but with a narrower scope since it does not try to account for high-level linguistic processing. It does however account for preview and context effects (Rayner 1988; Rayner and Serano 1999; Reichle et al. 2000). The model is unfortunately quite complex and the underlying assumptions are not as transparent as in the model presented by Just and Carpenter. Another processing model is the *attention-shift* model (Reilly 1993). It utilizes two connectionist back-propagating neural networks, one for word recognition and one for planning saccades. From a linguistic viewpoint the attention-shift approach seems a little too simple to be plausible, however the use a learning algorithm is appealing since individual differences in reading behavior are likely to be quite large.

2.2 The Concept of Readability

It is important to not mistake readability for legibility. Legibility has to do with the recognition of text items whereas readability has to do with the processing of continuous and meaningful text (Mills and Weldon 1987). Readability is typically rreferred to as the ease of "which the meaning of text can be comprehended" (Mills and Weldon 1987:331). This is of course a very vague definition, but the assessment of readability is also affected by a multitude of factors. First, there are many differences between texts, some are very comprehensive and well written whereas others can be totally unreadable. Second, there are differences between readers; some are very experienced whereas others cannot read at all. Third, there are differences between reading situations, reading reference literature before an exam differs a lot from reading a novel while waiting for the bus. Fourth, there are differences between the presentation formats, this thesis might be comfortable to read on paper but is likely to be strenuous to read on a flickering screen with low resolution. To summarize: There are so many factors that affect readability that it is impossible to account for them all.

Since readability is hard to enumerate the solution is to use approximate measures instead, the estimations may not be precise but they are at least likely to point in the right direction. The readability estimations used in this thesis can be categorized according to their use as either *ratings* or *measures*. Ratings are used to determine readability of text based on quantitative predictions whereas measures are used to evaluate readability based on actual reader performance.

2.2.1 Readability Ratings

Readability of text is usually rated by using *readability formulas*. Most readability formulas are quite simple and use a combination of word frequencies, word lengths and sentence lengths as a basis for the results. Although most formulas use purely quantitative measures they can give an indication of how hard or easy a text is likely to be to read. There are several readability formulas for English available (see Tekfi 1987 for an overview) but for Swedish there is only one that is widely known. LIX (Läsbarhetsindex in Swedish) is a quantitative readability formula for developed by Björnsson (1968). An estimated value of the readability of a text is calculated on basis of the percentage of long words, seven or more characters, and the average sentence length (Eq. 1).

$$LIX = (lwrd/nwrd) * 100 + (nwrd/nsnt) \qquad \text{(Eq. 1).}$$

The readability estimate (LIX) is reached by first dividing the number of long words with (lwrd) by the total number of words (nwrd), the percentage is then added to the average number of words per sentence, i.e. the number of words (nwrd) divided by the number of sentences (nsnt). The result is a value between approximately 1 and 100. The following table offers interpretations of what the LIX values stand for (Table 1):

Table 1. LIX value interpretations.

LIX value	Interpretation	Type
20	Very easy	Children books
30	Easy	Novels
40	Medium	News papers
50	Difficult	Official documents
60	Very difficult	Legal documents

Readability formulas like LIX have been criticized for generalizing too much. A long word is not necessarily harder to read than a short word and short sentence can also be less readable than a long sentences (Bruce et al. 1981; Redish 1981). However, readability formulas have also been found to be useful for many appliances, especially when detail level is not so important (Klare 1984). In the work with this thesis LIX has been used to determine the readability of the texts used in the evaluation (see web resources for an address to a web-based lix counter).

2.2.2 Readability Measures

Readability has mostly been evaluated in terms of *reading speed* and *comprehension* (Mills and Weldon 1987). Reading speed is often calculated as words read per minute (wpm) whereas comprehension is represented as percent of correctly answered multiple-choice questions, both these measures are *objective.* The reading speed results are mostly reliable when comparing results from different evaluations whereas the comprehension scores are a little unpredictable since they are highly dependant on the type of questions asked. The product of the reading speed and comprehension scores has been suggested as a composite measure for *reading efficiency* (Jackson and McClelland 1979; Rahman and Muter 1999; Castelhano and Muter 1999). The measure is used to avoid problems associated with assumed tradeoffs between speed and comprehension (Wickens 1992). However, since the comprehension scores are likely to be unreliable there is really no point in mixing them up with reading speed.

Although a high reading efficiency is likely be a good indication of good readability it is also common to use a few *subjective* measures. The most widely used measure of this type is the *attitude* inventory. It is especially common to use when different text presentation formats are compared against each other. Essentially attitude inventories are a set of questions about experience and preference. Unfortunately the questions often differ between evaluations so it is hard to compare the results, nonetheless they can be very illuminating for the evaluators. Another subjective measure used in evaluations, which actually is comparable, is the standardized *task load* inventory NASA-TLX (Task Load Index) (Hart and Staveland 1988). The inventory is composed of six factors denoting different cognitive demands that are rated by the subjects after completing a task. The NASA-TLX task load inventory was rewardingly used in the RSVP evaluation preceding the one presented in this thesis (Sicheritz 2000; Goldstein et al. 2001).

2.3 Screen Reading

Quite a lot of research in readability has focused on comparisons between reading on screen and reading on paper. It began with studies of readability on the first generation Cathode Ray Tube (CRT) screens, also commonly referred to as Visual Display Units (VDUs). In the majority of the early experiments on VDUs readability was found to be poor compared to paper. The average reading speed for an English text on paper is between 220-340 wpm (Kump 1999)**.** Reading speed on VDUs was around 20-30% slower although comprehension was roughly the same (Muter et al. 1982; Kang and Muter 1984). These findings are not too surprising since the first screens were primitive units with poor legibility due to low resolution and mediocre refresh rates. The designs of the early experiments have also been criticized mainly because the reading situations were quite unrealistic (Dillon 1992). For example, in the Muter et al. study the subjects were seated five meters from the display unit while reading and the text was presented in large white characters on a blue background.

2.3.1 Large Screen Readability

The Achilles' heel of the first-generation large screens seems to have been the low resolution. Screen technology evolved rapidly in the 80's and the second-generation CRT screens offered far better resolution and also color. However, the breakthrough in readability, and usability in general, came with the introduction of the Graphical User Interface (GUI). Later studies using computers with GUIs showed that there was in fact little or no differences between screen and paper, provided that attention was paid to such factors as screen resolution, refresh rates, anti-aliasing, text polarity, etc (Gould and Grischkowsky 1984; Osborne and Holton 1988; Muter and Maurutto 1991; Muter 1996). Although reading speed and comprehension does not differ much between high-quality screens and paper the users still seem to prefer reading on paper. This may be partially due to the fact that reading on a large screen requires the reader to view the text from a distance and in a fatiguing posture (Schneiderman 1998). An underestimated aspect is also that most readers are more used to reading on paper, with time however there might be people that prefer reading on a screen. However, the screen must not necessarily be seen as a successor to paper but rather as a complement. There are many things that can be done with a text on a computer that is hard, or impossible, to do on paper. A common situation is also to browse for documents on the screen and then print the selected document on paper for reading. In this light the readability on the large screens of today seems quite satisfactory. Readability on small screens is however another story.

2.3.2 Small Screen Readability

Most mobile devices utilize flat Liquid Crystal Display (LCD) screens. The early LCD screens were monochrome and offered poor resolution, a little like returning to a sized down version of the an early VDU unit. However, LCD technology evolved as well and today LCD's offer a resolution and colour depth that is comparable to second-generation CRT screens. The problem with readability on small screens is however not so much the resolution as it is the limitation in the screen space, which restricts the amount of information that can be presented at a time. This implies a higher the rate of interaction by the user to view the text. Reading a longer text on a small screen can thus be frustrating and, to complicate matters further, users of mobile devices do not always have access to printing facilities. Researchers have been studying the effect of display size on reading in order to determine how small it can become before problems occur.

Duchnicky and Kolers (1983) performed an experiment with varying window widths and heights and found that a height of 20 lines only increased reading speed by a mere 9% compared to using a height of 4 lines. Smaller window heights than 4 lines were however found to be significantly less efficient to use. The results also showed that a window width of 2/3 of a full page increased reading speed by 25% compared to using 1/3 of a full page. These widths are however much larger than what the average mobile device has to offer but the findings seem to suggest that a limited screen width decreases reading speed. Using a higher density of characters per line was also found to improve readability, using 80 characters compared to using 40 characters increased reading speed by 30%. This is not very surprising since a lower density implies less information in the perceptual span at a time and therefore also a lowered efficiency. The texts read in the Duchnicky and Kolers study were however quite short, ~300 words.

Dillon et al. (1990) investigated how the reading of longer texts, ~3500 words, was affected by using window heights of 60 and 20 lines. The results showed that neither reading speed nor comprehension differed. However, the subjects who read using the smaller window height were found to perform significantly more jumps and also altered the direction of reading much more often. Unfortunately, there are has been quite few evaluations on really small displays. The results from the evaluations reviewed here does however seem to support the assumption that a small screen space implies a higher interaction when the text is presented in the traditional way.

2.4 Text Presentation on Small Screens

There are a few different ways to present text on small screens. They can be divided into *traditional* and *dynamic* text presentation formats. The major difference between the formats is that traditional text presentation requires physical interaction when reading whereas the text proceeds automatically when dynamic text presentation is used. The text presentation formats presented here are however not applicable to small screens alone. Sometimes there is a need for squeezing a lot of information into a small area on large screens as well.

2.4.1 Traditional Text Presentation

The text on a screen is traditionally presented with a spatial layout in the same manner as on page. However, since a full page cannot be displayed on a small screen it is divided into smaller parts. The text can then either be presented as smaller pages that fits the screen or as a long page, i.e. a scroll, continuing outside the screen. In the page format turn-page keys are used to move between the pages and in the scroll format a scroll-bar is used to move in the text. Continuous scrolling has been found to be preferred compared step-by-step scrolling but the page format is still more popular to use (Mills and Weldon 1987). Both formats require the reader to interact physically in order to move forward in the text but the increase in interaction may be acceptable considered that these formats are more familiar to the readers.

2.4.2 Dynamic Text Presentation

Leading and *RSVP* are the two most common forms of dynamic text presentation. Both formats requires very little interaction from the reader since the text proceeds automatically, the obvious advantage of these techniques with respect to limited screen displays is however their ability to display text in one single line, leaving the rest of the space for other pieces of information. In Leading, or horizontal scrolling (also known as the Times Square Format), the text moves from right to left. Sekey and Tietz (1982) and Granaas et al. (1984) found that Leading was less effective than traditional text presentation in the page format. However, in these studies the text moved forward letter-for-letter. Kang and Muter (1989) evaluated leading that moved forward pixel-for-pixel and found that to be more effective. Chen and Chan (1990) evaluated leading with self-paced versus experimenter-controlled presentation speed. The results showed that the self-paced subjects read faster whereas the experimenter-controlled subjects remembered more. From these results they deduced that self-paced reading can be counterproductive and that experience is important when reading moving text.

RSVP originated as a tool for studying reading behavior (Forster 1970) but has lately received more attention as a presentation technique with a promise of optimizing reading efficiency (Joula et al. 1982; Masson 1983; Potter 1984; Joula et al 1995; Muter 1996; Rahman and Muter 1999, Sicheritz 2000; Goldstein et al. 2001). When RSVP is used the text is successively displayed as small chunks within a small area, each chunk typically contains one or a few words depending on the width of the text presentation window. When reading the text in this fashion it proceeds by itself and that makes the saccadic eye movements and the return sweeps superfluous. It also means that regressions, or the rereading of words and phrases, may be effectively prevented (Rahman and Muter 1999).

Comparisons between RSVP and Leading have so far been inconclusive. Kang and Muter (1989) found no significant differences between the two techniques whereas Joula et al. (1995) reported that Leading was inferior to RSVP. McCrickard et al (2001) used RSVP, Leading and ”fading” techniques to support secondary tasks on peripheral displays and concluded that Leading techniques are better for comprehension and memorability while RSVP and ”fading” techniques are better for rapid identification of items. However, when reading is the primary task it seems more *natural* to use RSVP. The eye processes information in fixed gazes and a format that moves the text successively rather than continuously is therefore assumed to better adhere to the reading process. This is the foremost reason for choosing RSVP as an alternative to traditional text presentation.

2.5 Rapid Serial Visual Presentation

The term RSVP was first introduced by Forster (1970) as a name for a technique used for studying text processing and comprehension. Later RSVP was introduced as a presentation technique for computer screens with the assumption that the reduced need for eye movements would reduce cognitive load and optimize reading (Joula et al. 1982; Masson 1983; Potter 1984). However, the term RSVP has come to label a wide variety of approaches for text presentation where the only common denomination has been that chunks of text have been successively displayed. The designs of most RSVP evaluations and implementations have differed so much that the findings from one evaluation are not necessarily applicable to another. Some have presented single words at a time whereas others have presented several words. A few have presented long texts where as most have presented only smaller paragraphs. The differences in the reading speed have also been quite large, some have used a very high presentation speed whereas others have let the readers choose their own.

A short review of the most influential evaluations where RSVP have been used follows next, although they differ a lot from each other there is something to learn from all of them. The review is followed by the description of the Reader version 1.0 evaluation, the first RSVP implementation evaluated on a PDA (Sicheritz 2000; Goldstein et al. 2001). The Reader evaluation is the one preceding the evaluation presented in this thesis and much has been learned from the experiences of that evaluation.

2.5.1 Previous RSVP Evaluations

Joula et al. (1982) presented shorter paragraphs of text on a CRT screen, either in the page format or in the RSVP format with text chunks of 5, 10 or 15 characters. Each text chunk was exposed for 200-300ms, which is equal to a reading speed of approximately 300 wpm. The results showed no significant differences in comprehension between the reading conditions.

Masson (1983) evaluated how the insertion of blank windows at sentence boundaries affected the RSVP format. Masson experimented with durations of 500 and 1000 ms and found that performance increased with blank windows regardless of duration.

Cocklin et al. (1984) compared RSVP with the text divided into either *idea units* or ad-hoc chunks. The idea unit segmentation was performed by hand and was based on clause and phrase boundaries as well as linguistic features. Each chunk averaged 13 characters and the reading speed was approximately 300 wpm. The results showed that the use of idea units increased comprehension a little but not significantly.

Muter et al. (1988) performed experiments with self-paced RSVP and RSVP that permitted regressions. The results showed that larger regressions yielded slower reading and regressions back to the beginning of the sentence were found to be more frequent than regressions two words back. Overall the results indicated that permitting reader control was feasible but permitting regressions resulted in lower performance.

Kang and Muter (1989) compared RSVP to word-by-word, letter-by-letter and pixel-by-pixel Leading. Except for word-by-word, comprehension was as high for Leading as it was for RSVP. The comprehension scores for pixel-by-pixel leading were also found to match RSVP at reading speeds ranging from 100 to 300 wpm. The subjects in the evaluation were also found to express a significantly higher preference for pixel-by-pixel Leading.

Fine and Peli (1995) evaluated how visually impaired and elderly subjects read using RSVP and scrolled text. They found that the visually impaired read at a similar speed using both formats whereas the elderly read faster using RSVP,

Rahman and Muter (1999) compared word-for-word RSVP and sentence-by-sentence presentation, with or without a completion meter, to traditional text presentation in the page format. No significant differences were found for comprehension and reading speed but the subjects liked the inclusion of a completion meter.

Castelhano and Muter (2001) evaluated the effects of using RSVP with or without punctuation pauses, variable word durations and a completion meter. They compared a few RSVP formats to traditional text presentation and sentence-by-sentence presentation. The results showed that pauses and variations made the RSVP format significantly more accepted. However, the sentence-by-sentence and traditional page format remained more popular although RSVP was just as effective.

2.5.2 The Reader Evaluation

Karin Sicheritz implemented the Reader version 1.0 in order to evaluate how it was to read using RSVP on a PDA compared to using a paper-book (Sicheritz 2000; Goldstein et al. 2001). The application was implemented on a Casio Cassiopeia E-105 PDA and offered a graphical user interface (Figure 2).

Figure 2. The Reader version 1.0 prototype, adopted from Sicheritz (2000).

In the Reader the user could select three different text presentation window widths (11, 17 and 25 characters), twenty-seven speed levels ranging from approximately 25 to 1000 wpm (low, 1–25 and fast) and five font sizes (8–16 points). The text presentation could be started, paused, continued and it was also possible to return to the previous text window. A page indicator showed the current and total number of pages. While presenting text the application searched for the nearest white space given the window width and segmented at that point. The text chunk was then exposed for a *fixed* time according to the selected speed. If the number of characters for a single word exceeded the chosen window width the whole word was presented anyway. The application also segmented the text when commas or other punctuation marks were detected. Blank windows of 250 ms were inserted between the sentence boundaries (Sicheritz 2000).

In a repeated-measurement within-subject experiment employing 10 subjects the Reader prototype was benchmarked against the paper-book. Two conditions with different window widths for RSVP, 11 and 25 characters respectively, and a *sonified* condition (i.e. where sounds were added to enhance the reading experience) were compared to a paper-book condition. The texts used in the experiment were the first six chapters from the novel "Röda Rummet" (in Swedish by August Strindberg), the chapters were between ~2700-6300 words long. The subjects read the first chapter in the paper-book and the following chapters using the Reader prototype. The subjects were instructed to read as *fast as possible* in all but the sonified condition. Readability was measured by reading speed, comprehension inventories consisting of 10 multiple-choice questions, the NASA-TLX (Task Load Index) (Hart and Staveland 1988) task load inventory and an attitude inventory consisting of five questions about Difficulty, Efficiency, Comprehension, Stimulation and Facilitation of the presentation (Sicheritz 2000; Goldstein et al. 2001). The results showed that neither reading speed nor comprehension differed between the conditions. The text presentation width of 25 characters resulted in the highest reading efficiency but the differences were not significant. The task load inventory did however reveal significantly higher task load ratings for all RSVP conditions for all factors but Physical demand. Also on the attitude inventory was the book condition significantly more favored for Difficulty, Comprehension and Stimulation (Sicheritz 2000; Goldstein et al. 2001).

The results obtained from the Reader evaluation were quite encouraging. At least they showed that reading using RSVP on a PDA was just as efficient as using a paper-book, although the subjects may not have agreed. The high task load rating and the lower attitude ratings for the RSVP format were disappointing. Many issues may have affected the discrepancy between the subjects' objective and subjective results. One problem with the Reader evaluation was that it compared screen reading to reading in a paper-book, which are two very different things. Although the resulting findings are interesting it would be fairer to see how RSVP performed compared to traditional text presentation on an actual PDA. There were also a few problems with the experimental design. All subjects read the first chapter in the paper-book but only the two RSVP conditions with different window widths were properly balanced. The problem is that both the text lengths and difficulty ratings (LIX) differed a lot between the chapters. Although the realism of reading chapters consecutively may be lost when balancing the conditions it is much easier to draw conclusions from the results if experimental control is kept. The texts chosen to be included in the experiment were also quite hard to read, the results may have looked different if easier texts were chosen. Another problem was that the subjects had to halt the experiment each time they made a change in reading speed (it was needed to keep an experimental record over the selected reading speeds). This may have caused the subjects to refrain from speed changes or at it least likely to have disrupted the reading experience. A probable cause for the high task load when using the Reader prototype might also have been that the exposure time for each text chunk was fixed. This does not seem to adhere very well to the reading process and Sicheritz and Goldstein also point that out on basis of the results from the Reader evaluation (Sicheritz 2000; Goldstein et al. 2001).

However, the last hypothesis is exactly what triggered the work with adaptive RSVP and the prototype presented in this thesis. This hypothesis also brings us to the end of the background part. The next part is about the design and that is where the hypothesis becomes a thesis.

3 Design

Unfortunately the Reader prototype (Sicheritz 2000) could not be improved further to include adaptive RSVP due to technical reasons. Instead an algorithm for Fixed RSVP that mimicked the performance of the Reader had to be designed. This section begins with a description of the fixed RSVP algorithm in section 3.1. In section 3.2 the design of Adaptive RSVP is introduced and explained. In section 3.3 the design of a distribution and processing model for RSVP is presented together with an example of a devised document format.

3.1 Fixed RSVP

The speed of the presentation when using RSVP is measured in words per minute (wpm). The exposure time of each text chunk is calculated on basis of the presentation speed and on how much that can be displayed in the text presentation window. Unfortunately there is little or no documentation in previous studies on exactly how the exposure times have been calculated (see Section 2.5 for an overview). What is known however is that the exposure times have generally been fixed. In this evaluation the following formula has been employed for calculating the fixed text chunk exposure times (Eq. 2):

$$time_0 = fchr/(wavg*wpm/60) \qquad \text{(Eq. 2)}$$

The average number of characters that can be displayed (fchr) is divided by the product of the average word length (wavg) for the current language and the presentation speed (wpm) divided by 60. The result is a fixed exposure time for each text chunk measured in seconds ($time_o$).

When presenting a text with RSVP the relation between reading speed and exposure time can be visualized in a speed-exposure plot where all the individual text chunks are represented (Figure 3).

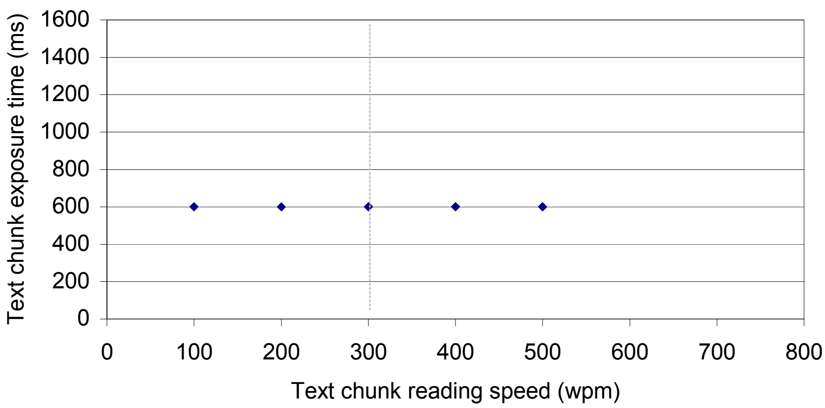


Figure 3. Variations in reading speed for individual text chunks containing 1, 2, 3, 4 or 5 words (the dots) presented at a set speed of 300 wpm (the vertical line) when using RSVP with fixed exposure times (data derived from the training text).

The plot is a result of presenting the training text used in the usability evaluation at a constant speed of 300 wpm using the fixed exposure time formula (Eq. 2). The width of the text presentation window was 25 characters (fchr=21) and the average word length was set to 7 characters. The obtained reading speed is a result of dividing a text chunks exposure time by the number of words it contains (1-5). This explains why there is a variation in reading speed (100-500 wpm) although the exposure time is fixed (600 ms). Only the text chunks with three words match the selected reading speed (the vertical line in Fig. 3). If more words appear in a text chunk speed increases and if fewer words appear speed decreases. The reading speed is thus *inversely* related to the number of words that each text chunk contains. One explanation to the high task load when using fixed RSVP may be attributable to the fact that the exposure time do not vary although reading speed actually does.

3.2 Adaptive RSVP

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 (Just and Carpenter 1980:330). The point with Adaptive RSVP is that it attempts to mimic the reader's cognitive text processing pace more adequately by adjusting each text chunk exposure time in respect to the text appearing in the RSVP text presentation window. By assuming the *eye-mind hypothesis* (Just and Carpenter 1980), i.e., that 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.

As mentioned earlier, the resulting reading model presented by Just and Carpenter is comprehensible but also quite complex since it tries to explain the entire reading process. Although the model has been accused for simplifying matters too much (Reichle 2000) it is still a too complex to utilize here on an as-is basis. The reason for keeping the adaptation simple is that the approach with varying presentation speeds is quite novel and its effect on readability unknown. It is better to begin with some small changes and observe how they affect the results since if a complex model is used it is much harder to identify the individual factors that affected the outcome. The adaptive algorithms presented here are therefore quite simple but they are still based on the assumptions made by Just and Carpenter.

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 (Just and Carpenter 1980; Just and Carpenter 1982). Further, most new information tends to be introduced late in sentences and therefore ambiguity and references tends to be resolved there as well. A shorter sentence is also usually processed faster than a longer one since it conveys less information (Just and Carpenter 1980; Just and Carpenter 1982). Thus, processing time differs both within and between sentences and the text chunk exposure times can therefore be adjusted according to this as well. On basis of these assumptions two adaptive algorithms were developed that were supposed to decrease cognitive demand (i.e. task load ratings). The first algorithm adapts the exposure times to the *content* of the text chunks whereas the second also looks to the *context* in the sentences. Both algorithms inserts a blank window between each sentence if there is not enough space to begin on the next sentence in the same window, otherwise a delay is added to the sentence boundary instead. This approach is assumed to generate a natural spacing between sentences and cause fewer interruptions while reading.

3.2.1 Content Adaptation

In content adaptive mode the exposure time for each text chunk is based on the numbers of characters and words that are being exposed for the moment. It is deliberately a very simple form of adaptation based on some general assumptions about word distribution. Longer words are assumed to be more infrequent and take longer time to read than shorter words. A higher number of words are also assumed to take longer time to read and should thus receive more exposure time. The following formula is used to calculate the text chunk exposure time for content adaptation (Eq. 3):

$$time_1 = (nwrd+nchr)/(davg*wpm/60) \qquad \text{(Eq. 3)}$$

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.

The effect of using content adaptation is probably best explained by using a speed-exposure plot (Figure 4).

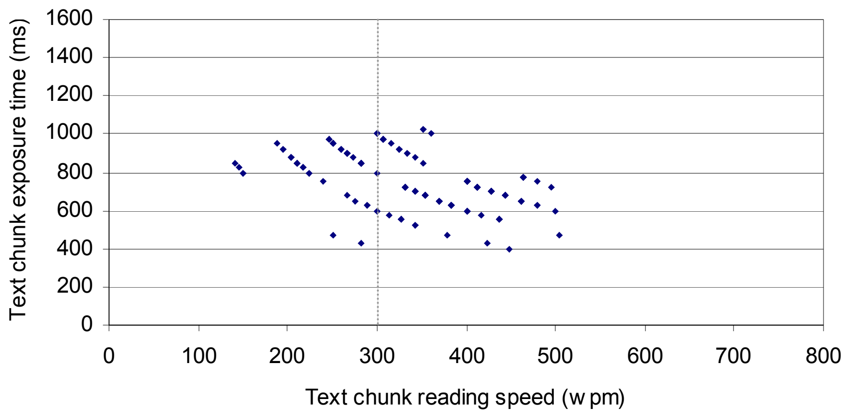


Figure 4. Variations in reading speed for individual text chunks containing 1, 2, 3, 4 or 5 words (the dotted lines) presented at a set speed of 300 wpm (the vertical line) when using content adaptation (data derived from the training text).

Again the plot is a result of presenting the training text at a constant speed of 300 wpm but this time by using the formula for content adaptation instead (Eq. 3). The average word length including delimiters was set to 7,8 whereas the other variables were the same as those used for Fixed RSVP. The speed-exposure plot illustrates the difference compared to Fixed RSVP.

However, even though the exposure time varies the variation in reading speed is actually smaller for content adaptation than for Fixed RSVP. When using content adaptation the exposure time for each text chunk is *directly* related to the number of words and characters it contains. Reading using this form of adaptation is supposed to decrease task load since the relation between what is being exposed and the time for exposure is more proportional.

3.2.2 Context Adaptation

In context adaptive mode the exposure time for each text chunk is based on the following: The result of content adaptation, the word frequencies of the words in the chunk and the position of the chunk in sentence being exposed. To begin with each word in the chunk is looked up in a lexicon with word frequencies. If the word is common it receives a weight lower than one and if it is rare or not in the lexicon it receives a weight higher than one. The following formula is used to calculate how the exposure time is affected by the word frequencies (Eq. 4):

$$time_2 = time_1 * ((wfrq_1+\ldots+wfrq_{nwrd})/nwrd) \qquad \text{(Eq. 4)}$$

The formula uses the exposure time for content adaptation ($time_1$) and the word frequency weights for the words in the chunk (wfrq) as a basis for the result. The word frequency weights are added and divided by the number of words in the text chunk (nwrd). The product is then multiplied with the content adaptive exposure time to get the weighted exposure time ($time_2$).

The next step is to give the chunk less exposure time if it appears in the beginning of a sentence and more if it appears in the end. The following formula is used to calculate the text chunk exposure time depending on the position in and the length of the current sentence (Eq. 5):

$$time_3 = (time_2+time_2 * \tanh(swrd/savg))/2 \qquad \text{(Eq. 5)}$$

The formula uses the intermediary exposure time reached earlier ($time_2$), the number of words in the sentence exposed so far (swrd) and the average sentence length (savg). In order to get a smooth drop-off in speed along the sentence, a mean of the previously calculated exposure time and its product with the hyperbolic tangent (tanh) of the division of the number of exposed words and the average sentence length is calculated. The result is a varying text chunk exposure time ($time_3$). The effect of using context adaptation is here illustrated by using a speed-exposure plot again (Figure 5).

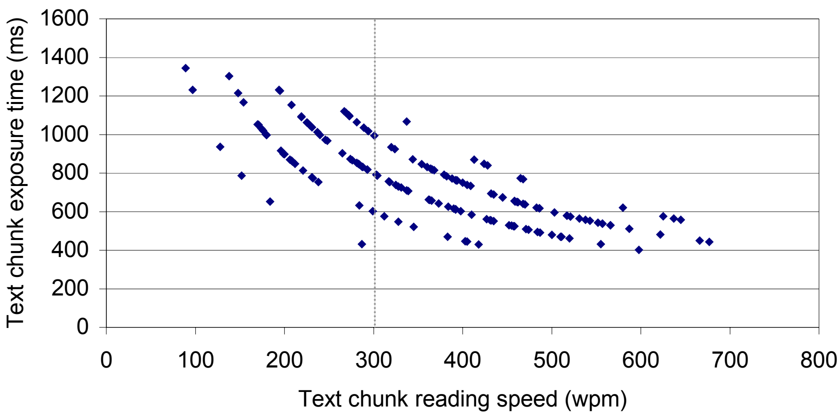


Figure 5. Variations in reading speed for individual text chunks containing 1, 2, 3, 4 or 5 words (the dotted lines) presented at a set speed of 300 wpm (the vertical line) when using context adaptation (data derived from the training text).

The plot is a result of presenting the training text at 300 wpm when using context adaptation (Eq. 5). The average sentence length was set to 11,5 words and the word frequency weights ranged between 0,6-1,2. A lexicon with frequencies for the 10.000 most common words in a corpus of 11,9 million words (Press 97) was used to assign the weights according to a lognormal distribution. The assumption with using a lognormal distribution was that the expected gaze duration on each word would decrease with the logarithm of the word's frequency. This captures the fact that there are small differences in frequency among infrequent words whereas the differences larger for common words (Just and Carpenter 1980).

When Context adaptive RSVP is used the variations in exposure time are larger than for both Fixed and Content adaptive RSVP. Task load is however still supposed to decrease since the variations are assumed to better match the cognitive demand while reading.

3.3 Distribution Model

The discussion around the distribution model originated from the question if the texts were supposed to be processed on the *client* or on the *server*. The reason for raising this question is that the mobile clients usually are quite thin (i.e. has limited processing power) and that linguistic processing can be quite demanding. If a server were used the adaptation could be made much more advanced including parsers for elaborate segmentation and linguistic processing. The server approach also makes it possible to update the processing capabilities continuously without having to reprogram the clients. On the other hand it is also appealing to have *independent* clients that can take a text and present it without being connected to a server. There are merits and pitfalls with both approaches and in the end a combination of both were chosen. In the distribution model applied here the linguistic processing can be done on both the server and the client. The server is supposed to perform advanced linguistic processing whereas the client is supposed to do simple linguistic processing if no server is available or needed. The client-server approach requires some form of intermediary document format in order to transfer the meta-information about the text. That format should ideally also be able to transfer information about the document structure and pointers to additional resources added to the text (i.e. sounds, images, bookmarks, annotations etc).

The RSVP formulas developed in this thesis are quite simple and can be considered as a form of client processing formulas. A document markup schema for RSVP was still designed in the eXtensible Markup Language (XML) format in order to facilitate the addition of external resources (i.e. sounds, see section 4.4 for motivation) and encourage further developments on the server side processing (see web resources specifications). The following excerpt from an XML tagged file (in Swedish) will serve as an example of what sort of information a RSVP document might contain (Figure 6):

```
<?xml><text title=”Glöd” author=”Anette Kullenberg”><chapter
number="1" title=”Väntan”> <paragraph> <sentence> Klockan
var <sound=”ticktwo.wav”> två på morgonen när Tomás Murillo
sköt ner rånarna. Han hade </sound> gömt sig bakom bardisken
och släckt ner och satt på huk och väntade. <bookmark
id=”1”> Bredvid sig hade han</bookmark> en flaska brandy,
men han drack inget. </sentence> <sentence>Han tände en
Kruger ibland och såg på klockan i tändstickans
sken.</sentence></paragraph> </chapter></text></xml>
```

Figure 6. Excerpt from an XML tagged RSVP file.

4 Implementation

In order to evaluate the RSVP algorithms they had to be incorporated into a prototype on a mobile device. Bailando is the resulting prototype that has been developed at the Ericsson Research's Usability & Interaction Lab by the author. Bailando is an acronym for: Better Access to Information through Linguistic Analysis and New Display Organization. Bailando was developed for the Compaq iPAQ 3630 Pocket PC, a small Personal Digital Assistant (PDA) with a touch sensitive high-resolution color display (see web resources for product specification). The iPAQ offers much more screen space than the average mobile device but it still shares many of the properties typical for smaller devices. The size and weight of the device was similar to a pocket book of approximately 250 pages (Figure 7).

Figure 7. The Bailando prototype running in reading mode on a Compaq iPAQ 3630.

The motivation for the choice of deployment is given in section 4.1 together with a brief overview of the structure of the implemented application. The Bailando prototype is demonstrated more in detail in section 4.2 where a walkthrough of the application is presented. In section 4.3 the RSVP characteristics used by Bailando are presented. Bailando was supposed to be used for experiments with sonification, although this is a little out of scope for this thesis, the functionality for this is described in section 4.4. Finally in section 4.5 the log functionality of the prototype is described.

4.1 Deployment

The Compaq iPAQ is a Windows 32-bit platform; the available programming languages were primarily Visual Basic and Visual C++. It might have been possible to use Java as well but there were just too many questions marks around the performance of Java on the iPAQ to make that choice feasible. It was important that the timing in the application was reliable since the differences in exposure time when using adaptive RSVP is quite small. In the end Embedded Visual C++ was chosen as a programming environment since it offered real-time capabilities (see web resources for product specification). Using C++ also made the program more extensible and easier to export the to other platforms.

4.1.1 Program structure

Bailando was implemented as multi-threaded application where the main processes communicated with each other by using semaphores (i.e. flags). Basically there was one thread to keep control over the graphical user interface, one thread to parse and process the text and one thread to play sounds (see section 4.4 for motivation). The widgets and the threads were implemented as objects whereas functions like redrawing the screen or calculating a text exposure were implemented as methods.

4.2 Graphical User Interface

It was important that the graphical interface was appealing and yet intuitive to use. It had to look as a professional application since it was supposed to be compared to other professional applications for traditional text presentation. Since ample screen space was available on the iPAQ all the application controls were implemented into the graphical user interface. This also makes the Bailando software easier to run on other PDA's since button assignments differ between devices. In order to get a feel for how Bailando turned out, a step-by-step walkthrough will be presented here with design choices and explanations given along the way.

4.2.1 Bailando walkthrough

The first that happens when Bailando is launched is that a start screen is shown (Figure 8, left). This screen is also showed if a text for some reason or another is not loaded properly, but then an advisory error message is shown as well. The main interaction control of the application is the toolbar. It is always located at the bottom of the screen so that the hand does not obscure the screen while it is used. It always has the same appearance although unselect able items are greyed out. As can be seen on the start screen the only item selectable on the toolbar is the Menu.

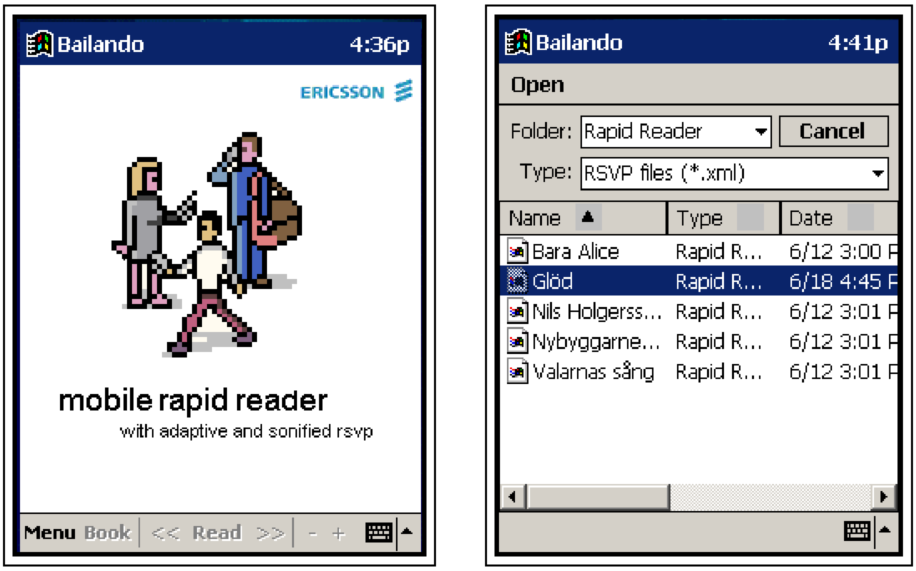


Figure 8. Bailando start screen (left) and the library menu (right).

When the Menu item is selected a pop-up main menu is shown. From the main menu it is possible to: exit the application, to get information about version, to change settings and to retrieve texts from the library. If the library alternative is selected the library screen is shown. The Bailando application supports files in either the XML format described earlier or in plain text. A text is selected from the library by selecting it in the list (The file "Glöd" is selected in Figure 8, right).

Once a text is selected in the library the processing of the file starts. First the file is checked to see if it can be viewed in the Bailando prototype at all. After that the data format is determined. If it is an XML file it is parsed and the meta-data is stored as variables and if it is a plain text file some of the file properties are picked up and stored (i.e. filename and type). The next step is to analyse the text, the number of characters, words, long words and sentences are calculated in order to determine the LIX value.

While processing the text the main reading screen is exposed (Figure 9), but not until the text is processed completely are all the toolbar items selectable. The complete toolbar contains the controls to start, pause and resume the presentation. The presentation can also be paused and resumed by touching anywhere on the text presentation area of the screen. It is important that it is easy to make pauses since the user is likely to need to do that fast when interrupted in order not to get lost. This important feature of the interface is introduced to the user by the following instruction when the text is ready to be viewed: "tap the screen to begin".

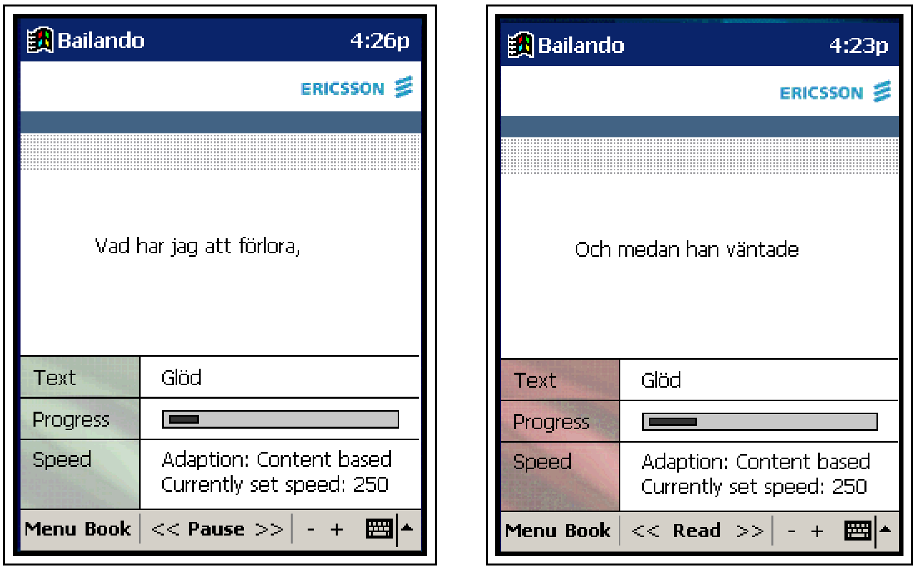


Figure 9. Bailando main screen in reading mode (left) and pause mode (right).

In Bailando the text is presented at an area located slightly above the half of the screen. The text is presented at one single line that utilizes 2:3 of the screen width and the vertical alignment is similar to the text presentation area as a whole. Above the text presentation area there is a border for aesthetical reasons. Below the text presentation area there is an information area displaying the text title, the progress bar and the current speed settings (Figure 9). The progress bar is included in order to support memory of spatial location while reading, as said earlier a completion meter has been found to increase the user preference for RSVP in a previous evaluation (Rahman and Muter 1999).

While reading the toolbar show the Pause button and the coloured part of the background is green (Figure 9, left). If the user feels he missed or did not understand a presented text window it is possible to go backwards (<<). It is also possible to skip text by going forward (>>). These orientation features are included to make it easier for the user to browse through the text. It is also possible to browse through the text step-by-step in paused mode. The reading speed can be easily increased (+) or decreased (-) with the speed control buttons in steps of 10 wpm. When the text presentation is paused the toolbar shows the Read button and the background is then also turned red as an extra affordance (Figure 9, right).

When a text is loaded the book item on the toolbar also becomes selectable. The idea was to add bookmark and table of content functionality to this menu but in the present version of Bailando this menu only lets the user choose to go to the beginning of the text, to the point furthest read in the text or to view text information. The text information screen contains the information selected during the initial processing of the text (Figure 10, left). Apart from showing the texts title, type and metrics it also displays a readability rating (based on the LIX value for Swedish). The suggested reading speed is calculated on basis of what and how fast the reader has read earlier. Bailando was supposed to keep a record of previous reading speeds for different text difficulties and match a text according to these data. However, in the present version it only relates the current reading speed to the LIX value.

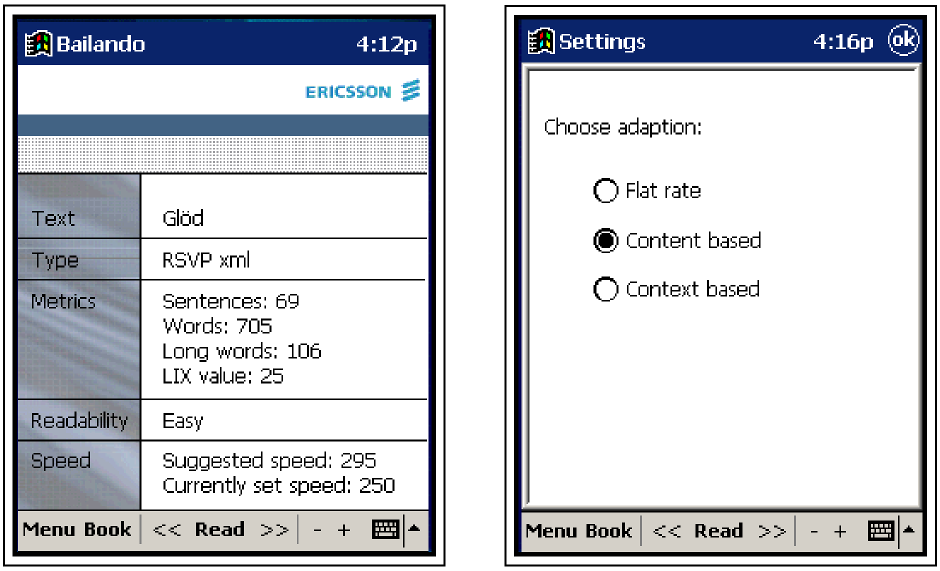


Figure 10. Bailando text information screen (left) and settings menu (right).

The last screen-shot from Bailando is the settings options available from the menu (Figure 10, right). The only setting variable that could be changed through the GUI in the present version was the adaptation mode (Flat rate is equivalent to Fixed RSVP). Ideally it would be better if more variables (i.e. average word length, blank window exposure time, window width etc) could be changed this easily. However, adaptation was the only setting needed to be changed during the evaluation and due to time constraints further improvements of the graphical interface were left out.

4.3 RSVP Characteristics

The window width of the RSVP display was 25 characters wide with the text presented left justified in a 10-pt. sans-serif typeface. The window width was chosen on basis of the results from the Reader evaluation where a width of 25 characters gave the best results (Sicheritz 2000; Goldstein et al 2001). Font-size has been found to have a minor effect on RSVP (Russel et al. 2001) and 10-pt seemed enough. Bailando supports all the three forms of RSVP that have been described in this paper. In all RSVP modes the text chunks that contained punctuation marks received an addition of 250 ms to their exposure times and all words longer than the display width were hyphenated. In Fixed RSVP mode a blank screen was inserted for 250 ms between each sentence. In content and Context adaptive RSVP mode a blank screen was inserted for 250 ms only if necessary, otherwise a delay of 250 ms was added to the exposure time of the text chunks exposing the sentence boundary. The following table offers a summary of the variables used by Bailando:

Table 2. Variables used by Bailando for RSVP.

Variable	*Fixed*	*Content*	*Context*
Average number of characters displayed	21	-	-
Average word length	7	-	-
Average word length incl. delimiters	-	7,8	7,8
Average sentence length	-	-	11,5
Default frequency weight	-	-	1,2
Blank window exposure time	250	250	250
Sentence boundary delay	-	250	250
Punctuation mark delay	250	250	250

4.4 Sonification Features

The Bailando prototype also supports Sonified RSVP. That means that it can play sounds at the same time as it presents text. This is accomplished by parsing XML tags in the text. When Bailando finds a *start-sound* tag (see Figure 6 p.19 for an example) in the text chunk to be displayed it plays the audio file linked to the tag. When a *stop-sound* tag is encountered, all playing sounds are stopped. Thus, exact synchronization between word and sound is attainable and works at any reasonable reading speed. The sonified features of Bailando have been evaluated in a usability evaluation performed by Mandana M-Bayat. The results will be published in a forthcoming article (Goldstein, Öquist and Björk 2002).

4.5 Log Functionality

In order to get reliable results from the usage of Bailando in the evaluation a few log functions were built in. Every time Bailando is initiated a new log file is created with a random name and a timestamp. Bailando logs all the user actions in the interface, i.e. speed changes, pauses and backward or forward movements. It also keeps track of the text presentation, what text it has displayed, for how long it was displayed, which variables that were used and what the individual and average reading speeds became. The log functions have proved to be very useful, both when debugging and when evaluating (see appendix A for an excerpt).

5 Evaluation

The aim of the evaluation was to see how traditional text presentation, Fixed RSVP, Content adaptive RSVP and Context adaptive RSVP affected the ability to read on a mobile device. This part of the thesis begins with a detailed overview of the usability evaluation in section 5.1. This is followed by an analysis of the results in section 5.2.

5.1 The Usability Evaluation

It was important that the same device was used for all conditions since the look and feel of the hardware was likely to bias the assessment. Long and short texts were also included in the evaluation since the experience of extended and brief reading was thought to differ.

5.1.1 Method

In order to assess the effects caused by reading long and short texts using the four presentation formats a repeated-measurement experimental layout was adopted. It also enabled us to see how repeated use of RSVP could affect the results. This is interesting since RSVP offers a novel way of reading and may take some time to get used to. The following null hypotheses were set for reading long texts, short texts and using RSVP repeatedly:

- 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.

5.1.2 Design

A balanced within-subject design was employed. Four experimental conditions were formed where each subject read one long and one short text using each presentation format. The combinations of long and short texts were fixed creating four text pairs Aa, Bb, Cc and Dd. The text pairs were balanced against condition and presentation order generating sixteen combinations. Each subject was randomly assigned to one of the sixteen combinations.

5.1.3 Subjects

Sixteen paid subjects participated in the experiment. They were all enrolled with the criteria that they were fluent in Swedish and had a self-reported interest in reading. The subjects had a mean age of 25 and half of them were male. All were computer literate and seven had some previous experience of using a PDA. Nine of the subjects had corrected vision and two were left-handed (see web resources for a link to subject registration page).

5.1.4 Apparatus

All experiments were performed on a Compaq iPAQ 3630. Bailando was used for all RSVP conditions. The initial speed of the text presentation was always set to 250 wpm but the subjects were allowed to alter the speed at any time. Two commercial programs were chosen for traditional text presentation. The Microsoft Reader version 1.0 for reading long texts and the Microsoft Internet Explorer for reading short texts (see web resources for links to product specifications). The reason for including two different programs was their intended *context of use*. The MS Reader is custom-made to present longer texts such as e-books whereas the MS Explorer is designed to present shorter web-content such as news articles (Figure 11).

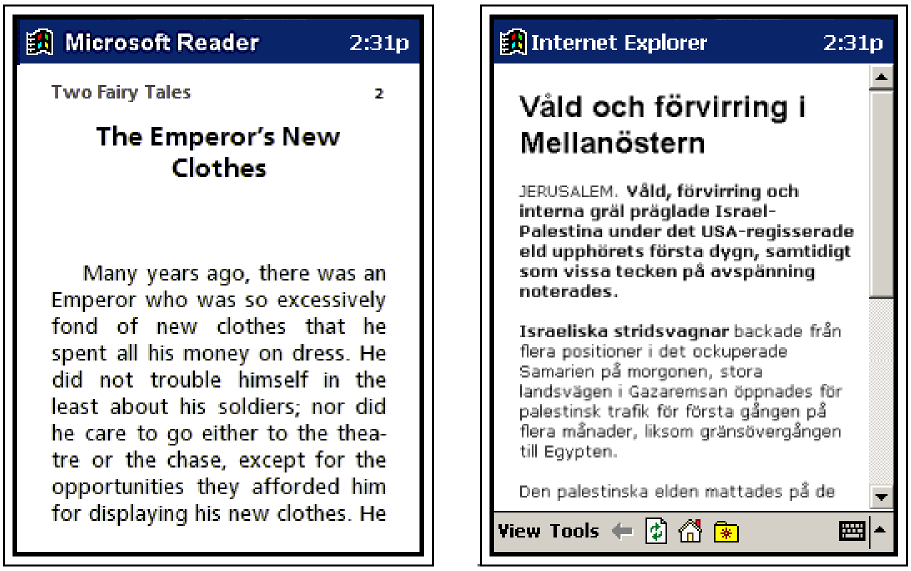


Figure 11. The interface for the Microsoft Reader (left) and the Internet Explorer (right).

There are also some important differences between the programs. The MS Reader uses page-turn buttons to move to between the pages whereas the MS Explorer uses both page-turn buttons and a scroll-bar. The MS Reader also utilizes a font-aliasing technique called ClearType (Hill 1999) in order to increase the legibility of the text whereas the MS Explorer uses regular fonts.

5.1.5 Texts

Four long fiction texts and four short news articles where chosen to be included in the experiment. One shorter fiction text was also used as a training text. Texts in Swedish with different readability ratings were chosen. The readability rating was measured with LIX (Björnsson 1968). The long fiction texts were of a similar length whereas the short texts differed some (Table 3).

All texts were formatted and processed to fit the different programs used in the experiment. In order to read the long texts in the MS Reader they were augmented with cover graphics and converted to the Microsoft e-book format. The short texts were converted to the HyperText Markup Language (HTML) in order to be viewed in the MS Explorer. All texts were also converted to plain text so that they could be presented with the Bailando program.

Table 3. Texts used in the evaluation.

Text type	*Title (in Swedish)*	*Author / Source*	*Words*	*LIX*
Training text	Glöd	Annette Kullenberg, chapter 1.	705	25
Long text A	Röda rummet	August Strindberg, chapter 3.	4272	37
Long text B	Nils Holgersson	Selma Lagerlöf, chapter 1 and 2.	4230	27
Long text C	Valarnas sång	Wally Lamb, chapter 1.	4326	31
Long text D	Bara Alice	Maggie O´Farrel, chapter 1.	4170	29
Short text a	Makedonien	Dagens Nyheter 2001-06-03	430	44
Short text b	Mellanöstern	Dagens Nyheter 2001-05-27	384	54
Short text c	Alacala	Dagens Nyheter 2001-06-02	628	40
Short text d	Lundin Oil	Dagens Nyheter 2001-05-18	365	49

5.1.6 Setting

The experiment took place in a dedicated usability lab equipped with audio and video-recording facilities (Figure 12). 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 setting.

Figure 12. Interior of the usability lab with the observation room (left) and the test room (right).

5.1.7 Instructions

Each subject received instructions before the experiment that pointed out that it was the applications and not the individual performance that were being tested. All were encouraged to ask questions whenever they wanted and also told that they could terminate the experiment at any time if they felt uncomfortable. Written instructions were administered before each session that described the principal features of the current user interface, what kind of text they were going to read and how long time it was likely to take. The subjects were particularly instructed to read at a pace as *comfortable* to them as possible (see Appendix B for instructions).

5.1.8 Training

To begin with each subject participated in two training sessions. In the first session the subject read the training text using the Microsoft Reader and in the second session the subject read the same text again using Bailando in Content adaptive RSVP mode. The subjects did not train on using the MS Explorer as all had prior experience of using it on desktop computers. During the training sessions the subject was encouraged to try out and get used to the user interface. The idea behind reading the same text twice was to give the subjects an early success experience and making them more willing to experiment with the interface. After the two training sessions were completed the subject was introduced to the questions in the inventories and filled them in.

5.1.9 Procedure

Four experimental conditions were administered and each condition was divided into two sessions. In the first session the subject read a long text and filled in the inventories, in the second session the subject read a short text and filled in the inventories. Between the first and second condition the subject had a 15-minute break and between the second and third condition the subject had a 45-minute lunch break. Between the third and fourth condition the subject had a 15-minute break again. The total participation time for each subject was around five hours.

5.1.10 Inventories

After each experimental session there were four inventories to fill in. The first inventory was a comprehension test with three-alternative multiple-choice questions. For long texts there were ten questions and for short texts there were five questions. The second inventory was the NASA-TLX Task Load Index (Hart and Staveland 1988), which was administered to check Mental, Physical, and Temporal demands, as well as Performance, Effort and Frustration levels. The third was a short attitude inventory with eight questions regarding the reading experience in terms of Ease, Effect, Comprehension, Stimulus, Immersion, Comfort, Naturalness and Focus. The attitude inventory had a bi-polar scale with a null point and two anchor points labeled High and Low (see appendix C for inventories). After the last experimental condition the subjects also participated in a follow up discussion on their experiences during the experiment. These inventories were selected to be included in the evaluation since the results would then be comparable to the Reader evaluation (Sicheritz 2000; Goldstein et al. 2001).

5.1.11 Caveats

Two problems occurred while performing the experiment that may have had a minor effect on the results. The first two subjects received a comprehension inventory where one question did not have an answer; this was however noted by the second subject and corrected after that. It also occurred an error while converting the texts; this caused the Bailando program to insert a few extra blank windows between some characters in Content and Context Adaptive mode. This error was corrected after the first three subjects had completed the experiment, the subjects did not note this error and it was discovered after an analysis of the log files.

5.2 Results

All subjects completed the experiment and there were few problems with understanding what to do or how to do it. Although RSVP was a novel way of reading for all the subjects none had any problems using the Bailando program. A few subjects did not know when the end of the text was reached when using the MS Reader and thus got confused, this was however only a minor problem. Those subjects that were left-handed had some problems when using the MS Explorer since the hand sometimes obscured the screen while using the scroll-bar.

The presentation of the results is divided into four sections: Reading speed, Comprehension, Task load and Attitude. Under each section the null hypotheses set for long texts, short texts and repeated use of RSVP is tested.

5.2.1 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 (pauses, regressions, speed changes etc).

5.2.1.1 Long texts

Adaptive RSVP improved reading speed some but the null hypothesis regarding no difference in reading speed between the conditions when reading long texts was kept (Table 4).

Table 4. Reading speeds for long texts.

Condition	*Avg. wpm*	*Std. dev.*
MS Reader	242	80,4
Fixed RSVP	249	58,5
Content adaptive	260	51,2
Context adaptive	258	79,5

5.2.1.2 Short texts

The null hypothesis regarding no difference in reading speed between the conditions when reading short texts was rejected since the main factor for reading speed was significant ($F[3,45]=8.4$, $p=0.040$). Pairwise comparisons revealed that all RSVP conditions increased reading speed significantly ($p\leq0.002$) compared to using traditional text presentation with the MS Explorer (Table 5).

Table 5. Reading speeds for short texts.

Condition	*Avg. wpm*	*Std. dev.*
MS Explorer	157	53,2
Fixed RSVP	212	46,5
Content adaptive	213	36,8
Context adaptive	203	43,9

5.2.1.3 Repeated use of RSVP

There were no significant differences in reading speed for either long or short texts between using RSVP for the first, second or third time compared to using traditional text presentation. The null hypothesis regarding no difference in reading speed caused by repeated use of RSVP was kept.

5.2.2 Comprehension

Comprehension was computed as percent of correctly answered multiple-choice questions. For long texts there were ten questions and for short texts there were five.

5.2.2.1 Long texts

The null hypothesis regarding no difference in comprehension between the conditions when reading long texts was kept. Content adaptive RSVP showed the best results but the differences between the conditions were very small (Table 6).

Table 6. Comprehension scores for long texts.

Condition	*Avg. % correct*	*Std. dev.*
MS Reader	73	19,6
Fixed RSVP	75	17,9
Content Adaptive	76	17,5
Context Adaptive	71	21,9

5.2.2.2 Short texts

The null hypothesis regarding no difference in comprehension between the conditions when reading short texts was kept. MS Explorer gave the best result but the differences between the conditions were small (Table 7).

Table 7. Comprehension scores for short texts.

Condition	*Avg. % correct*	*Std. dev.*
MS Explorer	70	21,9
Fixed RSVP	66	26,0
Content Adaptive	59	19,9
Context Adaptive	66	18,9

5.2.2.3 Repeated use of RSVP

There were no significant differences in comprehension for either long or short texts between using RSVP for the first, second or third time compared to using traditional text presentation. The null hypothesis regarding no difference in comprehension caused by repeated use of RSVP was thus kept.

5.2.3 Task Load

Task load was calculated as percent of millimeters to the left of the tick mark on a 120-mm scale. The factors were not rated within each other.

5.2.3.1 Long texts

The null hypothesis regarding no difference in task load between the conditions when reading long texts was rejected as all main factors except Physical demand became significant ($F[3,45] \geq 5.2$, $p \leq 0.014$). Pairwise comparisons revealed that the use of RSVP resulted in significantly higher ($p \leq 0.014$) task loads compared to using traditional text presentation with the MS Reader (Figure 13).

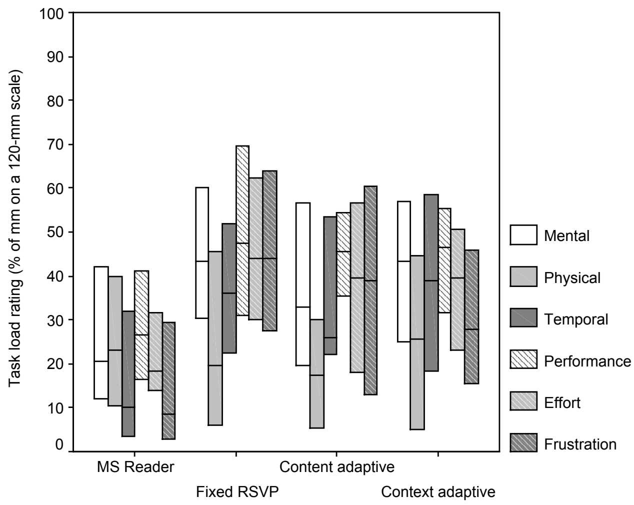


Figure 13. NASA-TLX (Task Load Index) ratings for long texts with median, 25- and 75-percentile represented. Lower ratings are better.

Content adaptive RSVP decreased task load ratings and the only factor that was rated significantly higher compared to the MS Reader was Frustration level ($p=0.002$). Context adaptive RSVP also decreased task load but in a different way. The only significantly higher factor compared to the MS Reader was Temporal demand ($p=0.001$).

5.2.3.2 Short texts

The null hypothesis regarding no difference in task load between the conditions when reading short texts was kept. All conditions showed a task load comparable to using RSVP when reading long texts (Figure 14).

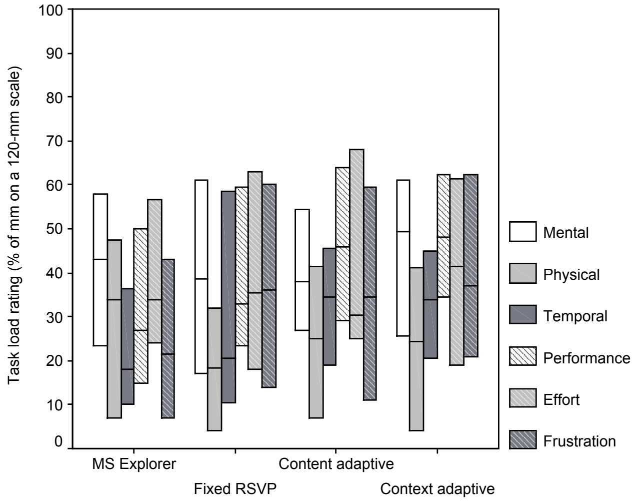


Figure 14. NASA-TLX Task Load Index ratings for short texts with median, 25- and 75-percentile represented. Lower ratings are better.

5.2.3.3 Repeated use of RSVP

The null hypothesis regarding no difference in task load caused by repeated use of RSVP when reading long texts was rejected ($F[3,45]\geq 4.5$, $p\leq 0.048$). When using RSVP for the first and second time all main task load factors except Physical demand became significant compared to the MS Reader. However, the third time RSVP was used Mental demand and perceived Performance and Effort levels decreased enough not to become significant compared to the MS Reader (Figure 15).

For short texts the pairwise comparisons between using RSVP repeatedly and using traditional text presentation did not result in any significant differences.

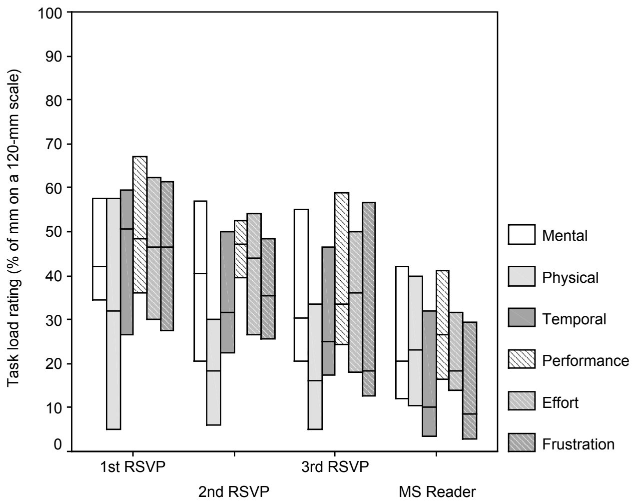


Figure 15. NASA-TLX Task Load Index ratings for repeated use of RSVP when reading long texts with median, the 25- and the 75-percentile represented. Lower ratings are better.

5.2.4 Attitude

Attitude ratings were calculated as percent of millimeters to the left of the tick mark on a 100-mm scale.

5.2.4.1 Long texts

The null hypothesis regarding no difference in attitude between the conditions when reading long texts was rejected. Five of eight main factors: Ease, Comprehension, Immersion, Comfort and Naturalness became significant ($F[3,45] \geq 8.3$ $p \leq 0.049$). Pairwise comparisons revealed that reading with RSVP resulted in lower ratings compared to using traditional text presentation with the MS Reader (Figure 16).

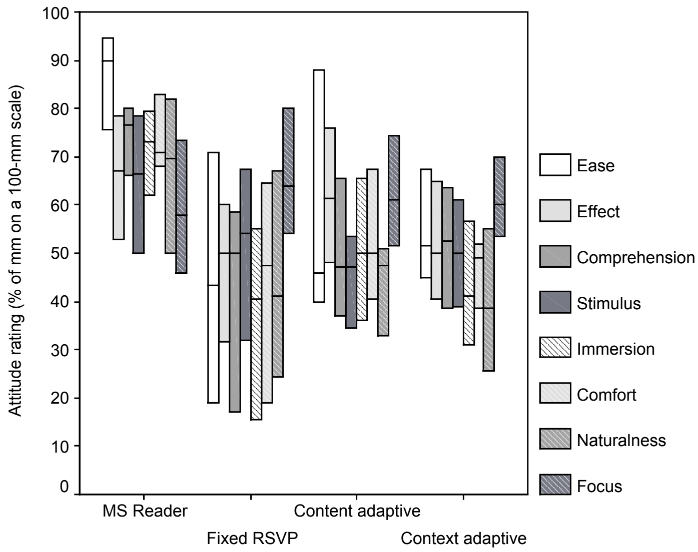


Figure 16. Attitude ratings for long texts with median, 25- and 75-percentile represented. Higher ratings are better.

Regardless of RSVP condition perceived Ease, Comprehension, Immersion and Naturalness was rated significantly lower compared to the MS Reader ($p \leq 0.049$). Perceived Comfort was however only rated significantly lower for Fixed RSVP and Context adaptive RSVP Compared the MS Reader ($p \leq 0.023$).

5.2.4.2 Short texts

The null hypothesis regarding no difference in attitude between the conditions when reading short texts was partly rejected as the main factor for Ease became significant ($F[3,45]=4.1$ $p=0.047$). Pairwise comparisons showed that the Ease of reading when using Context adaptive RSVP was rated significantly lower compared to using the MS Explorer (Figure 17).

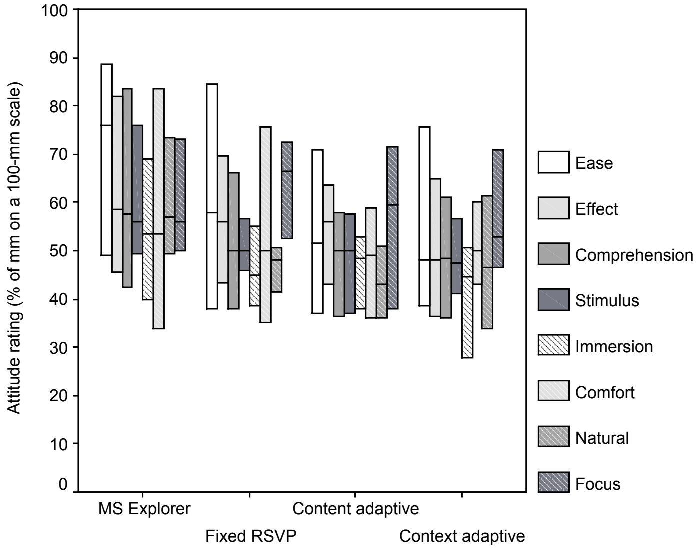


Figure 17. Attitude ratings for short texts with median, 25- and 75-percentile represented. Higher ratings are better.

5.2.4.3 Repeated use of RSVP

The null hypothesis regarding no difference in attitude caused by repeated use of RSVP was rejected (F[3,45]≥7.1 p≤0.047). Using RSVP when reading long texts for the first and second time resulted in significantly lower ratings for Ease, Comprehension, Immersion, Comfort and Naturalness compared to the MS Reader the. However, the third time RSVP was used Comprehension and Comfort ratings had increased enough to not become significant compared to the MS Reader (Figure 18).

For short texts the pairwise comparisons between using RSVP for the first, second or third time and using traditional text presentation did not result in any significant differences.

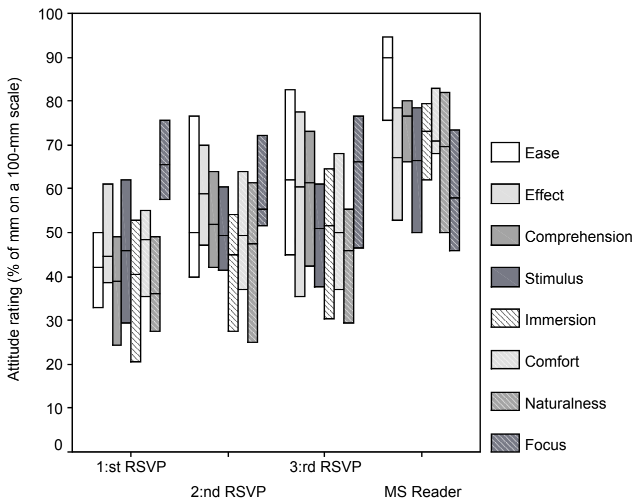


Figure 18. Attitude ratings for repeated use of RSVP when reading long texts with median, 25- and 75-percentile represented. Higher ratings are better.

6 Discussion

That no significant differences were found within the RSVP formats indicate that the effects caused by adaptation were quite small. Nevertheless, when the results obtained for RSVP were compared to those for traditional text presentation significant differences were found. The discussion will first be based on these findings and then some interface issues will be addressed. Finally, in the next section, some conclusions and directions for further work will wrap up the thesis.

6.1 Reading Speed

Since no significant differences in reading speed were found for reading long texts RSVP appears to be just as fast as traditional text presentation with the MS Reader. The lower reading speeds obtained for short texts is not very surprising as news articles are generally harder to read (Björnsson 1968; Tekfi 1987). However, the significant differences between using RSVP and the MS Explorer is very surprising since RSVP was found to increase reading speed by 33%. RSVP is usually not much faster than traditional text presentation but these findings indicate that RSVP can improve reading speed on a mobile device. That there were no significant differences in reading speed due to repeated use of RSVP might imply that training does not have any effect on reading speed. It may also indicate that the subjects selected their own natural reading speed at all times.

When reading long texts using RSVP the resulting reading speed in this evaluation, ~250 wpm, was considerably slower than what Sicheritz (2000) and Goldstein et al. (2001) reported from the Reader evaluation, ~320 wpm. The reason for this is probably that the subjects were now instructed to select a comfortable speed rather than a highest possible speed. In the author's view RSVP is a primarily a way of facilitating reading on small screens and not a way of optimising reading in itself. The average reading speed for an adult Swedish reader on paper is around 240 wpm (Björnsson 1968; Kump 1999) and in this light the obtained reading speeds are quite encouraging.

6.2 Comprehension

That no significant differences in comprehension were found is in line with findings from previous evaluations where blank screens were used (Joula et al. 1982; Masson 1983; Rahman and Muter 1999; Sicheritz 2000; Goldstein et al. 2001). The lack of differences here at least shows that adaptation does not affect comprehension in negative way.

6.3 Task Load

The task load ratings for Fixed RSVP and the MS Reader were close to identical to those obtained for Fixed RSVP and paper-book in the Reader evaluation (Sicheritz 2000; Goldstein et al. 2001). This is surprising since the subjects now selected a comfortable reading speed. This may imply that the size of the assumed trade-off between reading speed and cognitive task load is small for the RSVP format, quite contrary to the size of the well-established speed-accuracy trade-off (Wickens 1992).

Adaptive RSVP was supposed to decrease task load and it seems to have worked as expected when reading long texts. Compared to the MS Reader the only factor significantly higher for Content adaptation was Frustration level. Although the exposure times were directly related to the content some words were probably not exposed for a duration that matched the time needed for adequate cognitive processing. It is however encouraging that even the most straightforward form of adaptation actually decreased task load for many factors.

In Context adaptive mode the only significant factor compared to the MS Reader was Temporal demand. A probable cause for this is that the variations in exposure time were too large. However, the relation between what was exposed and the time for exposure was probably sound since the Frustration level decreased compared to Content adaptation. It seems that although the variations were too large they probably occurred at the right places.

Surprisingly there were no significant differences in task load when reading short texts. When RSVP was used the task load ratings were almost equal to using the MS Explorer although the reading speed was 33% higher. This indicates that traditional text presentation neither guarantees low task load nor high reading speed when screen size is limited. However, the lack of differences in task load when reading short texts may also imply that the NASA-TLX Task Load Index (Hart and Staveland 1988) does not give reliable ratings when the time for completing a task is small.

The results for repeated use of RSVP seems to support this claim since a significant decrease in task load was found for long texts but not for short. These findings imply that training only makes a difference when reading long texts and that seems contradictory. It seems more reasonable that repeated use should make it less demanding to use RSVP with time regardless of the text length. The implication of this is that the task load ratings obtained for the short texts may be unreliable and that repeated use of RSVP most likely decreases task load.

6.4 Attitude

The obtained attitude ratings in this evaluation reveal a pattern similar to what Sicheritz (2000) and Goldstein et al. (2001) reported from the Reader evaluation. The subjects performed objectively very well but rated their performance as lower than it really was. This is very likely to, at least partially, be a result of unfamiliarity with the dynamic presentation format. A good indication of this seems to be that both task load and attitude ratings improved with training.

6.5 Interface Issues Affecting Readability

The MS Reader performed much better than the MS Explorer even though both applications had interfaces tailored to fit their intended use. The sole purpose of the MS Reader is to present e-books (Hill 1999) and it seems to have an interface well customized for extended reading. The use of a scroll-bar may partially explain the lower reading speeds obtained for the MS Explorer but most likely the MS Reader just facilitated reading better.

Giving the user full control over the presentation, as the Bailando prototype does, has previously been shown to be counterproductive in terms of reading speed (Chen and Chan 1990). The obtained reading speeds in this evaluation were however faster than what both Rahman and Muter (1999), ~180 wpm, and Castellhano and Muter

(2001), ~95 wpm, reported. In both these evaluations the reader's control was restricted in order to increase speed. The results reached here do however indicate that giving the reader control over the text presentation might not be so counterproductive after all.

The text presentation window width used by Bailando was chosen on basis of the results from an earlier evaluation (Sicheritz 2000; Goldstein et al. 2001). Twentyfive characters might however still have been too much. According to eye movement research on English texts the eye can extract information in each fixation from approximately 4 characters to the left and 14 to the right from each fixation point (Rayner and Pollatsek 1989; Robeck and Wallace 1990; Rayner and Serano 1994; Rayner 1998). When Bailando was used an average of 21 characters were displayed, this may imply that the number of characters sometimes exceeded the eye's capability. In fact, as Castelhano and Muter (2001) point out, the effort required for suppressing eye movements may actually cause the increase in cognitive load when using RSVP. However, without proper eye movement recordings it is impossible to know if, why and when any unwarranted eye movements occurred.

7 Conclusions

The major drawback of RSVP appears to be the high cognitive demand placed on the subjects. An increase in task load may actually be inherent to the RSVP format since it remains high evidently independent of reading speed. Therefore, the most important findings in this evaluation are that task load actually can be decreased by using adaptation and through training.

Both adaptive algorithms were found to decrease task load for most factors. This means that better adaptations may decrease task load even further. That task load decreased with training seems to support that there is something about the RSVP format that takes some time to get used to. The high initial demand is however still likely to be a nuisance for the novel user. If adaptation can further lower initial task load, then RSVP can be considered as a serious alternative for some appliances.

There is really no reason to use RSVP when traditional text presentation can be used efficiently. In this evaluation RSVP was found to be just as effective as the MS Reader but also significantly more demanding. However, when traditional text presentation becomes ineffective it seems to become more demanding to use as well. The MS Explorer was found to be just as demanding to use as RSVP but significantly slower. It is probably when traditional text presentation, for one reason or another, becomes inefficient that RSVP can constitute a realistic alternative. Both formats are then likely to be demanding to use but compared to traditional text presentation RSVP will be more effective.

When screen space is so limited that traditional text presentation becomes ineffective, like it is on most mobile devices of today, RSVP appears to offer an improvement in readability. The inherent increase in task load may also be acceptable if it is compensated by an increase in efficiency.

7.1 Future work

Since even the most straightforward form of adaptation improved task load there is a lot of interesting work that can be done in order to further improve adaptive RSVP. Future studies on eye movements may offer some useful answers on the limits and optimal conditions for the RSVP format. It would also be interesting to see how RSVP works for different languages, especially for non-Latin languages like Chinese or Arabic. However, most exciting would be to see an evaluation where RSVP is compared to traditional text presentation on a mobile device with a smaller screen than the one used in this evaluation.

References

1. Björnsson, C.H. (1968). *Läsbarhet*, Stockholm, Liber.
2. Bruce, B., Rubin, A. and Starr, K. (1981). Why readability formulas fail. *IEEE Transactions on Professional Communication*, PC-24, 50-52.
3. Bruijn, O. Spence, R. (2000). Rapid Serial Visual Presentation: A space-time trade-off in information presentation. *In Proceedings of Advanced Visual Interfaces*, AVI2000, 189-192.
4. 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.
5. 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(3), 467-477.
6. Cocklin, T.G., Ward, N.J., Chen, H.C. and Juola, J.F. (1984). Factors influencing readability of rapid presented text segments. *Memory & Cognitio*n, 12(5), 431–442.
7. Dillon, A., Richardson, J. and McKnight, C. (1990). The effect of display size and text splitting on reading lengthy text from screen. *Behaviour and Information Technology*, 9(3), 215-227.
8. Dillon, A. (1992). Reading From Paper Versus Screens: A Critical Review of the Empirical Literature. *Ergonomics*, 35(10): 1297-1326.
9. 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.
10. Dodge, R. (1900). Visual perceptions during eye movement. *Psychological Review,* VII, 454-465.
11. 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.
12. Fine, E.M. and Peli, E. (1995). Scrolled and rapid serial visual presentation texts are read at similar rates by the visually impaired. *Journal of Optical Society of Americ*a, 12(10), 2286-2292.
13. Forster, K. I. (1970). Visual perception of rapidly presented word sequences of varying complexity. *Percep. Psychophys.* 8*,* 215-221.
14. Goldstein, M., Sicheritz, K. and Anneroth, M., (2001). Reading from a small display using the RSVP technique. *Nordic Radio Symposium,* NRS01, Nynäshamn, Sweden.
15. Goldstein, M., Öqvist, 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 2001.
16. Goldstein, M., Öquist, G. and Björk, S. (2002). Immersion does not Guarantee Excitement: Evaluating Sonified Rapid Serial Visual Presentation. Full paper submitted to *Designing Interactive Systems*, DIS2002.

17. Gould, J.D. and Grischkowsky, N. (1984). Doing the same Work with Hard Copy and with Cathode-Ray (CRT) computer terminals. *Human Factor*s, 26(3), 323-337.
18. Gould, J.D., Alfaro, L., Finn, R., Haupt, B. and Minuto, A. (1987). Reading From CRT Displays Can Be as Fast as Reading From Paper. *Human Factors*, 29(5), 497-517.
19. Granaas, M. M., McKay, T. D., Laham, R. D., Hurt, L. D., & Juola, J. F. (1984). Reading moving text on a CRT screen. *Human Factors*, 26(1), 97-104.
20. 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.
21. Hill, B. (1999). *The Magic of Reading*. Available at: http://slate.msn.com/ebooks/ (December 2001).
22. Huey, E.B. (1968). *The psychology and pedagogy of reading.* Cambridge, MA: MIT Press. (Originally published 1908).
23. Jackson, M.D. and McClelland, J.L. (1979). Processing determinants of reading speed. *Journal of experimental psycholog*y, 108, 151–181.
24. 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.
25. Juola, J.F., Tiritoglu, A., and Pleunis, J. (1995). Reading text presented on a small display. *Applied Ergonomics*, 26, 227-229.
26. Judd, C.H. and Buswell, G.T. (1922). *Silent reading: A study of the various types.* Chicago, IL: University of Chicago Press.
27. Just, M. A., and Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329-354.
28. Just, M.A., Carpenter, P.A. and Masson, M.E.J. (1982). *What eye fixations tell us about speed-reading and skimming.* (Eye-lab Technical Report) Carnegie-Mellon University.
29. Kang, T.J., and Muter, P. (1989). Reading Dynamically Displayed Text. *Behaviour & Information Technology*, 1989, 8(1), 33-42.
30. Klare, G. R. (1984). Readability. In *Handbook of Reading Research,* P.D. Pearson (Ed.), Longman Inc, New York, 681-743.
31. Kump, P. (1999). *Break-trough rapid reading.* New Jersey: Prentice-Hall Press.
32. Masson, MEJ. (1983). Conceptual processing of text during skimming and rapid sequential reading. *Memory and Cognition*, 11, 262-274.
33. McCrickard, D.S., Catrambone R. and Stasko, J.T. (2001). Evaluating Animation in the Periphery as a Mechanism for Maintaining Awareness. *In proceedings of INTERACT 2001*, 148-156.
34. Mills, C.B. and Weldon, L.J. (1987). Reading text from computer screens. *ACM Computing Surveys,* 19(4), ACM Press.
35. Muter, P., Kruk, R. S., Buttigieg, M. A., and Kang, T. J. (1988). Reader-controlled computerized presentation of text. *Human Factors*, 30, 473-486.

36. Muter, P. and Maurutto, P. (1991). Reading and skimming from computer screens and books: The paperless office revisited? *Behavior & Information Technolog*y, 10, 257–266.
37. 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.
38. Osborne, D.J. and Holton, D. (1988). Reading From Screen Vs. Paper: There Is No Difference. *International Journal of Man-Machine Studies*, 28, 1-9.
39. Paulson, L.J. and Goodman, K.S. (1999). *Influential Studies in Eye-Movement Research*. International Reading Association. Available at: http://www.readingonline.org (December 2001).
40. Potter, M. C. (1984). Rapid Serial Visual Presentation (RSVP): A method for studying language processing. In *New Methods in Reading Comprehension Research.* D.E., Kieras and M.A., Just (Eds.). Hillsdale, N.J: Erlbaum.
41. Proctor, R.W. and Proctor, J.D. (1997). Sensation and Perception. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics.* Second Edition, Wiley-Interscience, New York, 53-57.
42. 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.
43. Rayner, K. and Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, New York: Prentice Hall.
44. Rayner, K. and Sereno, S.C. (1994). Eye movements in reading: Psycholinguistic studies. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics*. San Diego, CA: Academic.
45. Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372-422.
46. Redish, J. (1981). Understanding the limitations of readability formulas. *IEEE Transactions on Professional Communication*, PC-24, 46-48.
47. Reichle, E.D., Rayner, K. and Pollatsek, A. (2000). *Comparing the E-Z Reader Model to Other Models of Eye Movement Control in Reading*. CogPrints electronic archive for papers. Available at: http://cogprints.soton.ac.uk/ (December 2001).
48. Reilly, R. (1993). A connectionist framework for modeling eye-movement control in reading. In G. d'Ydewalle & Van Rensbergen (Eds.), *Perception and Cognition: Advances in Eye Movement Research,* 191-212. Amsterdam: North Holland.
49. Robeck, M.C. and Wallace, R.R. (1990). *The Psychology of Reading: An Interdisciplinary Approach,* Second edition, Lawrence Erlbaum Associates, Hillsdale: New Jersey.
50. Russel, M., Hull, J. and Wesley, R. (2001). Reading with RSVP on a Small Screen: Does Font Size Matter? *Usability News,* Winter 2001, Software Usability Research Laboratory, Wichita State University.
51. Shneiderman, B. (1998). *Human-Computer Interactio*n, 3rd ed., Addison Wesley Longman, Inc, 412-414.
52. Sekey, A. & Tietz, J. (1982). Text display by saccadic scrolling. *Visible Language*, 16, 62-76.

53. Sicheritz, K. (2000). *Applying the Rapid Serial Presentation Technique to Personal Digital Assistants,* Master's Thesis, Department of Linguistics, Uppsala University. Available at: http://stp.ling.uu.se/ (December 2001).
54. Taylor, I. and Taylor, M.M. (1983). *The Psychology of Reading*. Toronto, Academic Press.
55. Tekfi, C. (1987). Readability formulas: An Overview, *Journal of Documentation,* 43(3) 261-73.
56. Wickens, C. D. (1992). *Engineering psychology and human performance*, 2nd edition, Chapter 8, Harper Collins Publishers Inc., New York.
57. Öquist, G. and Goldstein, M. (2002). Towards an improved readability on mobile devices: Evaluating Adaptive Rapid Serial Visual Presentation. Full paper submitted to *Advanced Visual Interfaces*, AVI2002.

Web resources

1. A web-based LIX counter developed by the author can be found at: http://stp.ling.uu.se/~gustav/lixcounter/ (December 2001).
2. The specifications of the eXtensible Markup Language (XML) can be found at: http://www.w3.org/XML/ (December 2001).
3. Product specifications for the Compaq iPAQ can be found at the following address: http://www.compaq.com/ipaq (December 2001).
4. Product specifications for Microsoft Embedded Visual C++ can be found at: http://www.microsoft.com/mobile/developer/ (December 2001).
5. Information about the Microsoft Reader and ClearType technology can be found at: http://www.microsoft.com/reader/ (December 2001).
6. Information about Microsoft Internet Explorer can be found at: http://www.microsoft.com/mobile/ (December 2001).
7. The registration page for participation in the usability evaluation can be found at: http://stp.ling.uu.se/~gustav/readability/ (December 2001).

Appendix

A Excerpt from a Bailando log file

System:	Bailando initierad							
Settings:	setwpm=250 avgword=7.8 avgsent=11.5 adaption=1							
Aktion:	Filen: Bara Alice öppnad							
Fildata:	Type=RSVPxml Rating=Easy Sent=310 Word=4172 Long=693 Char=18648 Lix=29 Rwpm=283							
Aktion:	Adaption satt till 2							
Aktion:	Läs							
Legend:	Text	Exposure	Act wpm	Avg wpm	Set wpm	Adapt.	Words	Chars
Visar:	Det enda Alice kan se av	695	517	517	250	2	6	24
Visar:	sin far är sulorna på	824	364	434	250	2	5	21
Visar:	hans skor. De är bleknat	1261	237	345	250	2	5	24
Visar:	bruna, repade av grus	689	348	345	250	2	4	21
Visar:	från de trottoarer där	778	308	339	250	2	4	22
Visar:	han har gått. Alice	1016	236	319	250	2	4	19
Visar:	brukar få lov att springa	711	421	331	250	2	5	25
Visar:	på trottoaren utanför	820	219	317	250	2	3	21
Visar:	deras hus för att möta	843	355	322	250	2	5	22
Visar:	honom när han kommer hem	970	309	320	250	2	5	24
Visar:	från arbetet på	675	266	316	250	2	3	15
Visar:	kvällarna. När det är	1126	213	305	250	2	4	21
Visar:	sommar springer hon	532	338	307	250	2	3	19
Visar:	ibland där i bara	542	442	313	250	2	4	17
Visar:	nattlinnet, och dess veck	1102	217	305	250	2	4	25
Aktion:	Pause							
Visar:	fastnar runt hennes knän.	1497	160	289	250	2	4	25
Aktion:	Gå bakåt							
Visar:	hon ibland där i bara	457	656	247	250	2	5	22
Aktion:	Gå bakåt							
Visar:	nattlinnet, och dess veck	1033	232	246	250	2	4	25
Aktion:	Gå bakåt							
Visar:	När det är sommar	346	693	211	250	2	4	18
Aktion:	Sänker till 240 wpm							
Aktion:	Läs							
Visar:	honom när han kommer hem	572	524	179	240	2	5	25
Visar:	från arbetet på	565	318	183	240	2	3	15
Visar:	kvällarna. När det är	1073	223	186	240	2	4	21
Visar:	sommar springer hon	567	317	190	240	2	3	19
Visar:	ibland där i bara	578	415	196	240	2	4	17
Visar:	nattlinnet, och dess veck	1176	204	197	240	2	4	25
Visar:	fastnar runt hennes knän.	1580	151	193	240	2	4	25
Visar:		250	0	191	240	2		
Visar:	Men nu är det vinter.	853	351	197	240	2	5	21

B Instructions

Välkommen

Du kommer under dagens försök att få läsa ett flertal texter under olika former och svara på frågor i anslutning till detta. Det är viktigt att du är medveten om att det inte är dig vi testar utan en applikation, du behöver därför inte känna någon press att göra något på ett särskilt sätt. Vi vill helst att du använder applikationerna och läser så som du skulle ha gjort i en helt vanlig vardagssituation.

Du får själv välja i vilken takt du ska läsa. Välj den hastighet som är så naturlig och behaglig som möjligt för dig. Du får närsomhelst ändra hastigheten under tiden att du läser. Du kan närsomhelst avbryta ett försök om du skulle uppleva någon form av obehag. Om du har några frågor eller stöter på problem är det bara att fråga närsomhelst under försöken.

Under experimentet kommer du att filmas och interaktionen mellan dig och applikationen att loggas, detta är för att vi senare ska kunna gå tillbaka och se på i detaljer hur det fungerat. Ingen annan än Ericsson Usability & Interaction Lab kommer att ta del av det videoinspelade materialet utan ditt uttryckliga medgivande.

Du kommer att få börja med att öva att läsa en text i två olika applikationer så att du kan vänja dig vid dessa. Sedan kommer du att få läsa fyra längre texter och fyra kortare artiklar under olika förutsättningar.

Nedan ser du ungefär hur dagen är upplagd:

Förberedelse / övning	9.00 - 9.30
Första betingelsen	9.30 - 10.30
Fika	10.30 - 10.45
Andra betingelsen	10.45 - 11.45
Lunch	11.45 - 12.30
Tredje betingelsen	12.30 - 13.30
Fika	13.30 - 13.45
Fjärde betingelsen	13.45 - 14.45
Uppföljning	14.45 - 15.00

Börja med att fylla i de formulär som finns i detta häfte och så kan vi börja med övningarna, vi hoppas att du kommer att få en trevlig dag hos oss!

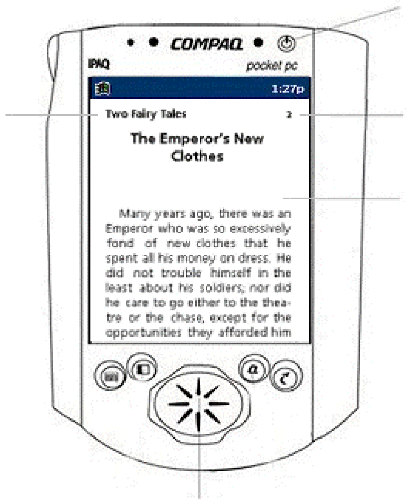
B1 Microsoft Reader:

Instruktioner

Du kommer nu att få läsa en längre text på ca. 4000 ord vilket tar ungefär 20 minuter. Efter att du läst texten ska du svara på frågorna på nästföljande sidor.

Kom ihåg att du ska läsa med den hastighet som du finner så naturlig och behaglig som möjligt.

Nedan ser du hur applikationen ser ut och vilka knappar som finns:



Börja inte innan du fått klartecken och får du problem eller undrar över något så är det bara att fråga.

Lycka till!

Betingelse: B1 Text: T2 Person: P7 Moment: M7

B2 Microsoft Internet Explorer

Instruktioner

Du kommer nu att få läsa en kortare artikel på ca. 500 ord vilket tar ungefär 2 minuter. Efter att du läst texten ska du svara på frågorna på nästföljande sidor.

Kom ihåg att du ska läsa med den hastighet som du finner så naturlig och behaglig som möjligt.

Nedan ser du hur applikationen ser ut och vilka knappar som finns:

Börja inte innan du fått klartecken och får du problem eller undrar över något så är det bara att fråga.

Lycka till!

Betingelse: B1 Text: A2 Person: P7 Moment: M8

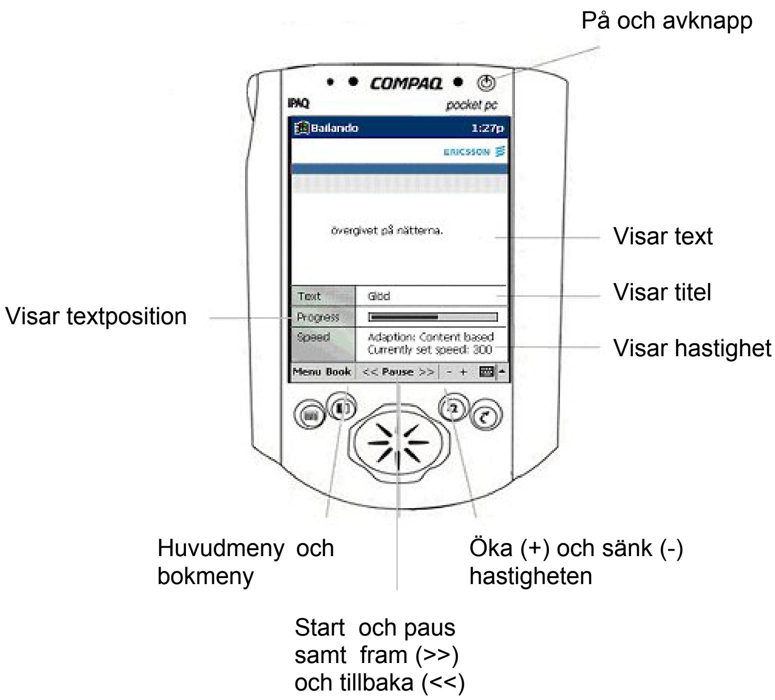
B3 Bailando

Instruktioner

Du kommer nu att få läsa en längre text på ca. 4000 ord vilket tar ungefär 20 minuter. Efter att du läst texten ska du svara på frågorna på nästföljande sidor.

Kom ihåg att du ska läsa med den hastighet som du finner så naturlig och behaglig som möjligt.

Nedan ser du hur applikationen ser ut och vilka knappar som finns:



Börja inte innan du fått klartecken och får du problem eller undrar över något så är det bara att fråga.

Lycka till!

Betingelse: B4 Text: T1 Person: P7 Moment: M5

C Inventories

C1 Comprehension

Förståelse

Instruktioner: Välj det svar som du anser är riktigt på var och en av de 5 frågorna nedan.

1. Vad var klockan på morgonen då Thomas Murillo sköt ner rånarna?
   - Tre
   - Fyra
   - Två

2. Hur gammal var Thomas Murillo när han sköt ner rånarna?
   - 46 år
   - 57 år
   - 68 år

3. Hur många inbrott hade han haft tidigare?
   - Fem
   - Tre
   - Två

4. Vad tände han medan han väntade på rånarna?
   - En Kuban
   - En Silviana
   - En Kruger

5. Hur många skott sköt Thomás mot rånarna?
   - Två
   - Tre
   - Fyra

Betingelse: B4 Text: T1 Person: P7 Moment: M5

C2 NASA-TLX (Task Load Index)

Instructions

Place a mark on each scale below that represents the magnitude of each factor in the task you just performed!

1. Mental demand

How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exact or forgiving?

Low ______________________________ High

2. Physical demand

How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Low ______________________________ High

3. Temporal demand

How much time pressure did you feel due to the rate or pace at which the task or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

Low ______________________________ High

4. Performance

How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

Good ______________________________ Poor

5. Effort

How hard did you have to work (mentally and physically) to accomplish your level of performance?

Low ______________________________ High

6. Frustration level

How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Low ______________________________ High

Betingelse: B4 Text: T1 Person: P7 Moment: M5

C3 Attittude

Värdering

Instruktioner: Markera för varje skala nedan dina skattningar för den nyss genomförda uppgiften.

1. Hur upplevde du svårighetsgraden när du läste på detta sätt?

   Låg ——————|—————— Hög

2. Hur upplevde du effektiviteten när du läste på detta sätt?

   Låg ——————|—————— Hög

3. Hur upplevde du läsförståelsen när du läste på detta sätt?

   Låg ——————|—————— Hög

4. Hur upplevde du stimulansen när du läste på detta sätt?

   Låg ——————|—————— Hög

5. Hur upplevde du inlevelsen när du läste på detta sätt?

   Låg ——————|—————— Hög

6. Hur upplevde du bekvämligheten när du läste på detta sätt?

   Låg ——————|—————— Hög

7. Hur upplevde du naturligheten när du läste på detta sätt?

   Låg ——————|—————— Hög

8. Hur upplevde du koncentrationen när du läste på detta sätt?

   Låg ——————|—————— Hög

Betingelse: B4 Text: T1 Person: P7 Moment: M5