Selina Sharmin

Eye Movements in Reading of Dynamic On-screen Text in Various Presentation Formats and Contexts

ACADEMIC DISSERTATION
To be presented with the permission of the School of Information Sciences of the University of Tampere, for public discussion in Pinni auditorium B1097 on April 17th, 2015, at noon.

School of Information Sciences
University of Tampere

Dissertations in Interactive Technology, Number 21
Tampere 2015
The originality of this thesis has been checked using the Turnitin OriginalityCheck service in accordance with the quality management system of the University of Tampere.
Abstract

Reading of on-screen text is one of the most prevalent interactions between humans and electronic devices. Eye movement data in reading can be a useful source of information about readers’ cognitive situation during processing of the text. On account of the rapid development of technologies and electronic devices, there is great diversity in the use of electronic text. The main aim with the dissertation is to observe, analyze, and discuss eye movements in reading of on-screen text within various reading circumstances, such as reading with diverse presentation formats and with variety in context.

The reading process is related to the overall design of the text, encompassing elements such as presentation format, display strategy, and font size. The text formats considered in the dissertation range from letter-by-letter presentation to a larger document that requires scrolling. The researchers examined eye movements in reading when the text was presented as a full paragraph, sentence by sentence, line by line with segmentation dictated by the width of the screen, and in chunks about 30 characters in length. Also eye movements were examined in reading of dynamic text rendered word-by-word and letter-by-letter as it was being typed. Furthermore, eye movements in reading of long pieces of text that required multiple instances of scrolling were studied with gaze-based automatic scrolling actions.

Along with various text presentation formats, the dissertation considers several reading contexts; it examined eye movements in reading for comprehension, print interpretation, and translation. Print interpretation is a means of communication for hard-of-hearing people wherein spoken utterances are typed in real-time and provided to the clients as text to read. Translation too generally involves typing text in a dedicated part of the screen. Accordingly, the text in both print interpretation and translation is created, not just rendered, dynamically.

The dissertation, a summarizing synthesis of six publications, indicates that eye movements are greatly influenced by the design of the text presentation format in various contexts. As the format unit shrank from the largest presentation unit (the paragraph) to the smallest unit, eye movement metrics showed significant changes. In addition, rendering format between word-by-word and letter-by-letter was found to have a significant impact on eye movement measurements when print interpreted text was read. Moreover, eye movement metrics varied also when the text was presented in different font sizes. Besides presentation
formats, translators’ eye movements affected significantly in translating text with different time pressure and varying level of text complexity involved.

The research confirmed that people have preferred reading regions on the screen. With long pieces of text, gaze-enhanced auto-scrolling did not affect reading strategy, an encouraging result. Reading of text within the preferred reading region with auto-scrolling was reasonably satisfactory for novice users.
Acknowledgements

This dissertation is an outcome of cumulative works that have been carried out in the Visual Interaction Research Group. I enjoyed the time while working in the EU project Eye-to-IT. Thanks to all the project members and Prof. Arnt Lykke Jakobsen from the Copenhagen Business School, Denmark for his tender cooperation. He was also a co-author in one of the articles included in this thesis. It was my pleasure to work with the colleagues from the University of Helsinki for SpeechText project. I like to say thank you to the project coordinator Prof. Liisa Tiittula from the University of Helsinki. I am glad to work with Mari Wiklund from the University of Helsinki, with whom we published one article.

Part of this work was conducted with the funding from the Graduate School in User-Centered Information Technology (UCIT). I am thankful for the funding which helped me to concentrate on my task. I would like to convey my thanks to TAUCHI professors and researchers with whom I spent quality time. Thanks to Yulia who sometimes provided useful tips in writing this dissertation and also for playing badminton occasionally. Thanks to Outi and Hanna for nice chatting. Thanks to Aulikki for her supportive words during my early research life. I also like to thank Päivi for her cooperation. I am grateful to our helpful administrative staff, specifically, thanks to Tuula and Minna from whom I got support in need every now and then.

I also acknowledge the reviewers of the dissertation Prof. Kenneth Holmqvist, Lund University, Sweden and Docent Roman Bednarik, University of Eastern Finland, Finland for their constructive and convincing opinion. It was my pleasure to work with my colleague and co-author Dr. Oleg Špakov. His super performance always inspired me to focus on my own.

I would like to express my profound gratitude for my supervisor Prof. Kari-Jouko Räihä for his endless support, friendly guidance, and valuable comments. This long process would not be ended without his encouragement, assessment, and patience.

All my friends and relatives all over the world inspired me in many different ways to go through this long journey. I am thankful to Sultana for being with me in crucial moments.

My deepest gratitude goes to my family members. I learnt the value of education from my parents Mr. Abdul Wahab and Mrs. Satara Begum.
Their valuable advices in phrases always inspired me and took me this far. Thanks to all my elder siblings Liton, Ripon, Annie, Jenny, and Sohel for being always beside me in joy and sorrow along this journey. I am thankful to my wonderful In-laws for their warm support.

I like to express my endless gratitude to my loving husband Miraj Mostafa who took every single pain of this long process and showed extreme patience. Thanks for being with me and providing emotional support and cooperation. My loving thanks to our twins Mimin Odri Mostafa and Simin Ratri Mostafa, and son Alif Rodro Mostafa, who taught me the meaning of life. Final thanks to the Almighty for everything.

Tampere; February 19, 2015

Selina Sharmin
# Contents

1 **INTRODUCTION** ................................................................................................................................. 1

2 **EYE TRACKING AND EYE MOVEMENTS IN READING** .......................................................... 7  
   2.1 Basic Concepts in Eye Tracking............................................................................................... 7  
   2.2 Eye Movements in Reading ................................................................................................. 10

3 **ON-SCREEN READING AND TEXT PRESENTATION FORMATS** ...................................... 17  
   3.1 Display Screens for Reading of Electronic Text ................................................................. 18  
   3.2 Dynamic Text Presentation Formats ............................................................................... 20  
   3.3 Previous Research on On-Screen Presentation Formats ................................................. 23

4 **READING IN VARIOUS CONTEXTS** ......................................................................................... 29  
   4.1 Reading of Print Interpreted Text ...................................................................................... 29  
   4.2 Reading for Translation ...................................................................................................... 32  
   4.3 Reading Long Documents and Scrolling .......................................................................... 36

5 **SUMMARY OF THE PUBLICATIONS** ................................................................................... 39  
   5.1 Paper I: “The Effect of Different Text Presentation Formats on Eye Movement Metrics in Reading” ................................................................................................................... 39  
   5.2 Paper II: “Dynamic Text Presentation in Print Interpreting – an Eye Movement Study of Reading Behaviour” .................................................................................................................. 41  
   5.3 Paper III: “Gaze Behaviour and Linguistic Processing of Dynamic Text in Print Interpreting” .................................................................................................................................... 43  
   5.4 Paper IV: “Where on the Screen Do Translators Look While Translating, and for How Long?” ........................................................................................................................................... 44  
   5.5 Paper V: “Reading On-Screen Text with Gaze-based Auto-Scrolling” ... 47  
   5.6 Paper VI: “Gaze-Contingent Scrolling and Reading Patterns” ..................................... 49

6 **DISCUSSION** ................................................................................................................................. 51

7 **CONCLUSIONS** ............................................................................................................................ 59

REFERENCES ................................................................................................................................................. 61
List of Publications

This thesis is composed of a summary and the following original publications, reproduced here by permission.


The Author’s Contribution to the Publications

All publications included in this thesis were co-authored with colleagues, my supervisor, or both. Each work was a collaborative effort of the authors. The research work presented in papers I and IV was conducted in the EU project Eye-to-IT. The articles were the outcome of joint work by the author of this thesis; Prof. Kari-Jouko Räähä; Dr. Oleg Špakov, from the University of Tampere; and Prof. Arnt Lykke Jakobsen from the Copenhagen Business School, Denmark. This author was responsible for conducting the empirical work, collecting and analyzing the research data, and writing the first draft of each publication.

Research included in papers II and III was conducted in the context of the SpeechText project in collaboration with colleagues from the University of Helsinki. Mari Wiklund, co-author of Paper III, is from the University of Helsinki. As in the other research work for the thesis, this author was responsible for carrying out the empirical research work, including data analysis, and writing the first draft of each article.

For papers V and VI, about reading of on-screen text in conjunction with auto-scrolling, this author contributed to the design of the empirical work, conducting of the experiment, collection of data, and analysis and reporting of the results. The present author wrote the first draft of Paper V.
1 Introduction

Reading is a part of our daily activities. We read text for various purposes like gaining knowledge, interest, translation, comprehension, correction, communication and so on. Eye movement data can provide a powerful source of information that aids in monitoring of the reading process.

Reading on-screen text is one of the most common interactions between humans and computers. The reading process is an activity involving integration, entailing the perception of form and content. It is especially the perception of form that is bound up with the overall design of the text, such as presentation format, font size, and display strategy. Text format has a huge effect on the quality of reading and the perception of text content. Recent rapid development of technology and electronic devices has brought great diversity to use of electronic text. For instance, it appears as short messages and lengthy documents, on Web pages and in video subtitles.

Eye movements in reading have been studied for a long time. It has been suggested that Helmholtz and Hering were among the first researchers who studied the understanding of eye movements in reading. Their work was conducted in the second half of the nineteenth century (Wade & Tatler, 2011). At roughly the same time, in 1879–1920, Javal also initiated research in eye movements, specifically looking at the role of eye movements in reading (Rayner, 1998).

To understand eye movements, one must track the eyes. In the very early years of this research field, Javal’s student Lamare used a tube placed over the eyelid from which the observer could hear a sound whenever the eyes moved (Wade & Tatler, 2011), while Huey’s (1908/1968) eye tracker used
a contact lens connected to an aluminum pointer that moved in response to the movement of the eye. All eye trackers in that era were intrusive and very different from those we use today. Work to develop non-intrusive eye trackers began in the late nineteenth and early twentieth century and allowed investigation of the fundamental nature of eye movements (Venezky, 1977). Since then researchers have discovered that eye movements are not as smooth and continuous as they seemed (Wade & Tatler, 2011). The development of eye tracking systems is discussed in Section 2.1.

Eye tracking provides eye movement data that reveal gaze behavior. In reading, the eyes make brief jumps along the line of text (Rayner, 1998). Rapid movements of the eyes are called saccades, while the stops between saccades are known as fixations. It is noteworthy that the movement within a line of text is not always forward motion. Saccades that move backward for rereading are called regressions. Research in Javal’s time uncovered many basic facts about eye movements in reading, including the crucial fact that no information is perceived during saccadic eye movements (Rayner, 1998).

Standard metrics in gaze data analysis are number of fixations and duration of fixations (Jacob & Karn, 2003). Other metrics can be derived from the fixation data, such as gaze duration (sum of several fixations), regression path reading time (the total duration of all regressions within a certain region), first-pass rereading time, and so on (Hyönä, Lorch, & Rinck, 2003). More detailed information on eye movement metrics in gaze data analysis is provided in Section 2.2.

The eye movements in reading can be influenced by the content of the text. The duration and length of both fixations and saccades depend mostly on the textual and typographic variables, just as much as on the overall ease or difficulty of understanding the text to be read (Rayner, 1998). Previous research has established that as text becomes conceptually more difficult to read, eye movement metrics change: fixation duration grows, saccade lengths decrease, and the frequency of regressions increases (Rayner & Pollatsek, 1989; Jacobson & Dodwell, 1979). At the same time, factors such as font size, line length, and letter spacing influence eye movements (Morrison & Inhoff, 1981; Kolers, Duchnicky, & Ferguson, 1981). Moreover, eye movement metrics are affected by the characteristics of the writing system, such as horizontal or vertical orientation of the text presented on the screen (Peng, Orchard, & Stem, 1983; Osaka, 1989; Sun, Morita, & Stark, 1985). Eye movement data have also shown that reading the same text several times produces faster reading, with a reduction in fixation duration, an increase in saccade size, and lower numbers of regressions (Raney & Rayner, 1995; Inhoff, Topolski, Vitu, & O’Regan, 1993; Hyönä & Niemi, 1990; Hyönä, 1995).
Clearly, the eyes’ behavior in reading varies: fixations differ in duration, regression counts vary, and so on. In addition to the complexity or simplicity of the text and its familiarity, gaze behavior in reading can be affected by time pressure and the structure of the text.

Reading from a screen tends to involve especially small and large display areas. Varied size of the display screens may require adaptation of the structure of the text that is presented for reading, with the adjustments based on the circumstances. Often we read static or dynamic text depending on the device from which we are reading and also on the purpose of reading. The reading device, specially its screen size, may have huge effects on eye movements. For example, we may expect eye movements with static text to be different from those in reading of text that appears dynamically on the screen. Issues of on-screen text presentation format are explored in depth in Chapter 3.

On-screen reading encompasses reading for comprehension, communication, and translation. In reading for communication, one reads generated or transformed text as a conveyed message of/about an utterance. In reading for translation, a translator reads the given text and produces appropriate rendering of that text in another language. Such different reading purposes may produce different eye movement patterns and thereby provide fruitful ground for investigation. Reading of on-screen text in different contexts is considered further in Chapter 4.

Although much previous research has been conducted to study eye movements during reading, considerable scope remains for examining how they differ between individual purposes for reading. Reading print interpreted text involves dynamic text. Print interpreting is a communication aid for hard-of-hearing and late-deafened people where spoken language appears as text and is displayed on a screen in real-time simultaneously with the speech. In print interpreting, letter-by-letter presentation format is the most familiar way for the text to be displayed on the screen: letters appear as the text is written. Reading for translation involves mostly static text, although reading the translator’s own emerging text involves dynamic rendering. In both cases, there is potential in studying the gaze behavior during reading of the dynamic text with different presentation formats. Also, on-screen reading often involves long documents that require scrolling to make new text visible. Since the text moving automatically might simplify the scrolling process, it may be useful to study how eye movements behave in reading with automatic gaze-based scrolling technique.

The aforementioned conditions motivated us to investigate eye movements in various cases, for greater insight into gaze behavior during reading. Eye tracking during reading to observe and investigate readers’ eye movements has already begun to contribute concretely to diverse
fields of applied research, such as clinical neuroscience (Schattka, Radach, & Huber, 2010), and training in elementary school education (Lehtimäki & Reilly, 2005; Radach, Schmitten, Glover, & Huestegge, 2009). The research carried out for this dissertation examined several on-screen reading phenomena in various contexts: reading dynamic text in diverse presentation formats, reading print interpreted text, reading during translation, and reading with scrolling that is performed either automatically on the basis of gaze position or manually.

The primary objective for the dissertation was to observe, analyze, and discuss eye movements in reading of on-screen text in various reading circumstances. Our general hypothesis was that the presentation format, the content read, and the purpose of the reading all influence eye movements significantly. To test this hypothesis, we set several research questions and ran multiple experiments accordingly. Therefore, the objective for the dissertation was to find answers to the following questions.

How do eye movements vary in reading of text with different presentation formats?

Improvement of text presentation formats especially for various kinds of display devices is an ongoing process aimed at facilitating day-to-day reading. One of our objectives was to verify the eye movement characteristics seen in reading of text with various presentation formats as the amount of text shown changes.

How does reading of dynamic print interpreted text affect eye movement metrics?

Print interpreting provides dynamic text to be read. We set out to compare eye movements in reading of print interpreted text with two types of presentation format and also to find feasible presentation formats for particular cases.

How do translators’ eye movements behave during reading for translation?

Translators read static text to be translated and also read the emerging, translated text produced. We analyzed how translators’ eye movements differ in reading under various time constraints and with several levels of text complexity.

Is there any preferred region of the screen in reading, and how does gaze-enhanced automatic scrolling affect eye movements?

Scrolling is often needed for reading of text that does not fit within the display area. We sought to find out how eye movements differ between automatic gaze-enhanced scrolling and manual scrolling. We also tried to determine whether people have a preferred reading region on the screen, information that could be applied as input to guide the automatic scrolling
to facilitate reading further. Finally, we investigated reading patterns with text that is scrolled automatically.

The research method employed multiple experiments addressing these four research questions. The experiments used several eye tracking devices: A Tobii 1750 remote eye tracking device tracked users’ gaze for the study of eye movements in reading for translation and in reading with various text presentation formats. Users’ eye movements in reading of print interpreted text were recorded with a Tobii T60 remote eye tracking device, which was also used to track eye movements in reading of longer pieces that require scrolling. The in-house software tool Sprintanium (Špakov, 2011) was used to prepare the stimulus with print interpreted text. Tobii Studio was used to record and collect eye movement data.

Each experiment design followed appropriate research methods. The necessary pilot tests were conducted before the proper actual experiment. Pre and post-test questionnaires were prepared for each experiment on the basis of the hypotheses set for the individual research questions. Numerical data were analyzed with the software MS Excel and SPSS tools.

The dissertation has been written to shed light on eye movement behavior during reading conducted for various purposes. It is hoped that the findings on the aforementioned research questions will ultimately lead to further analysis of human gaze behavior during reading and offer knowledge of diverse reading contexts that contributes to development of better tools, thereby supporting new technologies and methods that produce direct social benefit. The implications of the findings are discussed in more detail in Chapter 6.
2 Eye Tracking and Eye Movements in Reading

2.1 Basic Concepts in Eye Tracking

Eye tracking is a technique used to determine where a person’s visual focus is directed. It is a process of measuring either the gaze point or the movement of an eye relative to the head. Eye tracking does not disrupt our normal perceptual process as many other methods, such as application of think-aloud protocols, do. Hence, eye tracking frees the observers to examine any parts of the object or information displayed within the stimulus in any order without interruption by a secondary task.

A general indication of eye movements can be obtained by direct visual observation of the eye. Eye trackers reveal the features of eye movements in greater detail. Research examining psychology, product design, the visual system, usability, cognitive linguistics, user interfaces, and others uses eye trackers. Over the decades, various types of trackers have been used by researchers, employing a number of methods for measuring eye movements.

In the early stage, eye tracking research aimed at revealing eye movements, and several authors used mechanical methods to connect the eye and the recording system. Over a century ago, in the late 1800s, Javal used a mirror to observe eye movements. That mirror only enabled the measurement of large movements involving coordinated actions by the eyes, hand(s), and body. However, smaller eye movements are made with the eyes alone. Therefore, Javal’s method was unable to detect rotation of
the eye through 1° and corresponding movement on the retina through 0.2 mm (Yarbus, 1967).

The very first eye tracking devices providing accurate data were uncomfortable and were highly invasive (Majaranta, 2009). For example, Delabarre (1898) developed an eye tracking system that used eye cups (made of plaster of Paris or aluminum) with an extended lever to draw the eye movements on a smoked drum. The cup had a hole through which the test participants could see, and the cup was directly attached to the surface of the eye (Wade & Tatler, 2005). Another historical method consisted of wearing large sized contact lenses featuring a metal coil (Duchowski, 2007). Those lenses covered the cornea and sclera of the eye. As soon as the eyes moved, the metal coil also moved, and provided fluctuations in an electromagnetic field by means of which eye movements were measured.

These methods’ low accuracy, invasiveness, and complicated setup prompted developments that soon left them outdated. In the early 1900s, Dodge and Cline developed the first non-invasive eye tracking apparatus (Wade & Tatler, 2005; Holmqvist et al., 2011), which was based on photography and light reflected from the cornea of the eye. In recent decades, eye movement research has been revolutionized through developments in computer and video technologies (Duchowski, 2007).

After the early investigations, the latter developments meant that by the late 1960s, eye movement recording apparatuses had become much more sophisticated, although their basic principles of operation were similar to those of earlier equipment. Video images of the eye are used in the most modern eye tracking systems to determine where someone is looking. In addition, various distinguishing features of the eye, such as corneal reflections (also known as Purkinje images), the iris–sclera boundary, and apparent pupil shape can be used for inference of the point of focus (Duchowski, 2007). Advances in imaging sensors, video cameras, and image processing systems have made eye tracking systems one of the fields with the greatest promise for improved human–computer interaction. Nowadays, eye tracking systems capture the image of the eyes by using one or more cameras at a sampling rate of 30 to over a thousand frames per second (Feng, 2011). To facilitate tracking, newer eye tracking techniques increase the contrast between the pupil and the iris by means of mainly infrared or near-infrared light.

Often, eye trackers include a standard desktop computer with an infrared camera attached below or to the side of the monitor. The eye features and movements that are used for tracking are identified and located by means of image processing software. Infrared light is directed into the eye from an LED embedded in the infrared camera. It creates strong reflections in the target eye. A large proportion of the light reaches the retina and is
immediately reflected back. A corneal reflection, appearing as a small sharp sparkling area, is also generated by the infrared light.

Figure 1. a) A head-mounted eye tracker worn by a user (courtesy of Oleg Špakov) and b) a desktop (Tobii T60) eye tracker.

In the late 1940s and early ‘50s, individual researchers developed a number of eye tracking techniques. The situation changed in the mid-1970s: standardization emerged as companies driven by engineers started to build eye trackers and sell them to researchers (Holmqvist et al., 2011). There is still great variety in eye tracking setups, though. We can find head-mounted trackers (see Figure 1, pane a) and trackers that require the head to be stable (using, for instance, a chin rest). There are also devices that track the head remotely and automatically during its motion. The studies conducted for the dissertation used desktop Tobii eye trackers. Tobii 1750 and Tobii T60 remote eye tracking devices (see Figure 1 (b)) were used in various of the experiments to track the eyes.

Eye tracking devices measure eye positions and the eyes’ movement. The output of eye tracking is generated in the form of vectors of gaze direction. Each vector is presented as the X and Y coordinates (also sometimes Z) of the point where the gaze intersects with an observed surface (such as the screen of a computer or other device). Hence, the idea of eye tracking here is to track the movement and report the pupil activities that occur while the user is looking at a presented stimulus (Špakov, 2008). In our case, the tracking method is based on video-oculography.

In eye tracking, the mapping between eye position and the real world is achieved though calibration. There are several reasons which necessitate performing calibration for each test participant (Holmqvist et al., 2011). For instance, eye ball radius and shape differ from one person to the next, eyeglasses alter the size of the eye in video images, and reflections caused by certain eyeglasses influence the quality of eye movement data. The calibration process used for fine-tuning to the individual’s eye movements involves displaying dots on the screen or another 2D surface. Calibration points should cover the entire surface where the relevant stimuli will be
presented (Holmqvist et al., 2011; Goldberg & Wichansky, 2003). A 2-16-point grid pattern is used for accurate calibration (Holmqvist et al., 2011).

The reliability and validity of eye movement research were investigated by Tinker in his landmark research (1936a). Tinker was interested in knowing whether there is a significant change in reading strategies and procedures due to the artificial situation that arises with eye movement studies conducted in the laboratory. He concluded that eye movement research can reveal authentic reading behavior and, to some extent, allow application of the findings to situations outside the lab.

Eye trackers experience noise, which comes from imperfections in the system setup and also from interference created by the environment in which the measurement takes place (Chen, Tong, Gray, & Ji, 2008; Klingner, Kumar, & Hanrahan, 2008). On the other hand, minor eye movements such as tremors can be interpreted as noise (Duchowski, 2007). With high-frequency trackers, very small miscalculations of the gaze direction or eye movements may introduce substantial noise or amplify the noise in the velocity calculations in comparison to what one encounters with low-frequency trackers. At the same time, however, the noise reduction algorithms used by low-frequency eye trackers may introduce a greater risk of actual data modification.

To ensure that inaccuracies are not introduced and avoid data modification, one must employ utmost care in dealing with noisy data in the realm of low-frequency eye trackers. Studies conducted for this dissertation used low-frequency eye trackers. Investigations related to our research questions are examined with higher level interactions between eye movements and devices through a function of text presentation format, and low-frequency eye trackers are fully capable of collecting the required eye movement data in the relevant experiment conditions. While high-frequency eye trackers may provide more cognition related information, doing so is far from our core objective. Accordingly, although many eye tracking studies use high-frequency eye trackers, low-frequency devices suffice for addressing the research questions.

### 2.2 Eye Movements in Reading

Eye movements occur frequently when we look at a screen, read, or search for an object. Eye movements in reading are linked to visual processing of words. In reading of English text, the eyes move left to right and follow the words across the page. Again, the movement is not a smooth process: reading produces a series of discrete eye movements from one place to another. We scan from left to right to identify a section of text for cognitive processing. Usually, we sweep our eyes over the text and frequently stop
at some point for a short time. Then, the eyes continue their motion, moving forward to new text or back over text that we have previously gone through.

**Fixations, Saccades, and Regressions**

Therefore, in reading, eyes make brief jumps along the line of the text. Rapid movements of eyes are called saccades. Our peripheral vision is used to locate the next portion of text, upon which the eyes briefly rest. Stopping in between the saccades are called fixations. After reaching the end of the line, often the eyes make a return sweep, to the beginning of the next line, and repeat the same general sequence of saccades, fixations, and return sweeps (Huey, 1908/1968). Saccades which move backward to the text that the reader has already encountered are called regressions. Regressions are rereading a phrase, sentence, or entire passage. Thus, the reading process is a combination of fixations, saccades, and regressions of eye movements depicted in Figure 2.

*Figure 2. Visualization of a gaze path, with fixations, saccades, and regressions. Fixations are shown as circles, saccades are represented by straight lines between circles, and regressions are lines from right to left (e.g., the saccades from fixation 2 to 3 and from fixation 16 to 17 are regressions).*

The physiology of eye movements shows that we see nothing when the eye is in motion (Wolverton & Zola, 1983), and other research too has indicated that no new information is obtained during saccades. The eyes move so quickly across the area of visual stimulus during saccades that no clear perception is possible; instead, just a blur is apparent (Uttal & Smith, 1968). Only during fixations can a reader cognitively process information (Dodge, 1900). Therefore, during each fixation, the reader extracts the visual information that is then processed. After a fixation, one makes a saccade to relocate the position of fixation elsewhere in the text (Liversedge & Findlay, 2000).

Reading text from a screen involves eye movements along the lines of text with saccades of variable size and direction, connected through fixations of variable duration. Fixations account for 90% of the total reading time (Tinker, 1936b). The metric referred to as fixation duration refers to the amount of time for a single fixation. It lies within the range 100–500 ms, depending on the content of the target (Just & Carpenter, 1980). Average
fixation duration in reading of English text is about 200–250 ms (Rayner, 1998). The number of fixations in a given target area is expressed with the metric referred to as fixation count.

The average saccade length in reading of text in English is 7–9 characters (Rayner, 1998). Finnish text has a greater average word length than text in English, and the average saccade length for Finnish reader is 11 character positions (Hyönä, 1995). Considerable variability in saccade length is evident. Larger saccades can be over 20 characters long, while smaller ones involve the reader’s eyes moving by only a single character (Staub & Rayner, 2007).

As mentioned earlier, backward movement of the eye over information that it has previously encountered is called regression. More formally, whenever the eye moves forward to word \( n \) before activation of word \( n-k \) (with \( k > 0 \)) has been fully removed (i.e., lexical processing has been completed), regression is likely to occur (Engbert & Kliegl, 2011, p. 795). Eye movements in reading involve both within-word regression and inter-word regressions. Usually the eyes move forward from one word to an upcoming word. Sometimes they make an additional fixation on the currently fixated word before progression starts again which is known as within-word regression. Besides, eyes sometimes move back towards one of the previous words which is known as inter-word regression.

In addition to the existing eye movement metrics available, Hyönä et al. (2003) presented other measurements related to fixation duration, such as extended first-pass fixation time, first-pass rereading time, gaze duration and so on. Those measures are capable of reflecting several distinct attributes of global text processing. Global text processing refers to the process of integration of information that are not presented adjacently in the text.

Observations on Eye Movement Metrics

The length and duration of saccades and fixations depend mostly on the content of the text being read (Rayner, 1998) and these are related to the overall ease or difficulty of understanding the text (Rayner, 1998; Rayner & Slattery, 2009). Therefore, eye movements can point to problems in understanding the document being read. When the text becomes comparatively difficult, longer fixations are generated, and at the same time saccades get shorter. Existing reading process models generally assume that fixation duration and cognitive processing are directly related (Reichle, Rayner, & Pollatsek, 2000). Researchers have also examined whether fixations (indexed in terms of cognitive processing) are influenced by prior knowledge and reading perspective (Kaakinen & Hyönä, 2005).
In reading, a reader does not fixate on every word in the text, one after another, adjacently. In reality, foveal processing of each word is unnecessary. Overall, two thirds of the words are fixated and the other words are skipped (Buswell, 1922, 1937; Judd & Buswell, 1922; Huey, 1908/1968; Fisher & Shebilske, 1985). In particular, many short words are skipped, while some long words are fixated upon more than once (Weger & Inhoff, 2006). Earlier works have shown that in reading, function words are fixated only about 35% of the time, whereas the corresponding percentage for fixating content words is about 85% (Rayner & Duffy 1988; Carpenter & Just 1983). Content words are the words that have meaning and convey information in a text. In contrast, function words are the words with little or no meaningful content of their own (e.g. “the”, “that”, “of”). Function words convey relationships, while content words “put the meat on the bones” in the text. Nouns, verbs, adjectives, and so on are content words, whereas prepositions, conjunctions, pronouns, and auxiliary verbs are function words. Function words tend to be shorter than the content words and hence fixated less frequently.

The probability of fixation on a word increases with its length (Rayner & McConkie, 1976). The probability of skipping a word with one or two letters is 0.76, a figure that drops to about 0.42 for 4-letter words and to 0.05 for 9–10 letters (Vitu, 2011). Clearly, longer words almost always get a fixation, often more than one. The eyes usually fixate between the beginning and the centre point of a word (Rayner, 1979; Vitu, 2011), although there is evidence of fixating on more informative portions of words (O’Regan, 1981, 1990).

As they read individual words, readers develop a structural (or syntactic) representation of the sentences. Previous research has established that text with syntactic or semantic irregularities produces regressive eye movements with longer fixations (Ferreira & Henderson, 1990; Altmann, Garnham, & Dennis, 1992). Hence, detection of an irregularity in a word increases rereading of the text.

Rayner (1998) mentions in his survey that about 10–15% of saccades in reading of English text are regressions. On the other hand, another study found that 14% of saccades for adults are regressions while the figure for children is 25% (Starr & Rayner, 2001). For skilled readers, about 90% of the saccades move forward and the others move backward in the text (Staub & Rayner, 2007).

Several previous studies have established that regressions often occur on account of lack of concentration or a comprehension failure during the first pass over the material (Blanchard & Iran-Nejad, 1987; Hyönä, 1995; Just & Carpenter, 1980; Rayner, 1998; Ehrlich & Rayner, 1983). Readers are likely to look back more if the comprehension process does not go well. In addition, regressions can result from incomplete processing. Within-word
short regressive saccades may occur for processing of the word that is currently subject to fixation. On the other hand, longer regressions which could be more than 10 letter spaces in length, moving back along the line or to another line, may occur because of the ambiguity of the text (Rayner, 1998). Weger and Inhoff (2006), for example, have mentioned that regression sizes are influenced by the ease of sentence comprehension.

Accordingly, regressive eye movements can be a sensitive indicator of reading disruption. Sanders and Stern (1980) studied eye movement patterns that appear with atypical text formats and found that the number of regressions is a better predictor of reading disruption than is reading speed. Moreover, splitting sentences across the screen caused an increase in rereading of the text on the previous page (Dillon, Richardson, & McKnight, 1990). Splitting impedes the comprehension process because of the increased demand for working memory to hold the beginning of the current conceptual unit; therefore, rereading intensifies. In addition, there are strategic fixations, for look-back or rereading the important parts of the text (Hyönä & Nurminen, 2006; Hyönä, Lorch, & Kaakinen, 2002).

Rayner’s (1998) review also mentions that a word on the current line is a much more likely target of regression than are words located on previous lines (Ehrlich & Rayner, 1983; Duffy, 1992). On the other hand, a regression is more likely to occur when the reader encounters a word and recognizes that the prior interpretation of the sentence was in error. Good readers often focus their eyes very accurately on the critical part of text that caused them difficulty and confusion (Murray & Kennedy, 1988; Kennedy, 1983; Frazier & Rayner, 1982; Kennedy & Murray, 1987a, 1987b). On the other hand, poor readers engage in eye movements through the text that are more along the lines of backtracking (Murray & Kennedy, 1988). In addition, more difficult text or text in a second language turns out to prompt long fixation duration (Osaka, 1989) and significantly larger amount of regressions, since it is more likely that the reader did not understand the text. Regressions due to backtracking may slow the pace of reading; however, they aid some readers in comprehension of the passages read. As a result, rereading allows readers to improve in comprehension and detect misspellings (Levy, DiPersio, & Hollingshead, 1992).

Variation in Eye Movements

Variation in eye movements has been observed between general reading and searching for a target word within the text stimulus (Rayner & Fischer, 1996). Although fixation duration has been found to be longer when the readers looked at low-frequency words than at high-frequency words, the word frequency effect disappeared when the task was changed to search (Rayner & Raney, 1996). With the same stimulus, fixations were found to be longer in the search task than in reading (Rayner & Fischer, 1996; Vitu,
O’Regan, Inhoff, & Topolski, 1995). Though the cognitive demands differ between search and reading, Vitu and his colleagues stated that there were also similarities in global characteristics of eye movements (saccade length and the durations of fixations to some extent) between these tasks.

There is evidence that different readers read differently and that individual readers read differently in different circumstances (Judd & Buswell, 1922; Hyönä & Nurminen, 2006). Olson, Kliegl, Davidson, and Foltz (1985) identified two individual reading styles by examining disabled readers’ reading strategies. Those were plodder and explorer. With the plodder style, readers move their eyes forward along the text with relatively short saccades and produce fewer regressions between words or skipping over words. In contrast, explorer readers move their eyes by skipping over individual words more frequently, produce more regressions to previously encountered text, and therefore make fewer intra-word progressive movements.

Also, Hyönä et al. (2002) identified four types of reading strategies with distinct features while competent adult readers read a multiple-topic expository text. The reading strategies were varied in accordance with which text was being reprocessed by readers. The four categories identified were fast linear reader, slow linear reader, non-selective reviewer, and topic structure processor. No return fixation to previous text was made by fast linear readers. In contrast, slow linear readers made many re-inspections before moving on. Readers in the non-selective category performed many look-backs to previous sentences. Finally, readers in the fourth category (topic structure processors) were specific and intensively looked back to the headings that accompanied the text, producing a more precise summary.
Greater availability of computers and handheld devices, coupled with increasing use of distance learning and the availability of information via the Internet, has led to greater use of on-screen electronic text. Digital sales of books already exceed corresponding numbers for printed editions of paper copies sold by Amazon (Miller & Bosman, 2011). Several studies have been conducted to examine the advantages and disadvantages of reading printed vs. electronic material (Osborne & Holton, 1988; Dillon, McKnight, & Richardson, 1988; Muter, Latremouille, Treurniet, & Beam, 1982; Muter, 1996; Gould & Grischkowsky, 1984; Wilkinson & Robinshaw, 1987; Muter & Maurutto, 1991; Ziefle, 1998). Noyes and Garland (2008) summarized the early studies comparing computer- and paper-based tasks and found that paper showed better performance for reading speed, accuracy, and comprehension. On the other hand, studies of online assessment using computers and paper form provided mixed results. Some findings pointed to benefits relating to computers, some have favored paper, while most of the research studies have found no difference (Noyes & Garland, 2008). Moreover, there are many issues that are difficult or impossible to do on paper. Therefore, it has already been established that digitized text can present advantages over paper media (Egan et al., 1989) by providing a search option, the availability of updates to the information, interactivity, and so on. In other developments, reading of electronic text gained in flexibility and readability with the introduction of graphical user interfaces (GUIs), and it more recently became possible to read electronic text in mobile conditions instead of carrying paper versions.
Reading of on-screen material can feature both static and dynamic text. Static text usually appears on paper. On-screen static text can be defined as predetermined material positioned in a fixed location with defined boundaries. Static text can be used when one reads an article or book, browses a newspaper, and so on. If the static text extends beyond the boundaries of the presentation area of the computer or other device, readers manually move the text by scrolling or dragging, to bring new material into view.

In contrast, dynamic text appears almost exclusively on-screen. It moves on the screen automatically, and readers have little or no control during this action. Dynamic text has an unfixed and dynamic existence on the screen. Usually, only a fraction of the text can be seen at any one time. Often the extent of the text available is unknown to the reader. Dynamic text-rendering has become popular for displaying text on electronic billboards as advertisements, in sports galleries (for updates to the match information), with stock markets (to show the latest market values), for TV subtitling, and for other uses with computer screens or smaller electronic devices with limited screen area. In most cases, the text to be presented already exists and dynamic rendering is used to display the text in various ways. Translation and print interpretation, in contrast, provide another aspect of dynamicity where the text is created at the same time as it is being read.

In general, reading with a traditional text presentation format or static text requires substantial interaction whereas the text proceeds automatically with dynamic rendering. Electronically displayed static and dynamic text are both displayed on screens of various sizes. A cross-section of the range of display screens available for electronic text is provided in Section 3.1, followed by more details on dynamic text presentation formats.

### 3.1 Display Screens for Reading of Electronic Text

As technology has advanced, displays have improved and reading electronic text from various devices has become part of day-to-day life for many of us, involving activities with handheld devices, desktop monitors, wall-mounted large display screens, and so on. There is great diversity in the screens from which electronic text is read. They depend on the purpose and location. Office workers need a standard monitor, while a conference or seminar requires even larger screens, as in projection. At the other extreme are handheld devices such as mobile phones with somewhat smaller screens than tablet computers have, and even tiny screens of smart watches are used for presenting electronic text.
Projectors usually provide large displays and use larger fonts, to be read from a distance. In seminars, meetings, or classrooms, such large displays are needed to make the text or other material visible for the audience. Print interpreting too may require large-screen projectors, if the interpreted text is to be presented to a bigger audience (see Figure 3). Outdoors, one can find even larger displays, such as billboards for presentation of commercial advertisements or sporting events.

![Figure 3. A large screen displaying print interpreted text. Here, an interpreter types the text of a speech on a laptop and the text is projected on a bigger screen for a larger audience. Source: http://www.chhasudbury.com/programs-services/print-interpretation-services/](http://www.chhasudbury.com/programs-services/print-interpretation-services/)

More commonly used display types are seen with desktop and laptop computers, whose standard monitors come in various sizes. Relatively recently, tablets, with smaller screens, have flourished in the market. Finally, mobile phones have relatively tiny screens. Figure 4 presents various types of screens that are used daily.

Each breed of screen requires suitable text presentation formats: ones that satisfy the consumer. Many studies have been conducted to find appropriate ways of presenting text on the display screens of various electronic devices. On-screen text is traditionally presented with a layout similar to that of a printed page. The limited screen space of small displays restricts the amount of information that can be presented at any given time, however. Although smartphones can present full pages at a time, often text documents need to be divided into small parts for display. Therefore, more user interaction is needed to view the text in readable form and this may become frustrating in reading longer pieces of text.
3.2 Dynamic Text Presentation Formats

Electronic text facilitates the use of dynamic text-rendering with various presentation formats. On the basis of the reading context, computers and electronic devices allow different ways to present text dynamically over time. Rendering text dynamically on the screen is one possible option for presentation of text when small-screen interfaces are being used (Laarni, 2002).

Various formats to render text dynamically include scrolling, paging, leading, and rapid serial visual presentation (RSVP). Scrolling and paging follow traditional text arrangement as seen on paper. Leading and RSVP, on the other hand, are able to display text in a rather diminutive area.

In scrolling, the text is presented in a given area and the content of the document is larger than the screen. Readers need to scroll the page to continue reading. Manual scrolling is achieved via dragging the scroll bar, pressing keys on a keyboard, using a touchpad, or using a mouse button or wheel. With a keyboard, one can also use the Page Up and Page Down keys to go to the next screenful of text directly. A scrollbar indicates the amount of text displayed on the screen along with the current horizontal and vertical position. It is worth noting that Muter (1996) has pointed out that using both horizontal and vertical scrolling for text management is not efficient or suitable for users.

Usually, text on mobile devices is scrolled vertically, with the number of lines displayed at a time determined by the amount of text that can fit horizontally. Page-by-page scrolling is better than line-by-line, since it decreases the frequency of scrolling. Moreover, continuous scrolling was found to be preferable to discrete step-by-step scrolling (Mills & Weldon,
1987). Nowadays, smart electronic devices require only finger swipes or similar – in both horizontal and vertical ways, as needed – and also allow scrollbar use to move text into sight. Of the various solutions for scrolling on-screen text, Figure 5 presents examples of a scrolling implementation for small-screen (in pane a) and large-screen devices (in pane b).

Figure 5. Scrolling on a) a small smartphone screen and on b) a computer screen. Both horizontal and vertical scrolling using finger flips are allowed in the former, while only vertical scrolling is allowed in the latter. In pane b, the size of the scrollbar thumb indicates the amount of the text that is visible, and its position along the scrollbar shows where the visible text is situated within the piece.

Paging presents the text divided into pages that fit the display screen. Readers can move one page at a time to continue reading. A set of arrow keys, a joystick, or a mouse button can be used to move between pages. Also, the page number of the page that is currently visible is often displayed, along with the total page count for the content, to inform the user about the position of the text presented (Muter, 1996).

Nevertheless, paging can be confusing without further perceptual guidance. Wiping was introduced to improve reading on a small screen, such as that of a mobile device with perceptual guide (Melchior, 2001). Wiping is a form of perceptual guidance that hides or dims all the content apart from one line on the screen, at the bottom between the pages that will be scrolled off the screen. In this process, when the command is given to scroll the page, the material that is to be scrolled off the screen will be removed or grayed out on the display, and then scrolling occurs (the process is clarified in Figure 6).

Compared to scrolling, paging reduces the number of interaction steps. Paging displays the text in fixed amounts at a time (discrete screenfuls), while scrolling requires more interactions for seeing the same amount of text. Figure 7 (a) presents a layout with paging text presentation format on a small display screen. A progress bar or completion meter, along with the number of pages, indicate the location of the text shown within the larger piece.
In leading, dynamic text is scrolled horizontally from right to left on a single line across the screen. The text moves continuously at a certain speed, which sometimes can be controlled by the user. Moving the text in leading as character by character was found to be less efficient than moving it pixel by pixel (Kang & Muter, 1989). On small devices, the user can control leading with operations such as start, stop, and moving forward or backward, and the speed can be controlled by means of a joystick or arrow keys/buttons. The location within the text may be indicated by a completion bar (as in Figure 7, pane b). Leading requires little or even no interaction in reading when it starts automatically.

Finally, RSVP presents the text in a chunk of one or more words at a time in rapid succession in a fixed location on the screen at a predetermined rate. The chunk rendering rate can be set by the user (Muter, 1996; Potter, 1984). Similarly to paging and leading, this style of presentation can use a completion bar akin to a progress indicator to specify how far one has progressed through the text. This is seen as a beneficial feature for the RSVP format in particular (Rahman & Muter, 1999). Arrow keys/buttons or a joystick can be used to start and stop the RSVP rendering. Sometimes, these can also be used to change the speed or to move backward and forward in the text. Apart from controlling the speed and moving forward/backward, no user interaction is required in RSVP reading.
Figure 7 (c) presents the layout of RSVP text presentation on a small screened device.

RSVP is trying to ensure that every word is actually focused. It might be an efficient approach to searching for specific words, names, or ideas in lists or other text (Potter, Kroll, & Harris, 1980). Moreover, RSVP may help readers with impaired peripheral vision (Williamson, Muter, & Kruk, 1986). Also, a study by Potter (1984) suggests that reduction of eye movements by means of RSVP would be likely to contribute to the possibility of reduction in cognitive load.

### 3.3 Previous Research on On-Screen Presentation Formats

Improving efficiency in reading of electronic text on electronic devices has become an important issue. Much of the early research on this topic examined readability by focusing on different sizes of the text presented on screens of various sizes. Five distinct window heights (1, 2, 3, 4, and 20 lines) were used in one study of readability, which used three different line lengths and two different character densities (Duchnicky & Kolers, 1983). The authors found a significant effect of all three variables on reading rate. Specifically, lines of one-third screen width were read 25% more slowly than lines of full and two-thirds screen width. As for character density, text appearing at a density of 80 characters per line was read 30% more rapidly than text at 40 characters per line. Also, reading rate varied slightly with window height: text appearing in a 20-line window was read only nine percent more rapidly than text in one- or two-line windows. Moreover, comprehension did not vary much as a function of window height. The results of Dillon, Richardson, and McKnight (1990) were consistent with these findings: they did not find any difference in comprehension when they used window heights of 60 and 20 lines.

Although comprehension level was not influenced much by window size, researchers continued to focus on text presentation units: quite a few studies have been conducted to find the effect of various text presentation formats and satisfactory text-rendering for a variety of display screens. Radach, Huestegge, and Reilly (2008) have demonstrated that the setting of the reading material – such as single sentence presentation vs. the same sentences integrated into a longer text passage – significantly affected eye movements as measured by word viewing durations. So and Chan (2008) reviewed studies examining several display methods used in dynamic text presentation. The review focused mainly on suitable speeds for the various types of dynamic text-rendering (RSVP, leading, and scrolling). So and Chan also addressed specific design factors, such as text display or presentation direction (left–right or top–bottom), character types (for Chinese text), and text/background color combinations for a favorable
display and better reading performance. Table 1 provides an overview of several studies; the table briefly describes the measurement units and comparison parameters in each study.
<table>
<thead>
<tr>
<th>Author(s) and year</th>
<th>Display size(s)</th>
<th>Measurement</th>
<th>Comparison</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Juola, Tiritoglu, and Pleunis, 1995</td>
<td>Eight character horizontal display</td>
<td>Reading accuracy</td>
<td>RSVP and leading with these conditions: - Uppercase vs. lowercase letters - Slow (171 wpm) vs. fast (260 wpm) presentation</td>
<td>- The reading accuracy was greater with the RSVP format than the leading format, in all conditions</td>
</tr>
<tr>
<td>Lin and Shieh, 2006</td>
<td>Single line screen</td>
<td>- Recall efficiency - Satisfaction score</td>
<td>- Leading vs. RSVP - Word-by-word, with 120, 240, and 500 characters per minute</td>
<td>- Recall efficiency and the satisfaction score were better with leading than with RSVP - The layout was better when words rather than characters were the presentation unit - Performance deteriorated with high presentation rate for RSVP</td>
</tr>
<tr>
<td>Öquist and Lundin, 2007</td>
<td>Mobile device</td>
<td>Eye movements, reading speed, comprehension, and task load</td>
<td>Scrolling, paging, leading, and RSVP</td>
<td>- Paging was faster than scrolling and RSVP, but there was no comprehension difference - Compared to leading and RSVP, paging is less demanding to use - RSVP decreased eye movement, while leading increased it - Leading produced irregular eye movements - Scrolling was found to be the most frequently used format - Overall, paging offered the best readability for a mobile device</td>
</tr>
<tr>
<td>Öquist, Sågvall-Hein, Ygge, and Goldstein, 2004</td>
<td>Mobile phone</td>
<td>Eye movements, reading speed, comprehension, and task load</td>
<td>RSVP vs. paging</td>
<td>- RSVP decreased eye movements - RSVP increased regression but decreased the number of saccades</td>
</tr>
<tr>
<td>Bernard, Chaparro, and Russell, 2001</td>
<td>Small screen interfaces</td>
<td>- Reading comprehension - Satisfaction - Performance</td>
<td>RSVP, three lines, and 10 lines of text with three speed levels (slow, medium speed, and fast)</td>
<td>- A significant main effect of presentation speed was found - The slowest speed was favorable - RSVP and 10-lines of text presentation were preferred over the other option</td>
</tr>
<tr>
<td>Study</td>
<td>Display Screen</td>
<td>Presentation Formats</td>
<td>Usability Metrics</td>
<td>Findings</td>
</tr>
<tr>
<td>-------</td>
<td>----------------</td>
<td>----------------------</td>
<td>-------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Laarni, 2002</td>
<td>Laptop, palm type pocket computer, Communicator, and mobile phone</td>
<td>RSVP, vertical scrolling, paging, teletype, and leading</td>
<td>- Reading rate - Comprehension - Presentation methods' usability (satisfaction, preference, and comments)</td>
<td>- Vertically scrolled text was read faster than text in paging format - A Communicator screen is suitable with leading and teletype methods - Mobile phone screen is suitable for RSVP and vertical scrolling</td>
</tr>
<tr>
<td>Kang and Muter, 1989</td>
<td>16.5 × 10 cm screen dimension</td>
<td>RSVP, Times Square(^1)</td>
<td>- Comprehension - Preference</td>
<td>- Smoothly scrolling Times Square was preferred over other options - Comprehension for Times Square was as high as that for RSVP</td>
</tr>
<tr>
<td>Rahman and Muter, 1999</td>
<td>15-inch display</td>
<td>Normal page condition (23 lines per page), RSVP, sentence-by-sentence, RSVP with sentence-by-sentence with completion meter</td>
<td>- Reading efficiency (speed and comprehension) - Satisfaction - Performance</td>
<td>- Sentence-by-sentence reading was as efficient as normal reading - RSVP was less preferred and less efficient than sentence-by-sentence presentation</td>
</tr>
<tr>
<td>Harvey and Walker, 2014</td>
<td>1024 × 768 pixel CRT monitor</td>
<td>Horizontally scrolling text, static text</td>
<td>- Reading accuracy - Performance under eccentric viewing condition</td>
<td>- Reading accuracy improved with scrolling text - More errors occurred in static text than in scrolling - Performance in conditions of eccentric viewing strategies was also improved with scrolling text - Increased fixation stability was observed in scrolling</td>
</tr>
<tr>
<td>Bowers, Woods, and Peli, 2004</td>
<td>27-inch television monitor</td>
<td>RSVP, horizontal scrolling, vertical scrolling, page</td>
<td>- Reading rate</td>
<td>- Paging was faster than the other three formats - Half of the low-vision participants preferred horizontal scrolling format</td>
</tr>
</tbody>
</table>

**Table 1.** A few studies have reviewed various dynamic text presentation formats, for different display screens.

\(^1\)Another term for leading, named after New York City’s Times Square, where the most famous implementation of this technique was seen (Kang & Muter, 1989).
As the table shows, previous studies suggest that presentation formats affect reading performance with various sizes of screen, such as standard monitors (Bowers et al., 2004; Kang & Muter, 1989; Rahman & Muter, 1999) and small screen interfaces (Bernard et al., 2001; Lin & Shieh, 2006; Öquist & Lundin, 2007). In addition, a study of presentation format for Chinese text on a small screen such as that of a wristwatch established that reading comprehension was influenced significantly by presentation method (Chien & Chen, 2007). Continuing the line of investigation, the study by Harvey and Walker (2014) found increased stability of fixations with scrolling in terms of lower number of fixation. Authors suggested that scrolling text could benefit people with loss of central vision, holding potential as a reading aid.

Another subject of study has been whether local effects of eye movements exist in reading. The properties of currently fixated word and that of adjacent words both influence eye movements. Eye movements are also affected due to the difficulties in reading previous sentences. Huestegge and Bocianski (2010) investigated how global adjustment of eye movement control can be influenced by the syntactic properties of previously read sentences. The authors investigated eye movements in reading passages which were followed by identical target sentences. Sentence structure (embedded vs. non-embedded) and voice (active vs. passive) were both manipulated in the text passages. The results showed that the previous sentence’s structure affected gaze durations within the target sentence. More specifically, the mean gaze duration for words in the target sentence was found to be significantly greater after embedded sentence structures than after non-embedded structures.
4 Reading in Various Contexts

Several psycholinguistic studies (Rayner, 1998; McDonald & Shillcock, 2003) have shown that document characteristics influence the cognitive process in reading, an effect that is reflected in eye movements. Filik, Paterson, and Sauermann (2011), reviewing the influence of linguistic focus on eye movements during reading, have stated that the effect of focus was readily observed in the reader’s eye movement behavior. The focus varies, depending on whether one is processing individual words or sentence structure, and also depends on the nature of the mental representation constructed by the reader.

Other researchers have attempted to quantify the influence of the linguistic characteristics of text on reading behavior for certain types of reading tasks. The findings of Martínez-Gómez et al. (2012) indicate that, although linguistic features of documents can be used to explain eye movements and reading behavior, there might be other influencing factors.

Therefore, the cognitive strategies employed in reading for translation may differ markedly from those in reading for comprehension. Similarly, strategy may differ between reading for entertainment and what we have termed reading for communication. This chapter presents scenarios of reading in several contexts, which served as the primary basis for the studies described in the dissertation.

4.1 Reading of Print Interpreted Text

Interpretation facilitates smooth progress of necessary communication between two languages. Interpretation in cross-cultural communication
converts one language into another. In addition to cross-cultural circumstances, interpretation is needed to convey messages from one form of language into another. A good example is print interpreting, the text format described above that provides visual communication for the hard-of-hearing and people with acquired deafness, giving them access to spoken language. Print interpreting is also known as “typing/writing interpreting, captioning or real-time writing, and means the translation of spoken language and any accompanying significant audible information into written text simultaneously with the speech” (Tiittula, 2009). Print interpreting is a common tool for people with hearing disabilities.

In recap, print interpreting involves spoken information being typed in text form on a computer and simultaneously displayed on a screen. Generally, the text appears dynamically letter-by-letter on the screen as it is written by the interpreter. As new text is added, the lines scroll up the screen. When the display screen is full of text, the older text, at the top of the display, disappears and the new text appears at the bottom of the screen. Thus the writing is made visible to the client in a dynamic, real-time process (see Figure 8). A client reading print interpreted text will be able to see only the part of the text material that is visible at the time and fits the display screen.

![Figure 8. A print interpretation display screen (Tiittula, 2009; reproduced courtesy of Prof. Liisa Tiittula). When the screen is full of text, the top line disappears as soon as new text is typed, at the bottom of the screen.](image)

Print interpreting requires real-time captioning which is the word-for-word transcription of speech to text form in real-time. The task is not limited to typing as much of what was said as possible; it also includes conveying the emotion of the speech. Therefore, in print interpreting the interpreter also translates phonic symbols into caption form instantaneously. The client sees print interpreted text with errors, corrections, and other temporal features. In this communication, the client also experiences delay, since the interpreter needs to hear the words and then enter them as soon as possible on the computer. Complicating this
process is that normal and steady speech has been determined to be approximately three times faster than fast typing on a computer (cf. Salakari, 2008).

There are guidelines for print interpreting such as “everything shall be written, also dialects and slang” (Laurén, 2006, in Tiittula, 2009). However, research results have indicated that the content of the written text includes a large number of spelling errors and other mistakes. Furthermore, approximately 30–45% of the spoken words may be omitted – this depends on the speech situation and the speaker (Tiittula, 2006, in Tiittula, 2009). While speech recognition software could be used for avoiding errors generated in manual typing, there are several reasons such software is not used in print interpreting. Constant change of speakers, with a great variety of voices and speaker characteristics, can lead to poor recognition results (Wagner, 2005). Spontaneous oral language differs considerably from written language, and current speech recognition software does not handle elements of it such as false starts, repetitions, and unfinished sentences well. In addition, a verbatim transcript may be difficult to read. For greater readability, print interpreting not only compresses or condenses the original message but also adds information. For example, punctuation marks may indicate the speech segments, the meaning of relevant prosodic features (e.g., intonation indicating sarcasm), non-language sounds from the environment (e.g., laughing or applause), and speaker identity.

Therefore, print interpreting requires all relevant audible information to be rendered in visible and readable form (Tiittula, 2009). Figure 9 provides an example, which illustrates typing mistakes, errors, delay, and also reformulation. There is a great difference between the message conveyed through speech and that passed on by its interpretation. Clearly, conveying all information generated by a speaker as text for the client is still a challenge for print interpreters.

<table>
<thead>
<tr>
<th>Talk:</th>
<th>Print interpretation:</th>
</tr>
</thead>
<tbody>
<tr>
<td>first of all, today is the deadline for summaries / so may I have those first / also and those of you who haven’t given me the abstracts / please hand it in. remember I need two pieces of text / from each one of you.</td>
<td>This is deadline for summaries, may al have those first and those abstracts that are missing. I need 2 pieces of text @3ach one of you.</td>
</tr>
</tbody>
</table>

Figure 9. An example of print interpreting of a talk ((Tiittula, 2009; reproduced courtesy of Prof. Liisa Tiittula).

Print interpreting is used by the target group for various encounters, such as seminars and meetings. For larger audiences, the text output is likely to be presented through projection on a large screen (as shown in Figure 3). In recent years, use of print interpreting has increased among individual hearing-impaired people (Tiittula, 2009). Personal interpreting is often required at meetings or in lectures for a hearing-impaired client when
most of the other participants have normal hearing. Moreover, certain community interpreting settings demand personal interpreting for dealing with service encounters, legal and medical interactions, and so on. It is possible to have remote real-time print interpreting at a remote location. A voice connection is used to convey voice to the interpreter and the real-time interpreted text is transmitted via an Internet connection to the required location.

Since speech is much faster than writing/typing, advanced special keyboards have been developed that produce more text by means of key combinations to ease the interpreters’ burden and help them cope with the speed of the speaker. Finding an appropriate format for presentation of print interpreted text to the client is still an ongoing process.

By law, deaf people have the right to an interpreter, and print interpreting is the method used most among hearing-impaired people who have not learned sign language (Laurén, 2006). At the same time, print interpreting is quite a new communication method. Hence, the occupation is still taking shape, and efforts within the profession target better technology with well-developed methods, standards for the interpretation output, and advanced interpreter training. Finland has about 750,000 hearing-impaired persons, of whom the severely hard-of-hearing number approximately 10,000 and those with acquired deafness number 3,000 (Laurén, 2006). Therefore, there is a great need for research into this process in Finland, not just worldwide. Discussed in Chapter 5, papers II and III introduce our eye movement studies related to reading of print interpreted text. These results may be useful also for development of a better print interpreting software tool.

## 4.2 Reading for Translation

In translation, reading and writing processes occur simultaneously. On-screen translation generally uses two parts of the display. Usually, the text to be translated is displayed in one area while the text written by the translator is in another. To make the translation process less exhausting and to improve efficiency, one of the hardware setups that translators find best is a desktop computer with two monitors. If, instead, the source and the target text (the text to be translated and the translator’s emerging text, respectively) are on the same screen, the screen is usually divided into two parts (see Figure 10, below). For example, the upper part may be used to display the original text while the lower part presents the text produced in the target language. Alternatively, the right-hand part of the screen may display the original text, with the left part for typing the translation. The display screen can be distributed in the other way round according to users’ demand. Figure 10 presents a basic horizontal and a vertical layout
of translation windows for one monitor that are similar to presentation methods used in our studies. In real world translation, there may be, for example, many more panes and/or windows used.

**Figure 10.** A single screen divided into two parts for translation (Source: http://www.code-styling.de/english/development/wordpress-plugin-codestyling-localization-en, and http://questatlantisblog.org/wp-content/uploads/2011/02/qa-i18n.png).

In the circumstances of on-screen translation, each translator’s visual attention is likely to shift continuously between the area showing the source text and that displaying the target text. The given text to be translated is known as source text while the translator’s own emerging translated text is often called target text.

Readers’ eye movements during reading in parallel to writing have been studied using keystroke logging combined with eye movement
technology (Johansson, Johansson, Wengelin, & Holmqvist, 2008). They studied if the writer/reader read through the text produced so far during a pause, planned his/her next text, or simply looked at the monitor. Analysis of their video observation found inconsistent reading patterns. Frequently the writer/reader’s eye movements moved towards previous text and jumped back and forth. The underlying cognitive processes were unclear. However, the authors found variation in time spent while reading one’s own emerging text among different groups of participants.

Another study found that eye movements during the emerging text’s production are highly influenced by typing skill and the frequency of looking at the monitor(s) during writing (Johansson, Wengelin, Johansson, & Holmqvist, 2010). Johansson and his colleagues explored differences in text production between two types of writers, referred to as monitor gazer and keyboard gazer on the basis of the general focus of their gaze. Analysis of the participants’ reading behavior showed more regular forward reading by the monitor gazers than among keyboard gazers. Monitor gazers produced significantly longer fixation durations and were able to read in parallel with writing more of the time. Both groups produced longer fixations in reading while writing than when reading alone. Because more cognitive processing is associated with greater fixation duration (Rayner, 1998; Reichle, Pollatsek, Fisher, & Rayner, 1998), Johansson and colleagues suggested that reading in parallel with writing entailed more cognitive processing than reading during pauses in writing.

“Writing involves a combination of (at least) content determination, formulation, transcription, error monitoring, and reviewing,” state Wengelin et al. (2009, p. 337). In general and also in translation, writers review to accomplish the editing process. Often when processing the text already written, writers look back to detect errors and to correct previously detected errors if doing so is needed. Usually, writers use visual information from the text in reviewing and revising what they have written. Therefore, the differences between reading the source text and one’s own emerging text can be reflected in translators’ eye movements.

O’Brien (2006, p. 186) mentioned several previous studies related to eye tracking for translation where correlation between cognitive effort and pupil dilation was found. Text complexity affects the cognitive effort of translation related reading, as demonstrated by the speed with which the meaning of the text is determined. There is also an effect of the time pressure experienced while the translation is being carried out. Hansen (2006) conducted experiment to translate texts under time pressure (short time translation) and without any time limit (long time translation). The results showed that translators made more errors in short time translation than the other form although short time translation was more accurate compared to long time translation. Just and Carpenter (1978, 1980)
introduced the eye–mind assumption which leads us to presume that these effects are likely to be visible in translators’ eye movements.

In real-time interpreting, the effect of time pressure has been studied a great deal, although little research has been conducted into the time pressure effect specifically in translators’ experience while performing translation. Research has investigated whether an increase in time pressure leads to translators spending less time to orient themselves in the initial stage and not performing the final revision (Jensen, 1999, 2000). Experimenting with three groups of participants (non-translators, professionals, and novice translators), Jensen found that time pressure produced degradation in the text production procedure and also in how a translation strategy was chosen, for all of those groups. Participants’ translation quality deteriorated in similar experiments conducted by de Rooze (2003) when the task was to translate at a rate of more than 20 words per minute; however, 25% of those participants produced higher translation quality under this time pressure. Both Jensen and de Rooze used mainly keystroke logging that captured the translation output over time.

McDonald and Carpenter (1981) employed eye tracking in studies of translation from English into German. Investigating eye movements in mistranslations’ detection and correction, they examined the process of translating ambiguous phrases and how these phrases were parsed during comprehension and interpretation. The patterns of the eyes’ fixations on the phrases were different when there were two possible interpretations. The preceding context proved decisive.

In on-screen reading, iDict introduced a reading aid providing synonyms from dictionaries for words getting longer fixations during reading (Hyrskykari, 2006). A translation supporting system should be able to detect details of the translation process by observing the translator’s behavior and, on its basis, automatically retrieve the proper information from a large database and dictionaries. Takagi (1998) developed a translation support system that was able to detect the progress of a user’s translation process on the basis of eye movements. This tool could detect two characteristic patterns of eye movements, one indicating when the user was at a standstill in the translation process and the other revealing when too much information was being displayed. The system automatically retrieved information from the database at the appropriate time by detecting those patterns and, for example, removed information when the system determined that it was probably superfluous.

In translation, users move their eyes along the text to read and also to retrieve various bits of information. Eye movements are an aspect of user behavior that can illustrate the translation method applied and thereby provide valuable insight into the translation process (Takagi, 1998).
Moreover, it is necessary to accumulate new knowledge related to the method of visual attention to emerging written text. This kind of knowledge is important also, as noted above, from the perspective of use of gaze information in construction of translation support tools. Paper IV, presented in Chapter 5, introduces our study results related to translators’ eye movements during translation. It is an extended version of a paper published at the 2008 ETRA conference (Sharmin, Špakov, Räihä, & Jakobsen, 2008).

4.3 READING LONG DOCUMENTS AND SCROLLING

In addition to short pieces of text, we read long documents frequently, for various purposes. Screens such as a standard computer monitor or the display of an iPad or other mobile device often do not allow all the text to fit on the screen at once. Hence, scrolling is required for moving forward and continuing to read. Scrolling usually involves manual actions, such as clicking on a scrollbar, dragging the grip on the scrollbar, and/or clicking on arrows on the scrollbar. There are other than scrollbar tools that generate similar actions: we might press arrow keys or Page Up and Page Down keys, rotate a mouse wheel, use our finger(s) to interact with a touchscreen, sweep up and down over a touchpad, and so on.

Nowadays, smartphones come with touchscreens that allow scrolling by means of soft finger touches on the screen. No buttons or arrow marks are required. A single-touch motion of the finger up or down starts the desired scrolling. Furthermore, some new smartphones are claimed to be capable of scrolling the screen’s content without physical contact (Etherington, 2013).

Scrolling has become part and parcel of the day-to-day experience of reading from a standard screen and on small screen devices. Reading from the latter has increased significantly for handling of e-mail and other forms of text-based communication via the Internet. Sometimes in such processes, scrolling information or otherwise bringing new content onto the screen is likely to hinder reading by taking more attention.

It is useful to develop new scrolling techniques that take means of human visual information consumption and the characteristics of reading patterns into consideration (Beymer & Russell, 2005; Poynter & Eyetools, 2004; Rayner, 1998). Such techniques could use gaze information in automatic scrolling for controlling both the initiation and the speed of the scrolling action. Other techniques could use gaze information in a passive manner to support manual scrolling techniques.
Kumar and Winograd (2007) introduced several gaze-enhanced scrolling techniques that take gaze information as a primary input or an augmented input. They have implemented these techniques for both manual and automatic scrolling. In manual scrolling, to avoid the problem of many readers losing track of their position in the text once the page scrolls, the authors used a GazeMarker (shown in Figure 11) or highlighting to track the reader’s eyes in the text. The GazeMarker system shows eye position in the text throughout reading and also during page transitions. The process includes immediate highlighting of the region of the current location where the user was looking prior to pressing the Page Down key. The highlighted region moves to the uppermost part of the screen after the scrolling action. Since the highlighting was visible just before scrolling started, the reader’s gaze naturally follows the highlighting and ensures that the eyes remain on the part of the text from which reading should continue. After the scrolling is complete, GazeMarker slowly fades away.

Kumar and Winograd (2007) also introduced automatic smooth scrolling approach with gaze-based repositioning. The approach included multiple invisible threshold lines on the screen; see Figure 12. As the figure shows, scrolling starts slowly as soon as the reader’s gaze is under the start threshold line. Scrolling then terminates when the gaze reaches the stop threshold. When the reader’s gaze drops to a lower line beneath a faster threshold, the system interprets this as a sign that scrolling can occur more rapidly.
Before we can use gaze control as a scrolling technique, it is necessary to investigate which part of the screen a reader prefers to read. From eye movement actions in reading, it is possible to identify the preferred region of the screen for reading purposes. In an eye tracking study, Buscher, Biedert, Heinesch, and Dengel (2010) found individually preferred reading regions that vary in location on the screen and in size of the area. Heat maps from their studies showing different preferred reading regions on the screen are presented in Figure 13.

Gaze-enhanced automatic scrolling to allow reading within the preferred reading region has not been studied before. Paper V, presented in Chapter 5, introduces the results of our experiments with such a gaze-based automatic scrolling technique. At the core of that work is connecting the concept of an individual’s preferred reading region on the screen and gaze-based automatic scrolling. Further relevant investigation is described in Paper VI.
5 Summary of the Publications

Reading of electronic text has rapidly increased. At the same time, the display sizes of electronic devices have been changing, with many moving from bigger to smaller. Numerous studies have been conducted to find a suitable presentation format for text display on electronic devices. Part of that stream of research, the dissertation comprises several empirical studies of reading on-screen text, looking at various text presentation formats and reading contexts. This chapter summarizes all the published articles compiled for the dissertation.

5.1 Paper I: “The Effect of Different Text Presentation Formats on Eye Movement Metrics in Reading”

The first article looks at reading on-screen text with several presentation formats, including sentences as part of a full paragraph, sentences presented one by one, sentences presented line-by-line (with the lines based on the width of the computer screen), and sentences presented in chunks at a predefined rate (RSVP). The aim of the study was to investigate the effect of various text presentation modes on standard eye movement metrics such as fixation duration, number of fixations per minute, and regressions. The null hypothesis was that text presentation format would not have any effect on the typical metrics for eye movements.

We collected eye movement data from 16 participants. Each participant read eight pieces of text, each about 100 words long, from a computer screen. The participants read the first four pieces of text each displayed in a different presentation format. Presentation was paced automatically.
Later, they read the other four pieces of text, this time using manual pacing. The motivation with the latter technique was to understand whether the speed of the automatic pacing was appropriate and whether it was close to that of manual pacing.

The results showed that fixation duration and fixation count were affected significantly by changes in presentation format for both automatic and manual pacing. The longer presentation units resulted in a significantly higher number of fixations, of shorter duration, than the relatively short presentation units. In fact, all of the smaller units were associated with significantly longer average fixation duration and lower fixation count than the larger units. This showed that smaller presentation units were more likely to bring confusion or lack of clarity arising from the fact that the sentence was not complete yet and readers tried to make sense of the content before the end of the sentence appeared. Larger units of text, in contrast, represented wider context, and the viewer could read the text without paying more attention to low-level features and, consequently, pursue smooth reading.

The results showed also that another eye movement metric, regressions, was influenced by presentation format. We found significantly more regressions with the smallest unit (the chunk) than with the largest unit (the paragraph) and the other relatively large unit (the line) when manual pacing was used. This supported our observation that smaller units cause confusion in reading and readers produce longer fixations and higher regression counts in the process of clarification.

According to the post-test questionnaire, sentence and paragraph presentation were the preferred formats. Participants’ least preferred format was chunks. Most of the participants were of the opinion that there is less need to concentrate in reading of text in larger presentation units (paragraph and full-sentence) and encountered difficulties when the reading involved relatively small units (line and chunk).

We can conclude that the presentation format affects eye movement metrics. Participants’ opinions solicited in the post-test questionnaire aided in drawing the conclusion that the smallest units create confusion, which can be manifested in participants’ eye movements. Readers produced high fixation duration and more regressions to overcome the lack of clarity.

Further discussion related to the issue of text presentation formats is provided in Paper II, where we deal with even smaller units: word-by-word and letter-by-letter.
5.2 Paper II: “Dynamic Text Presentation in Print Interpreting - An Eye Movement Study of Reading Behaviour”

In the second study, eye movements were compared between two, alternative dynamic text-rendering formats, word-by-word and letter-by-letter. We studied eye movements in reading of dynamic print interpreted text, in the form of a spoken presentation that had been transformed into written format. The interpreted outcome was visible as dynamic text.

The aim was to gain knowledge from eye movements in reading of dynamic print interpreted text with two distinct presentation formats for generation of useful suggestions and to justify the use of supporting tools for print interpreting. We examined how eye movements varied between presentation formats and whether the presentation format had any influence on reading behavior.

We carried out two experiments. In the first experiment, eye movement data from 20 hearing participants were collected for analysis. Two pieces of Finnish-language print interpreted text were used in the experiment’s reading task. Each participant read one after another the text that appeared dynamically on the screen with the two presentation formats: word-by-word and letter-by-letter.

Our goal was to determine the effect of the two presentation formats on various eye movement metrics, such as fixation duration, fixation count, and number of regressions. On account of the nature of dynamic text, we were unable to collect appropriate fixation data by traditional methods. As mentioned earlier, the gaze in normal reading of static text jumps forward in steps of typically 7–9 characters (Rayner, 1998), which makes it possible to distinguish fixations from each other. The reading process is different with dynamic text presented in the letter-by-letter format, wherein the text appears by letter and the gaze follows the progress of its appearance. In this condition, two consecutive data points produced by the eye tracker are not sufficiently far from each other for distinguishing of the fixations. Therefore, we continued our analysis in this case with the number of regressions, a quantitative metric that has been used in previous studies (Sander & Stern, 1980; Perea & Pollatsek, 1998; Perea & Carreiras, 2003; Sharmin, Špakov, & Räihä, 2012).

A careful observational analysis of the video recordings of eye movements was carried out to find qualitative data for the reading process. We analyzed gaze data recordings for the first four and a half minutes, the time before the text required scrolling. We calculated the number of regressions and rereading actions.

We reviewed the eye movements in gaze path replay for each participant. Our observational gaze path analysis, justified by the k-means clustering algorithm (Lloyd, 1982), found three categories of reading behavior in
terms of fixation frequency and rereading strategies. These were Minor rereading, Moderate rereading, and Extensive rereading. Eleven of the 20 participants fell into the same category when reading text in the word-by-word and letter-by-letter presentation formats, while the others changed their reading strategy when the mode of presentation changed.

Comparing the word-by-word and letter-by-letter formats, we found that the percentage of reread words was significantly higher with the word-by-word format. Moreover, the share of words reread more than once was also significantly higher with the word-by-word format than letter-by-letter presentation. Consequently, the number of words from which the gaze started a regression was also significantly higher for the word-by-word presentation format than for the letter-by-letter format.

Readability of a dynamic print interpreted piece of text can be affected by pauses (the time before new text appears on the screen) between words. The letter-by-letter format has a rhythm that follows the interpreter’s typing speed. In contrast, the word-by-word format involves chunks of text appearing on the screen. Analysis of the eye movement data showed that the number of regressions during pauses between words was significantly higher with the word-by-word presentation format than the letter-by-letter format. A significant positive correlation was found between the length of the pause between two successive words and the number of regressions starting at the preceding word. Hence, a pause before a word appears affects the number of regressions starting from the preceding word.

We also found that the text’s content affects reading. There was a statistically significant difference between the number of regressions starting from the words that end a sentence and from any other words in the text. The result can be explained by two factors: the natural pause when the interpreter starts interpreting a new sentence and the reader’s tendency to understand the meaning of the preceding sentence. In addition, the frequency of rereading incorrect or abbreviated words was significantly higher than the frequency of rereading other words in the text again.

Responses from the post-test questionnaire expressed almost equal preference for the two presentation formats. Moreover, free-form comments from participants suggested that the letter-by-letter format was easy to read and make progress with, but that it did not necessarily lead to good understanding of the text.

A follow-up experiment with a hard-of-hearing group of participants was carried out. In addition to the test setup applied for the earlier experiment, a silent video of the speaker was displayed on an adjacent monitor, resembling a real-life print interpreting setup. Although there were 18
participants in the experiment, we were able to analyze data from only 10. Nevertheless, the results were in line with our earlier findings. Also, findings from this post-test questionnaire too suggested reasonably equal preference of the two presentation formats. Some participants indicated that their reason for preferring the letter-by-letter presentation format was its familiarity: the *de facto* standard applied by most of the software tools available today for print interpreting render text letter-by-letter.

In conclusion, this study emphasized the importance of providing print interpreters with more versatile tools. The word-by-word condition in the study was based on work with the Sprintanium (Špakov, 2011) tool, which enables buffering the rendering of dynamic text such that it appears a word at a time. We found that several print interpreting clients would prefer a presentation style different from the current *de facto* standard. For some who use print interpretation, a word-by-word presentation style might improve comprehension. In addition to the speed of interpretation, the importance of correctly typed full words became apparent: the study illustrated how abbreviations and spelling errors hinder the reading process. This work can be a message to the print interpreting profession.

Finally, the research uncovered clear differences in reading behavior between two dynamic presentation formats from the standpoint of understanding readers’ eye movements. Further studies are needed to compare reading of dynamic text and static text, for better understanding of the phenomenon.

5.3 Paper III: “Gaze Behaviour and Linguistic Processing of Dynamic Text in Print Interpreting”

As a continuation of the research conducted for Paper II, the study done for Paper III investigated gaze behavior by looking at regressive eye movements from a linguistic point of view. The study investigated the relationship of regression landing points to linguistic factors.

We analyzed eye tracking data from 20 participants. These were part of the dataset obtained from the experimental setup used for the first experiment described in Paper II. The linguistic analysis was based on the data from letter-by-letter presentation stimuli wherein we pinpointed the word(s) from which a regression started and also the word(s) where it landed.

Gaze behavior was found to differ between text presentation formats. Reading text with word-by-word rendering involved more rereading than did letter-by-letter rendering, while the two modes of rendering provided almost equal numbers of reread words per regression. For the linguistic
analysis, we analyzed the two words on which participants’ first and second regressions landed, which we refer to as the first and second landing points of regressions (see Figure 14). These two landing points were examined for which part of speech they represent and were classified as primary, secondary, or tertiary elements. We found that 74% of the first two landing points were content words even if the full piece of text had under 30% nouns. On the other hand, only nine percent of the first and second landing points were function words, whereas function words accounted for 43% of the words in the full stimulus (Figure 14). Therefore, the linguistic findings suggest a possible relationship between human gaze behavior in reading and the linguistic processing of dynamic text.

![Graph showing the distribution of word types at landing points](image)

**Figure 14.** Linguistic observation of regressions.

Analysis showed that participants were more likely to consider the primary and secondary elements of language in their regressions. Those elements, found at the landing points, usually assist the reader in constructing the meaning of the text being read. As is indicated in the discussion above, of the primary elements, the nouns are the most expected landing points. Therefore, our study suggested that regressions did not happen by chance and were related to linguistic processing.

Our eye movement-based analysis of reading of on-screen text in different contexts proceeded further by investigating gaze behavior during reading for translation. This work was reported upon in the following publication, Paper IV. As papers II and III do, Paper IV deals with reading text that is produced by a human while it is being read. However, the context is markedly different – in the earlier studies, the text-producer was not the same person as the reader, while in the work for Paper IV the two were the same individual. Further details about the investigation are given in the next section.

### 5.4 Paper IV: “Where on the Screen Do Translators Look While Translating, and for How Long?”

The study for Paper IV concentrated on examining translators’ gaze behavior in translation with source text of several complexity levels. Also,
three time constraints were imposed for the translation. Some of the participant translators had touch-typing skills, while others did not, and in translation the gaze was occupied by both the source and the target text. The aim of the study was to examine the distribution of eye movements between the source and target text. Also explored was how the translators managed their visual attention when subject to various of the constraints – such as time pressure and text complexity – that are part of most translators’ everyday experience.

There were 21 participants in the experiment, each of whom translated three short pieces of text, of three distinct complexity levels, from English to Finnish. One piece of text was considered easy, and the other two were more difficult. One of the latter was assessed as more complex than the simplest piece for its long and complex sentences, while the final one was judged to be difficult because it used more low-frequency lexical items than the easiest source material did. Moreover, the difficult pieces of text presented more lexical challenges to translation into Finnish than the simplest one did. Each piece of text for translation was about 70 words long. Each participant translated the three pieces of text given, with a six-, five-, and four-minute time limit. In the experiment, the upper half of the screen was used to display the text that had to be translated, and the lower half of the screen was used for the participants’ translation of the text.

Eye movement data from 18 students were considered in the analysis. As a part of human gaze behavior, fixation duration and fixation count (dependent variables) were analyzed for what they reveal of the effect of the independent variables, time pressure and text complexity. Furthermore, we studied translators’ attention to the two halves of the screen (the source and target sections) set up specifically for the translation task.

Firstly, we conducted an observational analysis of the recorded eye movement data and studied the gaze behavior displayed by the participants when they were translating. Although all of the participants were pursuing translation studies, their reading behavior during translation tasks showed inconsistency. Reading strategy ranged from reading the entire paragraph of source text at once to reading sentence by sentence or looking at even smaller segments. Most participants performed some revision of their target text after completing the translation task and corrected it as necessary. Occasionally, a participant consulted the source text and compared it with the translated text, whereas some participants worked on the fine-tuning by reading their own text only. Not all of the participants were able to complete every task within the set timeframe.

In general, touch typists, who need not look at the keyboard as frequently as non-touch typists kept more visual attention on the source window text.
Skilled touch typists were likely to be less affected by an increase in time pressure and also in text complexity than the less skilled typists.

In both the target and the source window, the average fixation count was higher and the average fixation duration was lower for a touch typist than for a non-touch typist. As expected, we found a longer total viewing time for all the touch typists than seen with non-touch typists. Touch typists made significantly more gaze transitions between the source and target window than did non-touch typists.

We found an influence of time pressure on viewing of text in the source window. Average fixation duration in the source window declined significantly as time pressure increased. However, for the target window the difference was not significant. This shows that the translators' reading strategy differed between reading of the source text and reading of their emerging text when a time constraint was imposed.

The variable of text complexity also affected translators' viewing strategy. As we predicted, the translators used significantly more time to look at complex text than simpler text during the translation process. On the other hand, complex text produced a significantly higher fixation count than the simplest text did in reading from the source window. Visual analysis with heat maps supported the statistical results, showing more intense fixation on the complex than on simpler text.

In analysis of overall gaze behavior in both the source and the target window, without regard for the effect of time pressure and text complexity, we found that the average total fixation duration in the source window across all three pieces of text read for translation was significantly lower than for the emerging text produced in the target window. Also, the average fixation count for the source window was higher than the equivalent figure for the target window. Longer fixation duration indicated paying greater attention to the target text relative to the source text. One can conclude that reading the text for translation was simpler than reading one’s own emerging text, for the form satisfying the translator best.

Finally, the write-up of this study specified several useful findings with respect to translators’ touch typing skills, on a limited scale. We saw how the translators’ visual attention was shared between the source and the target part of the screen. The touch typing translators’ gaze traveled more between the source and target areas, while non-touch typing translators’ gaze made more transitions to off-screen locations; this indicates that the touch typing translators had an advantage in being able to observe the text more than the others were. Moreover, we ascertained how time pressure and text complexity affect translators’ visual attention.
The ways in which translators manage their visual attention and its distribution between the source and the target window and also how they handle time constraints and text complexity led to ideas for continuation of the work presented, through investigation of precisely what advantages a translator with touch typing skills might have over one who lacks these skills. Further validation of our results pertaining to distribution of visual attention between the text in the source and the target window can be gained via manipulation of these windows’ position.

5.5 Paper V: “Reading On-Screen Text with Gaze-Based Auto-Scrolling”

The context of the next study differed from that of those presented above in that the movement of the text responded to gaze. In our first study, the text appeared at a fixed rate; for papers II–IV, the speed depended on the generator’s typing speed; and for Paper V, the speed depended on the reading speed. Hence, the paper explores yet another aspect of dynamicness.

The focus with this study was on finding the reader’s preferred regions of the screen for reading of on-screen text. Furthermore, an automatic scrolling technique based on the preferred reading region of each individual reader was implemented and applied in stimulus presentation. We compared eye movements in reading of text with manual scrolling and gaze-based auto-scrolling techniques. Then, we examined whether the manual and automatic scrolling affected reading behavior as measured by typical eye movement metrics, some of which are fixation duration and fixation count. The study also considered how different font sizes in presentation of the text affected the eye movement metrics.

We analyzed eye movement data from 24 participants. Six pieces of text were used in the stimuli, with different content topics. These were of comparable length. Each participant first read three pieces of text, using manual scrolling. Later, the other three pieces of text were read with automatic scrolling. Three distinct font sizes (100%, 125%, and 150% zoom) were used to present text within each scrolling condition. A brief questionnaire was given to the participants for filling in after each piece of text was read. Finally, the participants completed a post-test questionnaire when they had read all six pieces of text.

The results of the analysis did not point to any significant effect of text piece on the eye movement metrics. Therefore, we can assume in this particular case that the effect of the content of the text on the eye movement metrics was negligible.
One of the purposes of this study was to determine the preferred reading region of the screen in reading. Analyzing heat maps for each participant’s gaze data revealed that some readers tended to read from only a portion of the screen while others preferred to read by scanning the whole display. We calculated top and bottom parameters for the preferred reading region of each participant and found great variation between participants. In addition, we were interested in seeing where within the relevant reading range the participants did most of their reading. Results from analysis of the \(y\)-coordinate of fixations for all participants showed that reading of text with 100% zoom (the most common way of viewing on-screen text) displayed little variation between participants; they tended to read from the middle part of the screen.

The study introduced a tool that provided automatic scrolling on the basis of each reader’s preferred reading region on the screen and the eye movement observations made from each individual’s reading with manual scrolling. We compared eye movements during manual and automatic scrolling. Regardless of the font size, reading with automatic scrolling took less time than that with manual scrolling. Fixation count was significantly higher when the text was scrolled manually than with automatic scrolling. On the other hand, fixation duration did not differ significantly between the two scrolling conditions. Accordingly, the greater time for manual scrolling was caused solely by the larger number of fixations. Since insignificant differences in fixation duration reflected no difference in readers’ cognitive pressure in reading, we concluded that the participants read similarly no matter whether the text was scrolled manually or automatically.

As for the preferred reading region, automatic scrolling showed a higher average \(y\)-coordinate of fixation and a lower standard deviation of the \(y\)-coordinate of fixations than manual scrolling did. This was as expected, since the auto-scrolling algorithm tried to bring new text into the reading range with less need for the participant to read from the top of the screen and the reading was targeted at a shorter (and often relatively narrow) part of the screen. The automatic scrolling brought the text to a region where it was convenient to read.

In this study, we observed that fixation count increased and average fixation duration decreased in a linear fashion as font size grew. Font size also affected the standard deviation of \(y\)-coordinate of fixations. With incrementing of font size, the standard deviation of \(y\)-coordinates of fixations decreased linearly. Since all the words are not fixated when reading (because we are able to retrieve information through our peripheral vision), the amount of information gathered via the peripheral vision is reduced with increase in the font size. Hence, people tended to
fixate more often on the text with larger zoom factors than with smaller ones.

From the post-test questionnaire, 63% of the answers regarding participants’ opinion suggested that the automatic scrolling introduced in this study was completely or somewhat convenient during reading, whereas the corresponding figure for manual scrolling was 71%. It is notable that all participants were novice users with automatic scrolling. Being used to the system is likely to increase the level of perceived convenience.

The understanding of preferred reading regions can improve human interaction with on-screen documents. For example, we can use automatic scrolling avoiding the “calibration” phase involved in manual scrolling first when we already find the user’s preferred region of the screen. Moreover, information on which parts of the document have been read can be of use when the reader returns to the document later. Therefore, gaze data collection in general level and particularly with auto-scrolling approach may have several applications.

In future, the same method can be applied to reading from various devices with smaller screens, such as on tablet computers or even cellular phones. Moreover, gaze information can be used in a hands-free system for input to scrolling or to generate zoom-in/out actions and thereby create an appropriate font size for reading in the relevant conditions.

5.6 Paper VI: “Gaze-Contingent Scrolling and Reading Patterns”

As a continuation of Paper V, the final paper delved more deeply into reading patterns on the basis of gaze position during reading when the text is scrolled automatically.

The dataset used in the study was the same as for Paper V except that we excluded eight of the 24 participants. This exclusion was necessitated by a problem in data recording at the level of detail required for the relevant analysis. We used heat maps to get an overview of the distribution of gaze point coordinates. For analysis of gaze paths, we used temporal visualizations of gaze data wherein the x-axis was used to show progress of time, and the y-axis plots the y-coordinates of the fixations. The gaze path visualization suggested that automatic scrolling resulted in smooth reading activity while the gaze in manual scrolling jumped from the end reading position to the new position. Hence, manual scrolling showed a less stable pattern.

We looked at the graphical representation showing the y-coordinates of fixations plotted against timestamps and found two alternative sweeps in
the gaze data, down and up. We found a lot of variations in eye movements among the participants. Faster readers produced fewer sweeps including longer upward sweeps than downward sweeps. This justifies the conclusion that automatic scrolling brought new text into the preferred reading area and the reader managed to read more of it before the text reached the top threshold of the preferred area. The analysis prompted us to find the number of sweeps in automatic scrolling actions, their extent in both height and time, and finally the influence of font size on these metrics. It was found also that font size has a significant effect on the number of sweeps; the number of sweeps grew with font size.

For the previous paper, we analyzed gaze data at summary level. This shows only part of the picture: in addition to the summary metrics, gaze paths are important for revealing what happens during reading. Moreover, visual analysis of the reading patterns clearly shows the qualitative advantage provided by gaze-based automatic scrolling. Reading became more regular with automatic scrolling, and it became easy for the reader to keep track of the current reading point within the text. Detailed information on what goes on during reading is essential if we are to be able to develop auto-scrolling and other techniques further.
6 Discussion

The key challenge facing presentation of on-screen text is to find a presentation format that satisfies the reader. Displaying dynamic text on the screen requires more issues related to the presentation format to be taken into account than does static text. Dynamic rendering text formats may vary on account of the wide range of screen sizes as much as with the purpose and the context of the reading.

Text Presentation Formats and Dynamic Rendering

We have examined eye movements in reading of on-screen text with various presentation formats. The text formats ranged from a single letter to a larger document that required scrolling. Paper I considers four text presentation formats: text in paragraph form, displayed sentence by sentence, shown line by line (on the basis of what fits the width of the screen), and in chunks. Paper II looks at word-by-word and letter-by-letter rendering. Some of the analysis in Paper III deals with only a letter-by-letter presentation format, while, in contrast, the work for papers V and VI used large documents that required the reader to scroll multiple times. Moreover, the text units were presented with different font sizes.

Various text presentation formats have been used in many previous studies, mostly for evaluation of the user’s comprehension rate. The study described in Paper I focused on comparing presentation formats by systematically varying the size of the text pieces and evaluated the eye movement metrics. Results presented in Paper I indicate that gaze behavior differed on the basis of text presentation format during reading for comprehension. When the unit shrank from paragraph to sentence and then to the smallest text unit, chunks, the eye movement metrics too
changed. Specifically, fixation count increased and fixation duration decreased linearly as the size of the text presentation unit grew, from the smallest unit (the chunk) to the largest unit (a paragraph). Furthermore, the same trend in eye movement metrics was observed in the work for papers V and VI when the size of the text to be read increased from 100% zoom to 150% zoom. These findings are in line with several previous results (Kolers et al., 1981; Morrison & Inhoff, 1981), which have established that eye movements in reading of static text are influenced by print quality, variation in fonts, line length and letter spacing, and so on.

When comparing two text display methods, leading and RSVP, Juola et al. (1995) found that text was read aloud more accurately with the RSVP (i.e., chunks) format than when leading was used. Results reported in Paper I indicate that RSVP was the least preferred and found to be more uncomfortable in reading than larger text units. Our findings are in line with the results obtained in a study by Rahman and Muter (1999) wherein higher reading efficiency was found with a sentence-by-sentence (relatively large text unit) condition as compared to their word-by-word condition. In our study, we also found that chunks produced larger fixation duration and a larger number of regressions in reading than did the larger text units: line-by-line segmenting, sentences, and a whole paragraph. Confusion or lack of clarity created by the sentence not yet being complete may be a reason for the larger values of the eye movement metrics for the smallest text unit.

We also examined eye movements in reading of two types of dynamically rendered text: with word-by-word and letter-by-letter presentation (see Paper II and Paper III). In these particular cases, instead of analyzing fixation metrics, we dealt with regressions. As a whole, the findings presented in Papers II and III suggest that the word-by-word text presentation format produced more regressions to previous words and consequently more rereading of previously encountered text. On the other hand, Paper I also showed that reading with the smallest unit (chunks) yielded significantly more regressions than the largest unit (the full paragraph) did. Several prior studies with static text found a relationship between regressions and text complexity (e.g., Rayner, 1998). Usually readers reread more when there exists ambiguity. Therefore, one could assume that reading of chunks created more ambiguity and, consequently, readers showed more regressions. This thought is consistent with the findings presented in Paper I. However, those in Paper II may lead us to infer that, instead of ambiguity, the larger quantity of regressions with the word-by-word format gave readers an opportunity to reread more, which may have helped them understand the text better.

In addition, we considered eye movements when there was a pause in reading of dynamic text (see Paper II). Previous studies have found that
pauses took place during reading of content words, infrequent words, and thematically important words especially in normal reading progress (Just & Carpenter, 1980). Furthermore, longer pauses were found at sentence boundaries and at major and minor clause boundaries within a sentence (Rahman & Muter, 1999; Castelhano & Muter, 2001). Also, our eye movement results showed more rereading during pauses in rendering. Rereading of previously encountered words increased when there was a longer pause, especially in reading of text presented word by word. Furthermore, we found a positive correlation between the duration of the pause between two successive words and the number of regressions started at the preceding words. Pauses provide opportunities for more rereading.

When we focused on the structure and textual context of the material read, we found that the number of regressions starting at the final words of a sentence was significantly higher than that of regressions starting from all other words in the text (see Paper II). The reason might be the longer pause at the end of the sentence, which created a greater opportunity for rereading, for better understanding of the meaning of the preceding sentence, but similar behavior has been observed with reading of static text (Hyönä, 1995). Moreover, readers reread incorrectly rendered or abbreviated words significantly more than the rest of the words.

On the other hand, linguistic analysis of the first and second landing points of regressions (described in Paper III) showed that the lion’s share of those landing points was on primary elements while under ten percent of them were on function words, even though function words accounted for more than 40% of the text. These findings were in line with those of quite a few previous studies, wherein content and function words were fixated about 85% and 35% of the time, respectively (Rayner & Duffy, 1988; Carpenter & Just, 1983). The results of our study matched the lexical hierarchy that other studies have found for language acquisition (Furtner, Rauthmann, & Sachse, 2009; O'Grady, 1987) and reflect a cognitive approach to language processing.

**Reading Context**

In addition to various presentation formats for reading, we have analyzed eye movements in several contexts of reading. The reading context in Paper I was comprehension in a situation that resembles reading from a handheld device. Readers were told beforehand that there would be questions after the reading. In contrast, the context of the text used in the experiments for papers II and III was typical of communication: the stimuli used the output from print interpretation of text from speeches. The purpose of the reading for Paper IV was translation. We focused on the eye movements when someone was reading for translation and at the same time reading his or her own emerging text. Finally, Paper V and
Paper VI deals with reading long passages that required multiple scrolling actions. Together, the articles included in this dissertation cover several reading contexts that are part of many of our day-to-day activities. Some contexts are encountered by all of us, whereas others arise largely with special user groups—such as translators in their work and print interpretation with hearing-impaired people.

Studying translators' eye movements during translation (see Paper IV), we found that time pressure did influence fixation duration in reading of the source text. Although the effect was of only marginal significance, we found that the eyes did not fixate for a long time in one place under time pressure; instead, the average fixation duration in the source area decreased in the time pressure scenario. On the other hand, the fixation duration for text in the target area was not influenced by time pressure. Therefore, in conditions of time pressure, processing the static source text was easier than doing the same with one's own emerging text. This result is in line with the findings of Holmqvist, Johansson, Wengelin, and Johansson (2007) that there were significantly longer fixation durations in reading one's own emerging text as compared to someone else's text.

Schnitzer and Kowler (2006) found more rereading with difficult text than encountered in conditions of simpler text. Similarly, our findings in reading for translation show higher number of fixations in reading of complex text as compared to simple text. All of these findings reflect that complex text requires more fixations to overcome the complexity either in lexical items or in syntax.

Tools to Support Reading

Besides finding a suitable text presentation unit for the relevant screen and reading text in different contexts, we analyzed eye movements in connection with an automatic scrolling technique that may have potential for scrolling of larger text units at a convenient pace without breaking of the text into smaller pieces. Scrolling enables one to access virtually unlimited amounts of information within limited dimensions. Scrolling action is tied tightly to the ability to absorb information through the visual channel. Previous research has suggested the possibility of devising new scrolling techniques through an understanding of reading patterns and humans' reading techniques, along with the individual user's manner of visual information consumption (Beymer & Russell, 2005; Rayner, 1998).

Expanding from the work of Kumar and Winograd (2007) on gaze-enhanced scrolling techniques, we introduced an automatic scrolling approach (see Paper V) that utilized information on the preferred reading region, detected during manual scrolling. We then compared reading behavior, using eye movement metrics, between manual and automatic scrolling. The absence of a significant difference between the metrics for
the two techniques indicated application of similar reading strategies. Eye movements also illustrated that reading with auto-scrolling within the preferred reading region was convenient. Moreover, participants’ acceptance of auto-scrolling even though they were novices with the technique is an encouraging sign for its implementation in all screen based reading conditions. If the reader’s preferred reading region on the screen is known, scrolling can proceed without the need for a manual act, simply by following the eye movements between the boundaries of the preferred region. This approach could be especially helpful when one’s hands are occupied and there is a need to continue reading a piece of text that does not fit on a single screen.

In a continuation of the analysis presented in Paper V, the next paper introduces gaze data analysis on a more detailed level. Temporal visualization of gaze data analysis showed that automatic scrolling yielded smoother reading actions than manual scrolling did. Readers were still able to read during automatic scrolling, and, since the reading became more regular with automatic scrolling, it made it easier for the reader to keep track of his or her current location of reading point in the text.

Paper V also discusses user experiences of the manual scrolling and gaze-enhanced automatic scrolling techniques. Although the automatic scrolling was a new technique for them, many participants in the experiment expressed clear support for it. According to suggestions from some of them, automatic scrolling can be improved via a visual cue for keeping track of the focus position before scrolling starts: since there was variation among the participants in the preferred reading region, there are those (participants who read an entire screenful before wishing to scroll) who might find it convenient to read with a visual marker. Also suggested was a tactile cue to notify the user just before auto-scrolling starts. Tactile feedback may decrease the surprise effect of the auto-scrolling without any interruption of the reader’s flow. The first implementation of this idea (Käki, Majaranta, Špakov, & Kangas, 2014) showed that the tactile cue needs to be timed carefully if it is to be beneficial.

One key finding is that determining the user’s preferred reading region for purposes of guiding gaze-enhanced auto-scrolling can help us build a specific reading tool that can replace conventional manual scrolling with a keyboard or mouse. In future, we should be able to apply the same method in reading from smaller screens, such as those of tablets or even cellular phones.

In addition to the gaze-based tool for scrolling, the studies covered by this dissertation can contribute to development of supporting tools for various groups of users. For instance, Paper II looks at how eye movement metrics were affected in reading of print interpreted text with two distinct dynamic presentation formats, letter-by-letter and word-by-word. At
present, the *de facto* standard in print interpreting is production of text on the screen letter-by-letter, as it is typed. Our main motivation from the practical angle was to emphasize the importance of providing print interpreters with more versatile tools. Again, experiments conducted for Paper II used Sprintanium (Špakov, 2011), which made it possible to buffer the rendering of dynamic text such that it appeared one word at a time. Our findings suggest that both formats were accepted by the users even if the letter-by-letter format was more familiar. Getting used to the word-by-word format may affect its acceptability.

Another example of our research used for tool development is our work for Paper IV, which focused on translators and analyzed their eye movements during reading of the material for translation and of their own emerging, dynamic text. Thorough knowledge of translators’ visual attention and its distribution between the source and target spaces, alongside greater understanding of the role of time constraints and varying text complexity, is important if we are to create gaze-based translation support solutions.

**Individual-to-Individual Variations in Reading**

Previous eye movement research has provided evidence of variations in reading between readers and also that individual readers read differently in different circumstances (Judd & Buswell, 1922; Hyönä & Nurminen, 2006). Moreover, differences in eye movements have been observed between general reading and searching for a target word within the text given (Rayner & Fischer, 1996). In line with these findings, the study for Paper II found three reading strategies among the participants, who were reading print interpreted text. These strategies were defined on the basis of observations as to the frequency of regressive eye movements both visually and algorithmically.

Variations from reader to reader were found also in reading of long documents that involve scrolling (see Paper V). Some participants read a whole page of text before scrolling down to move to the next page, whereas others tended to read from only part of the screen. Paper V also points out that people have their own preferred reading region on the display; this is consistent with the findings of Buscher et al. (2010). Also, translators in their work read the target text and the emerging text differently. The results of the study reported on in Paper IV indicate that some participants read the source text in its entirety in preparation for generating the translation. Others started by reading the source text sentence by sentence and produced their translation accordingly. Differences in reading behavior as revealed by eye movements were correlated with participants’ typing skills during translation. Skilled typists’ gaze traveled more between source text (reading) and target text (writing) than did that of less skilled typists.
As a whole, the findings discussed above suggest that interactive tools making use of human gaze should have the flexibility of providing customization options for the users. An example is the auto-scrolling technique that proceeds on the basis of each reader’s manual scrolling practice. Besides, another tool “Sprintanium” provides flexibility to customize the print interpreted text according to clients’ demand.

In general, although eye movements connected with reading have been researched for quite some time, there is still need for further research. More ways to analyze eye movement data should be explored and refined, until a standard can be specified. Collecting data from reading of dynamic text is still particularly challenging. Furthermore, manual collection of eye movement data is always a questionable method. The results of our studies may show the way toward understanding how we can best deal with the sort of eye movement data seen in various particular cases. With the findings discussed in this dissertation as a guide, we can improve on our analysis for other, future studies.
7 Conclusions

The research presented in this dissertation was intended to observe, analyze, and discuss eye movements during reading of on-screen text for various purposes. For our gaze data analysis, the general hypothesis was that eye movement metrics are influenced significantly by the presentation format, the purpose of the reading, and the content read. Our detailed observational and numerical data analysis suggests that text presentation formats (discussed in Paper I and Paper II); the content of the text, specifically in terms of its difficulty (covered in Paper IV); the structure and textual context of the text (addressed in papers II and III); and font size (considered in papers V and VI) all affected readers’ eye movements.

To confirm the hypothesis in more specific terms, we set four research questions and ran multiple experiments accordingly. Analyses conducted for papers I and II collectively answered our first research question: How do eye movements vary in reading of text with different presentation formats? Results reported in those two papers indicate that gaze behavior differed with variations in text presentation format. In the work for Paper I, we found that eye movement measurements changed significantly as the format unit decreased in size from the largest presentation unit to the smallest text unit. Also, Paper II reported on the significant difference seen in eye movement metrics when text was read with two distinct dynamic rendering formats. In addition, eye movement metrics varied when the text was presented with different font sizes (see papers V and VI).

Our second research question, how does reading of dynamic print interpreted text affect eye movement metrics?, is answered in Paper II, which discusses the research issue at summary level for a specific reading context, print interpretation, for a specific group of users. We found significant
differences in regressions or rereading between two distinct presentation formats when the print interpreted text was read. Our findings indicate that a significant proportion of people reading print interpreted text would prefer a presentation style different from the current *de facto* standard. Besides presentation format, the results described in Paper III showed effects of structure and textual contexts on eye movement metrics.

In response to our third research question, *how do translators’ eye movements behave during reading for translation?*, the work for Paper IV showed that skilled typists’ or touch typists’ gaze traveled more between source text and target text than did the gaze of less skilled typists. In addition, we found that time pressure influenced eye movement metrics in reading of text from the source window. We also found a significant effect of text complexity on translators’ eye movements.

Papers V and VI deal with the eye movement analysis performed to answer our last research question: *Is there any preferred region of the screen in reading, and how does gaze-enhanced automatic scrolling affect eye movements?* The analyses revealed that people do have preferred reading regions. Reading behavior as manifested in eye movement metrics did not differ significantly between the two types of scrolling technique used in the study: gaze-enhanced automatic scrolling and manual scrolling. Hence, reading strategy was shown not to be affected by auto-scrolling. Although the automatic scrolling technique was new to the readers, it was met with sufficient user satisfaction.

This dissertation has explored various reading contexts, including reading for print interpretation, translation, and reading of text with small and large text units. Our findings indicate a significant effect on eye movements in reading with all of these reading contexts. We also found variations between individuals in reading.

Improvement of text presentation formats, especially for dynamically presented text, is an ongoing process. Improving the efficiency of reading electronic text from a normal computer screen is an especially important issue. Reading efficiency is influenced by text or font size, background and foreground color, presentation style (e.g., static or dynamic), and the screen (size and other factors). The results of our studies offer guidelines and document user experience, thereby contributing to presenting text conveniently in various formats.
References


Monty (Eds.), *Eye Movements and the Higher Psychological Functions*. Hillsdale, N.J., USA: Erlbaum.


Conference on Human-Computer Interaction (NordiCHI, Aarhus, Denmark), 217–220. New York, NY, USA: ACM.


url:http://www.respeaking.net/programme/Tiittula.pdf


Reprinted with permission.
The Effect of Different Text Presentation Formats on Eye Movement Metrics in Reading

Selina Sharmin
University of Tampere, Finland

Oleg Špakov
University of Tampere, Finland

Kari-Jouko Räihä
University of Tampere, Finland

Eye movement data were collected and analyzed from 16 participants while they read text from a computer screen. Several text presentation formats were compared, including sentences as part of a full paragraph, sentences presented one by one, sentences presented in chunks of at most 30 characters at a predefined rate, and line-by-line presentation fitting the width of the computer screen. The goal of the experiment was to study how these different text presentation modes affect eye movement metrics (fixation duration, fixations per minute, regressions, etc.). One-way repeated measures ANOVA revealed that differences in presentation format have a significant effect on fixation duration, number of fixations per minute, and number of regressions.

Keywords: Eye movements, reading, text presentation format, fixation count, fixation duration, regressions

Introduction

Eye movement data can provide a powerful source of information that is useful for determining viewers’ interest and intentions. Observing eye movements is an effective way to study visual perception. Visual perception may include reading of onscreen text. One of the most widespread interactions between humans and computers is the reading of onscreen electronic text. Study of eye movements provides a window to the cognitive processes of perception and comprehension that take place during reading.

For many of us, learning and working in day-to-day life involve a lot of reading. We read text for various purposes: desire to gain knowledge, interest, translation, comprehension, correction, copying, etc. There are many factors that may affect the readability of text presented on a computer screen. Mills and Weldon (1987) mentioned several elements that have potential to improve readability on the screen, among them the features (overall size, width, design, and case) of characters, the display’s formatting, the contrast and the color of both characters and background, and dynamic aspects of the screen.

With the development of technology, use of electronic text formats for reading has increased rapidly. In reading electronic text we have to deal with both static and dynamic text. For instance, reading for translation, i.e. reading text in order to create a translation of it, usually is based on static text. In contrast, reading during print interpreting (a communication mode for hard of hearing) means that text is read as it emerges dynamically while a print interpreter creates a transcript of spoken information. The various formats for presentation of dynamic text include scrolling, paging, leading, and RSVP (rapid serial visual presentation). In scrolling, the text is presented in a display area that could be larger than the screen. Readers need to “scroll” the page to continue reading. Paging presents the text divided into pages that fit the screen; the reader can move one page at a time to continue reading.

In leading, the text is scrolled horizontally from right to left in a single line across the screen. Finally, RSVP presents the text in successive chunks of one or more words at a time in a fixed location on the screen at a predetermined rate.

The study presented here involved analysis of eye movement data while participants read text on a computer screen, with several text presentation formats. These formats include sentences as part of a full paragraph, sentences presented one by one, sentences presented in chunks of at most 30 characters at a predefined rate, and line-by-line presentation fitting the width of the computer screen. One-way repeated measures ANOVA revealed that differences in presentation format have a significant effect on fixation duration, number of fixations per minute, and number of regressions.
chunks at a predefined rate (RSVP), and line-by-line presentation that matches the width of the computer screen. Studying the effect of different text presentation formats in reading from the computer screen could bear interesting fruit. The aim of our study was to investigate the effect of these various text presentation modes on standard eye movement metrics (fixation duration, fixation count per minute, regressions). Several studies have examined reading of onscreen text in different contexts and languages. Many studies were conducted mostly to evaluate users’ comprehension rate with different text presentation formats. Nevertheless, no study has considered the presentation formats used in this paper by systematically varying the size of the text block and studied their effect on reading using eye movement metrics. Moreover, most studies concerned with dynamic text have been done using mobile devices, whereas our interest was in reading from a full-size screen, either from a monitor or from its projection on the wall.

We set as the null hypothesis that the text presentation formats do not have an effect on the typical metrics for eye movements. Our results indicate that eye movement metrics are influenced by the text presentation formats. The shortest presentation format got significantly higher fixation durations and more regressions than the longer formats. Consequently, the fixation count was smaller for the shortest format than for the others. In the post test questionnaire, participants indicated the shortest format as their least preferred option. Our earlier study revealed that reading for translation differs significantly from reading one’s own emerging (translated) text (Sharmin et al., 2008). On the other hand, print interpreting requires the audience to read text emerging on the screen interpreted from the speech in real time. We can apply the eye tracking knowledge of this study to develop tools for various onscreen reading purposes, such as reading for translation and reading print interpreted text. In general, these results may help to determine suitable text presentation formats for reading emerging text.

Previous literature related to the topic is reviewed in the next section, after which we present the method of this study including detailed description of the participants, apparatus, procedure, and design. Analysis and results are addressed in the section following that, before final discussion and conclusions.

**Background**

Kang and Muter (1989) discovered that comprehension with the leading text presentation technique (also known as Times Square) was as good as that with RSVP. They also compared RSVP to three versions of Times Square format, with the only variation being in the size of the steps by which the display was scrolled. Kang and Muter found that readers preferred smooth scrolling Times Square over other conditions. However, there was no eye movement analysis in their study. Mills and Weldon (1985) argued that the difficulty, discriminability, and comprehensibility of text may be reflected in the eye movement data. Therefore, these measurements can be used as a method of assessing the cognitive effort involved in reading of text.

Some eye movement studies have been conducted to determine an appropriate reading format for a small display, such as that of a cellular phone. When displaying text on a cell phone’s display screen, Öquist and Lundin (2007) found that text presented in paging format was read significantly more rapidly than that in scrolling or RSVP. No significant differences in comprehension with these formats emerged. The results showed that paging offered the best readability in a cell-phone context and scrolling is the predominant method used nowadays. It was also found that RSVP significantly decreased eye movements, while leading was found to increase them. In RSVP, users’ eyes remain in a fixed position to view the text; hence, reading involved less eye movement than in other presentation formats. Also, Mills and Weldon (1987) reported no real difference between scrolling and paging where readability of text on a computer screen is concerned, though Schwartz et al. (1983) found that novices tend to prefer paging.

Chen and Tsoi (1988) examined the leading text presentation format using English language in terms of users’ comprehension. The results indicated that leading display speed has a significant effect on comprehension of information in English. In an experiment conducted by Juola et al. (1995), sentences were presented on a small display using RSVP and leading formats. Their results indicated that reading was more accurate for the sentences presented in RSVP format than those in leading format. Another study of presenting Chinese text for the small screen, of a wristwatch, by Chien and Chen (2007), found presentation method to be a significant factor for improvement of reading comprehension. Reading comprehension was significantly better with word-by-word for-
mat than with character-by-character. In a dual-task condition, participants had significantly higher reading comprehension scores with speed settings of 150 and 250 cpm (characters per minute) than with the faster setting, 350 cpm.

Three text presentation methods: word-by-word, sentence-by-sentence, and a full page on a 15-inch display were compared by Rahman and Muter (1999). Their findings indicated that reading efficiency was better with the sentence-by-sentence condition than with the word-by-word condition.

A study by Bernard et al. (2001) compared three RSVP presentation styles: a word-by-word, three-line, and 10-line format. Their results showed that the participants had significantly better reading comprehension with word-by-word and 10-line format than with three-line format.

Another study by Lin and Shieh (2006) investigated the effects of the presentation method by using character-by-character and word-by-word presentations for Chinese text. They found that on a single-line screen with leading, recall efficiency was significantly greater for the word-by-word format than for the character-by-character format. When examining the effects of the layout of the presentation method with single-line eight-character and word-by-word formats, they also found that the word-by-word format was significantly superior to the single-line eight-character format.

Regressive eye movements can be a sensitive indicator of reading disruption. Sanders and Stern (1980) studied eye movement patterns with atypical text formats and found that number of regressions is a better predictor of reading disruption than reading speed.

An empirical study by Dillon et al. (1990) found that splitting sentences between screens causes readers to return to the previous page significantly more often to reread text. Splitting was likely to disrupt the comprehension process by placing an extra burden on the limited capacity of working memory to retain the sense of the current conceptual unit. Furthermore, 10–20% of the eye movements found in reading in this condition were regressions to earlier fixated words. The subjective data too revealed a preference for larger screens and high awareness of text format.

**Method**

**Participants**

Seventeen participants took part in the experiment (three males and 14 females). Good data from 16 participants were used for the analysis. Data from one participant was discarded due to technical difficulties in achieving accurate calibration. The participants’ average age was 24.9, with an SD of 1.53 and a range of 22–27 years. All of them were students at the University of Tampere in at least the second year of their studies of English translation as a major or minor subject.

**Apparatus**

A Tobii 1750 remote eye-tracking device was used to track the users’ gaze on its integrated 17-inch TFT color monitor (with 1280 x 1024 resolution). The experiment was recorded with ClearView and Translog/GWM (gaze-to-word mapping). GWM (Špakov, 2008) was developed for mapping gaze coordinates to words on a text document, a functionality that is not routinely provided by eye trackers.

**Procedure and design**

Participants were informed about the test procedure at the outset. Then the eye tracker was calibrated for their eyes. The distance between monitor and participant was 50–60 cm. The participants were also told that there would be a questionnaire with a few questions related to the text at the end. The motivation here was to encourage them to read the passages carefully. Each participant then read eight pieces of text during the experiment. All the texts were about 100 words long and the topics were about global warming and climate change.

<table>
<thead>
<tr>
<th>Paragraph</th>
<th>Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line</td>
<td>Chunk</td>
</tr>
</tbody>
</table>

This table shows the breakdown of text into its constituent parts (paragraph, sentence, line, and chunk).

The text is an excerpt from a study on the effect of different text presentation formats on eye movement metrics in reading.
An example piece of text follows: The Major Economies process is seeking to find consensus among key countries and all industrialized countries must continue to take the lead in emission reductions, in accordance with the principle of common but differentiated responsibilities. The European Union has offered a very courageous commitment for international progress and a new international climate deal that addresses the interests of both developed and developing countries will make everyone a winner. We need a political response to what the scientists are telling us is necessary because a speedy and concerted international action can still avoid some of the most catastrophic projections.

The experiment was divided into two parts. In Part 1, participants read four passages and the changing of text was paced automatically. In Part 2, they read four other passages and the pacing was manual. The motivation behind using both automatic and manual pacing was to understand whether the speed of automatic pacing was appropriate and whether it was close to manual pacing. In each part of the experiment the pieces of text were presented each with a different presentation format (paragraph, sentence-by-sentence, line-by-line, or chunks) (see Figure 1) using a 4 x 4 Latin square format. The passages were organized in such a way that they were counterbalanced.

For the automatically changing text, we calculated approximate optimal timing for the exposure. Unfortunately, there is little or no documentation in previous studies as to exactly how exposure times have been calculated (Juola et al. (1982), Juola et al. (1995), Masson (1983), Muter (1996), and Rahman and Muter (1999)).

A study by Öquist and Goldstein (2003) used the following formula:

$$\text{time0} = \frac{fch}{(w_{avg} \times w_{pm}/60)}$$

Here the average number of characters that can be displayed (fch) is divided by the product of the average word length (wavg) for the current language and a sixtieth of the presentation speed (wpm). The result is a fixed exposure time for each chunk of text, measured in seconds (time0).

Our investigation used the following formula to calculate timing in seconds to present text during automatic pacing. It was derived especially for chunks, but the same was used for sentence, line, and even paragraph units.

$$\text{Time(chunk)} = \left\{ \frac{(60 \times \text{NCC})}{(\text{NCW} \times \text{WPM})} \right\} + \frac{1}{S},$$

where

NCC = number of characters in each chunk,
NCW = average number of characters in each word : 6,
WPM = expected reading speed in words per minute: 140, and
S = number of spaces.

In our study, chunks were constructed according to the rule that there should be at most 30 characters in a chunk, with spaces and special characters included, and no chunk extending beyond the period at the end of a sentence. The WPM parameter was determined through pilot tests.

At the end of the experiment, all participants filled in a background questionnaire. Their opinions were collected regarding readability and understanding of the text, and also on pacing and stress levels during the automatic and manual parts of the experiment. A few questions were asked about the passages.

Raw data with fixation durations was collected using the GWM tool (Špakov, 2008) which uses the dispersion based algorithm (see, e.g., Salvucci and Goldberg (2000)) for fixation computation. In our analysis the filter settings used 40 pixels for maximum fixation radius and 100 milliseconds for minimum fixation duration. For the analysis we considered three independent variables: average fixation duration, number of fixations or fixation count, and the number of regressions. MS Excel and the statistical tool SPSS were used for data analysis aimed at determining the effect of presentation format on fixation duration, fixation count, and number of regressions.

Results

Average reading speeds for different text presentation formats were calculated when the participants used manual pacing. Participants read the text in paragraph format at 136 words per minute whereas the reading speeds were 128, 156, and 137 words per minute for sentence-by-sentence, line-by-line and chunk presentation formats, respectively. No statistically significant difference was revealed between the text presentation formats in terms of reading speed, because of the high variation per participants. The median reading speeds varied between 134 and 144 words per minute, confirming that our setting for the automatic pacing was appropriate.

The main aim of this study was to see how fixation duration, fixation count, and the number of regressions
(dependent variables) would co-vary with the presentation formats for the passages (independent variable). Effects of the independent variable on dependent variables are discussed below, in two subsections, followed by an analysis of the participants’ subjective opinions.

**Effect of text presentation format on fixation duration and fixation count**

Figure 2: Average fixation duration for each participant with the various presentation formats (automatic pacing).

One-way repeated measures ANOVA showed that fixation duration was affected significantly by changes in presentation format ($F_{3,60} = 22.102, p < .001$ for automatic pacing and $F_{3,60} = 8.951, p < .001$ for manual pacing). In both automatic and manual pacing, the longer the presentation unit was, the shorter the average fixation duration observed (see Figure 2).

More specifically, Bonferroni post hoc tests for multiple comparisons found significant differences in average fixation duration in automatic pacing between lines and the other three formats (paragraph, sentences, and chunks) and also between chunks and the other three formats. All of these mean differences were significant at the .05 level. In manual pacing, average fixation duration differed significantly between the format with the smallest presentation unit (i.e., chunks) and the other three formats (paragraph, sentences, and lines) at $p < .05$.

One-way repeated measures ANOVA also revealed that fixation count per minute or normalized fixation count was affected significantly by differences in presentation format ($F_{3,60} = 22.938, p < .001$ for automatic pacing and $F_{3,60} = 9.681, p < .001$ for manual pacing). Predictably, the trend observed with normalized fixation count (see Figure 3) was opposite that seen with average fixation duration. Larger presentation units produced a higher number of fixations than did the smaller units.

Figure 3: Normalized fixation count with different presentation formats (automatic pacing). The error bars represent the standard error of the mean.

More specific analysis via Bonferroni post hoc tests for multiple comparisons found that the normalized fixation count in automatic pacing differed significantly within each pair in combinations of the four presentation formats. The mean differences were significant at the .05 level. On the other hand, with manual pacing the smallest format unit (chunks) showed a significantly lower fixation count than the other three formats did, with $p < .01$. The other small-unit format, lines, had a significantly lower fixation count than the paragraph only at $p < .05$.

Thus it was observed that all the smaller units were associated with significantly longer average fixation duration and lower normalized fixation count than the relatively large units. It was more likely that smaller presentation units for the text brought ambiguity and hence the fixation duration increased for retrieval of the content’s meaning. On the other hand, larger sections of text represented clearer meaning of the context: the viewer can read the text without any interruption. Therefore, instead of paying great attention to any other place in the text, the reader pursues smooth reading with relatively short fixation duration. As a result, in automatic pacing, since the time for reading text in different formats was the same, a higher numbers of fixations appeared, corresponding to shorter fixation durations.
**Effect of text presentation format on the number of regressions**

Presentation formats also influence the number of regressions. One-way repeated measures ANOVA revealed that the number of regressions was affected significantly by presentation format when the pacing was manual with $F_{3,60} = 2.843$ and $p < .05$ (see Figure 4). Post hoc tests for multiple comparisons found that in manual pacing the smallest unit (chunks) showed a significantly greater number of regressions than did the largest unit (paragraph) and the relatively large unit of lines. The mean differences were significant at the 0.05 level.

Thus, regressions were observed more over chunks than with the paragraph format. Readers went back and forth often while reading chunks. On the other hand, the eyes followed a pattern of smooth reading with the paragraph format. Usually, we reread more when there is ambiguity in the text. Using smaller portions of the text (small chunks at a time) does not provide enough information for understanding of the context. Users might feel more cognitive pressure when reading text in chunks. To overcome the ambiguity, participants tended to reread more. These findings also supported the results related to fixation duration and fixation count.

The post-test questionnaire administered to the participants at the end of the experiment comprised short questions to answer from the text. The motivation was to see that the participants had read the text carefully. The post-test questionnaire also gathered subjective opinions on the text presentation formats. There were five claims to be rated on a five-point Likert scale (1 = fully agree, 2 = agree somewhat, 3 = neutral, 4 = disagree somewhat, and 5 = fully disagree) for each text presentation format. The claims are presented in Table 1. Claim 5 used ratings from 1 = “too slow” to 5 = “too fast”. The box plot in Figure 5 presents, in compact form, the responses for the claims from all participants. Each box represents 50% of the data. Edges of the boxes represent lower and upper quartiles of the data, respectively. The bar in the box indicates the median. The dots denote outliers that are further away from the box than 1.5 times its length. Whiskers extend up to the maximum and minimum value of the data points that are not considered outliers. To simplify the figure, we have grouped the data, originally separate for each of the four presentation formats (paragraph, sentence-by-sentence, line-by-line, and in chunks), into two sets, as if there were only two formats. Data from the paragraph and the sentence-by-sentence formats are referred to as “multiple” as those formats consisted of text in multiple lines. Correspondingly, the line-by-line and chunk presentation formats are referred to as “single” (see Figure 5).

![Figure 4: Number of regressions with different presentation formats (manual), with error bars representing the standard error of the mean.](image)

![Figure 5: Box plot representing subjective opinions.](image)
Table 1
Claims on which participants’ opinion was collected

<table>
<thead>
<tr>
<th>Claim</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>q1</td>
<td>When the pacing of the text was automatic, I had to concentrate on being able to read all of the text before it disappeared.</td>
</tr>
<tr>
<td>q2</td>
<td>The automatic pacing of the text did not affect understanding of the meaning.</td>
</tr>
<tr>
<td>q3</td>
<td>Taking the text was stressful.</td>
</tr>
<tr>
<td>q4</td>
<td>The text was much easier to read when the pace was controlled manually instead of automatically.</td>
</tr>
<tr>
<td>q5</td>
<td>When the pacing was automatic, how did the pace feel in the different cases?</td>
</tr>
</tbody>
</table>

In automatic pacing, the text appeared on the screen for a constant duration. Out of 16 participants, 13 expressed the opinion that they did not have to concentrate more to read the text with a larger presentation unit (paragraph or full-sentence) when the pacing was automatic (claim q1 in Table 1). They encountered difficulties and had to concentrate more when reading with relatively small presentation units (line and chunks). In automatic pacing, the number of regressions was significantly higher for chunks than with paragraph format. With paired-samples t-test, the test statistics values were $t = -2.437$, $df = 15$, and $p < .05$.

On the other hand, the participants reported that automatic pacing affected their understanding of the meaning of the text while they were reading with a larger unit, more than when they were reading smaller units of text (claim q2). Consequently, larger text presentation units were more stressful to read than smaller units were (claim q3). Participants’ responses to these two claims in some respects contradict the responses to claim q1 and other findings.

Regardless of the presentation format, most of the participants thought that the passages were much easier to read when the pacing was controlled manually (claim q4). The speed of the automatic pacing was satisfactory for both the bigger and smaller units. For chunks, the respondents had mixed opinions. Six claimed that the pacing was slow, and another six claimed that it was faster.

Discussion

Our results indicated that smaller text presentation units (chunks and line-by-line) produced significantly higher average fixation duration and lower normalized fixation count than did the relatively bigger units (sentence-by-sentence and paragraph). According to Mills and Weldon (1985) eye movement data may reflect the difficulty, discriminability and comprehensibility of text. Therefore, we may assume that smaller units for presentation of the text created ambiguity, which led to increased fixation duration. In contrast, larger text presentation units represent clearer meaning of the context and elimination of ambiguity, since the viewer can read the text without any interruption. Therefore, instead of paying much attention to a single place in the text, the reader performs smooth reading with relatively short fixation duration. These findings are also in line with the studies of Mills and Weldon (1987) and Rudnicky and Kolors (1984), who claimed that larger text sizes are considered more readable than smaller ones.

Post hoc analysis for multiple comparisons found significantly more regressions with the shortest sections (chunks) than with the largest unit (paragraph) and a relatively long-unit format (lines). Often rereading occurs more when there is ambiguity in the text. Smaller snippets of text (i.e., short chunks at a time) do not provide enough information for understanding of the context. Users might feel more cognitive pressure while reading text in chunks. Our findings are consistent with those of Dillon et al. (1990), who found that splitting sentences causes a reader to reread more. Splitting was more likely to disrupt the comprehension process by placing an extra burden on the limited capacity of working memory.

The post-test questionnaire in our study revealed that the participants preferred the sentence and paragraph formats. Their least favored format was chunks. Most of the participants expressed the opinion that they did not have to concentrate more to read the text with a larger presentation unit (paragraph or full-sentence) when the timing to read was fixed or the pacing was automatic. They encountered difficulties and had to concentrate more when reading with relatively small presentation units (line or chunk). Correspondingly, statistical analysis showed that as the text unit shrank from paragraph to sentences and then to chunks, the eye movement metrics changed. The number of regressions was also lower for the paragraph than with chunks. Therefore, the null hypothesis that the presentation formats do not have any effect on the typical eye movement metrics can be rejected. We may summarize our findings as follows: There exists significant effect of text presentation format on the eye movement metrics (fixation duration, fixation count, and regressions) when we read onscreen text.
Our results confirm the findings of the previous studies discussed above and extends them in two ways. First, the length of the text unit to be read was varied systematically. Thus the results allow a quantitative analysis of the effect of the length of the unit, which should help in contrasting the results of previous studies that typically have covered only a part of this design spectrum. Second, our analysis was based on eye movement metrics, which most of the studies on reading dynamic text have not done. Again, although the conclusions concerning the preferred formats match those obtained previously, the eye movement data allows a quantitative analysis of the effects of the different formats.

The reading speeds obtained in our study were lower than those in the study of Öquist and Lundin (2007), who used a mobile device instead of the desktop monitor. This suggests that the width of the display can negatively affect reading speed. However, this would require further experimentation.

Our next step is to apply the current findings in developing a print interpreting tool which provides dynamic text to read for the hard of hearing people as a media of communication. The settings of the available print interpreting tool allow letter-by-letter and word-by-word presentation of the text. The findings of our study should provide useful knowledge to improve this tool with better text presentation formats.

Acknowledgements

This work was supported by the EYE-TO-IT project, part of the EU’s FP6 program, and also by the SpeechText project, which is a part of the MOTIVE program of the Academy of Finland, by means of academic grant funding. We thank the participants for their time and the reviewers for their comments.

References


Dynamic text presentation in print interpreting – An eye movement study of reading behaviour

Selina Sharmin *, Oleg Špakov, Kari-Jouko Räihä

Tampere Unit for Computer–Human Interaction, School of Information Sciences, Kärsämäentie 1, FI-33014, University of Tampere, Finland

A R T I C L E   I N F O

Article history:
Received 13 February 2014
Received in revised form 26 November 2014
Accepted 16 January 2015
Communicated by Duncan P. Brumby
Available online 7 February 2015

Keywords:
Dynamic text presentation
Eye movements
Regressions
Reading
Print interpreting

A B S T R A C T

Print interpreting supports people with a hearing disability by giving them access to spoken language. In print interpreting, the interpreter types the spoken text in real time for the hard-of-hearing client to read. This results in dynamic text presentation. An eye movement study was conducted to compare two types of dynamic text presentation formats in print interpreting: letter-by-letter and word-by-word. Gaze path analysis with 20 hearing participants showed different types of reading behaviour during reading of two pieces of text in these two presentation formats. Our analysis revealed that the text presentation format has a significant effect on reading behaviour. Rereading and regressions occurred significantly more often with the word-by-word format than with the letter-by-letter format. We also found a significant difference between the number of regressions starting at the words that end a sentence and that of regressions starting at all other words. The frequency of rereading was significantly higher for incorrectly typed or abbreviated words than for the other words. Analysis of the post-test questionnaire found almost equal acceptance of the word-by-word and letter-by-letter formats by the participants. A follow-up study with 18 hard-of-hearing participants showed a similar trend in results. The findings of this study highlight the importance of developing print interpreting tools that allow the interpreter and the client to choose the options that best facilitate the communication. They also bring up the need to develop new eye movement metrics for analysing the reading of dynamic text, and provide first results on a new dynamic presentation context.

© 2015 Published by Elsevier Ltd.

1. Introduction

Print interpreting is a method for making spoken language available for people with a hearing disability. In print interpreting, the spoken utterances and other significant audible information are translated into print in real time simultaneously with the speech. The process is also called typing/writing interpreting, captioning, and real-time writing (Tiittula, 2009). The deaf can use other communication methods, such as sign language or lip-reading, but these are rarely used among the hard-of-hearing and those with late deafness (Tiittula, 2009): they have acquired the language from hearing society and usually can speak. Therefore, sign language is not a suitable option for them; rather, they need interpretation that is as close as possible to the original speech.

Print interpreting is widely used in seminars and meetings for the deaf and hard-of-hearing group of people. In print interpreting, spoken language is typed on a computer and the text is displayed either on another computer screen or on a projected bigger screen for a larger audience. The most common presentation format for the text on the display screen is the letter-by-letter format, where letters appear at the rate the text is written. In this dynamic text presentation system, the lines scroll up from the bottom. As soon as the screen is full, the top lines disappear and the new text appears from the bottom of the screen. Similarly, when a word is typed towards the right edge of the screen, if the screen width is exceeded before the word is complete, the characters already typed disappear from the current line and reappear on the next line. Thus the reader sees the text dynamically in real time.

The letter-by-letter rendering of the text is the de facto standard used by professional print interpreters. However, reading text that appears letter-by-letter can be very different from the usual reading of static text where full sentences and paragraphs are in view. The print interpretation process demands that the text appears as soon as possible after the spoken utterance, so that those reading it can understand the mimicry of the speaker and the reactions of the hearing audience. Nevertheless, buffering the text slightly so that it is rendered word-by-word, not letter-by-letter, would probably not have a significant effect on the real time requirement, and reading text presented in such manner would be somewhat closer to the normal reading experience. In addition, although errors in the final interpretation remain visible independently of the rendering method...
and may create distractions in the reading, the word-by-word presentation format has the advantage that it hides those typing errors that the print interpreter corrects on the fly before completing a word.

This motivated us to create a tool, Sprintanium (Spakov, 2011), for studying the process of producing the interpretation and reading the resulting text. Of the many novel features included in Sprintanium, of particular relevance to this study is its ability to optionally produce the text letter-by-letter (the most common format) or one word at a time. Furthermore, it takes as input the real time keypress sequence of the print interpreter and can later render the text in either format at the original rate, allowing a comparative study involving several participants and different experimental conditions.

How does one study the effect of the text and its presentation format on the reading process? Reading requires the visual processing of words, and therefore eye movements provide a window to the cognitive process of perception and comprehension that take place during reading. Eye movements reflect difficulties in understanding the document being read and can also be used to automatically recognise the quality of the text by integrating gaze data from several readers (Biedert et al., 2012).

Reading on-screen electronic text is one of the most widespread interactions between humans and computers. A large number of previous studies have carefully analysed gaze behaviour in reading (Rayner, 1998). In reading, eyes make brief jumps along the line of the text. Rapid movement of eyes are called saccades. Stops in between the saccades are called fixations. Standard metrics in gaze data analysis are average number of fixations and average duration of fixations (Jacob and Karn, 2003). However, eyes do not move forward to read all the time. Often they also move backward for rereading. Saccades that move backward in the text that has previously been encountered by the reader are called regressions. Previous research has documented that regressions are an indicator of comprehension difficulties when reading static text (Rayner, 1998). In the context of print interpreting, where text is rendered dynamically and it appears at a rate that depends on the typing speed of the interpreter, this need not be the case anymore: the reader has time to review past text just to confirm the existing mental model, instead of resolving comprehension difficulties or ambiguities. Thus, the pauses introduced by word-by-word rendering may provide an additional advantage.

Several studies have been conducted focusing on using dynamic text on small display screens of devices like wrist watches, mobile phones, pagers, and desktop phones (Chien and Chen, 2007; Laarni, 2002; Brewster and Murray, 2000). Many studies were conducted mostly to evaluate users’ comprehension rate with different text presentation formats; we give more details in Section 2. However, no formal study has compared different presentation formats in print interpreting, where the text appears dynamically, at the rate of spoken speech, on a computer screen or projected on the wall.

Hence, we carried out an eye movement study where we compared eye movements in reading print interpreted text using two dynamic text presentation formats, word-by-word and letter-by-letter. Our study consists of two experiments. In the first experiment, presented in Section 3, we analysed eye movement data from 20 normal-hearing participants in reading two pieces of print interpreted dynamic text. The texts were presented on a computer screen with two presentation formats: word-by-word and letter-by-letter. The goal of that experiment was to investigate the effect of text presentation format on eye movements and reading behaviour. Specifically, the interest was in finding out whether there are any differences in eye movements during reading that are due to differences in text presentation format. In addition to standard metrics, we examined rereading of the preceding words or sentences, and regressive eye movements during the pauses, during editing, or when an incorrect word is typed. For reasons discussed in Section 4, regressions proved to be a more useful metric than the traditional fixation and saccade related metrics. After the first experiment we carried out a follow-up experiment where 18 hard-of-hearing participants took part. In addition to the stimuli used in the earlier experiment, here we showed the video of the speakers without any sound. The follow-up experiment is described in Section 5.

What did we expect to find in the study? Our main motivation was to compare the two presentation formats, word-by-word and letter-by-letter, in terms of regressions. The characteristic feature of dynamic text is that the pacing is not controlled by the reader. In print interpreting, in particular, there are many reasons why text does not appear at a regular pace, and it can be assumed that the pauses introduced between words that appear on the screen can have an effect on reading behaviour. In particular, with the word-by-word presentation format the pauses are longer, so it can be expected that this shows up in the regression data: it is not natural for the eyes to remain focused at the same spot if nothing happens there. Instead, the time waiting for the next word to appear could be used to review text that has been read already. This yields our two main hypotheses:

**Hypothesis 1.** The word-by-word format causes more rereading than the letter-by-letter format.

**Hypothesis 2.** The length of the pause before a word appears affects the number of regressions starting from the preceding word.

In addition to reading behaviour that is specific to print interpreting, we expected some known facts of reading to carry over to the print interpreting context. First, it is well known that people have different styles of reading static text (e.g., Wotschack, 2009), and there was no reason to assume otherwise for dynamic text. Similarly, it is known that end of sentence is a frequent trigger of rereading with static text (Hyöna, 1995), and we expected this to be the case with dynamic text as well. Moreover, because regressions are used to improve comprehension, we expected that incorrectly typed or abbreviated words would be reread more often than the other words. These considerations gave us three more hypotheses:

**Hypothesis 3.** The number of regressions varies considerably between the readers when reading dynamic text.

**Hypothesis 4.** More regressions start from the last words of sentences than from the other words.

**Hypothesis 5.** Regressions land more often on incorrectly typed or abbreviated words than on the other words.

In general, all hypotheses were confirmed by our analysis. We will discuss the implications of the findings in Section 6. Results from the post-test questionnaire suggested almost equal preference for both presentation formats. The follow-up experiment showed a similar trend for the hard-of-hearing participants. Taken together, the results indicate that the word-by-word presentation format is a viable alternative to be used by print interpreters and should be supported by the tools available for them. For eye movement research the study pinpoints problems with the usual analysis methods (fixation duration and number of fixations), suggests an alternative (number of regressions), and provides the first results concerning reading the specific type of dynamic text that is encountered in print interpreting.
2. Background

2.1. Reading dynamic text

In general, dynamic text – the presentation style used in print interpretation – comes in many forms. Common dynamic text presentation modes include scrolling, paging, leading, and rapid serial visual presentation (RSVP). In scrolling, the text is presented in the traditional form on a display area that may be larger than the screen. In vertical scrolling, text usually moves in a line-by-line manner. With paging also, the text is presented in the traditional form, but it is divided into pages that fit the screen area. In leading, the text is scrolled horizontally from right to left on a single line across the screen. Finally, in RSVP the text is presented successively in chunks of one or more words at a time in a fixed location on the screen at a predetermined rate.

Scrolling, paging, leading, and RSVP have been evaluated thoroughly. For instance, Öquist and Lundin (2007) determined which text presentation format is best to use on a mobile phone. They did not find any significant differences in comprehension between formats, though RSVP was found to decrease eye movements significantly. Their results also showed that paging offered the greatest readability. In studying the most suitable dynamic text presentation method for different types of screens, Laarni (2002) found that scrolling was the fastest method when people read from display screens such as a laptop, a PDA and a communicator (smartphone with horizontal screen layout). He did find an optimal screen type for each presentation method. However, none of the above conditions match those used in print interpreting. A fundamental difference is that in print interpretation the pace of rendering the text is neither constant nor controlled by the reader. Instead, the variations in the speed of the speaker and the interpreter add a new and interesting aspect to the dynamicity. In print interpreting the screen, too, is larger than what has been used in previous studies of dynamic text presentation methods. Finally, in most of the dynamic text presentation formats (leading, scrolling, and RSVP), when text appears from one side of the screen, older text disappears from the other side. In print interpreting, the text fills the available space to the extent possible, and only then does older text start disappearing. This, too, creates more possibilities for regressions, since the previous text is available longer.

Eye tracking, when applied to the study of reading, has generally been employed with reading of individual words, sentences, or pieces of text (cf. Rayner and Pollatsek, 1989; Radach et al., 2004). An eye movement study comparing different formats for presenting dynamic text (page-by-page, sentence-by-sentence, line-by-line, or in smaller pieces of text called chunks) revealed that differences in presentation format had a significant effect on various eye movement metrics, such as fixation duration, number of fixations, and number of regressions (Sharmin et al., 2012). Another study by Rahman and Muter (1999) compared three presentation methods: word-by-word, sentence-by-sentence, and a full page on a 15-in. display. Their results indicated that reading efficiency was better with the sentence-by-sentence condition than with the word-by-word condition. The full page format was on a par with the sentence-by-sentence format. Moreover, a study comparing another three text presentation methods: word-by-word, three-line, and 10-line formats showed significantly better reading comprehension with the word-by-word and 10-line formats than with the three-line format (Bernard et al., 2001). This is the only one of the studies that found a larger unit size (three lines at a time) to be inferior to a smaller unit size (word at a time), but even here word-by-word fared well. These results make one expect that the word-by-word format should have advantages compared to the letter-by-letter format that uses a smaller unit size.

The word-by-word format and letter-by-letter formats have not been compared in previous experiments for languages that use the Roman alphabet. However, for Chinese text Lin and Shieh (2006) found that on a single-line screen with leading, recall efficiency was significantly greater for the word-by-word format than for the character-by-character format. When examining the effects of the layout of the presentation method with single-line eight-character and word-by-word formats, they also found that the word-by-word format was significantly superior to the single-line eight-character format. In another study, for the small screen of a wrist watch, presentation method was a significant factor in improving reading comprehension of Chinese text (Chien and Chen, 2007). Reading comprehension was significantly better with the word-by-word format than with the character-by-character one. Both of these studies again point to the possible advantages of the word-by-word format, although the Chinese language is written and read so differently (top-to-bottom vs. left-to-right, and syllable-based structure) from languages using the Roman alphabet that the results cannot be expected to generalise as such.

2.2. Regressions and eye movements

Rereading a sentence, phrase or passage is known as a regression. It is often the result of a lack of concentration during the first pass through the material. Regressions can also be the result of incomplete processing. Whenever the eye moves forward to word \( n \) before activation of word \( n-k \) (with \( k > 0 \)) was fully removed (i.e. lexical processing was completed), a regression is likely to occur (Engbert and Kliegl, 2011, p. 795).

Regressive eye movements can be a sensitive indicator of reading disruption. Eye movement patterns with atypical text formats were studied by Sanders and Stern (1980), who found that number of regressions is a better predictor of reading disruption than is reading speed. It is likely that many regressions are due to comprehension failures (Blanchard and Iran-Nejad, 1987; Ehrlich, 1983; Hyönä, 1995; Just and Carpenter, 1980; Vauras et al., 1992; Shebilske and Fisher, 1983). A word on the current line is a much more likely target of a regression than the words on previous lines (Duffy, 1992; Ehrlich and Rayner, 1983). When readers encountered a word indicating that their prior interpretation of the sentence was in error, they often made a regression as soon as they encountered disambiguating information (Frazier and Rayner, 1982). Dillon et al. (1990) found that splitting sentences across screens caused an increase in rereading of the text on the previous page. Splitting impedes the comprehension process because of the increased demand for working memory to hold the beginning of the current conceptual unit.

We will base our analysis largely on regressions, so the above results are important for the theoretical foundation. However, the specific circumstances involved in print interpreting are such that the results cannot be assumed to hold without further investigation. The time pressure to keep in synchrony with the speaker has the effect on the print interpreter that the text produced is often grammatically incorrect, has a disproportionate number of abbreviations and misspellings, and is sometimes incoherent if the interpreter falls behind in producing the text. These are all reasons why it can be expected that the number of regressions is higher than when reading static text. On the other hand, the fact that new text appears at an unpredictable rate gives more opportunities for regressions and they are not necessarily caused by similar comprehension difficulties as with static text. We will discuss this dichotomy further in Section 6.
3. Method

3.1. Participants

In total, 24 native Finnish speaking participants took part in the experiment. Data from 20 participants were used for the analysis (four participants were rejected for the bad quality of the eye movement data). All participants had normal or corrected-to-normal vision. The average age of the participants was 28.4 years, with a standard deviation (SD) of 8.9 and an age range of 18–51 years. Seven of them were employed either at the university or in a company, with 1–20 years of working experience. Of the participants, 13 were university students, with various backgrounds, having computer science, environmental science, material science, English translation, biochemistry, information science, mathematics, and the German language as their degree subjects. Three of the student participants were also employees. The participants used computers 5 h per day, on average, with an SD of 2.1 h and a range of 1–8 h.

3.2. Apparatus

We used a Tobii T60 remote eye-tracking device to track the users’ gaze on its integrated 17-in. TFT colour monitor (with 1280 × 1024 pixels’ resolution). The experiment was recorded with Tobii Studio. Sprintanium (Špakov, 2011) was used to prepare the stimulus. We collected the eye movement data with Tobii Eye-tracking Analysis Software Tobii Studio. It was also used for the observational analysis of the eye movements.

3.3. Design and procedure

Two pieces of Finnish text were used in our experiment. The stimuli were produced before the experiment by two interpreters, from different videotaped speeches given by two speakers. The prerecorded speeches were projected on the wall in a room at the University of Tampere while a professional print interpreter transformed a spoken conference-like presentation into written format using Sprintanium (Špakov, 2011). Professional interpreters typed in their own typing style. The layout of the text produced by the interpreters is shown in Fig. 1. These interpreters had very different styles; the first interpreter (who produced Text1) used lots of paragraph breaks and empty space, whereas Text2 (from the other interpreter) forms a continuous stream of text and does not even use capital letters.

The experiment then took place in a university gaze lab. In the introduction, the participants were verbally instructed in the test procedure and informed briefly about print interpreting. They were told that there would be a post-test questionnaire at the end of the experiment including some questions regarding the text. The motivation was to convince the participants to read the text attentively. Participants were also requested to move their head as little as possible while reading. The distance between the participant and the eye tracker was about 60 cm. The tracker was calibrated for each participant before recording of the eye movement data.

Each participant read two pieces of text (Fig. 1), one after the other, one in each presentation format (letter-by-letter and word-by-word format) and one from each interpreter. In the letter-by-letter presentation format the text appeared at the rate it was typed by the interpreters while in the word-by-word format the text appeared as full words as soon as the interpreter pressed the space key or punctuation key. The order of the two pieces of text and the presentation formats were counterbalanced. Each piece of text was presented to the participants for about 5 min.

At the end of the experiment, the post-test questionnaire was given to the participants. It consisted of background questions and questions regarding the text and presentation formats. The total duration of the experiment was about 40 min.

4. Analysis and results

We carried out a careful observational analysis of the video recordings of eye movements to understand the reading process in the different cases. We analysed gaze data recordings for the first four and a half minutes for each of 20 participants reading each piece of text. In that time frame, neither piece of text required scrolling of the display. The total number of words was 160 in Text1 and 186 in Text2. Our analysis included only 155 words from Text1. Five words were discarded because the algorithm that rendered text word-by-word was too simplistic. For instance, a number containing a decimal point was split into two tokens, resulting in abnormal reading behaviour.

A survey was conducted among 11 professional print interpreters regarding the layouts used in our experiment (Fig. 1). Seven respondents said that they follow the layout in Text1, whereas two interpreters used the layout in Text2. Neither of the two layouts was preferred by two interpreters. Most interpreters noted that they tune their style according to the client’s
preferences. The layouts used in our experiment can be considered as extreme ends of the continuum. In addition, they are frequently used as such by the professionals.

4.1. Eye movement analysis method

In reading, properties of the fixated words, such as word frequency and word length, influence their fixation duration (Liversedge et al., 2011). Prolonged fixations are usually taken to indicate more demanding cognitive processing. Our original goal was, therefore, to compare differences in average fixation duration between the two presentation formats, to find indications of differences in cognitive demands. However, this approach did not prove feasible, because of the nature of dynamic text. A typical velocity-based fixation algorithm considers two gaze points to belong to the same fixation if their distance is below a specified threshold value. In reading of static text, gaze jumps ahead in steps of typically 7–9 characters (Rayner, 1998), then it is easy to distinguish fixations from each other. With dynamic text presented in the letter-by-letter format, however, the reading process is fundamentally different. Since the text appears little by little, the gaze follows the progress of letters appearing on the screen. Therefore, two consecutive data points produced by the eye tracker are seldom far from each other and it becomes impossible to distinguish the fixations from each other. One may even question whether reading is based on typical fixations at all anymore in this case; it more closely resembles smooth pursuit of the emerging text (Räihä et al., 2011).

The problem of using the standard fixation-based analysis in connection of dynamic text was also noted by Kruger and Steyn (2013). Therefore they suggested that a new metric, reading index for dynamic texts, could be used. While the metric seems to suit well their research area, which is subtitling, it still is not suitable for our purposes: it assumes that gaze data can be reliably classified into fixations and saccades.

This is why we adopted a different approach. The main quantitative metric used in this study is the number of regressions. This metric has been used in previous reading studies as well. Sanders and Stern (1980) used it to study the effects of text characteristics and Ashby et al. (2005) found it to reflect the reading proficiency of readers. For dynamic text, Specker (2008) used it as an additional metric to support the fixation-based analysis of eye movements in subtitles. Sharmin et al. (2012) studied several dynamic text presentation formats and found the results concerning regressions to be in line with those based on fixation duration and number of fixations.

Since we analysed raw gaze data and not fixations, we had to decide what to consider as a regression: eye trackers are typically noisy, and single gaze data points that deviate from the normal sequence cannot be taken into account. We settled on stipulating that 8 gaze points on a preceding word counts as a regression and rereading of the word. With a 60 Hz tracker this means that gaze stayed on the word for at least 135 ms, which is a reasonable threshold for eliminating chance glances and data anomalies.

Our main interest and within-subjects independent variable was text presentation format: letter-by-letter (lbl) or word-by-word (wbw). Between-subjects independent variables were the order in which the texts were presented, and the order in which the participants saw text produced by each interpreter. The dependent variables were number of regressions starting from a given word and landing on a given word. Significance was tested at the .05 level unless otherwise indicated.

4.2. Eye movements and reading behaviour with different presentation formats

In observation of the eye movements, different reading approaches among the participants were found. Occasionally, rereading a preceding word started when only one or a few letters had appeared in the current word. Gaze did not always return to where it started after rereading the preceding text. For some participants, the gaze often came back to the same position after rereading, and for others it landed on the next new word or letter that had appeared while the gaze was away. Thus regressions differ between participants.
We also found variations in rereading tendencies among the participants. By carefully reviewing the eye movements in gaze path replays, it seemed that participants showed three different reading behaviours on the basis of fixation frequency and rereading as follows (see Fig. 2):

- almost no rereading, focus of the eyes just following the typing of the interpreter most of the time (Minor rereading);
- moderate eye movements and frequent rereading of preceding words from the same line or from other lines (Moderate rereading); and
- extensive eye movements, a lot of rereading, rereading of almost every preceding word (Extensive rereading).

To justify our visual observation of three rereading styles, we used the k-means clustering algorithm (Lloyd, 1982) for the letter-by-letter and word-by-word formats separately. We experimented with k values 2, 3, 4 and 5. The algorithm got as input the percentage of reread words for each participant and condition. Table 1 shows characteristics of the clusters produced by the algorithm: the average distance between clusters (the difference between the maximum rereading percentage in one cluster and the minimum rereading percentage in the next cluster) and the average within-cluster standard deviation.

Ideally we would like the clusters to have large distance between them and small within-cluster standard deviation. We see that with k=2 the clusters are inhomogeneous and close to each other. With k=3 the clusters are well apart from each other and also reasonably homogeneous (have low internal standard deviation). For k=4 the internal similarity would be even better, but this is largely due to one cluster consisting of just one participant. With k=5 the distance between clusters is notably smaller than with k=3. Thus k=3 meets best our expectations for a clustering of the data, and matches our visual observations.

We also experimented with the Mean Shift algorithm (Comaniciu and Meer, 2002), which confirmed the above conclusion. The exact number of clusters is of less importance than the fact that the participants do exhibit different reading behaviour, and that this can be confirmed by algorithmic means.

Table 2 shows the total number of participants corresponding to each reading category and presentation format including the range and average percentage of reread words. Out of 20 participants, 11 belonged to the same reading category in reading text with the word-by-word and letter-by-letter formats. Note that the ranges for the lbl and wbw categories are partly overlapping in Table 2. We did the clustering separately for the conditions, as participants’ rereading activity could change when the presentation format changed. Six participants’ reading strategy belonged to the Moderate category in reading with the letter-by-letter format, while with the word-by-word format they belonged to the Extensive category.

4.3. Rereading and regressions in different presentation formats

In the previous subsection we found that participants had different reading strategies. In this section we report results from several statistical analyses on rereading for the two presentation formats. We take into account two independent between-subject factors, interpreter sequence (which text was shown first) and presentation format sequence (which presentation format was used first). The dependent variables are mentioned for each statistical test.

To begin with, we found that the average percentage of reread words was 52.1 (SD 16.55) for the word-by-word format and 31.3 (SD 11.94) for the letter-by-letter format. Results of three-way mixed-model ANOVA suggested that the percentage of reread words differed significantly between the word-by-word and letter-by-letter presentation formats ($F_{1,16}=35.53$, $p < .001$). Participants reread significantly more in reading with the word-by-word presentation format than with the letter-by-letter format. The between-subject factors, interpreter sequence and presentation format sequence, did not have any significant effect on the dependent variable. Moreover, there was no significant interaction between the independent variables.

The bars in Fig. 3 show the percentage of reread words for each participant in the word-by-word and letter-by-letter formats. Here we have numbered the participants so that participants 1–9 were those who preferred the word-by-word presentation format and participants 10–20 were those who preferred the letter-by-letter format at the end of the experiment. With the exception of participants 1 and 13, all reread preceding words significantly more while reading text in the word-by-word format than in the letter-by-letter format.

Perhaps a stronger indication of cognitive activity than just the share of reread words is the percentage of words reread more than once. Here a similar trend was observed. The percentage of words reread more than once was higher in the word-by-word format.
than in the letter-by-letter format for most of the participants. The average percentage of words reread more than once was 22.7 (SD 12.89) for the word-by-word format and 9.4 (SD 6.69) for the letter-by-letter format. According to three-way mixed-model ANOVA the difference was statistically significant ($F_{1,16} = 19.299, p < .001$). Between-subject factors interpreter sequence and presentation format sequence did not have a significant effect on the dependent variable. In addition, no significant interaction was observed between the variables.

As a continuation of our scrutiny, it was revealed that the percentage of words from which the gaze started a regression was higher for the word-by-word format (on average 24.5% of the words (SD 7.43)) than for the letter-by-letter format (13.3% of the words (SD 5.31)). Analysis with three-way mixed-model ANOVA also showed that the percentage of such words was significantly affected by different presentation formats ($F_{1,16} = 50.275, p < .001$). Although the effect of between-subject factors was not significant over the dependent variable, interaction between the two presentation formats and the difference between the formats was not statistically significant. Hence participants were likely to read onscreen dynamic text with similar regression length independently of the presentation format. Moreover, the between-subject factors neither had any significant effect on the dependent variable nor was there any significant interaction between the factors.

### Table 3

<table>
<thead>
<tr>
<th>First stimulus</th>
<th>Text1, lbl</th>
<th>Text2, lbl</th>
<th>Text1, wbw</th>
<th>Text2, wbw</th>
</tr>
</thead>
<tbody>
<tr>
<td>lbl</td>
<td>25.7 (4.11)</td>
<td>18.1 (6.70)</td>
<td>27.0 (8.15)</td>
<td>24.5 (7.40)</td>
</tr>
<tr>
<td>wbw</td>
<td>15.0 (4.11)</td>
<td>8.6 (4.70)</td>
<td>11.8 (2.77)</td>
<td>13.3 (5.3)</td>
</tr>
</tbody>
</table>

In addition to the percentages of reread words and words reread more than once, we analysed the average regression length in words. We define regression length as the average number of words reread per regression. For the word-by-word presentation format it was very close to that for the letter-by-letter presentation format. For most participants, the regression lengths were similar for the two presentation formats and the difference between the formats was not statistically significant. Hence participants were likely to read onscreen dynamic text with similar regression length independently of the presentation format. Moreover, the between-subject factors neither had any significant effect on the dependent variable nor was there any significant interaction between the factors.

#### 4.4. Effect of pauses on reading

Rereading occurs because of the natural behaviour of participants reading to comprehend the meaning of the text. Sometimes rereading or regressions may be increased because of a longer pause or lack of new text available on the screen.

In the word-by-word presentation format, the pause between two successive words appearing on the screen includes the time to type the word and waiting time between previous and current word. However, in the letter-by-letter presentation format, only the second of these elements, that is the pause between two consecutive words, is experienced by the participants. Otherwise, text appears on the screen in a continuous mode.

During presentation of dynamic print interpreted text on the computer screen, pauses may be caused by several factors, such as these:

- the speaker pausing,
- the interpreter’s failure to follow the speaker,
- the interpreter’s need to summarise the speech in some particular cases before typing,
- mistakes and editing, and
- natural slowing to type longer words.

While reading, the reader does not know what is causing a pause. The pauses can get longer for any or several of the above reasons. One can experience pauses due to mistakes and typing of longer words in the word-by-word format, though these are not visible in letter-by-letter presentation.

Viewing the eye movement video for individuals revealed varying behaviour during pauses. Participants reread preceding words, just stared at the last letter or word, or mixed the two actions during pauses. Sometimes the gaze was directed at the empty area where the new text was expected. We analysed the number of regressions during pauses (Fig. 4) for the two pieces of text (Text1 and Text2). Distributions of regressions are plotted separately for each presentation format and interpreter. One outlier was omitted for each format for Text1.

Fig. 4 shows a concentration of data points in the lower left corner of each graph with a positive relationship between the two variables. This means both that the number of regressions is higher when there is a long pause, and that there are not that many regressions if text appears fairly continuously. More formally, the length of a pause between two successive words is positively correlated with the number of regressions starting at the preceding word. The bivariate correlation is significant with level .01. The Pearson correlation coefficient was .648 for the word-by-word format and .699 for the letter-by-letter format. Moreover, comparing the two presentation formats, we found that the number of regressions during pauses between the words was significantly higher in the word-by-word presentation format than in the letter-by-letter format for both Text1 ($t=5.868$, $df=153$, $p<.001$) and Text2 ($t=7.92$, $df=185$, $p<.001$) (cd. Fig. 4).

It is illustrative to look at the causes of pauses and their effects at the level of individual words. Figs. 5 and 6 show excerpts from the two texts: Fig. 5 from Text1 and Fig. 6 from Text2. There are 40 words in each figure, which means 26% of Text1 and 22% of Text2. These excerpts were chosen so that they illustrate the effect of all the key elements: pauses, long words, abbreviations, spelling errors, and end of sentence.

The upper parts of Figs. 5 and 6 present the breakdown of typing duration. Words from the text are positioned on the x-axes. Light colour bars represent the duration of typing of a word, in milliseconds, while dark colour bars show the time between two successive words: the current word and the next word. The average time for typing a word was 854.6 ms for Text1 by the first interpreter and 931.1 ms for Text2 by the second interpreter, and the average time interval between words was 648.5 ms for Text1 and 424.8 ms for Text2. The words surrounded by rectangles indicate abbreviations or typing errors. A vertical line through the graphs indicates end of sentence.

The lower parts of the figures show the frequency of regressions starting from each word. The light colour portion of each bar shows the number of regressions that started during letter-by-letter presentation while the dark colour portion shows the equivalent for word-by-word presentation. This graphical representation of regressions and pauses indicates that rereading occurred most often at the end of a sentence.

The kind of detailed analysis illustrated in Figs. 5 and 6 is important for finding the reasons for increased regressions. For instance, in Fig. 5 the ninth word from the left (“hormoni”) is incorrect and therefore boxed. It is followed by three x’s, a common way for print interpreters to indicate that the preceding word was incorrect and readers should ignore it. This is often faster than backspacing through the incorrect word. The three x’s are then followed by the word that should have been typed in the first place (“hormoni”). From the lower graph in the same figure we see that this correction procedure caused readers to regress more than usual. The two other boxed words, “MPM” and “tiet.”, are abbreviations. The first is a medical term and an official abbreviation, whereas the latter is just shorthand for this sentence. The next word is typed soon after “tiet.” and it is also the last word of the sentence, so in this case the causes for the regressions can be manifold.

In Fig. 6 the last boxed word, “oli.”, is interesting: the full stop denotes the end of sentence, but it was pressed in error or too early. The next word, “pahimpia.”, is the one that actually ends the sentence. Inspecting the lower graph shows that the period after “oli” has sent many readers to reread preceding text. This is no wonder, as the sentence does not make sense without the last word. Such observations illustrate to the print interpreting professionals the concrete consequences of the errors in their text and can point to spots that need special care.
The figures also highlight the effect of long words, particularly in the word-by-word format. For instance, in Fig. 6 the third word of the second sentence, “tavanomaista”, takes a long time (more than 2 s) to type. Before the word appears on the screen, participants in the word-by-word condition are left with the previous word, “kuvailee”. The bottom graph in Fig. 6 shows that 5 out of 10 participants started a regression from that word. Similar, though less extreme, behaviour can be seen with the last two words in Fig. 6.

From visual inspection of the full texts using the techniques shown in Figs. 5 and 6 our attention is drawn to the effect of abbreviations, spelling errors, and end of sentence. We will next analyse their effect formally.

Fig. 7 shows how many regressions were started in different cases. For instance, for the leftmost bar in Fig. 7: there were 14 words that ended a sentence in Text1. From those words participants started between 6 and 24 regressions, i.e., the participant with the most regressions started a regression almost twice from every end-of-sentence word. On average there were 13.5 regressions, and the bar shows the share 13.5/14 = 96.4%.

Three-way mixed-model ANOVA found that both in word-by-word and letter-by-letter presentation formats regressions from the words that end a sentence were significantly more frequent than from the other words. The test statistics values were ($F_{1,16} = 119.619, p < .001$) and ($F_{1,16} = 219.817, p < .001$) respectively. The between-subject factors did not have any significant interactions in the word-by-word presentation format, but significant interaction was found for the letter-by-letter format ($F_{1,16} = 47.447, p < .001$).

We also found that the incorrectly typed or abbreviated words in the text were reread more often than the rest of the words (see Fig. 8). Three-way mixed-model ANOVA indicated that rereading strategies differ significantly for incorrectly typed or abbreviated words compared to the rest of the words in both word-by-word ($F_{1,16} = 16.633, p < .01$) and letter-by-letter ($F_{1,16} = 17.066, p < .01$) presentation formats.

4.5. Participants’ preference for particular presentation formats

The post-test questionnaire contained short questions based on the text. The motivation was to check that the participants had read the text carefully as instructed. The questions were not designed to be a comprehension test, and there was also not enough data to draw conclusions about the relationship between presentation format and comprehension.

The main goal of the post-test questionnaire was to gather opinions from the participants on the text presentation formats. There were five claims to rate on a five-point scale (‘fully agree’, ‘agree somewhat’, ‘neutral’, ‘disagree somewhat’, ‘fully disagree’). The claims are presented in Table 4. Claim 5 used ratings from ‘too slow’ to ‘too fast’.

Some observations can be made on the basis of the opinions collected from the post-test questionnaire (see Fig. 9). First, there was almost no difference in experienced understandability of the text in the two formats (claim 2). Second, text presented in the letter-by-letter format required less concentration (claim 1) and was easier to read (claim 4) than text presented in the word-by-word format. Participants also found the word-by-word format more stressful than the letter-by-letter format (claim 3). Finally, participants found both formats slow for their reading speed, with the word-by-word presentation experienced as extremely slow.
5. Follow-up study with hard-of-hearing participants

The first experiment was carried out in an eye tracking laboratory using participants with normal hearing. In our second experiment we used hard-of-hearing participants. In addition, the experiment was carried out in a normal meeting room, where we tried to create an atmosphere that resembled better the typical print interpreting context. Silent videos of the speakers were displayed on another screen placed next to the screen with emerging print interpreted text. Thus participants could follow the facial expressions of the speaker if they desired.

We used the same textual stimulus as in the first experiment. Eighteen participants took part in the follow-up experiment. We were only able to get good eye movement data from ten participants for the analysis. Rest of the data was rejected because of technical and physical problems with calibration. This was to be expected, as it is known that tracking elderly people is more complex than tracking younger participants, like those in our first test (Spooner et al., 1980). The average age of the participants was 62.1 years (SD 9.98).

5.1. Rereading and regressions with hard-of-hearing participants

The results obtained in the second experiment were comparable to those in our first experiment. The average number of words reread was 47.92 for the word-by-word presentation format and 41.33 for the letter-by-letter format. Again, the number of regressions that started from the words that end a sentence was significantly higher than the number of regressions that started from the other words both for the word-by-word ($t=2.283$, $df=9$, $p < .05$) and letter-by-letter ($t=4.581$, $df=9$, $p < .01$) presentation formats. Hence, participants were more likely to reread at the end of the sentences (see Fig. 10). Moreover, the difference was significant for Text1 with $t=4.495$, $df=9$, $p < .01$, but for Text2 the difference was only close to significant ($t=2.179$, $df=9$, $p = .057$) (Fig. 10).

These findings are in line with those from our first experiment: both normal-hearing participants and hard-of-hearing participants reread more from the words ending a sentence. However, comparing Figs. 7 and 10 we notice that the difference is less pronounced with the hard-of-hearing participants.

Significant difference was also found between rereading incorrect or abbreviated words compared to the other words in the text. Incorrect or abbreviated words were reread more than the other words with the letter-by-letter presentation format ($t=2.646$, $df=9$, $p < .05$). Although a similar trend was observed in the word-by-word presentation format, the difference was not statistically significant.

5.2. Speakers’ appearance and different presentation formats

The hard-of-hearing participants had the silent video of the speaker available on an adjacent monitor. This allowed them to follow the speaker in addition to the interpreted text, much like in a typical situation. The participants showed different tendencies in following the speaker. Four participants did not look at the speaker at all with either presentation format, or just glanced at the video very briefly. Of the remaining six participants, four looked at the video more in the letter-by-letter presentation format and two viewed it more in the word-by-word format. The average number of gaze moves to the speaker during reading with the letter-by-letter presentation format by all participants was 8.8, which was very close to the value of 9 for the word-by-word presentation format.

Although the word-by-word format provides longer pauses than the letter-by-letter format, which may allow readers to move their eyes more, in the current experiment the participants looked

---

Table 4

<table>
<thead>
<tr>
<th>Claims for rating by the participants.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
</tbody>
</table>
at the speaker almost equally long: 29.6 s with the letter-by-letter presentation format and 27.7 s with the word-by-word format. Even though reading with the word-by-word presentation format provided longer waiting time, participants neither moved their gaze to the speaker significantly more nor spent a longer time looking at the speaker than in the letter-by-letter presentation format. Rather, participants utilised their time in rereading previous text.

5.3. Participants’ preference for particular presentation formats

We used the same claims in the follow-up experiment as in the previous experiment (Table 4 in Section 4.5). Out of 10 participants five preferred the letter-by-letter presentation format and four preferred the word-by-word format. One participant stopped the test and did not give his opinion. Several participants indicated that the reason to prefer the letter-by-letter presentation format was because of its familiarity; existing software tools available for print interpreting produce text in the letter-by-letter format. In this light it is actually surprising that the word-by-word format fared so well.

When comparing the opinions provided by normal-hearing and hard-of-hearing participants, some interesting similarities and difference emerged. First, the need for concentration and stressfulness were rated similarly by both groups. The hard-of-hearing participants rated the word-by-word format higher in readability, contrary to the hearing participants who had preferred the letter-by-letter format. Most surprisingly, when the hearing participants had experienced the word-by-word presentation as far too slow, the hard-of-hearing participants found the pace too fast. We will discuss these findings more in the next section.

6. Discussion

We started our analysis by investigating the hypothesis that people read dynamic text in different ways. This was confirmed first by observing their individual gaze paths and then by applying clustering algorithms on the gaze data. Clustering, as the rest of our analysis, was based on the frequency of regressions back to text that had appeared before the current target of reading.

Analysis of the participants’ eye movements identified three categories of reading behaviour, on the basis of differences in the number of regressions. The categories were Minor rereading, Moderate rereading and Extensive rereading. Eleven of the 20 participants belonged to the same category in reading the text with the word-by-word and letter-by-letter formats. The rest changed their reading strategy when the presentation mode changed. Six participants belonged to the Moderate rereading category in reading with the letter-by-letter format but moved to the Extensive rereading category in reading with the word-by-word format.
We then drilled down on the causes of the regressions. Of particular interest to us was the performance of the word-by-word presentation format, since it is a novelty that is not yet supported by commercial print interpreting tools. We hypothesised (Hypothesis 1) that there would be more rereading with the word-by-word format than with the letter-by-letter format. As already indicated by the above observation on change of reading strategy, the hypothesis was confirmed. Three-way mixed-model ANOVA showed that the percentage of reread words was significantly higher with the word-by-word presentation format than with the letter-by-letter format. Further support for the hypothesis was found by analysing the share of words reread more than once, which was significantly higher in the word-by-word format than in the letter-by-letter format for most of the participants. Finally, the number of words from which the gaze started a regression was also significantly higher for the word-by-word presentation format than for the letter-by-letter format.

A fundamental difference between the word-by-word format and letter-by-letter format stems from the different pace at which new text appears. In the letter-by-letter format there is a fairly steady rhythm that follows the typing speed of the interpreter. With the word-by-word format new factors, especially the length of words, come into play. From previous studies on reading static text we already know that there can be pauses in reading even when new text would be available. For instance, the longest fixation durations and pauses (in normal progress) have been found in reading content words, infrequent words, and thematically important words (Just and Carpenter, 1980). Moreover, relatively longer pauses were made at sentence boundaries, whereas additional pauses were observed within sentences at major and minor clause boundaries (Rahman and Muter, 1999; Castelhano and Muter, 2001). Looking at the eye movements in our study revealed different gaze behaviour during pauses. The number of regressions during pauses between the words was significantly higher in the word-by-word presentation format than in the letter-by-letter format. A significant positive correlation was found between the length of the pause between two successive words and the number of regressions starting at the preceding word. This supports our second hypothesis that the pause before a word appears affects the number of regressions starting from the preceding word.

In addition to the dynamicity of the presentation, the textual content is another factor that affects reading. Considering first the starting points of regressions, there was a statistically significant difference between the number of regressions starting from the words that end a sentence and that of regressions starting from all other words in the text, confirming our Hypothesis 4. This can be for two reasons: the natural pause when the interpreter starts interpreting a new sentence, and the reader's desire to make sure that the meaning of the preceding sentence was understood.

Finally, Hypothesis 5 on the landing points of regressions was also supported by our data. We found that the frequency of rereading of incorrect or abbreviated words was significantly higher than the frequency of rereading the rest of the words.

In the post-test questionnaire, the participants rated the two presentation formats almost equally in terms of the understandability of the text. However, the letter-by-letter format was rated as requiring less concentration, less stressful, and easier to read. This sounds as bad news for the word-by-word format, but the free-form comments by the participants shed some light on this. The positive comments on the letter-by-letter format seem to have been caused by the mechanical act of reading, not on reading for comprehension. One participant commented: "Reading text that appeared word-by-word felt surprisingly bumpy; letter-by-letter felt rather fluent, almost as if I had typed the text myself based on what I had heard." Another participant continued: “I would rather read text that appears letter-by-letter, which might be due to my better concentration on the reading; however, I remember better what I read when the text appeared more slowly word-by-word that loosened the grip on the gaze.” These two participants preferred the letter-by-letter format. A comment by a participant that preferred the word-by-word format ends in a similar observation: “It was more tiring to read text that appeared letter-by-letter, since it was much more difficult to keep in mind the context of the words.” Thus, all these comments point to the fact that it was easy to feel that one was reading and making progress with the letter-by-letter format, but it did not necessarily lead to a good understanding of the text. This makes it understandable that when asked which format the participants would have chosen for themselves, the opinions were almost equally split between the two alternatives.

There is some past research that brings up similar observations. Granas et al. (1984) found that static presentation resulted in a significantly better reading comprehension than a leading dynamic format did. Experimenting with different leading formats, they observed that text appearing one or two characters at a time produced poorer reading comprehension than text appearing from four to ten characters at a time. This is very similar to letter-by-letter and word-by-word presentation formats. Cognitive demands with the letter-by-letter format may be similar to those with the leading format. The word-by-word format, by contrast, entails more time between the words appearing on the screen, and readers can use this time to reread the previous words to improve comprehension. Similarly, for Chinese text recall efficiency was significantly greater for the word-by-word format than for the character-by-character format (Lin and Shieh, 2006). Our findings are in line with these results.

With static text, regressions have been linked to the need to reread ambiguous words or sentences (Rayner, 1998). The discussion above indicates that with dynamic text this may not be the case: regressions can be beneficial for comprehension also when there is no ambiguity in the text. A similar result has recently been shown regarding static text, too (Schotter et al., 2014).

It is illustrative to review the rereading behaviour shown in Fig. 3. Participants 1–9 who in the end preferred the word-by-word format used in the letter-by-letter condition an almost equal number of regressions as those participants (10–20) who preferred the letter-by-letter format: the averages for the letter-by-letter condition were 35.4 and 34.4 regressions. The difference comes from the frequency of regressions in the word-by-word format: those preferring that format made more use of it (54.7 regressions on average) than those preferring the letter-by-letter format (44.5 regressions on average). When regressions took place, they were similar in both presentation conditions: for instance, average regression length (number of words reread per regression) was similar with both presentation formats. Thus it seems that it was the increased number of regressions that was linked to increased preference for the word-by-word format. Interestingly, the interaction effect discussed in Section 4.3 indicates that this advantage was only obtained when the participants did not start with a condition that encouraged them to keep their eyes stationary.

One may also ask if the regressions really reflect reading behaviour, or whether they are simply the consequence of more spare time for the eyes to wander in the text (Vitu et al., 1995; Rayner and Fischer, 1996). First, we controlled this by demanding at least eight gaze points to land on a word before a regression was recorded. Second, analysis of the word classes of reread words (Sharmin and Wiklund, 2014) shows a distribution that strongly resembles the distribution in normal reading: content words (nouns, verbs, and adjectives) are reread much more often than their share of the text. For instance, the first or the second landing point of a regression is a noun in 74% of the cases, but only 30% of all words in the texts are nouns. This is a strong indication that the landing points of regressions were not random.

Our first experiment was carried out with hearing participants. More important than their ability to hear (since there was no...
audio in the experiment) was the fact that they were all novices to print interpreting. We therefore arranged a follow-up experiment with hard-of-hearing participants. We were able to collect data from a smaller group than in the first experiment. Nevertheless, the results confirmed our earlier findings.

The second experiment differed from the first one in that a silent video of the speaker was displayed on an adjacent monitor, resembling a real-life print interpreting setup. It turned out that this did not affect the findings. Participants viewed the video relatively little even when the word-by-word format provided more opportunities for it. The viewing time was very similar for both formats.

Again, findings from the post-test questionnaire suggested almost equal preference of both presentation formats. Some participants indicated that the reason to prefer the letter-by-letter presentation format was because of its familiarity: most of the software tools available for print interpreting today render text in the letter-by-letter format. In this light the word-by-word format was received surprisingly well. The main difference to the results from the first experiment was that the participants now considered the pace with the word-by-word as too fast. This, too, may be because the format was unfamiliar to them.

7. Conclusions

Today's de facto standard in print interpreting is production of dynamic text on the screen letter-by-letter as it is typed. Sprinta- nium makes it possible to buffer the rendering of dynamic text such that it appears word-by-word. We wanted to find out the effect of this alternative presentation format on the reading process and users' opinions concerning it. An eye movement study of print interpreting was carried out to gather information on how people read dynamic text in the print interpreting context.

From a practical point of view, our main conclusion is the importance of providing print interpreters with more versatile tools. A significant portion of their customers would prefer a different presentation style than the current de facto standard. There are indications that the word-by-word presentation style can improve comprehension, especially after customers become familiar with it. Another message to the print interpreting profession is to bring up, in addition to the speed of interpretation, the importance of correctly typed full words, as abbreviations and spelling errors impede the reading process.

Methodologically our main contribution is the discussion of the problems in using standard eye movement metrics and the possibilities of using an alternative metric. Clearly, this is a topic that needs further attention from the eye tracking community. New metrics have already been proposed for analyzing the reading of dynamic text, but they need to be further developed to cope with the level of dynamicity (letter-by-letter) in print interpreting and other dynamic presentation formats.

Finally, from the point of view of understanding eye movements, the paper shows the difference in reading behaviour between two dynamic presentation formats. Further studies need to dig deeper into this theme to understand the phenomenon better and to contrast reading dynamic text and static text in a controlled study.

Also in the print interpreting context there are more conditions left for further study. For instance, our experiment used stimuli where the speeches were shorter than 5 min, and thus the entire interpretation could fit on the screen. The effects of new text always appearing at the bottom of the screen and old text disappearing from the text need another experiment. In English-speaking countries it is common to use automatic speech recognition for producing the print interpretation. Then it would be interesting to study a setup that is reverse to ours: since words a recognised one at a time, what would be the effect of making them artificially appear letter-by-letter?

In summary, print interpreting provides a rich context for the study of the profession, of print interpreting tools, and of eye movements alike.

Acknowledgements

This work was supported by the SpeechText project, part of the MOTIVE programme of the Academy of Finland. We thank the participants for their time and Sirpa Laurén for making the hard-of-hearing participants available. We also thank project partners Liisa Tiittula, Mari Wiklund, Anukaisa Alanen, and Maija Tanhua-mäki for their comments. The detailed comments of the reviewers and the editor that helped to improve the presentation are also appreciated.

References


Schotter, E.R., Tran, R., Rayner, K., 2014. Don’t believe what you read (only once); comprehension is supported by regressions during reading. Psychol. Sci. 25 (6), 1218–1226.


Selina Sharmin has been working as a researcher in Visual Interaction Research Group in Tampere Unit for Computer–Human Interaction (TAUCHI) at the University of Tampere. Her area of interest is to study human gaze behaviour and analyze eye movement data. She got her Bachelor of Science and Master of Science in Statistics from the University of Dhaka, Bangladesh in 1998 and 2000 respectively. She also obtained her Master of Science in Interactive Technology at the University of Tampere in 2004. Currently she is pursuing her post-graduate studies in Human Gaze Behaviour in Reading at the University of Tampere.

Oleg Špakov is a member of the interaction by gaze community. The development of applications with gaze used as the only or additional input modality, and studying and improving various object selection methods used in gaze-controlled interface is his main interest. He received his Ph.D. degree from University of Tampere in 2008, and the dissertation was focused on studying various aspect of visualization of gaze data. Currently he is a post-doc researcher at the University of Tampere and developing a ‘native’ gaze-contingent interface.

Kari-Jouko Räihä obtained his Ph.D. in Computer Science at the University of Helsinki in 1982. Since 1985 he has been a full professor of computer science at the University of Tampere. He has done research in applied eye tracking for more than 15 years. His primary interest is using eye gaze for computer control, both for users with motor impairments, and also in the context of attentive interfaces.

Gaze Behaviour and Linguistic Processing of Dynamic Text in Print Interpreting

Selina Sharmin
School of Information Sciences
University of Tampere, Finland
{firstname.lastname}@uta.fi

Mari Wiklund
Department of Modern Languages
University of Helsinki, Finland
{firstname.lastname}@helsinki.fi

Abstract

Print interpreting is a form of communication that allows deaf and hard of hearing people to get access to speech. We carried out an eye tracking experiment where twenty participants read print interpreted text presented dynamically on a computer screen. We compared regression landing points on reread words between two dynamic text presentation formats: letter-by-letter and word-by-word. Then we investigated the gaze behaviour from a linguistic point of view in order to discover whether the dynamic presentation has an effect on linguistic factors. In particular, we have examined the parts of speech of the first and the second landing points of regressions. The findings suggest significant difference between the presentation formats. There is also a relationship between the gaze behaviour and the linguistic processing of dynamic text. Being conscious of this lexical hierarchy may help to develop supporting print interpreting tools and consequently may also help print interpreters to improve the presentation of dynamic text to the user.

CR Categories: H.5.2 [Information Interfaces and Presentation]: User Interfaces – Evaluation/methodology; Input devices and strategies

Keywords: regression, gaze behaviour, linguistic processing, print interpreting, eye tracking, eye movement, dynamic text

1 Introduction

Gaze behaviour has been used in many studies to investigate human cognitive processes. Analysis of eye movements along with linguistic features of the text can be used to recognize indirectly the current mental state of the reader.

Dynamic text presentation formats are often used in various large and small screen electronic devises. Text proceeds automatically in dynamic text presentation format. Bernard et al. (2001) compared three text presentation methods: word-by-word, three-line and 10-line text formats. Reading comprehension was better with the word-by-word and 10-line formats. Studies by Chien and Chen (2007) and Lin and Shieh (2006) have shown that the reading comprehension is better with the word-by-word format than with a letter-by-letter format in reading Chinese text. None of the previous studies have compared dynamic text presentation techniques for print interpreting. Print interpreting is a form of communication that allows deaf and hard of hearing people to get access to speech. In print interpreting, the interpreter translates the speech including significant audible information into written format in real-time. In this study, we compare the gaze behaviour in reading print interpreted dynamic text between two presentation formats, word-by-word and letter-by-letter.

Moreover, the current study investigates the gaze behaviour from a linguistic point of view. The paper focuses on the regressions of eye movements and on their relationship to linguistic factors. The aim of this study is to discover if regressive eye movements vary between the mentioned two dynamic text presentation formats, and if there is any relationship to linguistic factors. The analyses are theoretically based mainly on O’Grady’s (1987) cognitive approach to the acquisition and to the use of language.

In O’Grady’s (1987) cognitive approach, nouns are “primary” elements of language, because they are characterized by autonomous meaning and function. Their referents are perceptively distinct and coherent. Verbs, for instance, have a more fragmented meaning. The referents of verbs are not “present” in the perceptive field as concretely as the referents of nouns (Gentner 1982; Maratsos 1991; Caselli et al. 1995, Furtner et al. 2009). Verbs and adjectives are “secondary” elements. They depend on a relationship to at least one primary element. Function words, in turn, are “tertiary” elements. They depend on a relationship to at least one secondary element. Because of this lexical hierarchy, children tend to learn first most of all nouns, then verbs and adjectives, and the function words are the last ones to be acquired (Furtner et al. 2009).

Gaze behaviour in reading has been analysed in several previous studies (Rayner 1998). Typical fixation duration in reading is 100–500 ms. However, readers do not fixate all the words. Foveal processing of each word is not necessary. Especially many short words are skipped (Weger and Inhoff 2006). Previous studies have shown that content words are fixated about 85% of the time, whereas function words are fixated only about 35% of the time (Carpenter and Just 1983; Rayner and Duffly 1988, Furtner et al. 2009).

In reading, eyes do not move forward persistently all the time; they also move backwards for rereading. Opposite movements of eyes from right-to-left along the line or movements back to previously read words and lines are called regressions.

The first landing point is the first word where the regression lands at. The second landing point is the second word where the regression lands at. Short within-word regressions may occur when the reader has difficulty in processing the currently fixated word (Carpenter and Just 1983; Rayner and Duffly 1988). Longer regressions back along the line or to another line may occur because the reader did not understand the text (Rayner 1998).
In this study, we have examined the parts of speech of the first and the second landing points of regressions. Then we have classified them into primary, secondary and tertiary elements. The objective is to discover if the same lexical hierarchy that prevails in the acquisition of language can also be found here.

We examined eye tracking data from an experiment where twenty test participants read a dynamic text presentation on a computer screen. In our data, the text is a spoken presentation that has been transformed into written format. This may play a role in the gaze behaviour. Results of this study reveal that gaze behaviour differs due to different presentation formats. More rereading occurs in the word-by-word format compared to the letter-by-letter format although the number of reread words per regression was almost equal for both presentation formats. From the linguistic point of view, the findings suggest that there is a relationship between the gaze behaviour and the linguistic processing of dynamic text. Understanding this lexical hierarchy may help to develop supporting print interpreting tools and consequently may also help print interpreters to improve the presentation of dynamic texts to the user.

2 Background

Eye-tracking is a sensitive method to language processing without interfering with it. It allows non-disruptive observations in natural experimental settings. Previously many studies have tested the participant’s ability to maintain information from the perspective of reading span tasks. It has been shown that elaborative encoding strategies – such as chaining, mental imagery and semantic elaboration – are more beneficial than simple rehearsal (Daneman and Carpenter 1980; Friedman and Miyake 2004; Kaakinen and Hyönä 2007; Turner and Engle 1989; McNamara and Scott 2001). The development of general theories of language processing also made it possible to use eye movements for examining cognitive processes underlying reading (Rayner 1998). Tanenhaus and Trueswell (2006) gave an overview of research that uses eye movements to investigate spoken language comprehension. The study was about spoken language, but nothing was discovered about transcribing spoken language into written format.

Different factors, such as a word’s frequency, length, predictability, and ease of integration into the sentence are believed to influence eye movements on the particular part of the text. Those factors influence whether and for how long the eyes flit on a word (Just and Carpenter 1980; Rayner 1998; Reicher et al. 2003). Readers develop structural (syntactic) representation of sentences incrementally in reading each word. Usually detection of syntactic or semantic irregularities in a word evokes longer fixations, regressive eye movements, and rereading (Altmann et al. 1992; Ferreira and Henderson 1990).

Rereading is a natural human behaviour of eye movements in reading. It can indicate an active process to serve a useful function, such as allowing readers to improve text comprehension or fill in gaps in memory about the content of the text (e.g. Levy et al. 1992). Past research has also shown that look backs or rereadings are often an indicator of comprehension difficulties (Rayner 1998). If the comprehension process does not go well, readers tend to look back more. On the other hand, look back fixations to the most important segments of the text are strategic in nature (Hyönä et al. 2002; Hyönä and Nurminen 2006). Thus, different eye movement behaviours in reading, such as looking at the text for a long time or creating longer fixations or looking back, could be caused by different cognitive mechanisms.

3 Methods

3.1 Participants

Twenty native Finnish speaking participants took part in the experiment. All participants had normal or corrected-to-normal vision. The average age of the participants was 28.4 years, with a standard deviation (SD) of 8.9 and an age range of 18–51 years. All of them were either members of the university staff or students.

3.2 Apparatus

A Tobii T60 remote eye-tracking device was used to track the users’ gaze on its integrated 17-inch TFT colour monitor (with 1280 x 1024 pixels’ resolution). Eye movement data were collected with Tobii Studio. It was also used for the observational analysis of the eye movements.

3.3 Procedure and Design

First, participants were informed about the test procedure. Then the eye tracker was calibrated for the participants’ eyes. The distance between the monitor and the participant was about 60 cm. Stimulus consisted of a text (160 words in length), which was the output of a print interpreting process, where a professional print interpreter was transforming a spoken conference-like presentation into written format in Finnish. The interpretation was first produced in a live situation, and afterwards rendered at real-time pace on the screen either in the word-by-word presentation format or in the letter-by-letter format. Half of the participants read the text presented in word-by-word format while the other half read it in the letter-by-letter presentation format.

A background questionnaire was delivered at the beginning of the test. There was also a post-test questionnaire regarding user experience about reading the text. Participants were informed about the post-test questionnaire in order to motivate them to read the text carefully.

4 Results

Videos of the eye movements of each participant were used in the analysis. A careful observational analysis was carried out to spot the words from which the rereading or the regressions started as well as the words where the regressions landed at.

![Figure 1: Percentage of words from which rereading started. The error bars denote the standard deviation.](image-url)
presentation format. Moreover, the percentage of regression landing points in the word-by-word presentation format (83.25) was significantly more than in the letter-by-letter format (50.63) with \( p = .05, F_{1,10} = 4.434 \). Thus, participants were more likely to start rereading and consequently there were more regression landing points in the word-by-word presentation format than in the letter-by-letter presentation format. On the other hand, it is interesting that although there were more regression landing points in the word-by-word format, regression length or percentage of average number of reread words per regression was almost equal for both presentation formats (average values for the letter-by-letter and the word-by-word formats were 3.44 and 3.73, respectively).

We continued our analysis from the linguistic point of view by observing the landing points per regression in the letter-by-letter presentation format. We examined the words at which the first two regression points landed. Our data consists of 109 regression clusters. As all the first landing points are followed by second landing points, the total number of the first and the second landing points is 218 (= 109 + 109). We have examined the parts of speech of these first and second landing points of regressions. Then we have classified them into primary, secondary and tertiary elements. Our linguistic analysis is theoretically based mainly on O’Grady’s (1987) cognitive approach to the acquisition and to the use of language.

The results show that the first landing point is a noun (that is, a primary element) in about 52% of the cases. For instance in Example 1, the first landing point is the word ‘ryhmittymä’ (‘groups’), which is a noun. The text in Example 1 is presented in Figure 2 with landing points as clusters of circles. Numbers inside the circles indicate the ascending sequence of gaze points.

Example 1:

‘In America there are even religious groups that champion this cause.’

The first landing point: ryhmittymä (noun)

![Figure 2: Example of a noun as the first landing point](image)

When we studied the landing points in more detail, we discovered that the first landing point is the closest noun in 33% of the cases.

Example 2:

‘It refers to developmental disorders of the central nervous system, including autism and aspegren syndrome [sic].’

The first landing point: aspegren (the closest noun)

The first landing point is a content word (that is, a primary or a secondary element) in 76% of the cases.

Example 3:

‘There has been a lot of discussion about MPM [sic] -vaccines (?) that they would cause autism.’

The first landing point: tuottais, ‘would cause’ (a verb → a content word)

In Example 3, the first landing point is the verb tuottais (‘would cause’), which is a content word. The first landing point is also the closest content word here. This is the case in 43% of the regressions.

When we considered the first and the second landing points of the regressions, we obtained the following results. The first or the second landing point of the regression is a noun in 74% of the cases, even if less than 30% of all words of the data are nouns. This is in line with the findings of Furtner et al. (2009) according to which readers recur their fixations to nouns more than to words of other parts of speech in order to enhance the comprehension of the surrounding words. When the first or the second landing point is not a noun, it is most likely a verb or an adjective: indeed, the first or the second landing point is a content word in 91% of the cases. This is interesting, because only 57% of all words of the data are content words. Function words, in turn, are rare in this position: the first or the second landing point is a function word only in 9% of the cases, even if as much as 43% of all words of the data are function words.

5 Discussion and Conclusions

Earlier studies have shown that the comprehension is improved in the word-by-word format compared to the letter-by-letter format. On the other hand, Rayner (1998) has documented that look backs or rereading can be indicators of comprehension difficulties. In this study, observing gaze behaviour revealed significantly more regressive starting points and landing points, i.e., participants reread previous words more in the word-by-word format than in the letter-by-letter format. The increase in number of regressions can have two causes: poor comprehension on first reading, or desire to strengthen comprehension when there is a better chance for it because the new text appears with longer breaks between words than between letters. In either case regressions should help and result in better understanding. In addition, regression length or percentage of reread words per regression was almost equal for both presentation formats. Hence, in parallel to existing default letter-by-letter presentation in print interpreting, this study suggests the use of the word-by-word format as an alternative.

As already mentioned, previous studies have shown that content words are fixated about 85% of the time, whereas function words are fixated only about 35% of the time (Carpenter and Just 1983; Rayner and Duffy 1988, Furtner et al. 2009). This result falls in line with our findings according to which the first and the second landing points of regressions are generally (90.8%) content words.

Indeed, our results show that the test subjects look for primary and secondary elements of language in order to construct the meaning of what they have just read. Nouns, which are primary elements, are the most likely landing points of regressions (Furtner et al. 2009).

The fact that the lexical hierarchy that can be found here is the same as the one typically observed in the acquisition of language (O’Grady 1987; Furtner et al. 2009) reflects the cognitive processing of language by which the meaning is being constructed. This, in turn, suggests that there is a relationship between the gaze behaviour and the linguistic processing of dynamic text. In addition, since regressions were more common for the word-by-word presentation format, rendering the text word-by-word should be supported by print interpretation software. Sprintanium (Špakov 2011), the tool used in our experiments, does this, whereas many professional print interpreters use simply Microsoft Word which does not have this option.

Being conscious of this lexical hierarchy may also help print interpreters to improve the presentation of dynamic texts to the
user. This could be done for example by highlighting the primary elements and by reducing only tertiary elements of language.

References


Where on the screen do translation students look while translating, and for how long?

Selina Sharmin, Oleg Špakov, Kari-Jouko Räihä, and Arnt Lykke Jakobsen

Abstract

Eye movements were tracked of translation students, with and without touch-typing skills, while they translated three short texts of different levels of complexity under three different time constraints. The aim was to chart the distribution of visual attention between source and target texts, and to study how visual attention was affected by text complexity and time pressure. Participants with touch-typing skills were found to attend more to on-screen text than those without. Further it was shown that time pressure affected average fixation duration on the source text, while text complexity affected the number of fixations on the source text. Also, touch typists made more direct, on-screen transitions from the source text to the target text and back than non-touch typists. Overall, average fixation duration was consistently longer in the target text area than in the source text area.

1. Introduction

Eye tracking is only just starting to be applied to translation research, but promises to yield much new insight. Our exploratory and essentially naturalistic experiment was designed to study the effects of time pressure and text complexity – separately and in combination – on the fixations of a number of translators drawn from a group of subjects with supposedly comparable translation skills. It was hypothesised that touch typists would have an advantage over less skilled typists in that they would be able to devote more constant visual attention to text on the screen. As a result of this, it was expected that skilled typists might be less affected by the combination of increasing time pressure and text complexity than students
who were less skilled, but we were uncertain to what extent this would emerge from their gaze behaviour. Our main interest was to see how visual attention would be distributed and managed under constraints that are part of many translators’ everyday experience.

After the background to our study (section 2), we outline our research design and the technological apparatus used to collect the data (section 3). In section 4 the results and statistical analyses are presented, interspersed with comments mainly on observational and methodological issues. In our discussion and conclusion in sections 5 and 6, we comment on our interpretation of the data and on how the findings add to our knowledge about translators’ visual behaviour. In the conclusion the main results are summarised and ways are suggested in which this exploratory study could be followed up by further studies.

2. Background

Eye tracking, when used to study reading, has generally been applied to reading of individual words, phrases, or sentences (Rayner & Pollatsek 1989, Rayner 1998, Radach et al. 2004), but also to longer texts, e.g. by Hyönä et al. (2003) and Schnitzer and Kowler (2006). In the case of on-screen translation, the translator’s visual attention is constantly shifted between two texts, a source text and the translator’s emerging target text. This is equally true of touch-typist and non-touch typist translators. The possibility of tracking the translator’s gaze pattern across the source and target texts has opened up an exciting new research field, which can draw on existing reading research up to a point but needs to develop its own body of knowledge in view of the specific constraints that apply to translation. Similarly, there is a need to build new knowledge about the way in which emerging written text is monitored visually. Knowledge of this kind is critically important for constructing translation support tools based on gaze information.

Text complexity has frequently been shown to affect the cognitive effort of readers and translators, the speed with which meaning is constructed and represented, and the experience of time pressure (e.g. Davison & Green 1988, Shreve & Diamond 1997). The eye-mind
Where on the screen do translation students look?

assumption launched by Just & Carpenter (1978, 1980) led us to assume that these effects would be reflected in translators’ eye movements.

While the effect of time pressure on simultaneous interpreters has long been studied, little research has been done on the effect of time pressure on translators working in the written modality. Jensen (1999, 2000) found that increasing time pressure caused translators to shorten the time spent on initial orientation and to abandon end revision rather than change the pace of translation drafting. She also found that time pressure resulted in degradation both with respect to the manner of text production and the choice of translation strategies for all of her groups of experimental subjects (professional translators, professionals (but non-translators), and translation novices). In similar experiments, de Rooze (2003) found that the quality of his subjects’ translations deteriorated when they had to translate more than 200 words per 10 minutes. Surprisingly, the quality of translation was higher for 25% of his participants when translating under time pressure. Martin (2006) demonstrated similar time pressure effects. The studies by Jensen, Martin and de Rooze were all based on keystroke logging, which captures the time course of translation output and is well suited for studying degradation effects on output. However, this was not the aim of our study, which focussed on tracking translators’ gaze behaviour as a means of giving us insight both into the mental effort that goes into comprehending a source text (the translation input) and into how a translation is produced.

3. Method

3.1 Participants

Twenty-one participants took part in the experiment (6 males and 15 females). All were second-year students (or higher) at the University of Tampere studying English translation as their major or minor subject. Participants were asked to produce translations that were as good as possible within the given time limit.
3.2 Apparatus

A Tobii 1750 remote eye-tracking device was used to track the participants’ gaze on its integrated 17 inch TFT colour monitor (with 1024 x 768 pixels’ resolution). The experiment was recorded with ClearView and Translog. For the present report only ClearView data were analysed. The combined heat maps showing the areas in the source text that received most visual attention, in terms of overall fixation duration, were produced by means of the iComponent tool (Špakov 2008).

3.3 Procedure and design

First, participants were informed about the test procedure; subsequently, the eye tracker was calibrated for the participants’ eyes. The distance between monitor and participant was 50 to 60 cm. No head support, chin rest or bite bar was used.

Each participant then translated three short texts with different complexity levels – one easy and two more difficult – from English into Finnish. Text 2 was judged to be more complex than Text 1 mainly because of the long and complex second sentence. Text 3 was assessed as being more difficult mainly because it had more low-frequency lexical items than Text 1. Texts 2 and 3 were also judged to pose greater challenges to translation into Finnish than Text 1. Each of the three texts was about 70 words in length. Flesch-Kincaid reading scores for the texts were 57, 24 and 33, respectively. (On this scale, higher scores indicate that the passage is easier to read while lower scores designate more difficult passages.) Grade-level readability scores also showed a clear distinction between Text 1 on the one hand (11.5) and Texts 2 and 3 on the other (18.3 and 14.8). (On this scale, higher scores indicate increased difficulty.) As far as translational difficulty is concerned, our experience told us that T1 would be a good deal easier than T2 and T3.

1 The exact number of words were 74 (T1), 68 (T2), and 73 (T3), with 4.92, 5.63, and 5.33 characters per word on average. The percentage of words with very high frequency (K1 words) was 79, 79, and 75, respectively, while the percentage of less frequent words (K2-K20) was 19, 21, and 22. Translation students were familiar with the low-frequency word forms in Text 1 (MEPs and EU’s), but less familiar with low-frequency words like ensuing and unalienable in Text 3. Full texts in Appendix A.
The experiment began with a warm-up copying task of a text, also about 70 words long. The time frame for the translation of each text was 6, 5 and 4 minutes. The sequence of time was constant for each participant, but the sequence of texts was different. Texts were presented in the order T1-T3-T2, T3-T2-T1 and T2-T1-T3, where T1 was the easy text and T2 and T3 the more difficult texts. Each text was translated seven times under all three time constraints. During the experiment, text appeared in the upper (source-text) half of the screen and participants wrote their translation in the lower (target-text) half of the screen (Figure 1).

After translating, all participants filled in a background questionnaire. In addition, they were subjected to a language level test (DIALANG) to ensure that they were at more or less the same level (C1) of proficiency in English. The whole experimental procedure took about an hour. The total time to translate the texts for each participant was 15 minutes.
4. Results

Data from 18 participants were used for analysis (7 touch typists, 11 non-touch typists; 3 male, 15 female). Data from the remaining three participants were discarded owing to poor eye-tracking quality. Our 18 participants were distributed across three groups, each consisting of six participants, who translated the three texts in the same sequential order (groupwise) and under the same time conditions (groupwise), but with unequal numbers of touch typists and non-touch typists in the groups. Our main interest was to see how average fixation duration and the fixation count (our dependent variables) would co-vary with time pressure, text complexity and typing skill (our independent variables). We analysed participants’ visual attention to the source and target text areas separately, and also the participants’ eye movements on (and off) the screen, in order to correlate our findings with typing skill. For this analysis, normalised fixation counts and transition counts were used.

The statistical analysis of test data involved paired samples $t$-test, independent samples $t$-test, one-way ANOVA and repeated measures ANOVA.

4.1 Participants’ behaviour during translation

An observational analysis of the recorded data was conducted to study the participants’ gaze behaviour during translation. This was done by watching the gaze replay of each participant in ClearView. Though participants had been selected from a well-defined population of translation students, their behaviour in the translation tasks showed little consistency. Six participants read all three texts before embarking on the translation. Two participants read two texts, but not the third, and two participants read only one. The remaining eight participants either read (and translated) the source text sentence by sentence or in smaller segments. During translation, all participants read or re-read the source text segment by segment whether or not they had previously read the whole paragraph or sentence.

After completing the translation, most participants undertook some revision of their target text, correcting it where necessary. Sometimes they read the whole sentence from the source text and compared it with the translated text; sometimes this process was performed on smaller chunks.
Some participants fine-tuned their translation by reading their own text only.

Often participants made corrections while drafting their translation even if such revision might jeopardise their ability to finish the translation in time. Nine participants completed all the tasks within the set time frame; four were only able to finish two tasks; and five only one. All participants were able to finish Text 1 in time under all time constraints, but only ten were able to finish their translation of Texts 2 and 3 in time. One participant did not finish Text 2 within 6 minutes. Three failed to complete either Text 2 or 3 within 5 minutes, and six failed to complete either Text 2 or 3 within 4 minutes. Texts 2 and 3 were both left unfinished in five instances and therefore appeared to be equally difficult. Participants who were able to perform their translation quickly enough to leave them time for revision did not always spend the remaining time revising, but just looked randomly at or off the screen. No correlation was found between completion and typing skill. Touch typists left six translations unfinished; non-touch typists four.

4.2 Difference in visual screen attention between touch typists and non-touch typists

Based on their observed typing skills, participants were divided into two groups, touch typists and non-touch typists. The group of touch typists consisted of seven participants, the other of eleven. Figure 2 shows the distribution of total viewing time on the screen (both halves) during translation by all the participants.

The total on-screen viewing time for the group of touch typists averaged 12.94 minutes and (as expected) was higher than that of non-touch typists, which was 9.29 minutes (Figure 2). The difference is highly significant with $p < .0001$, $t = 7.56$, $df = 16$. It is likely that the group of touch typists did not spend as much time looking at the keyboard and therefore were able to devote more visual attention to the screen.
Figure 2. Distribution of total time spent on the screen by all the participants. The striped bars represent touch typists while the solid bars indicate non-touch typists.

The average fixation count on both the target and source screens for all the tasks (and normalised per person per minute) was higher for touch typists than for non-touch typists (Figure 3).

Figure 3. Average number of fixations per minute by touch and non-touch typists on Texts 1, 2 and 3 (separately for source and target areas). Error bars with 95% confidence intervals.

According to the independent samples $t$-test, the difference was significant for the target screen (with $p < .01$, $t = 3.424$, $df = 16$), but not for the source screen.
With respect to the average duration of fixations on both halves of the screen (Figure 4), a paired samples $t$-test showed that touch typists had significantly lower average fixation durations than non-touch typists (with $p < 0.05$, $t = -3.732$, $df = 5$). The difference for the target screen was close to being statistically significant, with $p = .06$, $t = -1.997$, $df = 16$.

Touch typists had significantly more overall on-screen gaze time, significantly shorter average fixations and significantly more fixations per minute than non-touch typists. One would expect this to give touch typists an advantage over non-touch typists so that they would be better able to complete tasks under time pressure, but in our study touch typists completed fewer tasks than non-touch typists.

![Figure 4. Average fixation duration on source and target areas under different time constraints by touch and non-touch typists. The average fixation duration of non-touch typists on the screen was 260 milliseconds (ms); for touch typists it was 207 ms.](image)

### 4.3 Difference in transitions by group

#### 4.3.1 Transitions by touch and non-touch typists between source and target screens

According to independent-samples $t$-test, touch typists had significantly more transitions per minute from source to target screen than non-touch typists in all the tasks (with $p < 0.01$ and $df = 16$). For 6, 5 and 4 minute tasks, the test statistics values were $t = 3.252$, $t = 3.189$ and $t = 3.699$, respectively.
Touch typists also made more transitions per minute from the target to the source screen than non-touch typists in all the tasks. According to independent-samples $t$-test the results are significant with $p < 0.01$ and $df = 16$. Test statistical values for 6, 5 and 4 minute tasks were $t = 3.423$, $t = 3.428$, and $t = 3.722$, respectively.

When averaging all the tasks (Figure 5), we found the same trend for results in both transitional directions. The independent-samples $t$-test showed that touch typists had significantly more transitions per minute than non-touch typists in both the directions S-T ($t = 3.711$, $df = 16$, $p < 0.01$) and T-S ($t = 3.888$, $df = 16$, $p < 0.01$).

![Figure 5. Number of transitions by touch and non-touch typists between source and target screens (both directions) under three time conditions.](image)

4.3.2 Transitions with off-screen intervals by touch and non-touch typists
Transitions involving off-screen intervals for a minimum of 500 ms were counted. There was no upper threshold. Four directions of transitions were taken into account: S-O-S, S-O-T, T-O-S, and T-O-T, where ‘S’ stands for source screen, ‘T’ for target screen and ‘O’ for outside the screen. Thus S-O-S means a transition of the gaze that travels from the source text area to outside the screen and then returns to the source text screen again.

The count of transitions with off-screen intervals (Figure 6) showed that non-touch typists made many more transitions than touch typists, in all four directions, and that both groups made about twice as many T-O-T transitions (from target text to keyboard and back to the target text) as in
any of the other directions. According to the independent samples t-test, the differences were significant in the 6-minute task for directions S-O-S \((t = -2.720, df = 16, p < 0.05)\) and T-O-S \((t = -3.354, df = 16, p < 0.01)\); in the 5-minute task for directions S-O-S \((t = -2.151, df = 16, p < 0.05)\) and T-O-T \((t = -2.769, df = 16, p < 0.05)\); and in the 4-minute task for directions S-O-S \((t = -2.895, df = 16, p < 0.05)\), T-O-S \((t = -2.639, df = 16, p < 0.05)\) and T-O-T \((t = -2.253, df = 16, p < 0.05)\).

![Figure 6. Number of transitions between text areas with off-screen intervals by touch and non-touch typists under three time conditions.](image)

In all the directions of transitions, the off-screen duration per minute was higher for non-touch typists than for touch typists. One-way ANOVA found that the differences were significant for all four directions: S-O-S \((F_{1,16} = 7.757, p < 0.05)\), S-O-T \((F_{1,16} = 17.044, p < 0.01)\), T-O-S \((F_{1,16} = 8.334, p < 0.05)\), and T-O-T \((F_{1,16} = 17.994, p < 0.01)\) (Figure 7).

The obvious main reason why non-touch typists spent more time outside the screen was that their gaze travelled more frequently to the keyboard.

A paired samples t-test found that the total duration of transitions per minute for all the tasks in the direction S-O-S was significantly lower \((t = -3.117, df = 17, p < .01)\) than that in the direction T-O-T. We do not have observational data to help explain why this was the case, but it is probable that the Os in S-O-Ss were instances when participants’ gaze went off-screen to allow the person to think about a word, and that the Os in T-O-Ts were intervals of visual attention to the keyboard.
Figure 7. Total duration in ms per minute of time spent off-screen (by transition type) for all the tasks by touch and non-touch typists.

The total duration of transitions in the direction S-O-T was found to be significantly higher than in the direction T-O-S, $t = 2.242$, $df = 17$, $p < .05$. Evidently, the distance (in time) the eyes travelled after leaving the source text screen until they returned to the target text was longer than the distance covered in the opposite direction. This makes good sense, intuitively, since – as hinted above – many participants would have gone from reading the source text (S) to looking at the keyboard while typing the translation (O), and then looking at what they had typed on the screen (T). After monitoring their typing, participants either returned to reading the source text directly (T-S) or, occasionally only, looked at the keyboard before returning to the source text (T-O-S).

4.4 Effect of time pressure

Figure 8 presents the average viewing time per minute on the screen during each task under different time-pressure constraints. Though the time pressure was different for each task, the viewing time per minute (for all participants) was almost the same.

Analysis of the effect of time pressure on average fixation duration, such that more time pressure correlated with shorter average fixation
duration, was close to reaching statistical significance for the source text area (with \( t = 1.874, \text{ } df = 17 \) and \( p = 0.078 \)). (The degree of closeness to significance here is illustrated by the fact that after removal of one participant’s data from this analysis, a paired samples \( t \)-test on the data from 17 participants showed that the average fixation duration on the source text during the six-minute task (216.8) was significantly higher than for the four-minute task (203.5) with \( p < 0.01, \text{ } t = 2.908, \text{ } \text{and } df = 16 \).

Time pressure was not found to affect the duration per minute of off-screen attention; nor did it influence the fixation count per minute.

![Figure 8](image.png)

**Figure 8.** Viewing time per minute on the screen for each task

### 4.5 Effects of text complexity

Figure 9 presents the average viewing time per minute on each text. We found that the viewing time was longer for the complex texts than that for the easier text.

![Figure 9](image.png)

**Figure 9.** Viewing time per minute on the screen for each text
Repeated measures ANOVA found a significant effect of text complexity on normalised fixation count of the source text with \( p < .05, F_{2.34} = 3.439 \). More specifically, the average fixation count per minute was significantly higher in the complex texts than in the simpler text. A paired samples \( t \)-test showed that the source text with complex vocabulary (Text 2) had a higher fixation count than the easy text (Text 1) with \( p < .05, t = -2.212, df = 17 \). For the structurally complex source text (Text 3) the fixation count was also significantly higher than for Text 1 with \( p < .05, t = -2.690, df = 17 \).

Table 1 presents the average fixation count per minute and the average fixation duration in milliseconds on the source text in each of the three texts.

<table>
<thead>
<tr>
<th>Fixation count</th>
<th>Text 1</th>
<th>Text 2</th>
<th>Text 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text 1</td>
<td>49.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text 2</td>
<td>64.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text 3</td>
<td>67.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixation duration</td>
<td>213.09</td>
<td>210.42</td>
<td>212.88</td>
</tr>
</tbody>
</table>

It appeared that while reading the source text for translation, translators had more fixations in complex text (whether it was the result of complex structure or difficult vocabulary) than in simple text. The fixation duration times, however, were almost identical.

Figure 10 presents ‘heat maps’ of the three source texts for one group of participants who translated the same sequence of texts (3, 2, 1) under the same time pressure conditions (6, 4, 5 minutes). Each map visualises the total fixation time (the “sum of gaze duration and regression time”, Hyönä et al.: 331) for this group of participants. The upper text in Figure 10 is the lexically difficult one (Text 3), the middle text is the structurally difficult Text 2, and the bottom one is the easier Text 1. The heat maps were produced with the iComponent tool (Spakov 2008) with the same sensitivity level, brightness and hiding level for each screen. The white blotches represent the screen areas that received the longest overall duration of fixations. Visually the heat maps support the statistical results showing that the complex texts were fixated more intensely than the simpler text. The time allowed for translating the complex texts for this group was six (Text 3) and four minutes (Text 2). Five minutes were
allowed this group to translate the simple text. Though the participants had more time to look at the simple text, it received fewer fixations than the structurally more difficult text above it. Evidently, the more complex texts required readers to make more regressions in order to grasp the meaning and produce a translation.

Figure 10. Heat maps showing intensities of area fixations for one group of participants. From the top: Text 3, Text 2, Text 1.

4.6 Gaze behaviour on source and target screen

Figure 11. Distribution of fixation duration (in ms) on source and target text areas by all the participants.
Figure 11 shows the frequency distribution of the fixation duration by all participants on the source and target text areas. Fixation durations less than 100 ms were omitted from the analysis, and the upper limit was set at 1000 ms. It can be seen that the frequency of fixations with relatively short duration (100-200 ms) was higher in the source area than in the target area.

Overall, the average fixation duration for all participants and all tasks was higher on the target screen (266 ms) than on the source screen (212 ms). This difference was found to be highly significant with $p < .01$, $t = -3.454$, $df = 17$ (Figure 12).

![Figure 12. Average fixation duration on source and target area for all the tasks](image)

More specifically, fixation duration was significantly higher on the target screen than on the source screen for all the tasks regardless of time pressure with $p < 0.01$ and $df = 17$ (Figure 13). Paired samples $t$-test values for the six, five and four-minute tasks were $t = -2.955$, $t = -3.280$, and $t = -3.360$, respectively.

Furthermore, the average fixation count per minute for all the tasks on the source screen was higher than for the target screen with $p < .05$, $t = 2.291$, $df = 17$. There were differences under all three time constraints, but the paired samples $t$-test only found the difference in the six-minute task to be significant, $t = 2.305$, $df = 17$ and $p < 0.05$. 
Thus, in our population of Finnish translation students, we found that, independently of time pressure and text complexity, the target text received fixations with longer duration than did the source text, though there were numerically fewer fixations (per time unit) than on the source text.

5. Discussion

The main effect found of the group difference in typing skills was as expected. Touch typists attended visually to text on screen – whether source text or emerging target text – more than did non-touch typists. The fact that touch typists did not have to look at the keyboard allowed them to make more direct and more rapid transitions between the source and target text areas on the screen. By contrast, non-touch typists frequently had to move their eyes off the screen to the keyboard. Somewhat surprisingly, however, we found no evidence that the difference resulted in touch typists being significantly less affected by time pressure and text complexity. For this reason, only the overall effects of time pressure and text complexity have been reported.

The only (marginally) significant effect of time pressure on fixations was on the average duration of fixations in the source text area, which decreased as time pressure increased, whereas average fixation duration in the target text area was unaffected. This indicates that it is easier for
translators to adapt their reading-for-comprehension to variable time constraints, whereas it is more difficult for them to adapt their reading-and-monitoring of the target text, possibly because this process has to await text being typed. Translators may not be able to adapt their typing speed as flexibly as they are able to adjust the speed with which they read for comprehension. In a translation study comparing reading of someone else’s text versus reading of one’s own emerging text, Holmqvist et al. (2007) found that fixation durations were significantly longer while reading one’s own emerging text than when reading somebody else’s text. This result is in line with our findings that fixation duration is longer on the (own) target text than on the source text.

Text complexity was found not to affect fixation duration, which remained constant irrespective of time constraint. However, there was a significant difference between the number of fixations on the simple text and the two more complex texts. It would appear that complexity requires more fixations regardless of whether it is the result of lexical items or syntax. Schnitzer & Kowler (2006) similarly found that difficult text was read and reread with a higher frequency of regressions than simpler texts.

Our finding that fixations on the target text were longer than on the source text was very systematic. The fact that our participants had more fixations on the source text than on the target text is consistent with the findings by Jakobsen and Jensen (this volume) and probably reflects that, unlike what seems to be the case with professional translators, translation students struggle rather heavily with comprehending source text in a foreign language.

6. Conclusion

Our study resulted in several robust findings concerning

- the (limited) usefulness of touch typing,
- the extent to which translation students’ visual attention was affected by time pressure,
- the extent to which fixation frequency was affected by text complexity,
- the ways in which visual attention was distributed across source and target text areas,
how transitions were made between the two on-screen texts, and
the various directions participants’ gaze travelled when it went off-
screen.

One line of research we would like to pursue in continuation of the
work presented here is to test more exactly what advantages, if any, a
touch-typing translator has over one who is not touch-typing. Translators
who dictate their translations might be included for comparison in such a
study. This research would require more careful pre-experimental screening
of participants than undertaken for the present study. We would also like to
further validate our findings concerning the relative distribution of visual
attention to the source and target texts, e.g. by manipulating the on-screen
position of the text windows. The concept of text complexity is very
slippery indeed and also needs to be more carefully controlled in future
research.

Nevertheless, we are convinced that there is much valuable insight to
be gained from experimenting at the level of granularity such as in the
present experiment, i.e. with continuous reading and translation of
authentic texts by participants performing their tasks in an environment that
is simultaneously a fairly naturalistic setting and a high-tech lab with
intensive monitoring of behaviour.

Our findings add insight into the special kind of reading that
translators engage in as they create and align an emerging target text with a
source text. Such insight is critically important both for creating efficient
gaze-based translation support applications of the kind envisaged in the
Eye-to-IT project and for a better understanding of translation.

Acknowledgements

We thank Poika Isokoski for his useful advice. This work was supported by
the Eye-to-IT project in EU’s FP6 programme.
References


Appendix A

First text item (easy)
Cost in translation: EU spends €1bn on language services.
The Independent, 1 March 2007,
http://news.independent.co.uk/europe/article1178569.ece

MEPs are not elected on their linguistic ability and many speak only their native tongue. In a TV age they argue that it is important for them to be seen to be addressing their constituents in a language they understand, hence the need for full translation and interpretation. But the growing demands have put a massive strain on the EU’s interpretation and translation services, which have struggled to recruit speakers of the new languages.

Second text item (harder, structure)
Commission launches cooperation with universities in translator training.
EU press release, 18 October 2006.

The European Master’s in Translation degree should focus primarily on the translation component, and not on the language skill/language acquisition aspect of the training. The programme is, however, flexible in order to take into account the requirements of the European institutions, including the need for specialists in a number of policy areas with language skills, as well as specific national conditions and developments in the translation profession.


The most problematic intercultural relationships at this end of the twentieth century are associated with disputes over borders. They often ensue from notions of cultural sovereignty, from beliefs that a culture has some kind of unalienable right to some kind of specificity. Such beliefs become most problematic when expressed as claims to a particular territory, to a cultural homeland, spreading across the space of one particular color on a map of world cultures.

Reading On-Screen Text with Gaze-Based Auto-Scrolling

Selina Sharmin  Oleg Špakov  Kari-Jouko Räihä

School of Information Sciences  University of Tampere, Finland
{selina.sharmin}@uta.fi

ABSTRACT
Visual information on eye movements can be used to facilitate scrolling while one is reading on-screen text. We carried out an experiment to find preferred reading regions on the screen and implemented an automatic scrolling technique based on the preferred regions of each individual reader. We then examined whether manual and automatic scrolling have an effect on reading behaviour on the basis of eye movement metrics, such as fixation duration and fixation count. We also studied how different font sizes affect the eye movement metrics. Results of analysis of data collected from 24 participants indicated no significant difference between manual and automatic scrolling in reading behaviour. Preferred reading regions on the screen varied among the participants. Most of them preferred relatively short regions. A significant effect of font size on fixation count was found. Subjective opinions indicated that participants found automatic scrolling convenient to use.

Categories and Subject Descriptors
H.5.2 [Information Interfaces and Presentation]: User Interfaces – Evaluation/methodology; Input devices and strategies

General Terms
Algorithms, measurement, experimentation, human factors

Keywords
Eye movements, reading, manual scrolling, automatic scrolling, reading region, fixation duration, fixation count

1. INTRODUCTION
Human-device interaction can be improved through systems making use of eye contact, movement, or position. One of the most common interactions between humans and computers is the reading of on-screen electronic text. Scrolling is often needed for continuing to read longer documents, such as Web pages. Manual scrolling is achieved via pressing of keys on a keyboard, use of a mouse button or wheel, or use of a touchpad. For avoidance of frequent manual actions, automatic scrolling techniques (auto-scrolling, for short) have been implemented in various ways, some of which involve gaze tracking.

Scrolling is related to our ability to absorb information through perceptual organs. As a powerful source of information about the focus of viewers’ attention during reading, eye-movement data have proven to be valuable in the study of reading and other information-processing tasks [12]. Visual information from eye movements is also a useful source of contextual information for enhancement of scrolling during reading. Gaze information can be used to enable auto-scrolling when the reader wishes to move the text continuously on the visual display without having to press the relevant control on the scrolling device repeatedly.

On the basis of the characteristics or reading patterns and how readers consume visual information, Kumar and Winograd [7] presented several scrolling techniques that use gaze information either as primary input to automatic control of the onset and speed of scrolling or passively, to augment manual scrolling. Their study indicated that all participants preferred the gaze-enhanced Page Up / Page Down technique over the normal Page Up / Page Down.

Readers initiate scrolling actions to move the text into their preferred reading region on the screen for convenient reading. An eye-tracking study by Buscher et al. [2] found that people differ in their preferred reading regions. The on-screen location of the reading regions and its size vary between readers. In previous research, the ideas from the aforementioned studies on preferred reading regions and gaze-enhanced scrolling techniques have not been brought together. Here we present a study of whether automatic scrolling based on gaze information allows readers to read within their preferred reading regions.

We have implemented an auto-scrolling technique and conducted an empirical study to find preferred reading regions in reading with manual and automatic scrolling. We examined whether manual and automatic scrolling have an effect on reading behaviour on the basis of eye movement metrics, such as fixation duration and fixation count. We studied whether people have preferred reading regions and, if so, what the location of these is on the screen. We also examined whether there were different reading styles – e.g., using a small factual reading window vs. reading everything before scrolling. Moreover, we studied how different font sizes affect eye movement metrics.

Gaze-tracking software was implemented to track gaze data on a Web page. It first recognized the reading range by collecting gaze data in reading with manual scrolling and then applied an algorithm to scroll the page automatically. We conducted an experiment with 24 participants using manual and automatic scrolling in reading of six pieces of text. Our results indicated that there was no effect of manual or automatic scrolling on reading behaviour. Most of the participants read from their preferred, relatively short region of the screen. There was also a significant effect of font size on one of the eye movement metrics: fixation count. Subjective opinions indicated that participants found auto-scrolling comfortable.

2. BACKGROUND
Reading of electronic documents is integrated deeply into our everyday activities. We read documents on the World Wide Web, electronic journals, electronic files in our professions, and material in electronic form as part of recreational activities. In this reading, we might need to interact with a computer or some other device by scrolling from time to time with a mouse, keys, or a touchpad. Scrolling is needed for viewing larger amounts of information on electronic displays with limited screen space.

Copyright © 2013 by the Association for Computing Machinery, Inc

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions Dept, ACM Inc., fax +1 (212) 869-0481 or e-mail permissions@acm.org.

Copyright 2013 ACM 978-1-4503-2110-5/00/0010...$15.00
As mentioned above, an exploratory study by Buscher et al. [2] using eye tracking revealed that people have individual preferences for certain reading regions on the computer screen. Results from that study indicated that the gaze was concentrated vertically around the middle of the screen during a reading task. Large variation across readers was found with respect to vertical preferred reading location and vertical spreading of the gaze around that preferred location. Buscher and colleagues also analysed the relationship between scrolling behaviour and the distribution of visual attention over the screen. The findings indicated that participants scrolling once in a while for one full screen height produced greater vertical gaze spreading. On the other hand, scrolling frequently after a few lines produced more restricted vertical gaze spreading on the screen.

Buscher et al. [2] used goal-oriented reading tasks in their study. One was a shopping task that required finding specific components in a very long catalogue. Another was a fact-finding reading task with two narratives. The latter produced a mixture of actual reading and quick scanning for moving ahead in the document. It was found, for instance, that the scrolling distance was different for reading behaviour as compared to scanning behaviour. In our study, we are interested in normal reading and have chosen the test setup accordingly. The only goal-related requirement was that the text be read linearly in its entirety, which should result in a more homogeneous data set.

Another difference is that Buscher et al. [2] found a bias toward the left edge of the screen, which is understandable in view of the nature of their task. Since we were interested in the vertical position of the preferred reading window, we used text that was centred in the window.

Hornbæk and Frøkjær [4] presented an exploration of reading patterns and analysed reading behaviour with long documents. They explored the effect of different document visualization techniques. Looking at scrolling behaviour, they observed that documents were sometimes read linearly, sometimes parts were skipped, and sometimes portions were read multiple times. Again, this is related to the task in the experiment (writing an essay about the content of the text). We concentrated on linear reading, expected to resemble reading for pleasure.

Hsieh et al. [5] conducted an experiment to study the effects of scrolling speed in text-error search tasks. They considered three experimental factors: scrolling speed, error type, and article length. The results indicated that scrolling at high speed affects the quality of reading by decreasing identification of errors.

Russell and Chaparro [13] examined the effect of three font sizes (12, 20, and 28 points) on reading comprehension when text was presented in RSVP. No effect of font size was found in reading for comprehension, though participants preferred the 20-point font size over 12-point size. In our study, we also varied the font size to see whether it has an effect on the size of preferred reading regions or on eye movement metrics.

Reading behaviour is one of the main themes of this paper, and supporting reading by automatic gaze-based scrolling is another. This is an old idea: in 1991 Jacob [6] had already proposed the use of gaze-activated buttons to accomplish scrolling. Automatic scrolling without explicit activation was suggested by Tognazzini [15] in an insightful patent, where he described most of the key features: visible or invisible buttons, automatic scrolling rate based either on fixed areas on the screen or on distance from the centre point, need to avoid abrupt changes by adjusting the scrolling rate smoothly, and various functions for computing the scrolling rate. His patent has been cited in many later patents. For instance, Mackay and Peck [9] describe a variant wherein the anchor point of scrolling is not necessarily in the centre of the window and, rather, varies with the user; however, the patent does not give details of how the anchor point is determined. Lemelson and Hiett [8] present a version in which scrolling can take place in four directions (up and down plus left and right).

Patents provided the ideas, but the first implementations of auto-scrolling were several more years in coming. Kumar and Winograd [7] implemented three scrolling techniques that start and stop scrolling automatically on the basis of the reader’s gaze position. Activation of the auto-scrolling feature, however, requires an explicit command issued by the user. We apply the same idea as in one of the techniques implemented by Kumar and Winograd, scrolling speed controlled by discrete regions of the screen, but with dynamic thresholds for the regions customized for each reader.

Most suggestions for supporting a flexible scrolling rate use the natural choice of making the rate depend on the gaze position: for example, the closer to the bottom of the window the gaze point is, the more rapid the scrolling. Porta and Ravelli [11] suggest an interesting alternative wherein a transparent widget for scrolling-rate control (both up and down) is placed on top of the text. Users can then explicitly control the rate and rapidly scroll even long distances, while making use of the underlying text display to stop the scrolling when they have reached the desired position. This is useful for goal-oriented tasks.

New smartphones are developing that will be capable of scrolling the screen’s content without physical contact with the screen [3]. Although the technology was dubbed “Eye Scroll,” it is currently based on face tracking, not eye tracking [10]. While mobile eye-tracking technology is still in its early stages, the progress is so rapid that basic results such as the ones found in our study may soon find use in everyday interaction techniques.

3. METHOD AND TEST SETUP

3.1 Participants

In all, 33 people took part in the experiment. Good data from 24 participants were used for the analysis. The mean age of these participants was 31.5 years, with a standard deviation (std) of 7.93 and a range of 18–51 years. Most participants were researchers or students at two Finnish universities. They used computers daily, for, on average, 7.2 hours, with a range of 4–12 hours. All participants had normal or corrected-to-normal vision.

We had to reject data from nine participants. In one case, this was because there was an error in the process wherein the parameters for auto-scrolling were manually entered in the software. For the remaining eight participants, the reason was that the tracker was unable to collect proper x- and y-coordinate information on eye movements for the full duration of the experiment. Although this seems to be a high percentage of rejections, one must consider too that each participant read six pieces of text. We chose a cautious policy and rejected all data for any participant who did not produce valid data for all six tasks. Accordingly, eight participants’ rejection was caused by eight recording problems, for 198 cases, an acceptable rate.

3.2 Stimuli

We conducted four separate pilot tests prior to the experiment discussed here, to determine the optimal test setup and stimuli. In total, 15 pilot users took part in those tests. The stimuli...
selected consisted of six pieces of text, differing in their topics. The mean length of these pieces was 802.5 words, with std 18.56 and range 779–834 words. All text was in the Finnish language, and the topics are presented in Table 1. The goal was to have text that was reasonably interesting and motivating to the participants without specific fact-finding tasks.

Table 1. The six pieces of text used in the stimuli

<table>
<thead>
<tr>
<th>Text code</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>A report on how computer science education started in Finland</td>
</tr>
<tr>
<td>T2</td>
<td>Early history of motion pictures</td>
</tr>
<tr>
<td>T3</td>
<td>An essay on global warming</td>
</tr>
<tr>
<td>T4</td>
<td>A trip report to Italy</td>
</tr>
<tr>
<td>T5</td>
<td>A humoristic short story by a Finnish novelist</td>
</tr>
<tr>
<td>T6</td>
<td>A brief history of games and game industry</td>
</tr>
</tbody>
</table>

Text was displayed to the participants centred on the screen. The three font sizes used were actually produced by varying the zoom factor of the instrumented Internet Explorer browser: 100%, 125%, and 150% zoom (see Figure 1). Consequently, the line length grows with font size, which makes our test setup different from that used by Hornbæk and Frøkjær [4]. The pieces of text were about 5 pages long each in 100% size.

3.3 Apparatus

A Tobii T60 remote eye-tracking device was used to track the users’ gaze on its integrated 17-inch TFT colour monitor (with 1280 by 1024 pixels’ resolution). The experiment was recorded with Tobii Studio. The stimuli were presented through Internet Explorer and the ETU driver [14].

3.4 Procedure

At the beginning, each participant was briefly instructed about the test procedure. The test consisted of reading text, using manual and automatic scrolling. Each participant read three pieces of text using manual scrolling, and later three other pieces with automatic scrolling. The manual scrolling condition had to be applied first, because user-specific data on preferred reading region were collected thereby and used as input for the auto-scrolling algorithm.

In each scrolling condition, the pieces of text were presented in three distinct font sizes (100%, 125%, and 150% zoom). Participants were asked to read the text linearly and use the scroll wheel of the mouse to scroll the pages in reading with manual scrolling. They were told about filling in a post-test questionnaire including one comprehension question per piece of text; the motivation here was to get the participant to read carefully and to obtain good eye movement data. Participants were allowed to quit the test whenever they felt frustrated or exhausted. They were also instructed in how to stop the moving text in automatic scrolling in case this became necessary.

The sequences of presentation of the six pieces of text in three distinct font sizes were counterbalanced. A 3 × 3 Latin-square array was used to counterbalance the three font sizes, and part of a 6 × 6 Latin-square array was used for the pieces of text.

After reading of each piece of text, a brief questionnaire was immediately delivered to the participant to fill in. The questionnaire had one comprehension question and two questions related to the reading experience. After reading of all six pieces of text, another post-test questionnaire was delivered at the end, containing background questions and other items related to reading experience during the test.

The total duration of the test was about an hour per participant. We compensated each participant with two movie tickets. The experiment was conducted in a peaceful place in the gaze lab of our university.

4. RECOGNITION OF THE PREFERRED READING RANGE

Each scrolling action defines a new reading range. When reading of a Web page is complete, the reading ranges are analysed, outliers removed, and the average values computed for determination of the preferred reading range. The details of this algorithm are explained next.

While the tracking is active, our algorithm detects fixations (the threshold is 50 pixels, with noise filter) and assigns them to either the “reading” or the “non-reading” pool. A fixation is assigned to the “reading” pool if:

- it lies to the right of the previous fixation,
- the distance between it and the previous fixation is less than 170 pixels, and
- the absolute angle of the saccade is less than 45 degrees.

When a user starts scrolling, the algorithm creates a new range and stores the y-coordinate of the last “reading” fixation (this is the bottom parameter of the range). After scrolling is finished (the algorithm considers it finished if no scrolling event occurs within 500 ms), the algorithm checks whether the page was scrolled down. If so, the algorithm waits until the first “reading” fixation is detected, then completes creation of the range by setting the y-coordinate of this fixation as the top parameter of the range.

After reading is finished (the “Stop” button in the title bar is clicked), the algorithm determines the average range. The estimation function requires at least two ranges to be in the list. It first finds mean values of the y-coordinate for the top and bottom parameters; also, it calculates their standard deviations. Next, it detects ranges that have either their top or

![Figure 1. Text displayed in the three font sizes. On the left is 100% size, in the middle 125% size, and on the right 150% size.](image-url)
bottom parameters beyond two standard deviations from the average. The algorithm removes any such ranges found from the list of ranges, re-computes the means and standard deviations, and repeats the outlier-detection procedure. It is finished when no more outliers are found. The estimated values or the preferred reading ranges ($\mu_T$, $\mu_B$), with standard deviations ($\sigma_T$, $\sigma_B$), are analysed in Section 5 and used in the auto-scrolling algorithm, discussed in Section 6.

5. RESULTS FOR READING WITH MANUAL SCROLLING

Our analysis treated font size and scrolling type as independent variables, while average fixation count, average fixation duration, total reading time, average $y$-coordinate of fixations, and std of $y$-coordinates of fixations were the dependent variables. We used six different pieces of text as stimuli. There was no statistically significant effect of text piece on the dependent variables of eye movement metrics. Hence, we continued our analysis without considering text as a variable.

Two-way within-subjects ANOVA was carried out to analyse the main effect of the independent variables on the dependent variables of the eye movement metrics mentioned earlier. Here the independent variable, or within-subjects factor, font size consisted of three levels: 100%, 125%, and 150% zoom. The other within-subjects factor, scrolling type, consisted of two levels: manual scrolling and automatic scrolling.

Participants first read three pieces of text with manual scrolling. Text was displayed in three font sizes. We then analysed the reading region of each participant. Figure 2 shows heat maps for the two participants with the most extreme reading regions. We can easily see the differences in reading patterns between these participants. Some people tend to read from only a portion of the screen, which is their preferred reading region, and others may prefer to read by scanning the whole page.

For more detailed analysis, we calculated the top and bottom parameter ($\mu_T$ and $\mu_B$) of the preferred reading regions for each participant. There was much variation among the participants. The preferred reading range, averaged over the three font sizes, was very short for one participant (306–439 pixels) and much taller for another participant (265–917 pixels). Figure 3 presents data from the participants with the shortest (A) and tallest (B) $\mu_T$ and $\mu_B$ for reading range. It also presents the average of $\mu_T$ and $\mu_B$ for each font size over all participants (C). Though the preferred reading range of the participant in Figure 3 (B) appears to grow with increasing font size, for others (such as the participant in Figure 3 (A)) this was not the case, and overall there was no significant effect of font size on the height of the preferred reading region.

In addition to the preferred reading ranges, an interesting question is where within the reading range the participants did most of their reading. This is analysed next.

Figure 2. Heat maps for participants 23 (on the top) and 22 (below) with the lowest and highest std of the $y$-coordinates of fixations, respectively, in reading with manual scrolling. On the left is 100% size, in the middle 125%, and on the right 150%.
The average and std of the y-coordinates of fixations averaged over all participants are presented in Table 2. We can see that the smallest font size (100%) showed a slightly lower average of y-coordinates and higher std of y-coordinates of fixations than did the bigger fonts, for which the average y-coordinate of the fixations was almost identical.

Table 2. Average and std of y-coordinates for each font size

<table>
<thead>
<tr>
<th>Font size</th>
<th>Average y-coordinate</th>
<th>Std of y-coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td>503.4</td>
<td>205.73</td>
</tr>
<tr>
<td>125%</td>
<td>519.6</td>
<td>194.65</td>
</tr>
<tr>
<td>150%</td>
<td>520.6</td>
<td>181.64</td>
</tr>
</tbody>
</table>

The std of the y-coordinates of fixations for all participants differed significantly among the font sizes. Paired-samples t-test values for three font-size pairs are presented in Table 3.

Table 3. Test statistics values for the combination of fonts

<table>
<thead>
<tr>
<th>Font size comparison</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% and 125%</td>
<td>2.069</td>
<td>23</td>
<td>.05</td>
</tr>
<tr>
<td>100% and 150%</td>
<td>3.730</td>
<td>23</td>
<td>.001</td>
</tr>
<tr>
<td>125% and 150%</td>
<td>2.131</td>
<td>23</td>
<td>.05</td>
</tr>
</tbody>
</table>

Figure 4 presents the distribution of the average y-coordinate of fixations on the screen in reading of text with the three font sizes. The trend was clear, as shown in Table 4. To find out why, we analysed the eye movement metrics.

6. AUTOMATIC SCROLLING

We have implemented a tool that provides automatic scrolling on the basis of eye movements observed individually during manual scrolling. The auto-scrolling algorithm is implemented as a JavaScript module in the browser that has been instrumented with a plugin of the ETU driver [14]. On loading, it adds an “Auto-scrolling settings” button to the eye-tracking panel. Thereby, a user (or, in our case, the experimenter) may change its setting and switch between the modes (recognition of the preferred reading range in manual scrolling or use of the detected range in auto-scrolling). For starting in either mode, the eye tracker is calibrated first. The parameters collected during the manual scrolling phase were entered manually as input for auto-scrolling by the experimenter.

In auto-scrolling, the algorithm waits until a “reading” fixation with a y-coordinate lower than $\mu_B$ is recognized, and then it starts scrolling down a Web page at a speed of one pixel per 40 ms. The speed increases to one pixel per 20 ms, if $y > \mu_B + 2\sigma_B$ for the y-coordinate of a detected “reading” fixation. The scrolling stops if the y-coordinate of a newly detected fixation (either “reading” or “non-reading”) is less than $\mu_B$.

The algorithm can also scroll a Web page up, if allowed. Scrolling up starts when a fixation with $y < \mu_T - 2\sigma_T$ (i.e., above the top of the reading range) exceeds two seconds and ends when a user glances below this threshold. This option was not used in our experiment.

7. RESULTS IN READING WITH AUTOMATIC SCROLLING

On average, regardless of font size, reading text with manual scrolling took more time than with automatic scrolling. Though the differences were not statistically significant, the trend was clear, as shown in Table 4. To find out why, we analysed the eye movement metrics.
Table 4. Average time (in seconds) to read a piece of stimulus text with the various font sizes

<table>
<thead>
<tr>
<th>Scrolling</th>
<th>100%</th>
<th>125%</th>
<th>150%</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual</td>
<td>336.4</td>
<td>334.1</td>
<td>339.4</td>
<td>336.6</td>
</tr>
<tr>
<td>Automatic</td>
<td>316.2</td>
<td>330.7</td>
<td>322.4</td>
<td>323.1</td>
</tr>
</tbody>
</table>

According to repeated-measures two-way within-subjects ANOVA, the two scrolling conditions had a significant main effect on fixation count, $F_{1,23} = 6.246, p < .05$. The fixation count was significantly higher when the text was scrolled manually than in automatic scrolling (see Figure 5).

The lines in Figure 6 show average fixation duration in reading with manual and automatic scrolling by all the participants in the three individual font sizes. No significant difference between the scrolling conditions was found. Had the fixation durations differed for the same individual between the two scrolling conditions, it might reflect a difference in the readers’ cognitive pressure in reading. Since this was not found, we conclude that the participants read in a similar way no matter whether the text was scrolled manually or automatically.

In summary of the above findings, reading with manual scrolling took more time than with auto-scrolling. This was solely caused by the higher number of fixations, since there was no difference in the duration of the fixations between the scrolling conditions.

To continue the analysis, we then looked at the reading regions. The scrolling condition had a significant main effect on the average y-coordinate of fixations, with $p < .05$, $F_{1,23} = 4.894$. Automatic scrolling showed a higher average y-coordinate of fixations than the manual scrolling. This is as we expected, since the algorithm tries to bring new text into the reading range with less need for the participant to read from the top parts of the screen.

Repeated-measures two-way within-subjects ANOVA also showed a significant main effect of automatic and manual scrolling on the std of y-coordinates of fixations ($F_{1,23} = 52.089, p < .001$). Manual scrolling showed a higher std of y-coordinates of fixations than automatic scrolling did (see Figure 7). Again, this is as expected, since with auto-scrolling the reading is targeted at a shorter part of the screen.

8. EFFECT OF FONT SIZE ON EYE MOVEMENT METRICS

To analyse the effect of font size on the eye movement metrics, we first checked whether use of ANOVA is valid. The outcome of Mauchly’s test was significant ($p < .05$) for the font sizes of the text. Therefore, we used Greenhouse-Geisser-corrected $F$-values for the effect of different font sizes on fixation count.

Within-subjects contrasts test of linearity proved significant for font size, with $p < .001$, $F_{1,23} = 59.623$. Hence, there was a significant trend in the data. Fixation count increased in a linear fashion with growth in font size (see Figure 5). In general, we do not fixate on all words while reading. Our peripheral vision helps us to see and retrieve information near the fixated location. Therefore, increasing the font size reduces the amount of information gathered through our peripheral vision. Consequently, people tend to fixate more in reading with bigger font sizes than with smaller ones.

Similarly, for analysis of the effect on average fixation duration, the outcome of Mauchly’s test was significant ($p < .001$) and we used Greenhouse-Geisser-corrected $F$-values. Within-subjects contrasts test of linearity was significant for font size, with $p < .001$, $F_{1,23} = 57.385$. Thus, average fixation duration decreased in a linear way as the font size of the text
There was also a significant main effect of differences in font sizes on the std of y-coordinates of fixations, with $p < .001$, $F_{2.46} = 21.798$. The within-subjects contrasts test of linearity was significant for font size, with $p < .001$, $F_{1.23} = 36.755$. Thus, the std of y-coordinates of fixations decreased in a linear way as the font size used for the text increased (see Figure 7).

9. SUBJECTIVE OPINIONS

Two types of questionnaires were delivered to the participants: a brief questionnaire after reading of each piece of text and a longer post-test questionnaire. With the post-test questionnaire we collected background data and general comments on the experiment. In particular, participants reported a variety of options on how they usually scroll. Many of them described using some or all of the options available (clicking on the scroll bar, dragging the scroll box in the scroll bar, clicking on the arrows on the scroll bar, using keyboard arrow keys, pressing Page Up and Page Down, operating the mouse wheel, and using a touchpad). No trend that would have triggered clustering of the participants into different groups by their scrolling behaviour was found.

The questionnaires after each text were used to collect opinions about how interesting and convenient the text was to read with the current setup. The setup included different font sizes and scrolling options. Participants rated their answers on a five-point Likert scale ($1 =$ “Fully agree”, $2 =$ “Agree somewhat”, $3 =$ “Neutral”, $4 =$ “Disagree somewhat”, $5 =$ “Fully disagree”). There was one comprehension question to answer about the content of the text.

Figure 10 shows the distribution of responses from all of the participants to the claim as to whether the reading was convenient. Answers were gathered for all pieces of text. About 63% of the answers suggested that reading with automatic scrolling was fully or somewhat convenient while the equivalent value for manual scrolling was 71%. The automatic scrolling was new to all participants, and the speed was convenient for most of them. Being used to the system might increase the level of convenience in long-term use. Most participants were impressed and liked the auto-scrolling techniques.

![Figure 8](image-url) Heat maps of participants 23 (on top) and 22 (below) with the lowest and highest std of y-coordinates for fixations, respectively, in reading with automatic scrolling. On the left is 100% size, in the middle 125% size, and on the right 150% size.

![Figure 9](image-url) Average fixation duration during reading of text with manual and automatic scrolling in different font sizes, with standard deviation error bars.

![Figure 10](image-url) Participants' responses to the claim “Reading was convenient.”
10. DISCUSSION AND CONCLUSIONS
We started our analysis by finding preferred reading regions on the screen in reading of text in different font sizes. We found variation among the participants, in line with the findings of Buscher et al. [2]. Our results also showed that the standard deviation of the y-ordinates of fixations decreased linearly with font size. Hence, participants’ reading regions tend to get shorter as font size increases.

The fixation count fell when the pieces of text were read with automatic scrolling. On the other hand, no significant difference was found in fixation duration between the two types of scrolling. Consequently, it took less time to read text with automatic scrolling than with manual scrolling. Participants read in a similar fashion with manual and automatic scrolling. Hence automatic scrolling may improve readability by reducing reading time and will allow the reader to be a relatively fast reader.

Eye movement data were significantly affected by differences in font sizes. Fixation count increased and fixation duration decreased linearly with an increase in font size. People tend to fixate more and with shorter duration in reading with bigger fonts than with smaller ones. It was also found that the std of the y-ordinates decreased in a linear fashion with increased font size. These results can be applied for improving on-screen presentation of text. We ran our experiment with a 17” monitor, but the findings can be useful for smaller screens as well.

We observed some variation in participants’ opinion on reading with automatic scrolling. Many provided affirmative support for automatic scrolling. Some suggested an improved version with a visual cue to keep track of the focus position before scrolling starts. Some suggested a tactile cue to notify the user just before auto-scrolling begins, in order to decrease the surprise effect.

Determining the preferred reading region to guide auto-scrolling is just the first step. The eye movement data could also be used to control the pace of the scrolling. Although most participants found the current fixed pace convenient, paces set individually for each user could better accommodate the needs of fast and slow readers alike. Similarly, those users who prefer reading a screenful before scrolling might find it more convenient if in auto-scrolling the text is scrolled rapidly, with a visual marker, so that they can continue reading at the top of the screen.

In our study, we evaluated the experience of manual and automatic scrolling in reading of text on a computer screen. In the future, we could apply the same method in reading from smaller screens on tablets or even cellular phones. A hands-free system can use gaze information to operate scrolling, or zoom-in/out actions to create an appropriate font size for reading. On the other hand, it is worth pointing out that the mere use of an eye tracker in the experiment may affect reading behaviour. The distance between the participant and the screen was kept constant (about 60 cm), and the angle of the screen was such that the eye tracker could work without problems. In office use, people might want to read with a different setup.

If we can find the user’s preferred region on the screen, automatic scrolling could be used without the need for a “calibration” phase based on manual scrolling first. Knowledge of the approximate preferred reading region will improve interaction with documents. In addition, information on which parts of a document on the screen have been read can be informative when the reader returns to the document later (see, for example, [1]). Thus, collection of gaze data in general and the auto-scrolling approach in particular show several potential applications, and it is encouraging that the latter was, overall, found to be convenient by the participants and even to improve the reading experience.

11. ACKNOWLEDGMENTS
This work was supported by the Finnish Doctoral Programme on User-Centered Information Technology and by Academy of Finland grants 129335, 138908, and 266285.

12. REFERENCES
Paper VI


Gaze-Contingent Scrolling and Reading Patterns

Kari-Jouko Räihä
School of Information Sciences
University of Tampere, Finland
kari-jouko.raiha@uta.fi

Selina Sharmin
School of Information Sciences
University of Tampere, Finland
selina.sharmin@uta.fi

ABSTRACT
An automatic technique that scrolls the window content while the user is reading the text in the window has been implemented. Scrolling is triggered by gaze moving outside the reader’s preferred reading zone. The reading patterns instigated by automatic scrolling are analyzed both quantitatively and using gaze path visualizations. Automatic scrolling is shown to result in smooth reading activity.

Author Keywords
Gaze-based scrolling; analysis of reading.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI)

INTRODUCTION
Text is increasingly read on large displays and multi-monitor setups [13], where controlling the mouse and scrolling can be tedious, and on small display devices, such as tablets. In particular, in hands-busy situations where one hand is occupied with holding the device it would be an advantage if scrolling could be done without engaging the other hand. Various techniques for doing this have been developed, e.g. by tilting the device. Eye tracking offers an alternative to previous techniques for hands-free scrolling. Eye trackers have become smaller and cheaper, and tablet prototypes equipped with eye trackers already exist.

Traditional scrolling techniques have been sped up by automatic zooming based on scrolling rate [5], and gaze has been used to facilitate zooming [10]. The idea of using gaze to scroll the content in the display window while text is being read is not new, either [9, 17]. However, implementations of this idea are rare.

Such automatic scrolling technique was developed by us [15]. We found that on a general level, the technique works well: it makes reading faster, it does not change the basic reading parameters (average fixation duration), and it is liked by users. Here we continue the analysis by digging deeper into the reading patterns to see what, in detail, goes on during the reading process when text is scrolled automatically based on gaze position.

PREVIOUS WORK
Gaze has been used to control computers in a number of applications [8, 16]. In our earlier work we implemented a system for gaze-based scrolling [15]. The process had two stages: in the first one, readers read pieces of text and scrolled the web page manually using the scroll wheel of the mouse. Their preferred reading region (from the point where reading resumed to the point where the content was scrolled) was observed. In the second stage eye tracking was employed to see when gaze reached the bottom of the preferred region, and automatic scrolling of the content was initiated at the rate of 40 ms/pixel. Scrolling stopped when gaze had reached the top of the preferred reading region.

Our previous study found that a majority of the 24 participants in the experiment found reading convenient in the auto-scroll condition. Important findings were that reading was on a general level similar in both conditions (the average fixation duration did not change), but with auto scrolling the time to read the text was somewhat smaller, because the number of fixations was smaller.

We ran our experiment with a 17-inch TFT monitor with a resolution of 1280×1024 pixels. The browser window used in the experiment occupied the full screen (see Figure 1). We used three different font sizes, obtained by different zooming factors of the default text size (100%, 125%, and 150%), and six pieces of text, so that each text was read only once by each participant. Font size was found to have a significant effect: an increase in font size resulted in an increase in the number of fixations and a decrease in their average duration. The texts were carefully selected through pilot testing to be interesting. They were of reasonably similar size, ranging from 6841 to 8448 characters. Text did not have an effect on the metrics.

However, in [15] we only analyzed gaze data on a summary level. As in [4], heat maps were used to get an overview of the distribution of gaze point coordinates (Figure 1). Eye movements during reading were not analyzed. In addition to the summary metrics, gaze paths are important for finding out what happens during reading on a detail level [7, 12]. Analyzing the gaze paths and reading patterns during auto-scrolling is the topic of this paper.
READING PATTERNS

Several metrics and algorithms can be used for analyzing gaze paths [6]. We will start with the temporal visualizations of gaze data [11], where the x axis is used to show progress of time, and the y axis plots the y coordinates of the fixations. Thus the x coordinates of fixations are ignored in the analysis. This technique has been used for revealing individual differences between users of search engines [3], and it is useful for analyzing reading patterns in our study.

We use the same data set as in [15]. We had to exclude 8 of the 24 participants because of problems in data recording at the level of detail required for this analysis, or because they did not fully follow the task instruction of reading the text once, without trying to memorize it by rereading sentences. We chose to omit all data from participants whose data was not good for all three texts read with auto-scrolling.

Figure 2 shows an example of the temporal visualization for the same participant and font size that was used for the heat map in Figure 1. The pattern that is evident in Figure 2 is typical for most of the 48 cases (16 participants, 3 font sizes for each). The gaze path consists of alternating sweeps [1] down and up. The first sweep is longer: it starts from the top of the screen and continues down until the auto scroll starts to move text up. Then the path oscillates, with text being stable during the sweeps down and text scrolling up during the sweeps upward. The last sweep down is longer, as the reader has reached the end of the text at the bottom of the screen (no scrolling is possible as the text is exhausted).

In Figure 2, the upward sweeps vary from 250 to 300 pixels in height and last about half a minute each. We can also see that the participant was reading the text all the time even when it was being scrolled. In other words, the slow pace of scrolling did not disturb the reading process. On the other hand, scrolling speed was still faster than reading speed; a slower speed would have resulted in gaze remaining even more stable, closer to the bottom of the reading area.

Note, however, that the sweeps here are different from the ones in [1], since the x coordinates are not shown. The traditional two-dimensional gaze plot would form a zigzag pattern, with gaze returning to the beginning of the next line after finishing one line. What Figure 2 tells us is that while text is scrolled, the lines that are being read move slowly upward. The small dips within the upward sweeps mean that the start of the next line is still somewhat below the end of the previous line when reading continues.

The text that was read for Figure 2 was the longest of the six. It took about 2 minutes to scroll from top to bottom at the 100% font size. Thus, participant 1 read for slightly more than 2 and a half minutes while text was stable and for 2 minutes while it was scrolling. The speed of scrolling was the same for all participants; thus the difference in reading times was caused solely by the difference in reading while text was stable. Some rereading of the text that had been scrolled can still have occurred after the text had stabilized.

It is illustrative to compare the pattern in Figure 2 to the pattern produced by the same participant reading another text with manual scrolling. This is shown in Figure 3. We can notice a less stable pattern, with text being scrolled varying distances upward. Several fixations on the way up are used to keep track of the line that was most recently read, to be able to continue reading after scrolling. This is a typical pattern for almost all participants when using manual scrolling.
There is, of course, a lot of variation between participants. Figure 4 shows data from a reader that is faster than the one in Figure 2, taking only 221 seconds to read the text, as opposed to 278 seconds for the one in Figure 2. Consequently, there are fewer sweeps, and the upward sweeps take longer than the downward sweeps (as new text is brought into the preferred reading area, and the reader manages to read more of it before text reaches the top of the preferred area and scrolling stops).

Consequently, there are fewer sweeps, and the upward sweeps take longer than the downward sweeps (as new text is brought into the preferred reading area, and the reader manages to read more of it before text reaches the top of the preferred area and scrolling stops).

Figure 5 shows the gaze path of a reader reading text at the largest font size. This is also a slower reader than the ones in Figures 2 and 4. Both facts contribute to the number of sweeps, which is higher than in the others (13 upward sweeps, as opposed to 4 and 3 in Figures 2 and 4, respectively). Upward sweeps are of particular interest, because they correspond to the automatic scrolling sequences.

The gaze paths shown above are, just individual examples. More generally they raise the following questions:
1. How many upward sweeps, i.e. automatic scrolling actions, are there on average?
2. What is their extent both in space (height) and time?
3. How are these metrics affected by font size?

ANALYSIS OF READING PATTERNS

Table 1 summarizes the key properties of the sweeps. The height columns denote height of the main reading area, or more exactly, the difference between the bottom and top y coordinates. For instance, in the case of Figure 2, the bottom coordinate is 684 (start of the second upward sweep) and the top coordinate is 414 (end of the third upward sweep). The duration was estimated from the beginning of the first upward sweep to the end of the last upward sweep, divided first by the number of sweeps and then by 2 (since there was a downward sweep for each upward sweep).

As a whole, ANOVA shows a significant effect of font size on the number of sweeps with \( p < .01 \) and \( F_{2,47} = 5.628 \). We see from Table 1 that the number of sweeps increased with font size. This is natural, as there was less text on one screen. More specifically, the difference was significant between two pairs of font sizes, 100% vs. 125% and 100% vs. 150%, with pairwise \( t \)-test values \( t = 2.285, p < .05 \) and \( t = 3.271, p < .01 \) respectively. For the 125% vs. 150% pair the difference was not significant.

<table>
<thead>
<tr>
<th>Size</th>
<th># of Sweeps</th>
<th>Height</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>avg</td>
<td>std</td>
<td>avg</td>
</tr>
<tr>
<td>All</td>
<td>7.9</td>
<td>3.8</td>
<td>371</td>
</tr>
<tr>
<td>100%</td>
<td>5.7</td>
<td>3.0</td>
<td>359</td>
</tr>
<tr>
<td>125%</td>
<td>8.3</td>
<td>3.6</td>
<td>367</td>
</tr>
<tr>
<td>150%</td>
<td>9.8</td>
<td>3.6</td>
<td>387</td>
</tr>
</tbody>
</table>

Table 1. Summary of the reading pattern metrics. Height is in pixels and duration in seconds.

Font size did not significantly affect the average height of the reading area. Since with the increase in font size the number of sweeps grows and their height does not change much, their average duration should decrease. This is indeed the trend, but here, again, the difference was not statistically significant.

Figure 6 shows the number of sweeps plotted against total reading time. Since each participant read six different texts of slightly varying size, the times have here been normalized according to text length to allow comparisons.
Each data point in Figure 6 represents one participant reading one text at the indicated font size.

Figure 6 reveals the smaller number of sweeps for the basic 100% font size, as indicated by the statistical analysis. For larger font sizes individual variation plays a bigger role, both as regards reading time and the number of sweeps.

CONCLUSIONS AND FUTURE WORK
Visual analysis of the reading patterns shows the qualitative advantage provided by gaze-based automatic scrolling. Comparing, for instance, Figures 2 and 3 for the same participant, we see that reading becomes more regular. In particular, it becomes easy for the reader to keep track of the current point of text. This is highlighted in Figure 3 with the several intermediate fixations while moving from the bottom towards the upper part of the screen.

The technique works best with smooth, progressive reading, as in reading literary texts. However, for exploratory reading it could be augmented with other recent suggestions, such as automatic earmarking in the scrollbar [2] of places where reading has digressed non-sequentially, in the style of “Pick up where you left off” of MS Word 2013. Making such earmarks gaze-responsive would add another dimension to gaze control in reading. Furthermore, if the point of reading is indicated with a suitable cue, such as underlining, pauses in reading could be used to scroll the text to the top of the preferred reading region while the reader is occupied with other tasks.

It would also be interesting to experiment with different scrolling speeds. We saw that scrolling rate is now convenient, but still somewhat faster than reading. With a slower rate text could be made to remain in an even narrower window than now. But is that good? With gaze moving alternatingly down and up the reader gets a feeling of making progress in the text. A slower scrolling rate would start to resemble dynamic presentation formats where text does not move vertically at all [14]; comparing these cases in terms of comfort and reading rate would be interesting. In general, while the lab experiment gives promising results, a long-term study in continuous use is needed to study, e.g., the effects on eye fatigue, or the stability of the reading region in daily reading practice.

ACKNOWLEDGEMENTS
We thank Oleg Špakov for implementing the auto-scroll technique. This work was supported by the Academy of Finland and the UCIT doctoral programme.

REFERENCES
Publications in the Dissertations in Interactive Technology series

Details of the dissertations are available at http://www.uta.fi/sis/tauchi/dissertations.html.

1. **Timo Partala**: Affective Information in Human-Computer Interaction
2. **Mika Käki**: Enhancing Web Search Result Access with Automatic Categorization
3. **Anne Aula**: Studying User Strategies and Characteristics for Developing Web Search Interfaces
4. **Aulikki Hyrskykari**: Eyes in Attentive Interfaces: Experiences from Creating iDict, a Gaze-Aware Reading Aid
5. **Johanna Höysniemi**: Design and Evaluation of Physically Interactive Games
6. **Jaakko Hakulinen**: Software Tutoring in Speech User Interfaces
7. **Harri Siirtola**: Interactive Visualization of Multidimensional Data
8. **Erno Mäkinen**: Face Analysis Techniques for Human-Computer Interaction
9. **Oleg Špakov**: iComponent – Device-Independent Platform for Analyzing Eye Movement Data and Developing Eye-Based Applications
10. **Yulia Gizatdinova**: Automatic Detection of Face and Facial Features from Images of Neutral and Expressive Faces
11. **Päivi Majaranta**: Text Entry by Eye Gaze
12. **Ying Liu**: Chinese Text Entry with Mobile Phones
13. **Toni Vanhala**: Towards Computer-Assisted Regulation of Emotions
14. **Tomi Heimonen**: Design and Evaluation of User Interfaces for Mobile Web Search
15. **Mirja Ilves**: Human Responses to Machine-Generated Speech with Emotional Content
16. **Outi Tuisku**: Face Interface
18. **Joel S. Mtebe**: Acceptance and Use of eLearning Solutions in Higher Education in East Africa
19. **Jussi Rantala**: Spatial Touch in Presenting Information with Mobile Devices
20. **Katri Salminen**: Emotional Responses to Friction-based, Vibrotactile, and Thermal Stimuli
21. **Selina Sharmin**: Eye Movements in Reading of Dynamic On-screen Text in Various Presentation Formats and Contexts