A longitudinal study of text entry by gazing and smiling

Titta Rintamäki

University of Tampere
School of Information Sciences
Interactive Technology
M.Sc. thesis
Supervisor: Outi Tuisku
February 2014

University of Tampere
School of Information Sciences
Interactive Technology
Titta Rintamäki: A longitudinal study of text entry by gazing and smiling
M.Sc. thesis, 52 pages, 2 appendix pages
February 2014

Face Interface is a wearable device that combines the use of voluntary gaze direction and facial activations, for pointing and selecting objects on a computer screen, respectively. In this thesis a longitudinal study for entering text using Face Interface is presented. The aim of the study was to investigate entering text with Face Interface within a longer period of time. Twelve voluntary participants took part in an experiment that consisted of ten 15-minutes long sessions. The task of the participant in each session was to write text in fifteen minutes with Face Interface and an on-screen keyboard. Writing was done by pointing at the characters by gaze and selected by smiling. The results showed that the overall mean text entry rate for all sessions was 5.39 words per minute (wpm). In the first session the overall mean text entry rate was 3.88 wpm, and in the tenth session 6.59 wpm. The overall mean minimum string distance (MSD) error rate for all sessions was 0.25. In the first session the overall mean MSD error rate was 0.50 and in the tenth session 0.05. The overall mean keystrokes per character (KSPC) value for all sessions was 1.18. In the first session the overall mean KSPC value was 1.26 and in the tenth session 1.2. Subjective ratings showed that Face Interface was easy to use. The rating of the overall operation of Face Interface was 5.9/7.0 in the tenth session. Subjective ratings were positive in all categories in the tenth session.

Key words and terms: Text entry, eye tracking, facial activity detection, multimodality, longitudinal study

Contents

1. Introduction

The use of mouse and physical keyboard are the most common ways to interact with the computer. However, the field of human-computer interaction (HCI) studies alternative ways to communicate with computers and tries to model human communication in developing the new methods.

Gaze is a natural way to communicate, since people usually look at other people or target when they are interacting with them. Gaze is also a very fast way to communicate, as the attention of the person is often already there where he or she is looking at [Jacob, 1991; Ware and Mikaelian, 1987; Surakka et al, 2003]. Due to these benefits, gaze has been used in HCI for modeling the functionalities of the computer mouse, that is, for pointing and selecting targets on the computer screen. When the gaze is used for selecting targets, the selection is often done with dwell time, which means that the person focuses his or her gaze on target for a predefined period of time (e.g., for 500-1000 ms) [Majaranta et al., 2002; Surakka et al., 2003; Majaranta, 2009]. If the dwell time is too long, it can be tiring and slow down performing the task [Isokoski, 2000; Majaranta et al., 2009]. On the other hand, if the dwell time is too short, it can cause the so-called Midas touch problem, which means that whenever the person looks at something, it gets activated and selected [Jacob, 1991; Isokoski, 2000].

Due to these problems with gaze-only selection, it would be useful if pointing targets could be done by gazing, but the selection could be done by means of some other interaction method [Surakka et al., 2003]. Using human face for selecting targets is one option. Face has an important role when people are communicating with each other. The facial expressions are a part of non-verbal communication [Rinn, 1984; Surakka et al., 2004], but the facial movements can be produced voluntary as well. Those voluntary movements, for example, frowning or smiling, can be measured and used for selecting targets on computer screen [Barreto et al., 2000; Rantanen et al., 2010; Rantanen et al., 2011].

Human to human communication is typically multimodal, which means that people use two or more interaction methods simultaneously when they are communicating with each other (e.g. vision and hearing). In order to develop natural and versatile interaction methods, combining different ways to interact into one might be useful also in HCI. Combining makes it possible to utilize the advantages of the methods and compensate the disadvantages of them [Surakka et al., 2003; 2004].

The use of voluntary gaze direction for pointing and facial activations for selecting have been combined into one wearable device, Face Interface. Face Interface is an eye-glass like wearable device that combines a video-based wearable eye tracker and a capacitive facial movement detector in one device. Pointing and selecting experiments have been conducted with Face Interface [Tuisku et al., 2011; Tuisku et al., 2012] and based on the results, Tuisku et al. [2012] suggested that Face Interface could be used in more advanced tasks, for example in writing with an on-screen keyboard [Tuisku et al., 2012]. Following this, the functionality of Face Interface for entering text has also been experimentally tested. The results showed that Face Interface could be used for entering text [Tuisku et al., 2013].

Longitudinal text entry studies have been made earlier with gaze only methods [Tuisku et al., 2008; Majaranta et al., 2009]. Now there was a need to conduct a longitudinal study also with a multimodal method, Face Interface. In this thesis, a longitudinal study for entering text with Face Interface was presented. Twelve participants took part in an experiment that consisted of ten 15-minutes long sessions. The aim of the study was to investigate, how people are able to write with Face Interface. Also, subjective ratings and an interview were conducted to find out, how the participants felt the writing with Face Interface.

The chapter 2 of this thesis describes eye tracking, and facial activity detection is presented in the chapter 3. The chapter 4 introduces multimodal combinations of gaze interaction and facial activity. Face Interface is described in the chapter 5. Chapter 6 explains the methods of this study, and the results of the study are presented in the chapter 7. The discussion is in the chapter 8, and the conclusion of the study is presented in the chapter 9.

2. Eye tracking

2.1 About eyes

Gaze is a natural way to communicate, since people normally look at other people when they are interacting with them. Also, people get information about items or objects by looking at them [Just and Carpenter, 1976]. When person is looking at something, the gaze is fixated on it. Fixation is a short period of time (typically 200-600 milliseconds) when the gaze is relatively still [Jacob, 1995]. During the fixations, the eyes do not remain completely still but make small movements. Only during the fixation the object can be seen and information about it can be received. When the fixation is over, the gaze moves rapidly from one point to another. This move is called saccade, and it lasts about 30-120 milliseconds [Majaranta and Räihä, 2002]. Saccade is a ballistic movement, which means that when a saccade begins, it can neither be interrupted nor its direction can be changed. During the saccade the person cannot see the object, that is, the vision of the person is suppressed. Fixations and saccades form the gaze path of a person [Jacob, 1991]. During the gaze bath the eyes move unevenly, that is, the gaze “jumps” due to fixations and saccades. Only during smooth pursuit the eyes move smoothly. Smooth pursuit is a slower and less sudden eye movement that occurs when a moving object is viewed [Jacob, 1991; 1995].

2.2 Eye tracking technology

For detecting and identifying eye movements, eye trackers are used. Eye tracker is a device for finding out where the person is looking at. In video-based eye trackers, a video camera images the eye and transfers the information to the computer. The gaze points, that is, the fixations are identified from the video image by utilizing the pupil detection and corneal reflection detection. Corneal reflection means reflection that is generated on the cornea of the eye by infrared light sources of the eye tracker. Gaze direction is then calculated by measuring the changing relationship between the pupil and the corneal reflection [Surakka et al., 2003; Hansen and Majaranta, 2012].

In order to estimate the individual user's gaze, the eye tracker needs to be calibrated for each person separately. The calibration is performed by asking the user to gaze at calibration points (e.g., nine equally spaced points) that appear on the computer screen. The calibration points appear one at a time on different locations of the screen. When the user looks at the calibration point, the camera images the eye. Then the images of the eye are analyzed in order to estimate the persons individual gaze direction on the computer screen. Based on the estimation, the eye tracker is optimized for the user [Majaranta and Räihä, 2007].

Eye trackers can be quite accurate, that is, about 0.5 degree visual angle from the user. In practice, however, their accuracy may be less because of drifting which means that over time the measured point of gaze drifts away from the actual point of gaze. Thus, a calibrated eye tracker may be accurate in the beginning but become more inaccurate when used longer. Drifting occurs due to for example changes in pupil size. The accuracy with about 0.5–1 of drifting equals approximately 1–1.5 cm on the computer screen at a normal viewing distance. Drifting can be somewhat corrected dynamically [Majaranta, 2009]. Stampe and Reingold [1995] developed a method of dynamic re-centering for correcting drifting. They assumed that on average the gaze of a person is directed at the center of the target. Based on this assumption, they calculated the drift to be the mean compensation between the target and the gaze position at each selection. Thus, they corrected drifting dynamically by realigning the measured gaze position to the center of the target [Stampe and Reingold, 1995]. Some newer eye trackers have built-in techniques for preventing drifting [Majaranta et al., 2006; Majaranta, 2009]. The technique can be for example using information gathered from both eyes, not just one [Tobii Technology, 2013].

An eye tracker can be either remote or wearable. In a remote eye tracker the camera is placed near the computer screen, whereas in a wearable eye tracker the camera is attached to a separate headgear of the user (e.g., to a helmet or a pair of eye glasses). A typical set-up in remote eye tracking is shown in Figure 2.1. The person sits in front of a computer screen which has a video camera integrated. Infrared light sources are directed to the person's eye and they create reflections on the surface of the eye. Infrared light does not disturb the person [Majaranta et al., 2009; Hansen and Majaranta., 2012].

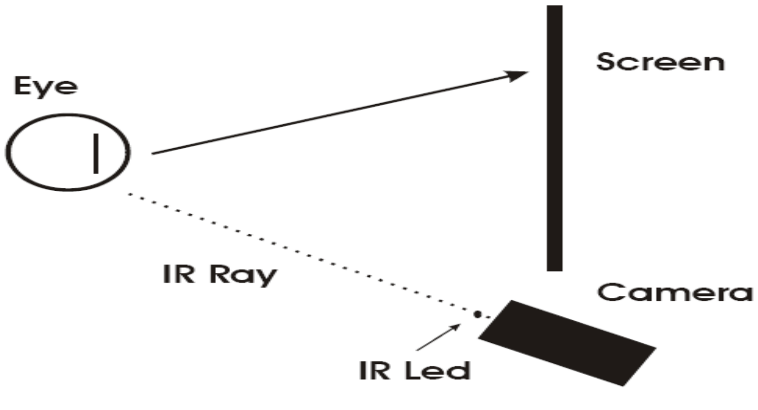


Figure 2.1 Remote eye tracking system. [Goni et al., 2004]

A typical set-up in wearable eye tracking is shown in Figure 2.2. An eye camera is integrated on a wearable headset and it is imaging the left eye of the person. Wearable eye trackers are used in mobile situations where the user can move freely, since the camera is attached to a headgear [Villanueva et al., 2012].

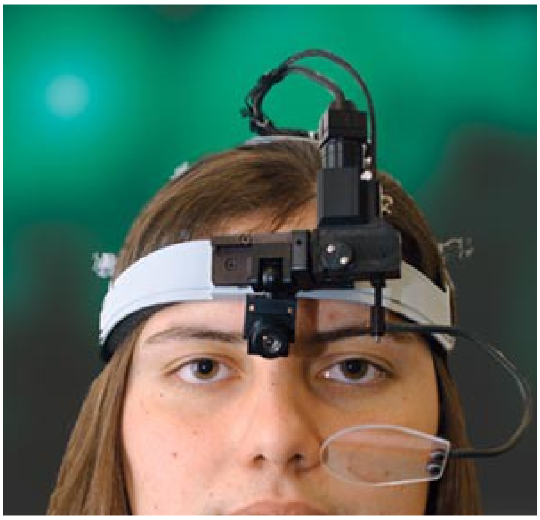


Figure 2.2 Wearable eye tracking system [http://commons.wikimedia.org]

When using a remote eye tracker, the person needs to sit quite still, so that the head does not move much. The head movements can prevent tracking the eye movements properly. That is, if the head moves, the direction (and position) of the user's gaze might change. Earlier, separate head rests have been used for compensating the head movements [Hansen and Majaranta, 2012]. Today, most remote eye trackers have set limits, on which distances and angles they track the eye accurately and how much head movements they tolerate [Surakka et al., 2003; Hansen and Majaranta, 2012]. A wearable eye tracker needs to stay firmly on the head so that the eye camera does not slip. On the other hand, if the user has to wear the headgear for a longer period of time, the wearable eye tracker needs to be light enough to make it comfortable to wear [Villanueva et al., 2012].

2.3 Pointing and selecting by gaze

The use of voluntary eye movements is utilized in HCI. Gaze can be used as an input method since it is a fast interaction method [Ware and Mikaelian, 1987; Surakka et al, 2003]. Gaze pointing means that the user points at the target on the computer screen by looking at it. Since in gaze pointing, the gaze has the same functions as the computer mouse has (i.e., pointing and selecting), it is quite easy to learn [Stampe and Reingold, 1995].

One method for selecting targets on the screen with gaze is dwell time. After pointing the target by gaze, the user continues looking at it for a predefined period of time. When the dwell time has elapsed, the target is selected. The selection can be interrupted by looking away from the pointed target before the dwell time runs out. The duration of dwell time is typically 500-1000 ms, but the optimal dwell time may depend on the task and the user [Majaranta et al., 2002; Surakka et al., 2003; Majaranta, 2009]. If the dwell time is too long, the user's eyes might get tired because he or she needs to keep the gaze focused for the time needed. It can also be frustrating to wait for a confirmation of the selection, if the user wants to proceed faster with the task. Thus, too long dwell time can slow down performing the task and disturb user's concentration on the task [Isokoski, 2000; Majaranta et al., 2009]. If the dwell time is too short, it can cause the so-called Midas touch problem, that is, whenever the user looks at something, it gets selected. The user might choose wrong targets because he or she is just studying the target, not intending to select it [Jacob, 1991; Isokoski, 2000]. In the systems where the dwell time is adjustable, experienced users may use shorter dwell time and proceed faster than novices [Majaranta et al., 2002].

Ware and Mikaelian [1987] combined the use of gaze as a pointing method with three different selection techniques. Dwell time, on-screen selection button and a physical keyboard button were used for selecting targets on the computer screen. The dwell time was set to 0.4 seconds. The on-screen button was a rectangular area in the computer screen. It was used so that the user first pointed at the target by gaze, and after that he or she pointed at the on-screen button by gaze in order to make the selection. With the physical button, the user pointed at the target by gaze and pressed the button. .Based on the viewing distances and the target sizes, the visual angles were calculated. The results showed that the selection times were below one second for all selection methods. The use of the dwell time and the use of the physical button were equally fast, and they were faster the use of the on-screen button. The fastest selection times with the dwell time and the physical button were about 0.6 seconds, whereas they were about 0.7 seconds with the on-screen button. Ware and Mikaelian [1987] further continued the studies with the dwell time and the physical button. They wanted to find out how the size of the target affects the selection. The dwell time was again set to 0.4 seconds. The targets on the computer screen were square menu items. The target sizes were 7.2 mm, 12 mm, 36 mm and 48 mm on a side. The results showed that it was faster to select the target when the target size increased. The use of the physical button was faster than the use of dwell time for selecting objects. However, the use of dwell time produced fewer errors than the use of the physical button. Ware and Mikaelian [1987] stated that the use of gaze is fast, if the size of the target is large enough. The use of dwell time leaves the hands free for other activities and could also be a suitable selection method for disabled persons. Ware and Mikaelian [1987] also noticed that pointing the target by gaze has an advantage over the pointing by mouse, since the target is already fixated, without the need of finding the cursor and dragging it to the target with the mouse [Ware and Mikaelian, 1987].

Sibert and Jacob [2000] conducted two experiments where they compared the use of gaze with the use of computer mouse for pointing and selecting targets on a computer screen. The aim was to study, whether gaze is faster than mouse in selecting targets on computer screen. In the first experiment the users had to select a highlighted circle among a grid of twelve circles as fast as possible. The task measured selection speed, that is, the user needed to detect the circle on the screen and quickly point at it. When the target was selected, it was non highlighted and the next circle became highlighted. For the use of gaze, the dwell time was set to 150 ms. The overall mean selection time was 503.7 ms for the gaze and 931.9 ms for the use of mouse. The results showed that the use of gaze was significantly faster than the use of mouse in target selection. The task in the second experiment was otherwise the same, but now a letter was added inside each circle. The user was told through an audio speaker, which letter he or she needed to select. This task now included a cognitive load for the user, that is, the user needed to listen to the speaker, look for the correct letter on the screen and then point at it. When the correct letter was selected, it was highlighted, and the next letter was presented. Also this experiment showed that the use of gaze was faster than the use of mouse. The overall mean selection time was 1103 ms for the use of gaze and 1441 ms for the use of mouse. Sibert and Jacob [2000] also stated that the distance to the target had an effect on the speed of the pointing technique. Farther the user needed to point, the greater the speed advantage of gazing was. This means that the difference between the gazing speed and the mouse speed increased (i.e. gazing being even faster) when the distance to the target increased [Sibert and Jacob, 2000].

Even though using dwell time can be fast for selecting objects, it is not an optimal selection method in all situations. For example, some people have difficulties in focusing the gaze for a prolonged time due to their physical restrictions (i.e., they are not able to keep their gaze still for the period of dwell time) [Majaranta and Räihä, 2002]. Beside the dwell time, other eye-based solutions for selecting targets are, for example, eye switches and gaze gestures. Using eye switches means that the selection of the target is done by means of, for example, voluntary blinking or winking. In this case, the user first points at the target with the gaze and then selects it by blinking or winking his or her eye [Majaranta and Räihä, 2007]. Gaze gestures can be gaze movements to a specific direction or area in order to make the selection [Majaranta and Räihä, 2007]. For example, a target on the screen is first pointed with the gaze and then selected by moving the gaze to a separate area of the screen (e.g., a part of the screen reserved as a selection area) [Yamada and Fukuda, 1987].

2.4 Text entry by gaze

Along with pointing and selecting, text entry is the most common task in HCI [Surakka et al., 2003]. Thus, different techniques for entering text have been studied and developed for gaze-based interaction methods. The most common way of writing with gaze-based methods is eye typing, that is, entering text with an on-screen keyboardThe first virtual keyboards were modeled after physical keyboards. Because the traditional QWERTY keyboard layout is familiar to most users [Majaranta and Räihä, 2002], it has been commonly used also for on-screen keyboards. Figure 2.3 shows a typical setup of eye typing. In eye typing, the user points at a character by looking at it and selects it by using dwell time or a separate switch. The eye typing system might give feedback on the pointing, for example by highlighting the pointed character (e.g., by changing background color or drawing a border around the key) or showing the cursor on the character. When the character has been selected, the system may, for example, perform a click sound, to give a confirmation of the selection. The character might also change visually (e.g., by shrinking or changing the color). After a successful selection, the character appears in the input field [Majaranta et al., 2006]. Eye typing has been an especially important communication method for severely disabled people who are not able to communicate otherwise [Majaranta and Räihä, 2002].

Figure 2.3 An on-screen QWERTY keyboard and an eye tracking device. The eye tracker follows the user's gaze and the computer analyzes the gaze behavior. The pointed character is shown highlighted on the keyboard layout [Majaranta et al., 2006].

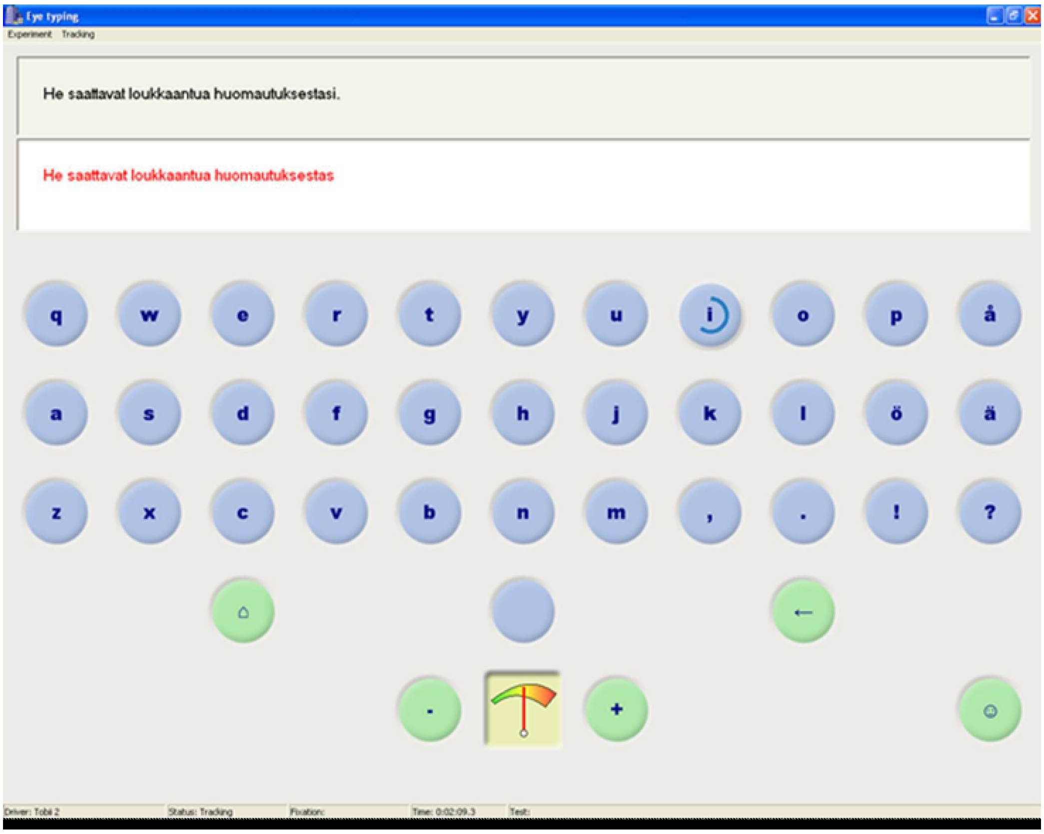
Majaranta et al. [2006] performed three eye typing studies in one experiment, where they used a QWERTY on-screen keyboard layout (see Figure 2.3). The aim of the experiment was to study the effect of visual and auditory feedback on eye typing. The task of the participants was to write short sentences one at a time by pointing at the desired characters and selecting them with dwell time. Dwell time was used together with different combinations of visual and auditory feedback. The first study investigated the effects of auditory and visual feedback on eye typing with a long dwell time (900 ms). The feedback combinations were visual only, visual and click, visual and speech and speech only. Visual feedback was highlighting and shrinking the key (along with the dwell time progressed) for pointing, and changing the key color and “pressing” the key down for selecting. Auditory feedback was a click sound or synthetic speech (i.e., the selected key was spoken). The results showed that the overall mean text entry rate of the study (i.e., all combinations) was 6.97 wpm. The fastest combination was to have both visual feedback and a click sound. The overall mean text entry rate of that combination was 7.55 wpm. The participants also preferred the click sound to visual feedback only. The overall mean MSD error rate was 0.54. The highest overall mean MSD error rate 0.95 was produced when there was only visual feedback available. The overall mean KSPC value was 1.09 [Majaranta et al., 2006].

The second study investigated more closely the effects of animated feedback (shrinking effect of the key) on eye typing with long dwell time (900 ms). The overall mean text entry rate of the study was 6.83 wpm. For shrinking, the overall mean text entry rate was 7.02 wpm, and when there was no shrinking feedback, it was 6.65 wpm. The overall mean MSD error rate was 0.43 and the overall mean KSPC value was 1.09 [Majaranta et al., 2006].

The third study investigated the effects of auditory and visual feedback on eye typing with a short dwell time (450 ms). The feedback combinations were speech only. 2-level visual (highlighting for pointing, and changing the key background color for selecting) and 1-level visual (only changing the key background color for selecting). The overall mean text entry rate of the study was 9.89 wpm. With the result of 10.27 wpm, the fastest combination was the 2-level visual, that is, to have visual feedback both for pointing and selecting. However, many participants commented that they would like to have the click sound supporting the visual feedback. The overall mean MSD error rate of the study was 1.20. The results showed that when the dwell time is shorter, text entry is faster, but the accuracy is lower. For the overall mean KSCP values, the differences between feedback types were significant in the third study. When only speech was used, the KSPC value was 1.28. For 1-level visual feedback KSPC value was 1.17, and for 2-level visual feedback KSPC value was 1.19 [Majaranta et al., 2006].

Based on the results, Majaranta et al. [2006] stated that adding a short click sound improved both the eye typing speed and accuracy. The duration of the dwell time had an effect on what kind of feedback should be used in eye typing. If the dwell time is short, the feedback should be short and clear [Majaranta et al., 2006].

Further, Majaranta et al. [2009] performed a longitudinal study on eye typing with a QWERTY on-screen keyboard using an adjustable dwell time. After pointing at the character, the progression of the dwell time was visible (see Figure 2.4). Majaranta et al. [2006] aimed to find out, how fast novice users could learn to write by gaze, if they had the possibility to adjust the duration of the dwell time. The study consisted of ten 15-minutes long experimental sessions. The task of each session was to write Finnish sentences as fast as possible within 15 minutes. [Majaranta et al., 2009].



He saattavat loukkaantua huomautuksestasi.

He saattavat loukkaantua huomautuksestas

Driver: Tobii 2 Status: Tracking Fixation: Time: 0:02:09.3 Test:

Figure 2.4 The QWERTY on-screen keyboard used in the experiment. The dwell time visualization (here the character ‘i’) included an animation of a closing circle, which indicated the progression of dwell time. The minus and plus keys were used for adjusting the dwell time [Majaranta et al., 2009].

When the dwell time elapsed and the circle closed, both visual and auditory feedback were given to the participant. The selected character was depressed and a click sound was played. The duration of dwell time was first set to 1000 ms for all participants. Participants were instructed to adjust the dwell time between sentences but they were able to adjust it whenever they wanted to. The adjustment was done with plus and minus keys. The plus key increased the typing speed by decreasing the dwell time, that is, if the user wanted to type faster, he or she pointed at the plus key with the gaze. Accordingly, the minus key slowed down the typing speed by making the dwell time longer. The minimum duration of the dwell time was 150 ms, and the maximum duration was 2000 ms. A speed indicator was placed between the minus and plus keys. When the hand of the indicator was in the middle, the dwell time was 600 ms. The results were based on ten participants. The results showed that the text entry rate was 6.9 wpm in the first session and 19.9 wpm in the tenth session. The overall mean dwell time was 876 ms in the first session and 282 ms in the last session. Thus, the writing speed increased and the dwell time decreased along the study. The overall mean MSD error rate was 1.28 in the first session and 0.36 in the last session. The MSD error rate decreased even though the writing speed increased. The overall mean KSPC value was 1.09 in the first session and 1.18 in the last session. When the writing speed increased, the participants probably needed to correct more errors, and that increased the KSPC values [Majaranta et al., 2009].

Even though QWERTY keyboard layout is familiar to most users, the keys of it need to be big enough in order to point at them accurately [Majaranta and Räihä, 2002]. As one solution for having big enough keys, Špakov and Majaranta [2008] developed a scrollable keyboard. The keyboard saves screen space by showing only part of the keyboard at the time. The full keyboard is modeled after common QWERTY keyboard layout (see Figure 2.5: on top). The two-row keyboard (Figure 2.5: in the middle) shows only two rows of keys at the same time. The special scroll keys for up and down are located on the left. These keys are selected when the user wants to reach the third row. The one-row keyboard (Figure 2.5: on bottom) has only one row visible at the same time. The scroll keys are located on the sides of the keyboard. For both one-row and two-row keyboards, the invisible rows can be reached by selecting either of the scroll keys, as the scrolling is cyclic [Špakov and Majaranta, 2008].

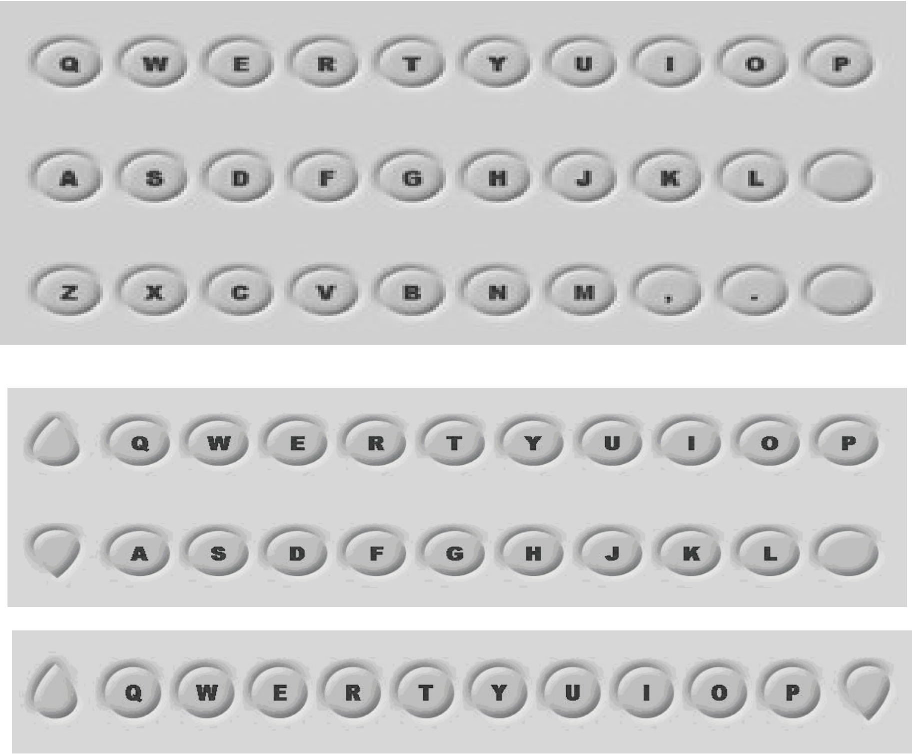


Figure 2.5 A scrollable keyboard: full keyboard on top, two-row keyboard in the middle and one-row keyboard on bottom [Špakov and Majaranta, 2008].

Špakov and Majaranta [2008] conducted an experiment with the scrollable keyboard. The task was to write English sentences as fast as possible. Participants were instructed to ignore mistakes, since there was no backspace key on the keyboard. One-row keyboard, two-row keyboard and full keyboard were used in the experiment. There were eight separate testing sessions. Each session included six sentences for each keyboard. The results showed that in the last session the overall mean text entry rate was 15.06 wpm for the full keyboard, 11.12 wpm for the two-row keyboard, and 7.29 wpm for the one-row keyboard. The error percentages varied between 1% and 5% during the experiment, and in the last session the average error percentages were below 2% for all three conditions. Since the use of the scroll key produced at least one extra keystroke (due to the scroll key), the overall mean KSPC value was more than 1. The overall mean KSPC value was 1.2 for the two-row keyboard and 1.64 for the one-row keyboard [Špakov and Majaranta, 2008].

Based on the results of the scrollable keyboard study, Špakov and Majaranta [2009] optimized the keyboard layout by grouping the most frequent characters together, in order to reduce the scrolling. The most frequent characters were placed in the topmost row, and the least frequent characters in the last row (see Figure 2.6).

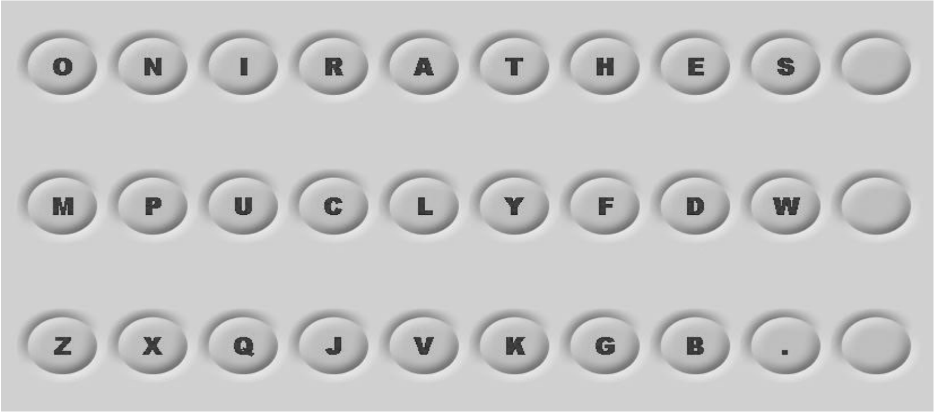
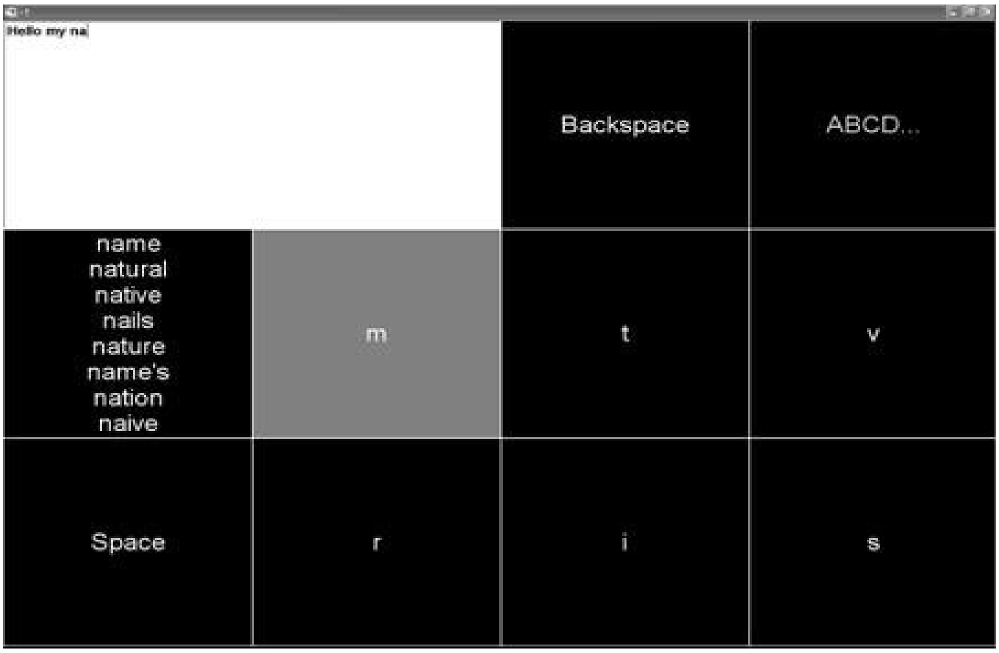


Figure 2.6 An optimized on-screen layout (full). The most common characters are placed on the top row [Špakov and Majaranta, 2009].

The optimized layout was experimentally tested. The method was the same as in the previous experiment [Špakov and Majaranta, 2008]. Only the full keyboard (three rows) was left out from the experiment. The results showed the overall mean text entry rate was 12.18 wpm for the two-row keyboard and 8.86 wpm for the one-row keyboard. The overall mean MSD error rate was approximately 2 for both keyboards. The overall mean KSPC value was 1.11 for the two-row keyboard and 1.49 for the one-row keyboard. The results also showed that the scroll button was used less with this optimized layout than with the reduced QWERTY layout in the earlier experiment. Špakov and Majaranta [2009] stated that the scrolling keyboards could be improved further by utilizing word prediction or word completion. The text entry system could, for example, display a list of predicted words based on the characters written so far [Špakov and Majaranta, 2009].

MacKenzie and Zhang [2008] compared word and character prediction in an eye typing method using a QWERTY on-screen keyboard. The method showed the next probable words on separate buttons that were located below the text input field, above the character buttons. Also, when the user typed a character, the system highlighted the three most probable next characters on the keyboard. MacKenzie and Zhang [2008] conducted an experiment where they compared the word and letter prediction. In the experiment they used two button sizes (large and small). Ten participants took part in the experiment. The results showed that the text entry speed varied from 10.8 wpm to 12.3 wpm. When small buttons were used, the letter prediction was about 10% faster than the word prediction. When large buttons were used, there was only little or no improvement in text entry speed. Word prediction was about 10% faster with large buttons than with small buttons. MacKenzie and Zhang [2008] stated that the letter prediction was as good as word prediction, sometimes even better [MacKenzie and Zhang, 2008] .

An example of a text entry system utilizing word prediction is GazeTalk software [Hansen et al., 2001]. GazeTalk uses a dynamic, non-QWERTY keyboard. The screen is divided into a 3 x 4 cells (see Figure 2.7). The word being entered is shown in the two top left cells, and six most likely characters to continue the word are shown on the six bottom right cells. The leftmost cell on the middle row shows a list of the next probable words, predicted on the basis of the text written so far. If the user selects the cell with the list of words, the two bottom rows are filled in with those words. If none of the predicted characters or words are correct, the user can choose the button “ABCD…”, which fills in the bottom row with the next options (i.e., the next suggested continuations for the typed word). The buttons are selected by using dwell time [Hansen et al., 2001].

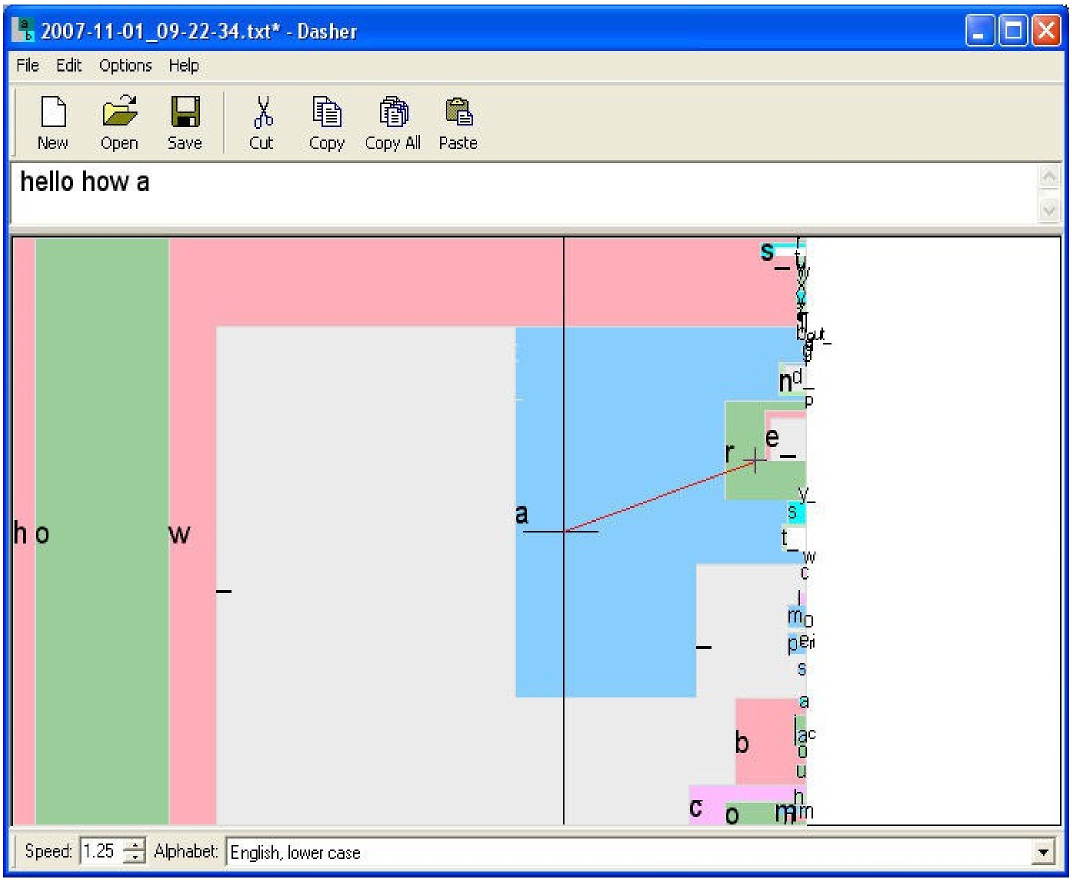


Hello my na

Figure 2.7 GazeTalk text entry keyboard. The user has written “Hello my na”. On the basis of “na”, GazeTalk suggests the character “m” to be the next character to be written. GazeTalk also presents the list of suggested words: “name”, “natural”, “native” etc. [Hansen et al., 2001].

Hansen et al. [2004] conducted a study with the Danish and Japanese versions of GazeTalk. The study consisted of two testing sessions. The task was to write predefined sentences as quickly and accurately as possible. The sentences were shown to the participants on a clipboard that was mounted by the top left corner of the computer screen. The results showed that the overall mean text entry rate was 6.26 wpm for the Danish version and 11.37 characters per minute (cpm) for the Japanese version. Based on the learning curve calculations, Hansen et al. [2004] estimated that after 1000 written sentences the maximum text entry speed would be 9.4 wpm for the Danish version, and 29.9 cpm for the Japanese version. The overall mean MSD error rate for Danish version was 1.09. For the Japanese version no MSD rate was obtained, since MSD method does not apply to the Japanese character system (due to different writing system, e.g., in Japanese one character represents one concept, and no spaces are used between the words) [Hansen et al., 2004].

Dasher [Ward and MacKay, 2002] is a text entry system using continuous pointing gestures. Writing is done by zooming the characters. When starting the writing, the characters are on the right side of the screen ordered alphabetically. In the case of gaze, user starts to move the cursor by looking at the desired character. Then the interface zooms in and the area of the selected character starts to grow and move left, towards the center of the screen (see Figure 2.8). At the same time, the language model of the system predicts the most probable next characters. The areas of the next most probable characters start to grow within the chosen area. This brings the most probable next characters closer to the current cursor position and minimizes the time and the distance to select them. The character is written when it crosses the horizontal line at the center of the screen. The entered character can be canceled by looking to the left. Then the interface zooms out and the character moves back to the right side of the screen [Ward and MacKay, 2002].



hello how a

Speed: 1.25 Alphabet: English, lower case

Figure 2.8 Dasher text entry system. The user has written "hello how a". Based on "a", the system predicts the next characters "re" and zooms into the area, where they are located [Tuisku et al., 2008]

The first text entry experiment with gaze-controlled Dasher was conducted by the inventors of Dasher, Ward and MacKay [2002]. They compared entering text with Dasher to entering text with an on-screen QWERTY keyboard. They had four participants in the experiments, two novices and two experienced users. The task was to write English sentences which were told to the participants by a recorded voice. After one hour of writing with Dasher, the text entry speeds varied from 10 wpm to 25 wpm. The highest text entry rate, 25 wpm, was reached by an experienced user of Dasher. The highest text entry rate reached with the on-screen keyboard was 15 wpm [Ward and MacKay, 2002].

Tuisku et al. [2008] conducted a longitudinal study on writing with Dasher. Twelve participants took part in an experiment that consisted of ten 15-minutes long sessions. Thus, the participants wrote text with Dasher 2.5 hours in total. The sentences were taken from the 500 sentence set originally published by MacKenzie and Soukoreff

[2003] and translated into Finnish by Isokoski and Linden [2004]. Each participant completed ten sessions by writing with the gaze and one session by writing with the mouse in the last gaze session. The results showed that the overall mean text entry rate for the gaze was 2.49 wpm in the first session and 17.26 wpm in the last session. The highest text entry rate reached individually was 23.11 wpm. The text entry speed for the mouse was 20.69 wpm. The overall mean MSD error rate for the gaze was 10.72 in the first session and 0.57 in the last session. The overall mean MSD error rate for the mouse was 0.93 [Tuisku et al., 2008].

3. Face in HCI

3.1 About face

Human face has an important role when people are communicating with each other. In communication, people can express their intentions and emotions via facial expressions [Rinn, 1984; Surakka et al., 2004]. The facial expressions are a part of non-verbal communication.[Rinn, 1984]. Physically, facial expressions result from muscle contradictions. The facial muscles contract and cause movements of facial skin and the connective tissue, *fascia*. Due to contractions, for example, wrinkles appear in the skin, and eyebrows or mouth corners move. The facial muscles respond to emotions, and in face the emotions can be differentiated from each other at the clearest [Rinn, 1984]. When developing new interaction methods, HCI field tries to model human communication, and using facial system could be a potential solution also for HCI. For example, voluntary facial movements can be used for interaction by combining the facial movements with gaze interaction [Surakka et al., 2004; Rantanen et al., 2012].

3.2 Facial activity detection technology

One way to investigate the possible facial solutions is to measure signals that originate from human facial expression systems [Surakka et al., 2004]. Electromyography (EMG) is a technique for measuring the changes in electrical activity of muscles. For EMG measurement, electrodes are placed on the region of the face above the muscles to be measured [Surakka et al., 2004]. Because the electrodes are attached to the face of the researched person (see Figure 3.1), the measurement needs special preparations. First the skin is cleaned properly with ethanol, and then it is abraded with electrode paste and cotton sticks to lower the impedance. Then the electrodes are attached firmly to the face [Barreto et al., 2000; Surakka et al., 2004].

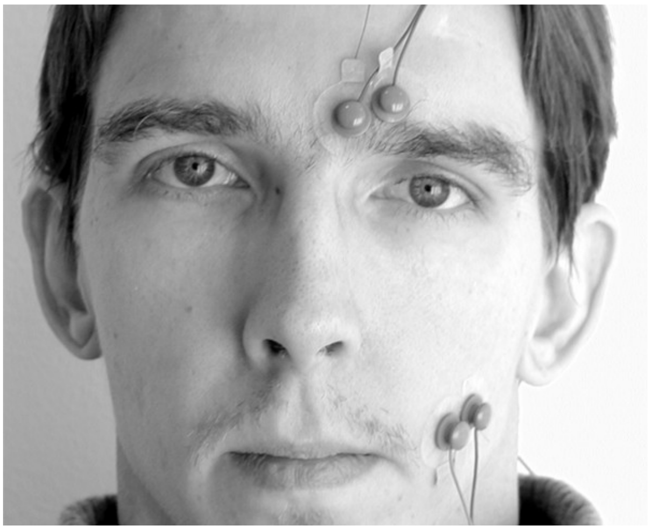


Figure 3.1 Facial EMG measurement. Electrodes are attached on the region of the muscles from which the activation is measured [Surakka et al., 2005].

EMG has been used for measuring facial expressions, for example, in several studies on emotion research area [Surakka and Vanhala, 2011]. One example of the studies is the experiment by Partala and Surakka [2004]. They conducted an experiment where psychophysiological effects of interventions were investigated. Eighteen participants took part in the experiment. The participants needed to perform an interactive problem-solving task with a computer, but during the task they faced pre-programmed mouse delays (the mouse was stopped, i.e., the mouse did not react and the participant could not continue performing the task).. After the mouse delay, the participants got either positive or negative interventions (via speech synthesizer) or no intervention at all. Facial responses were measured with EMG above the facial muscles *zygomaticus major* (i.e., the muscle activated when smiling) and *corrugator supercilii* (i.e., activated when frowning). The EMG activity was measured both during and after the intervention. The results showed that the smiling activity was significantly higher during the positive interventions than negative interventions or without interventions. Smiling activity was also significantly higher after the positive interventions than when there were no interventions. The frowning activity decreased significantly more after the positive interventions than when there were no interventions. When the participants got positive interventions, their problem solving performances were significantly better than when they did not get any interventions. Partala and Surakka [2004] stated that both positive and negative interventions were more beneficial than getting no interventions at all [Partala and Surakka, 2004].

Recently, another option for measuring facial activations has been introduced, a capacitive detection method [Rantanen et al., 2010]. The capacitive method measures physical movement of the facial skin that results from the activations of the facial muscles instead of the electrical activity. The capacitive method is contactless, that is, no electrodes are attached to the face. The sensor is placed in front of the muscle to be measured (see Figure 3.2). Thus, the detection method does not need any preparations for skin. A pair of electrodes is used to form electric fields. Facial movements result in changes in the electric fields, because the facial movements change the distance of the facial skin from the electrode (e.g.,when a person frowns, the tissue of the forehead area changes its form which changes also the distance to the capacitive sensor). The changes are then detected by measuring capacitances between the electrodes. [Rantanen et al., 2010; Rantanen et al., 2012].

Figure 3.2 Capacitive facial measurement. Capacitive sensors of the wearable device measure the movement of the muscles that are activated when the person is frowning or raising eyebrows [Rantanen et al., 2010].

Rantanen et al. [2010] conducted a test to evaluate the feasibility of the capacitive method. In the test, ten participants performed frowning and raising eyebrows movements (see Figure 3.2). A speech synthesizer produced starting indications with words ”frown” and ”lift”, and the participants acted accordingly. One test consisted of 25 indications. Each indication was chosen randomly to be either for frowning or for raising (*frontalis* facial muscle used for raising). Each participant performed four sessions. The participants were instructed to produce the movement smoothly, without any hurry. First, the participant needed to tense the muscle quickly, and then, relax it immediately. This way distinguishable movements were produced. The participants did not get any visual feedback during the movement. The signal from capacitive sensor was recorded and analyzed offline. The results showed that the capacitive method could be utilized in detecting facial movements [Rantanen et al., 2010]. Rantanen et al. [2012] improved the capacitive measurement method by extending the channels of facial movements. Beside frowning and raising eyebrows, also closing the eye, opening the mouth, raising and lowering the mouth corners, and relaxing the face could be detected. The improved measurement device was constructed as a headset with 22 electrodes (see Figure 3.3). This new prototype was tested by ten participants. Similarly to the earlier test of Rantanen et al. [2010], a speech synthesizer produced the starting indications, and the participants performed the required voluntary facial movement. Each movement was performed ten times in a randomized order. The participant could see himself or herself in the mirror during the test, that is, visual feedback was given [Rantanen et al., 2012].

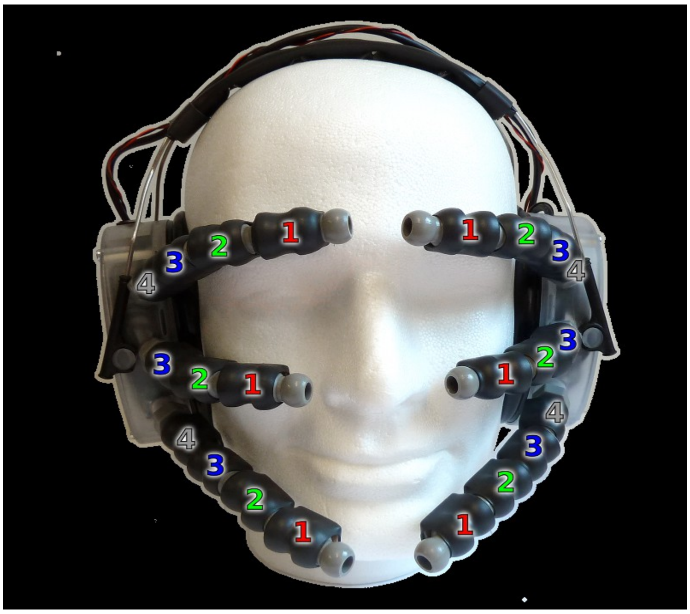


Figure 3.3 Wearable capacitive facial measurement device. The device includes 22 electrodes in total, 11 on both sides of the face. The extensions on top and bottom have 4 electrodes each, and the extensions in the middle have 3 electrodes each [Rantanen et al., 2012].

3.3 Pointing and selecting by face

Barreto et al. [2000] utilized facial EMG for pointing and selecting targets on a computer screen. In their method the signals measured from the facial muscles were transformed into two-dimensional cursor movements. The cursor was moved up by raising eyebrows up and moved down by lowering the eyebrows. Moving right jaw caused cursor moving to the right and moving left jaw caused cursor moving to the left. The full jaw clinch activation performed the mouse click, that is, the selection. The method was tested with an experiment, where six participants performed pointing and selecting tasks. In the task there was a a start button in one corner of the screen. The size of the start button was always 8.5 x 8.5 mm. There also is a stop button, which is always located at the center of the screen. The sizes of the stop button were 8.5 x 8.5 mm, 12.5 x 12.5 mm, 17 x 17 mm, and 22 x 22 mm. Each test session consisted of 20 trials with each size of stop button, that is, there were 80 trials in total. The participant first needed to select the start button, then move the cursor toward the stop button, and finally select the stop button. The timer was started when the start button was selected, and stopped, when the stop button was selected. Before the next trial began, the cursor was placed at the starting point. Five trials were started at each corner. The results showed that the overall mean task time of performing a task was 16.3 seconds [Barreto et al., 2000].

4. Multimodality

4.1 About multimodality

According to Surakka et al. [2003] interaction between humans or between a human and a computer can be either unimodal or multimodal. Unimodal interaction uses only one interaction channel, whereas multimodal interaction uses two or more interaction channels simultaneously. Combining different interaction methods into one makes it possible to utilize the advantages of the methods and compensate the disadvantages of them. Pointing and selecting are essential tasks in HCI, and different multimodal methods have been developed, where pointing could be done with one interaction method and selecting with another [Surakka et al., 2003; 2004]. For this thesis, it is essential to concentrate on methods that combine gaze interaction method with facial activity detection method. Thus, only gaze and facial based multimodal interaction methods are introduced.

4.2 Pointing and selecting by multimodal interaction

Partala et al. [2001] conducted a study where voluntary gaze direction was used for pointing and voluntarily activation of a facial muscle *corrugator supercilii* for selecting targets on computer screen. The use of this new technique (i.e., gaze pointing and facial muscle activation selection) was compared with the use of the mouse. The data was gathered with an eye tracker and an EMG recorder, and it was analyzed offline. The task was to point and select first a home square and then a target circle on the screen. The sizes of both objects were 32 pixels. Three pointing distances (50, 100, and 150 pixels) were used in the study. The results showed that the new technique was statistically significantly faster to use than the mouse at medium and long distances. The overall mean pointing times for the new technique were about 0.6 seconds at all distances. Thus, with the new technique, the distance did not have significant effect. The error percentage was 34% for the new technique and 0.9% for the mouse. With the mouse, the task times increased significantly as the distance to the target increased. The overall mean pointing time for mouse was about 0.6 seconds at the short distance, about 0.8 seconds at the medium distance and about 0.9 seconds at the long distance. Subjective ratings showed that the users preferred the use of the mouse to the use of the new technique [Partala et al. 2001].

Surakka et al. [2004] continued developing the technique of using voluntary gaze direction for pointing and voluntarily produced facial muscle activations for selecting targets on computer screen. They conducted an experiment where simple pointing and selecting tasks were performed both with using the new technique and with a computer mouse. Remote eye tracker was used for measuring the gaze direction, and facial EMG was used for measuring activations of *corrugator supercilii* facial muscle. Fourteen participants took part in the experiment. The task was as follows. First, a home square and a target circle appeared simultaneously on a computer screen (see Figure 4.1). The task of the participant was first to point at the home square by gazing at it and then to perform a click by frowning. Then he or she pointed at the target circle and performed another click. The target circle became highlighted when the gaze was inside it, and after a successful click, the target circle disappeared. The participant was not able to click the target circle before a successful click of the home square. After a successful click of the target circle there was a pause of 2000 ms and then the home square and target circle appeared again on the screen. The pointing task time measured was the time between the selection of the home square and the selection of the target circle. The experiment included three different pointing distances. The distances were short (60 mm), medium (120 mm) and long (180 mm). Also three different target circle sizes were used: small (25 mm), medium (30 mm) and large (40 mm). The size of the home square was always 30 mm. The target circles appeared in one of eight different angles around the home square. There were three different pointing distances, three different target circle sizes, and eight different angles, that is, 72 different combinations in total.

The results showed that the overall mean pointing task time was about 0.7 seconds for the new technique and about 0.6 seconds for the mouse technique. The mouse technique was statistically faster only at the shortest pointing distance. For other distances there were no statistically significant differences between the mouse and the new technique. The overall mean error percentage was about 17.5% for the new technique. Statistically, fewer errors were made with the new technique when the size of the target circle increased. The overall mean error percentage was about 3% for the mouse technique. Statistically, fewer errors were made at all sizes and distances with the mouse technique than with the new technique. The participants rated the pointing techniques after the experiment. The ratings showed that the mouse was rated as more accurate and easier to use than the new technique. However, the new technique was rated faster to use than the mouse technique [Surakka et al., 2004].

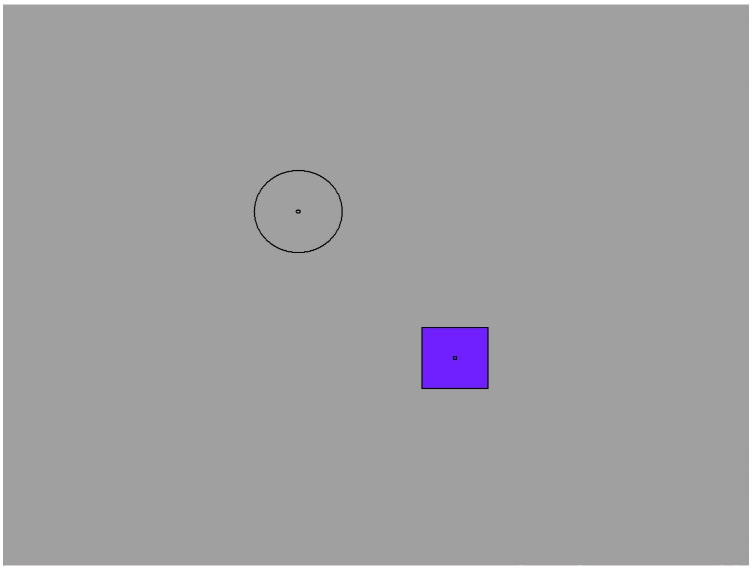


Figure 4.1: Pointing and selecting task. The home square is highlighted when the participant points at it with the gaze [Surakka et al., 2005].

In a follow-up study, Surakka et al. [2005] extended the technique to two facial EMG channels, that is, *corrugator supercilii* facial muscle and *zygomaticus major* facial muscle.. The experimental task was the same as in the previous study [Surakka et al., 2004]. Gaze direction was used for pointing, and selecting was performed by frowning or smiling techniques. Results showed that the smiling technique functioned faster than the frowning technique. The overall mean pointing task time was 0.5 seconds for the smiling technique and 0.8 seconds for the frowning technique. The error percentage was 16.7% for the smiling technique and 27.1% for the frowning technique. Subjective ratings showed no significant differences between the techniques [Surakka et al., 2005].

Based on Surakka et al. [2004; 2005] others have developed techniques for controlling computers with gaze direction and facial muscle activations. San Agustin et al. [2009] compared the use of two pointing and two selection techniques. In their experiment, mouse and gaze were used for pointing, and mouse click and voluntarily facial EMG movements (frowning and jaw clenching movements) were used for selecting targets. All four possible pointing and selection combinations were tested in simple tasks, where participants pointed and selected targets. Three pointing distances (200, 250 and 300 pixels) and three target sizes (100, 125 and 150 pixels) were used. Results showed that the overall mean task time of all combinations was 0.4 seconds. The fastest of the pointing and selection combinations was gaze combined with facial

EMG (task time approximately 0.35 seconds). The error percentage of all combinations was 22.25%.

Chin et al. [2008] used facial EMG for correcting the inaccuracies of the eye tracker and for selecting targets. First, the user gazed at the target to be selected on the computer screen. Then, if the cursor was not inside the object, the user moved the cursor by using facial movements and continued gazing at the target at the same time. Left and right jaw clench made the cursor move left and right, respectively. Eyebrows up and eyebrows down movements made the cursor move up and down. Finally, the user selected the target by clenching the whole jaw. Chin et al. [2008] conducted two experiments with their technique. In their experiment, the new technique was compared with the gaze only technique (dwell time) and the mouse technique. Three pointing distances (286, 578 and 778 pixels) and three target sizes (48, 66 and 96 pixels) were used. The overall mean task time was 4.7 seconds for the new technique, 3.1 seconds for the dwell time technique and 1 second for the mouse technique. However, the error rate was lower with the new technique than with the dwell time technique. The error percentage was 0.14% for the new technique, 3.98% for the dwell time technique and 0.01% for the mouse.

In comparison to the gaze-based interaction techniques, no text entry experiments have been conducted with a multimodal combination of gaze interaction technique and facial activity interaction technique. However, as the pointing and selecting results of both techniques show such a promise, the techniques could be used for entering text as well.

5. Face Interface

5.1 About Face Interface

Face Interface is an eye-glass like wearable device that combines a video-based wearable eye tracker and a capacitive facial movement detector in one device [Rantanen et al., 2010; Tuisku et al., 2011].

Face Interface consists of two units: the head-mounted unit and the wireless unit (see Figure 5.1). The head-mounted unit was built on the frames of protective glasses. It includes two cameras: one for imaging the eye and the other for imaging the computer screen. There is an infrared light emitting diode for illuminating the eye and to provide the corneal reflection. The capacitive sensors were placed in front of both eyebrows and cheeks, and one was placed in front of the forehead. The used cameras were commercial, low-cost cameras. The eye camera was a greyscale camera with a resolution of 352 × 288 pixels and it was modified to image infrared wavelengths. The scene camera was a color camera with a resolution of 597 × 537 pixels. The eye camera was placed near the user's left eye and the IR light source was placed right next to it. The scene camera was placed in front of the user's forehead. [Rantanen et al., 2012; Tuisku et al., 2013].

The wireless unit includes 4 AA batteries, power supply, and two wireless transmitters for the video signals. The computer used with the prototype needs two video receivers and two frame grabbers for the video signals [Rantanen et al., 2012].

Figure 5.1 Face Interface. A person wearing the head-mounted unit on the left. The wireless unit on the right.

Estimating the eye orientation is done as follows. To find the locations of the pupil and the corneal reflection for the estimation, a feature-based approach similar to the one described in [Rantanen et al. 2011] is used. First the image is preprocessed, then the pupil is searched from the darkest pixels of the image, and the corneal reflection from the lightest. From all possible candidates two candidates are identified, and the eye orientation is considered to be the 2D vector between their centers [Rantanen et al., 2011].

Estimating the head orientation is done by detecting the computer screen from the scene camera image. The screen detection finds the frames of a dark computer display from the scene camera image. The screen is usually brightly illuminated and lighter than the surroundings. Contrasts between the illumination of the screen and the surroundings, and between the dark screen edge and the background enable the detection. Both the screen and the screen frame are typically rectangular, which means that they have four corners. The corners of the outer border of the screen frame are close to the corners of the screen. The translation and rotation of the scene camera are calculated based on the found screen [Rantanen et al., 2012].

The capacitive detecting of facial movements is the one described in Rantanen et al. [2011] (see Chapter 3).

5.2 Previous studies

The first version of Face Interface (see Figure 5.2) had a wired commercial USB web camera for imaging the eye. The capacitive sensor was used to detect the facial movement of *corrugator supercilii* facial muscle [Tuisku et al., 2011].

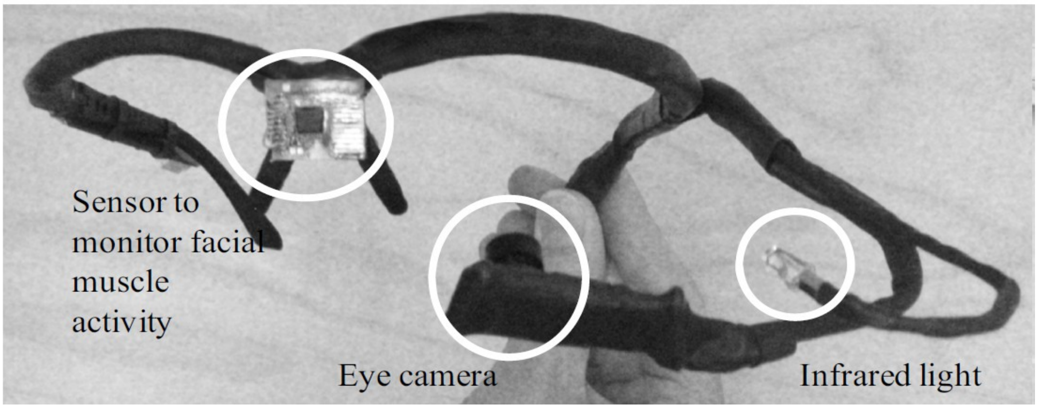


Figure 5.2 The first prototype of Face Interface [Tuisku et al., 2011].

Tuisku et al. [2011] tested the functionality of the first version of Face Interface with pointing and selecting tasks. The prototype had no compensation for head movements, but a chin rest was used for preventing involuntary head movements. Tuisku et al. conducted two experiments. In the first experiment, the participants used frowning as the selection technique. The experimental task was the same as Surakka et al. [2004; 2005] had used earlier (see Chapter 4.2 and Figure 4.1). Also, the target sizes, pointing distances and angles were the same as for Surakka et al. [2004; 2005]. At the end of the experiment, the participants rated the method with six nine-point bipolar scales. The scales varied from -4 (e.g. bad experience) to +4 (e.g. good experience). 0 represented the neutral value (e.g., not bad or good). The scales were: general evaluation, difficulty, speed, accuracy, enjoyableness and efficiency. The results showed that the overall mean pointing task time was 2.5 seconds. The error percentage was 28.5%. The subjective ratings showed that the participants liked using Face Interface and rated it as fast and accurate to use [Tuisku et al., 2011].

In the second experiment the wearable eye tracker was substituted by a commercial desktop eye tracker, in order to find out how the technique would function with a high quality eye tracker. The task and the procedure were identical to the first experiment. The results showed that the overall mean pointing task time was 1.1 seconds. The error percentage was 9.6%. The subjective ratings were somewhat lower than after the first experiment [Tuisku et al., 2011].

In the second version of Face Interface (see Figure 5.3), a scene camera was added to compensate the head movements. The device was also made wireless. The capacitive sensor was detecting both the movement of *corrugator supercilii* facial muscle and *frontalis* facial muscle (i.e., the muscle activated when the raising the eyebrows) [Tuisku et al., 2012].

Figure 5.3 The second prototype of Face Interface [Tuisku et al., 2012].

Tuisku et al. [2012] tested the second version of Face Interface prototype with pointing and selecting tasks. The experimental task was similar to the one conducted with the first version of Face Interface, but seven different distances were used in this experiment (60, 120, 180, 240, 260, 450 and 520 mm). The selecting was done either by frowning or raising the eyebrows. Twenty participants took part in the experiment, and half of them performed the selection by frowning and the other half by raising. At the end of the experiment, the participants rated the method with ten nine-point bipolar scales. The scales varied from 4 (e.g., bad experience) to +4 (e.g., good experience), and 0 represented the neutral value (e.g., not bad nor good). The scales were: general evaluation, difficulty, speed, accuracy, enjoyableness, efficiency, usefulness, naturalness, amusement, and interestingness. The results showed that at the pointing distances from 60 to 260 mm the overall mean pointing task time was 2.4 seconds for the frowning technique and 1.6 seconds for the raising technique. The error percentage was 28.9% for the frowning technique and 25.7% for the raising technique. The subjective ratings showed that the two selection techniques did not differ statistically significantly from each other. The ratings of the technique were not depended on the selection technique. The participants gave positive ratings of the prototype on the scales of general usability, usefulness, naturalness, entertainment and interestingness. Based on the results, Tuisku et al. [2012] stated that Face Interface could be used in more advanced tasks, for example in writing with an on-screen keyboard [Tuisku et al., 2012].

The next step in the research of Face Interface was to investigate the functionality of the prototype for entering text. In earlier studies of Surakka et al. [2005] smiling proved to be a faster selection method than frowning. Thus, smiling was chosen as the selection technique for the experiment. The aim of the experiment was to find the most optimal keyboard layout and to compare writing with Face Interface to writing with the mouse [Tuisku et al., 2013].

The aim of the first experiment was to compare three on-screen keyboard layouts that differed from the traditional QWERTY layout. The order of the characters on the layouts was different than on QWERTY. Also, all three layouts were designed so that the sizes of the keys were larger in the edges of the keyboard than in the middle of the keyboard (see Figure 5.4). This design solution was due to the results of the previous study [Tuisku et al., 2012] that showed that the target selection was more accurate in the middle of the computer screen than in the edges. The designed layouts were tested to see which of the layouts would be the most promising to be used for entering text with Face Interface. The task of the participant was to write one word “aurinko” (i.e., “sun” in English) ten times with each of the three keyboard layouts. “Aurinko” is also a quite long word and each character of it appears only once in the word. The writing was done by pointing the gaze at the character and selecting it by smiling. The places of the characters were randomized each time a word was entered. Each participant wrote the word "aurinko" ten times with each layout. The results showed that the overall mean text entry rate was 14.5 cpm for Layout 1, 14.9 cpm for Layout 2, and 16.2 cpm for Layout 3. The overall mean MSD error rate was 0.09 for Layout 1, 0.27 for Layout 2, and 0.12 for Layout 3. The overall mean for KSPC was 1.3 for Layout 1, 1.18 for Layout 2, and 1.22 for Layout 3. Subjective ratings showed that the clearest and the most functional and enjoyable layout was the one with large keys on the edge and small keys near the center of the keyboard [Tuisku et al., 2013].

Figure 5.4 The keyboard layouts evaluated in the first experiment. Layout 1 on top, Layout 2 in the middle, and Layout 3 on the bottom [Tuisku et al., 2013].

In the second experiment aim was to compare entering text with Face Interface to entering text with computer mouse [Tuisku et al., 2013]. The keyboard layout that proved to be the most functional in the first experiment was used in this experiment (i.e., Layout 2 in Figure 5.4). The experimental task was the same as in the first experiment. Participants performed the task ten times, then there was a short pause and they performed the task ten times again. After performing the task with one pointing device he or she rated the experience with six nine-point bipolar scales. The scales varied from -4 (e.g., bad experience) to +4 (e.g., good experience), and 0 represented a neutral experience (e.g., not slow nor fast). The scales were general evaluation, difficulty, speed, accuracy, enjoyableness and efficiency. Then the same procedure was repeated with the other pointing device. The results showed that the overall mean text entry rate was 19.4 cpm for Face Interface and 27.1 cpm for the mouse. The overall mean MSD error rate was 0.12 for Face Interface and 0.0 for the mouse. The overall mean KSPC value was 1.1 for Face Interface and 1.0 for the mouse. The participants rated the use of mouse more accurate, faster and easier than the use of Face Interface. The ratings did not reveal significant differences between the two techniques. Tuisku et al. [2013] stated that even though writing with the mouse was faster than writing with Face Interface, difference in text entry speed was relatively small considering that Face Interface was a new input method for the participants [Tuisku et al., 2013].

6. Methods

Based on the experiment of Tuisku et al [2013], Face Interface proved to be functional for entering text. The aim of this experiment was to investigate writing with Face Interface. The method of the experiment followed the methods of the longitudinal studies by Tuisku et al. [2008] and Majaranta et al. [2009]. The experiment was conducted in the laboratory premises of Tampere Unit for Computer-Human Interaction (TAUCHI) at the University of Tampere in spring 2013.

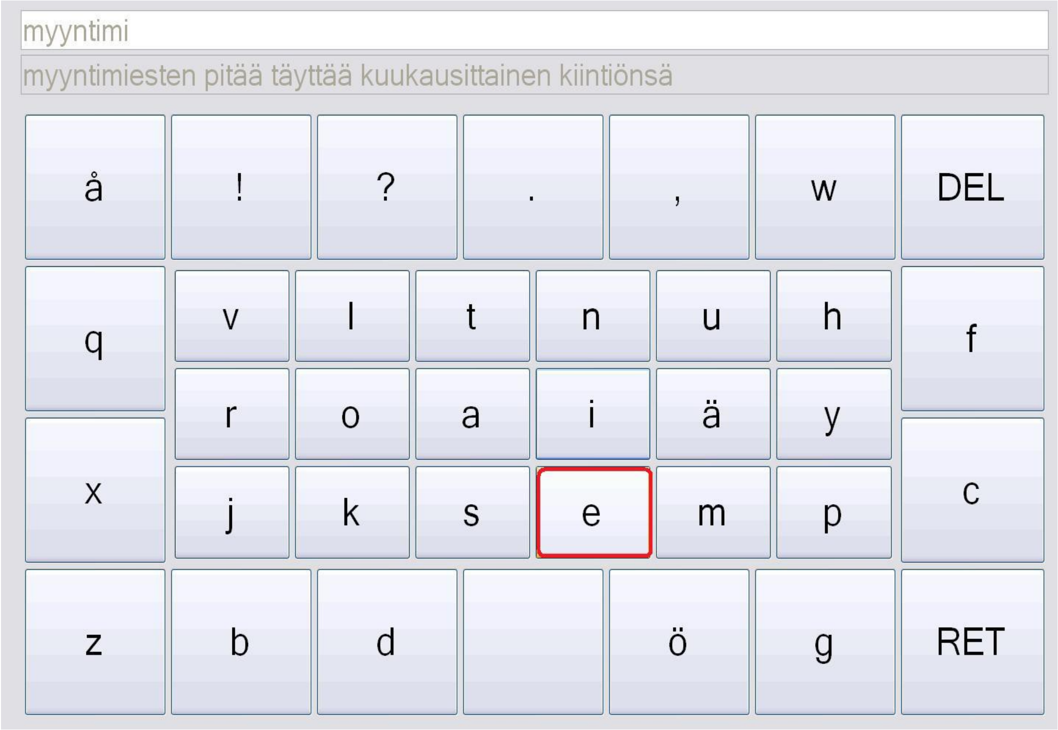
6.1 Participants

Twelve voluntary participants (2 male, 10 female) took part in the experiment. Their mean age was 27 (range 19-37). All participants were native Finnish speakers. All reported to have normal or corrected-to-normal (i.e., with contact lenses) vision. All participants were novices with eye tracking and facial activity detection methods, and none of them had attended any text entry experiments before. Each participant attended ten 15-minutes long experiment sessions. He or she was rewarded with four movie tickets after the last session.

6.2 Apparatus

Face Interface was used in the experiment (see Figure 5.1). Display was 24″ widescreen display, and the viewing distance was approximately 60 cm. The software for online processing of data was Microsoft Visual C++ 2008 [Rantanen et al., 2011].

The keyboard layout used in the experiment is seen in Figure 6.1. The most frequently used Finnish letters were placed in the three middle rows of the keyboard layout. The less frequently used characters were placed on the edges of the keyboard. Delete (DEL) and enter (RET) keys were placed on the most distant corners (upper right and lower right) of the layout. The punctuation marks were arranged on the top row, and the space character was in the middle of the lowest row.



myyntimi

myyntimiesten pitää täyttää kuukausittainen kiintiönsä

Figure 6.1 The keyboard layout used in the experiment. The pointed character is highlighted, and the entered characters appear in the text box above the phrase to be written. For illustration purposes, the pointed character has been highlighted stronger than in the actual keyboard.

6.3 Experimental task

The task in one experiment session was to write pre-defined phrases for 15 minutes. The phrases were from a 500 phrase set of Mackenzie and Soukoreff [2003], and they were Finnish translations by Isokoski and Linden [2004]. The phrases were easy to remember, everyday sentences. All phrases were written in lower case letters, and no punctuation or other special characters were used.

Participants entered characters by pointing at the characters by gaze and selected them by smiling. When the participant's gaze was on target, the character was highlighted. After selecting the character, a click sound was played and the character appeared in the input field above the phrase to be written (see Figure 6.1). When the participant had written the phrase, he or she selected the enter key. The written phrase disappeared and a new phrase appeared. The participants were instructed to write the phrases as fast and as accurate as possible. If the participants detected an error immediately while entering text, they were instructed to correct it with the delete key. If the participants had written further when noticing an error in the text, they were instructed to continue writing.

The text entry software closed after fifteen minutes of writing. If a phrase was unfinished after the fifteen minutes had passed, the text entry software closed after the participant had finished the phrase.

6.4 Procedure

When the participant arrived in the laboratory, he or she was introduced the laboratory premises and the equipment. Then the experiment and its purposes were described. The participant was told that the purpose of the experiment was to study writing text with Face Interface. The participant was asked to sign a consent form and to fill in a background information form.

Next, the prototype was introduced to participant. The eye camera, scene camera and sensors were introduced. Then the participant wore the prototype and he or she was shown live videos from the eye camera and the scene camera. The participant was shown the pupil of his or her eye. He or she was instructed to try different head orientations to see how large head movements were possible to keep the computer screen still visible in the scene camera image. After this, the participant was instructed to try and produce clicks by smiling. After a few successful clicks, the eye tracker was calibrated.

Before the actual task there was a practice task so that the participant could practice the pointing and selecting with Face Interface. The task was similar to ones used in earlier studies [Surakka et al., 2004; 2005; Tuisku et al., 2011; 2012].

After the practice task, the actual experiment task was introduced to the participant. He or she was shown the keyboard and explained, how pointing and selecting were done. Next, the participant was asked, whether he or she had any questions. If not, the eye tracker was calibrated and the actual experiment task started. This procedure was conducted only in the first session. In the sessions 2-10, the participant wore the prototype immediately after arriving in the laboratory. After the calibration the writing started.

If there was a need for a re-calibration of the eye tracker during the writing, the text entry software was paused during calibration. If the re-calibration had to be performed in the middle of a phrase, that phrase was left out from data analysis. In the end of the first, fifth and last sessions the participants filled in a rating form. After the last session, the participants were interviewed briefly.

6.5 Data analysis

The analysis of a phrase started from the first character and ended to the enter character. For the analysis metrics, text entry and error rates were measured. Text entry rate was measured in wpm, where one word is defined as five characters, including spaces and punctuation marks. However, wpm does not take into account errors (i.e., whether the written word is correct or not), only the end result [Mackenzie and Soukoreff, 2003].

Error rates were measured in two different metrics: MSD and KSPC error rates. MSD error rate was calculated by comparing the text that was written by the participant with the presented text, using minimum string distance. MSD indicates how well the written text matches with the presented text. However, MSD does not take into account, how many corrections the participant has made during producing the text, only the end result. Whereas MSD error rate only compares the transcribed text to the presented text, KSPC value takes into account the errors produced. KSPC indicates how often the participant has cancelled characters during writing. When KSPC is 1.00, every key selection has produced a correct character. If a participant makes a correction during writing (i.e., presses delete key and selects another character), the value of KSPC is more than one. Thus, KSPC measures the accuracy of the text input [Mackenzie and Soukoreff, 2003].

One-way repeated measures analysis of variance (ANOVA) was performed for session as within-subject factor. Bonferroni corrected t-tests were used for post hoc pairwise comparisons.

At the end of the first, fifth and the last session, the participants rated the method with eleven seven-point Likert scales [ISO 9241-9], (see Appendix 1). The scales were: overall operation of input device (i.e., from very easy (to use) to very difficult (to use)), smoothness (i.e., from very rough to very smooth), operation speed (i.e., from unacceptable to acceptable, mental effort required for operation (i.e., very high to very low), physical effort required for operation (i.e., very high to very low), accuracy (i.e., from very inaccurate to very accurate), target selection (i.e., from very uncomfortable to very comfortable), general comfort (i.e., from very uncomfortable to very comfortable), eye fatigue (i.e., very high to none), facial muscle fatigue (i.e., very high to none) and neck fatigue (i.e., very high to none).

7. Results

7.1 Text entry rates

The text entry rates of all sessions are presented in Figure 7.1. The overall mean text entry rate ± standard error of the means (S.E.M.s.) of all sessions was 5.39 ± 0.65 wpm. The overall mean text entry rate for the first session ± S.E.M was 3.88 ± 0.39 wpm and in the last session 6.59 ± 0.37 wpm. One-way repeated measures ANOVA showed a statistically significant effect of the session F(9. 99) = 5.935, $p < 0.001$. Post hoc pairwise comparisons for the session were not statistically significant.

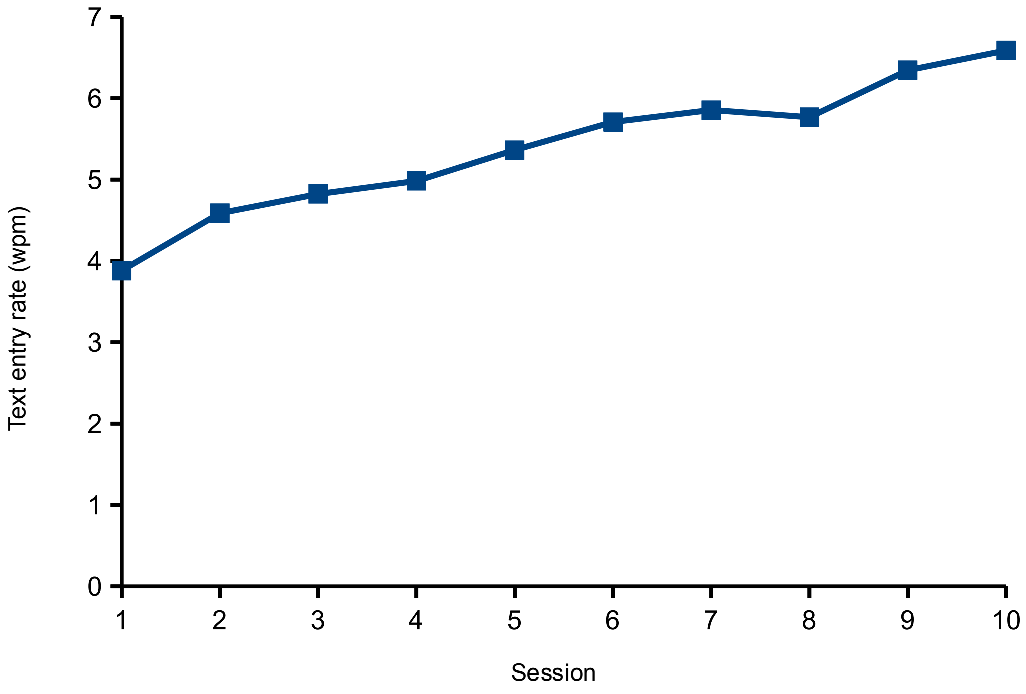


Figure 7.1 Overall mean of the text entry rate (wpm) by session.

Figure 7.2 presents the text entry rates for all participants. The highest text entry rate was 8.38 wpm, reached in session 10.

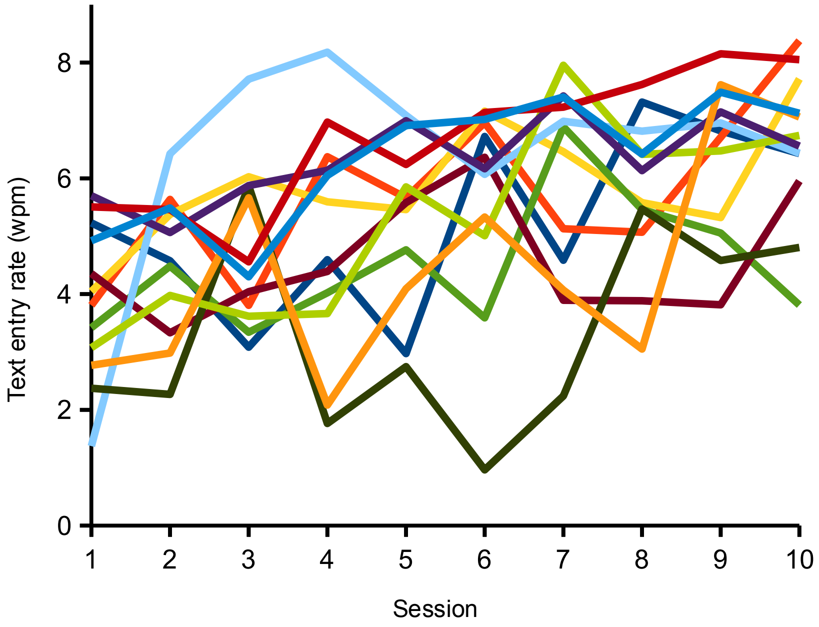


Figure 7.2 Text entry rates (wpm) of all participants divided by session.

Figure 7.3 presents individual maximum text entry rates of all participants.

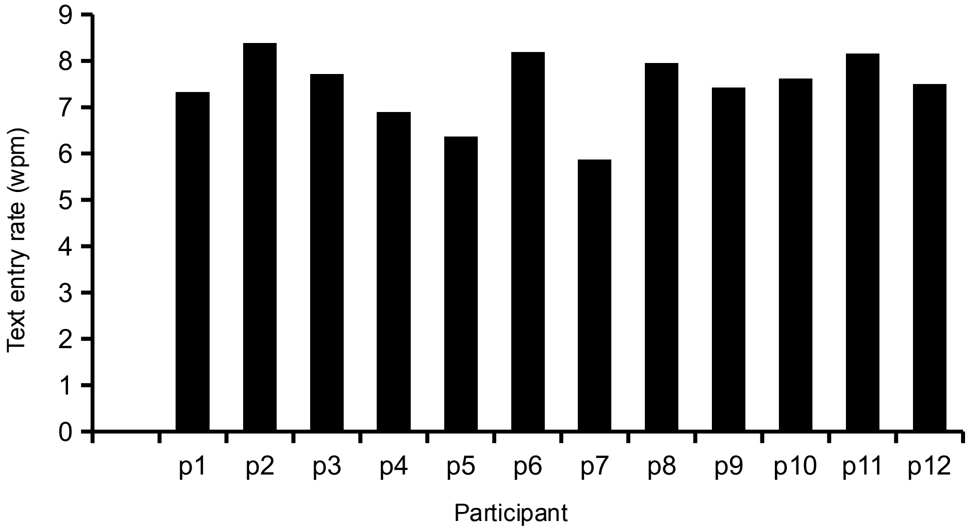


Figure 7.3 The maximum text entry rates (wpm) of all participants.

7.2 Error rates

MSD error rates of all sessions are presented in Figure 7.4. The overall mean MSD error rate ± S.E.M was 0.25 ± 1.24. In the first session the overall mean MSD error rate ± S.E.M was 0.50 ± 0.31 and in the last session 0.05 ± 0.03. One-way repeated measures ANOVA for the session was not statistically significant.

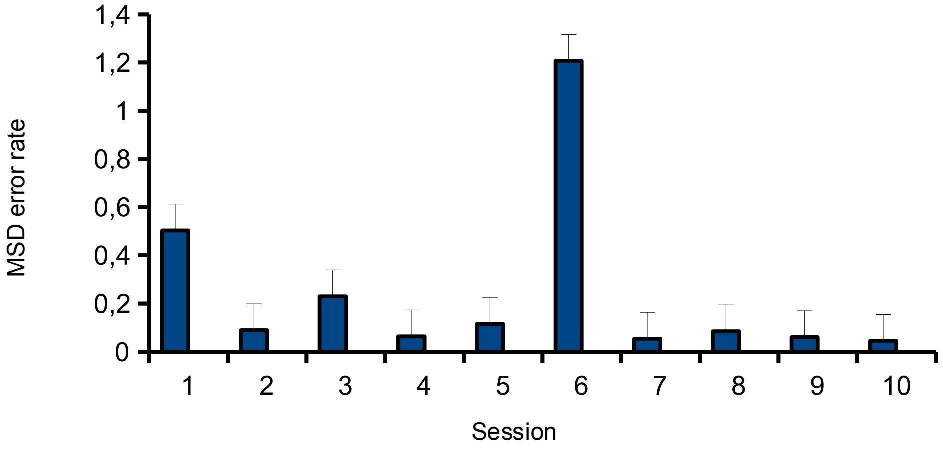


Figure 7.4 The overall mean of the MSD error rate by session.

Figure 7.5 presents the MSD error rates of all participants. The lowest individual mean MSD error rate of all sessions was 0.006.

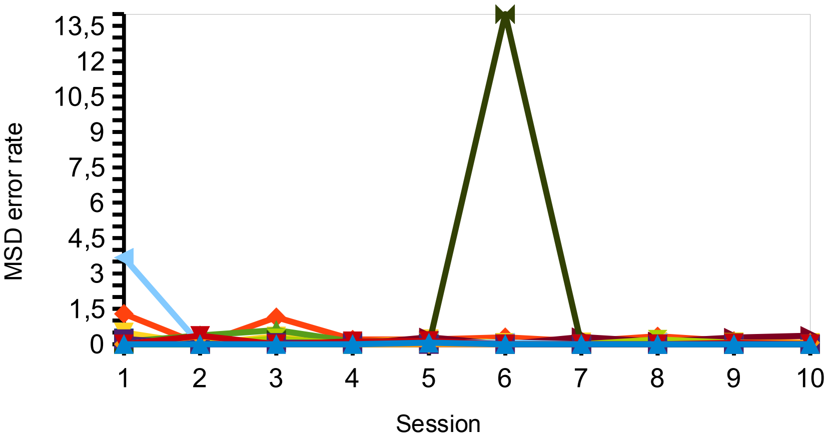


Figure 7.5 The error rates (MSD) of all participants divided by session.

The overall mean KSPC values of all sessions are presented in Figure 7.6. The overall mean KSPC rate ± S.E.M was 1.18 ± 0.16. In the first session the overall mean for KSPC ± S.E.M was 1.26 ± 0.16 and in the tenth session 1.2 ± 0.35. The lowest overall mean for KSPC was reached on session 9 and it was 1.12. One-way repeated measures ANOVA for the session was not statistically significant.

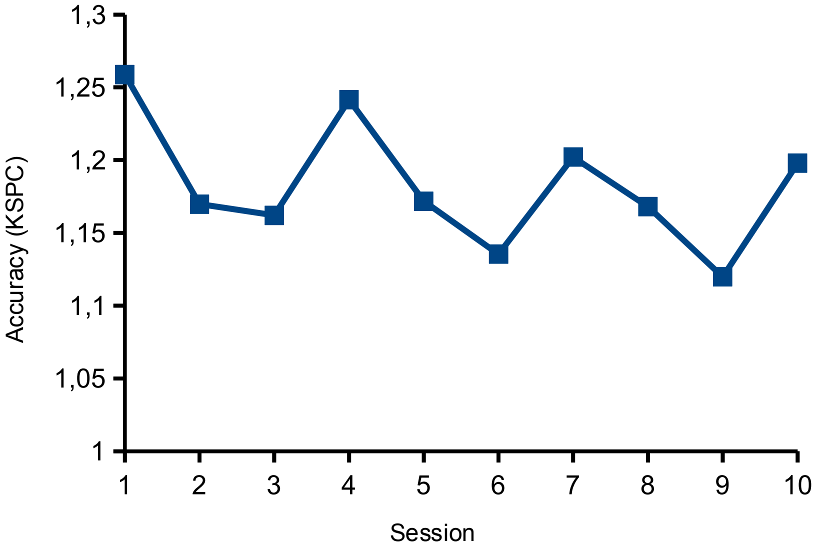


Figure 7.6 The overall mean of KSPC values.

Figure 7.7 presents the KSPC values of all participants. The lowest individual mean KSPC value of all sessions was 1.1.

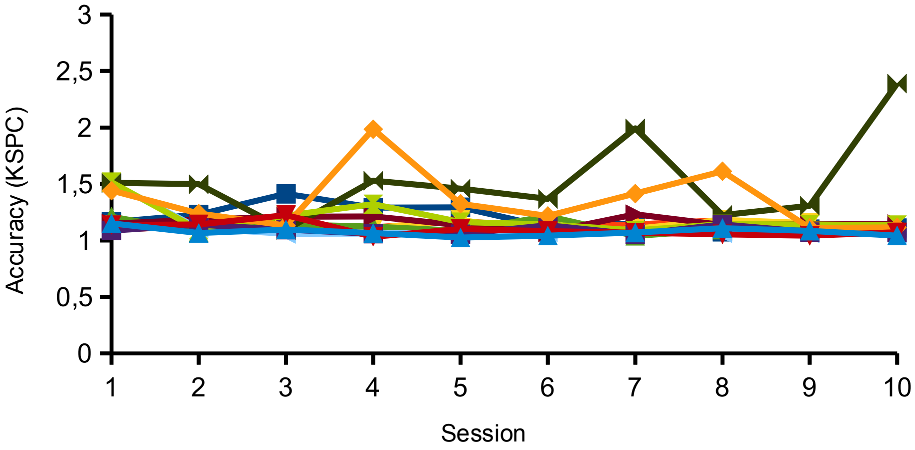


Figure 7.7 The KSCP values of all participants divided by session.

7.3 Subjective ratings

Figure 7.8 presents the overall means of the subjective ratings after the first, the fifth and the tenth session.

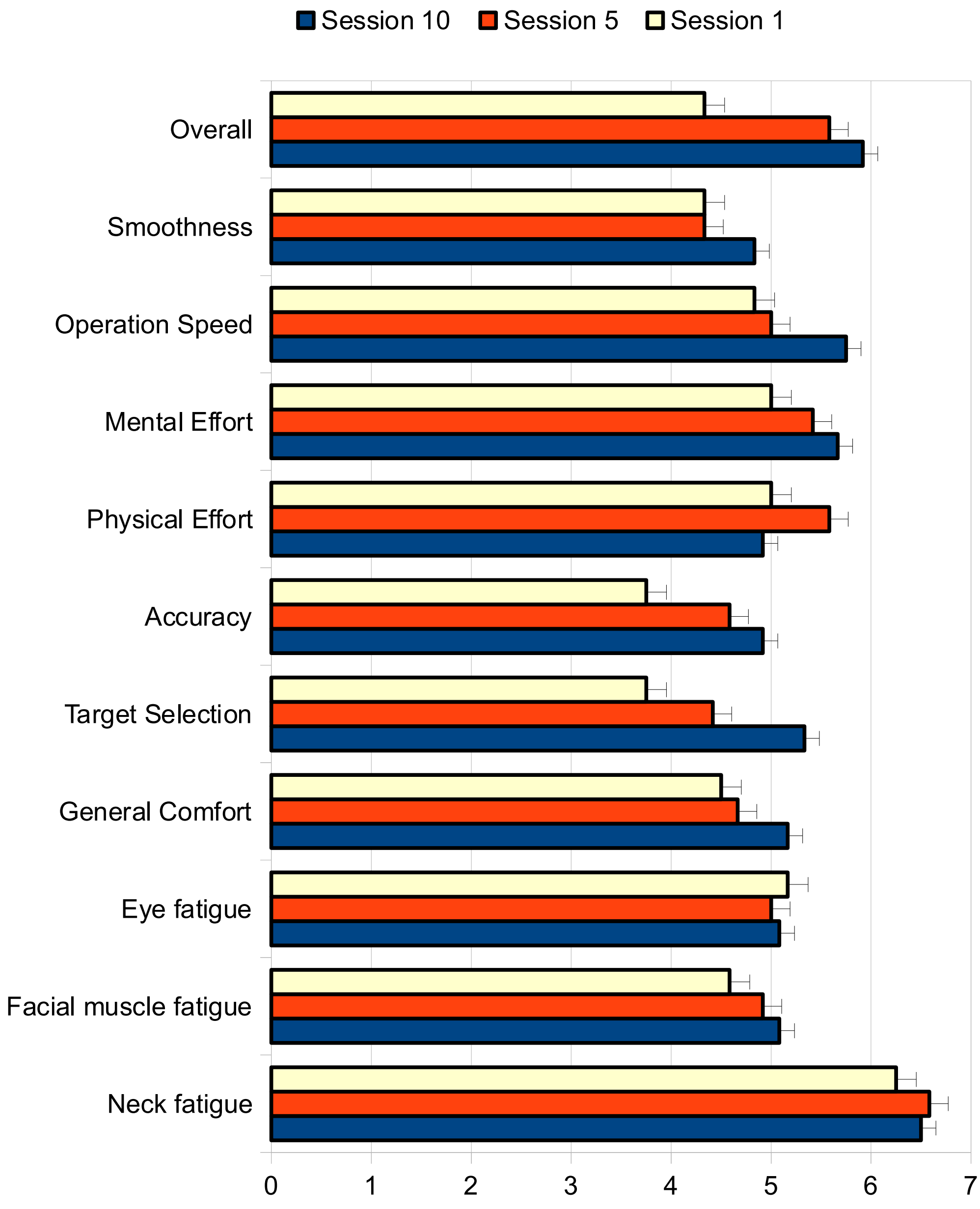


Figure 7.8 Subjective ratings.

7.4 Interview

The participants were interviewed in the end of the last session. The interview questions are in the Appendix 2.

The participants were first asked what was their general feeling about writing with Face Interface. Nine participants had a positive general impression about Face Interface. All participants mentioned something positive about Face Interface in general: easy and fast to use, easy to learn, useful, fun, different, "*this might be a big thing someday*". Few participants told that they would probably not use Face Interface themselves, but could imagine someone else using it for writing. One participant mentioned that the headset prototype was not that convenient to be worn and that Face Interface requires a lot of preparations (calibration etc.) in order to use it.

Almost everybody thought that combining gazing and smiling was easy to learn and that the combination did not stress too much, since two different methods were used and the smiling movement was so small. Two participants felt that writing with this technique soon became very automatic: "*you did not need to think about it at all*". One person thought that smiling was stressful and suggested that for example voice commands could be used for selecting targets instead of facial activities. Also, one participant mentioned that this method is not suitable for situations where you need to speak.

When the participants were asked to compare their first and last sessions and tell, whether there were differences, almost everybody answered that there was a big difference between the beginning and the end of the study. Ten participants felt that their writing with Face Interface had improved during the study. The participants also listed several issues that they thought had improved: writing speed, use of Face Interface in general, face-eyes interaction, concentration, finding characters, adjusting his or her own position, "learning" calibration.

Almost all participants gave positive feedback, when they were asked about the keyboard layout used in the experiment. Ten participants told that they had learned the positions of the characters on the layout during the study. The participants thought that the placements of the characters were logical or smart, and they could be learned by heart quickly. The sizes of the keys were big enough and the participants thought that it was good to have larger keys on the edges of the layout. Two participants answered that they had not really learned during the study, where the characters were located on the keyboard. One participant was wondering herself, why she never learned the places of the characters, but had to look for them each time again. She thought that she continued looking for QWERTY characters and that is why the learning of the new layout was difficult. One participant did not like the layout at all, but thought that QWERTY should be used instead.

When asking about the possible situations where Face Interface could be used, a lot of interesting ideas came up. Ten participants mentioned that this technique would suit for disabled people, that is, for people that can not use their hands or communicate otherwise (but still have facial movement capabilities). Three participants thought that Face Interface could be used instead of mouse when interacting with the computer. Using computer or playing games came up as suggestions for Face Interface. One participant thought that maybe this technique could be used at the gym. Nowadays entering gym can happen via fingerprint recognition, and maybe Face Interface method could be used there, too. Further suggestions included Face Interface to be used as a remote control device, or in a car, army or under the water. According to the participants, some situations where Face Interface would probably not be suitable could be for example cash machine (due to data security) or teaching situations (not optimal for a bigger group of people).

Finally, when the participants were asked, whether they had any improvement ideas for Face Interface, the following were mentioned: possibility to choose the selection method or the layout used, possibility to store the calibration points, double-clicking the same character, a lighter and more convenient wireless headset, the pre-defined sentences to be placed in the middle of the screen.

8. Discussion

The results showed that the overall mean text entry rate of all sessions was 5.39 wpm. There were no outliers in the study, but all participants reached the overall mean text entry rate of the whole study. The overall mean text entry rate was 3.88 wpm in the first session and 6.59 wpm in the tenth session. Thus, the overall text entry speed of the participants was faster in the end of the study than in the beginning of it.

The method and the procedure of this longitudinal study followed the ones used in longitudinal studies of adjustable dwell time typing [Majaranta et al., 2009] and writing with Dasher [Tuisku et al., 2008]. The text entry rates of adjustable dwell time (19.9 wpm) and Dasher (17.3 wpm) are not directly comparable since they used a gaze only technique whereas this study used a multimodal technique.

The text entry rate of the first session is comparable with the results of the first writing study of Face Interface, where the overall mean text entry rate was 4 wpm (19.4 cpm) [Tuisku et al., 2013]. That study included only one session. The overall mean text entry rate of the last session of this study (6.59 wpm) is comparable with the study of Majaranta [2006]. Majaranta [2006] studied the effect of visual and auditory feedback during eye typing, and the results showed an overall mean text entry rate of 6.97 wpm. This study included both visual and auditory feedback, too. The pointed character was highlighted and a successful selection of the character produced a click sound.

The characters of the keyboard layout used in this study were grouped together based on the character frequency, as was in the study of Špakov and Majaranta [2009]. They received text entry rate of 12.18 wpm for the two-row keyboard and 8.86 wpm for the one-row keyboard. The fastest writers of this study reached 8 wpm. The keyboard layout in this study not being the traditional QWERTY layout required special concentration on writing in the beginning and made the writing slower. However, most participants learned the layout fast and remembered by heart, where the characters were located. This refers to the same conclusion to which Tuisku et al. [2011] came in their study. They stated that the accuracy is better in the middle of the screen and that the targets need to be big enough in order to select them. Based on the results of Tuisku et al. [2011], the traditional QWERTY layout is probably not the optimal keyboard layout for this kind of interaction method, even though it is familiar to everybody. The fast learning of the keyboard layout in this study would support those earlier statements.

The overall mean MSD error rate of this study was 0.25. The overall mean MSD error rate was 0.50 in the first session and 0.05 in the last session. When compared to other text entry methods, the MSD error rates of this study are very low. For example, the overall mean MSD error rate for Dasher was 10.72 in the first session and 0.57 in the last session [Tuisku et al., 2008]. The overall mean MSD error rate of adjustable dwell time study was 1.28 in the first session and 0.36 in the last session [Majaranta et al., 2009]. Both in their study and this study, the MSD error rate decreased even though the writing speed increased. That means that the participants were very accurate and produced only few errors in the text output. The participants in this study produced error-free text, that is, they concentrated on performing the writing without errors and corrected actively the errors they made during writing. The corrections may have an effect on the text entry speed. Ten out of twelve participants were women, and that might have had effect on the writing style, since women might be more accurate than men. In general, the speed was probably not the main issue for all participants but avoiding errors instead. The MSD error rate also proved to be lower in the end of the experiment, which would indicate that learning the text entry system and the keyboard layout prevented making errors or enabled them to be corrected quickly. The accuracy on calibration also improved towards the last sessions, and that prevented erroneous selections. One participant had one unusual difficult session, where she made many errors and left them uncorrected (see Figure 7.4).

The KSPC values reveal that the difference between the first session and the last session is not that big: 1.26 in the first and 1.2 in the last session. For the adjustable dwell time study [Majaranta et al., 2009], the results were similar. There the overall mean KSPC value was 1.09 in the first session and 1.18 in the last session [Majaranta et al. 2009]. For this study, the reason for the result might be that the participants corrected errors actively during writing. In general, the text entry speed was faster in the last sessions, and due to faster speed more errors were probably both produced and corrected.

The subjective ratings of the participants show that the ratings were positive in almost all categories. The overall mean of the subjective rating was lower than middle value (4) only in target selection and accuracy in the first session rating: both got 3.57 points out of 7. That is probably due to the new interaction method used for the first time: both pointing with gaze and selecting with facial muscles was new to all participants. Neck fatigue was the only scale receiving overall mean rating higher than 6, and that every time when rated. Ratings showed that the facial muscles and eyes got somewhat more tired than neck when using Face Interface, but also their ratings stayed on positive side all the time. An interesting point is that the most tiring session for eyes was rated to be the fifth session, not the last session. On the other hand, the subjective ratings showed that the last session had required more physical effort than the first or fifth session. This might be due to longer use of Face Interface in total, that is, the participants may have thought about the whole experimental process when rating the last session.

The ratings became more positive during the study in almost all categories. The most improved ratings during the experiment were the following: general comfort, target selection, operation speed and overall operation of the device. These could indicate that when the participants learned to use Face Interface and write faster with it, the general impression of the device also got more positive. The rating of the overall operation was 5.9 in the last session, which means that the participants rated Face Interface easy to use.

Some participants faced technical problems during the experiment, which naturally affected the text entry. Calibration was not always accurate enough for some parts of the layout or some characters, which caused slowness in pointing. It might happen that the participant moved herself or himself during the session, and the accuracy decreased. Also lifting or moving the headset caused inaccuracy. Problems with face detection occurred for some participants: characters could not be chosen right away or they were chosen accidentally. That was probably due to the strength of the signal received from the sensors: sometimes it was weaker (i.e. the selection could not be done right away when performing the smile movement) and sometimes very strong (i.e. a very small facial movement of the participant caused the selection even though not intended).

Because the pre-defined phrases to be written were displayed on top left corner of the screen, it required the gaze shifting there every now and then, if the participant did not memorize the phrase in the beginning. Many participants thought that it would be easier to have the phrase presented, for example, in the middle of the screen. In that way the participants would not need to move their gaze and possibly also head. This could be something that might be improved in the Face Interface system. Many participants mentioned that a double-clicking possibility of characters could had enabled faster text entry. Now the participant had to wait for a while until selecting the same character again. In addition to double-clicking, some other possibilities to adjust the operation speed of the text entry program might improve the text entry speed speed.

The learning curve was still rising when the study was finished (as can be seen in Figure 7.1). In general, this could indicate that the participants could had reached even higher text entry speed if the experiment had been longer. Some participants reached their maximum speed of this study early and stayed approximately on the same level until the end of the study. As the learning curves of the participants were still rising in the end of the study, it would be interesting to see, how high they could rise. One development idea could be conducting a study that would last longer and include more writing than the 2.5 hours that this study included. On the other hand, if could be feasible to study people writing free text with Face Interface, that is, the users could choose what to write instead of pre-defined sentences. That way they could concentrate on writing, not needing to memorize a phrase or shift gaze back to the phrase when not remembering it by heart.

The participants of this study were quite young, and most of them were students. It would be interesting to have a wider group of participants in the study, for example, some older people or disabled people who could use Face Interface as a communication method. The experimental method could be developed so that the subjective ratings of the participants would be gathered after each session, not only after the first, the fifth and the tenth session. That way the ratings would be gathered in-line with the other results that are gathered after every session (text entry speed etc.). Some shorter interview could also be done during the study, not only in the end of it. That way more subjective opinions and development ideas could be gathered, and they would be gathered real time.

For further studies, Face Interface could also be utilized for some different purpose than entering text, for example, for browsing the Internet or making electronic home work for school or studies.

9. Conclusion

In this thesis a longitudinal study for entering text with Face Interface was presented. The aim of the study was to investigate entering text with Face Interface in a longitudinal study. Twelve voluntary participants took part in an experiment that consisted of ten 15-minutes long sessions. In each session, the participants wrote with an on-screen layout so that they pointed at the characters by gaze and selected them by smiling. The results showed that the overall mean text entry rate of all sessions was 5.39 wpm. In the first session the overall mean text entry rate was 3.88 wpm and in the tenth session 6.59 wpm. The overall mean MSD error rate of all sessions was 0.25, and the overall mean KSPC value of all sessions was 1.18. In the first session the overall mean MSD error rate was 0.50 and in the last session 0.05. In the first session the overall mean KSPC value was 1.26 and in the last session 1.2. Subjective ratings showed that Face Interface was easy to use and that the ratings got more positive along the study. The results showed that Face Interface can be used for writing within a longer period of time. In the future, Face Interface could also be utilized in some other area than entering text, for example, browsing the Internet or for electronic study material (e.g. by pointing at the text and selecting links on it in order to find more information or to perform a task).

References

[Barreto et al., 2000] A.B. Barreto, S.D. Scargle, and M. Adjouadi, A practical EMG-based human-computer interface for users with motor disabilities, *Journal of Rehabilitation Research and Development* **37**(1), 2000, 53-64.

[Chin et al., 2008] C.A. Chin, A. Barreto, J.G. Cremades, and M. Adjouadi, Integrated electromyogram and eye-gaze tracking cursor control system for computer users with motor disabilities, *Journal of Rehabilitation Research and Development*, **45**(1), 2008, 161-174.

[Goni et al., 2004] S. Goni, J. Echeto, A. Villanueva, R. Cabeza, Robust algorithm for pupil-glint vector detection in a video-oculography eyetracking system. In: *Proc. of the 17th International Conference on Pattern Recognition (ICPR '04)*, **4,** 2004, 941-944.

[Hansen et al., 2001] J.P. Hansen, D.W. Hansen and A.S. Johansen, Bringing gaze based interaction back to basics. In: *C. Stephanidis (Ed.)*, Universal Access in HCI, Lawrence Erlbaum Associates, 2001, 325-328.

[Hansen et al., 2004] J.P. Hansen, K. Tørning, A.S. Johansen, K. Itoh, and H. Aoki, Gaze typing compared with input by head and hand, In: *Proc. of the Symposium on Eye Tracking Research and Applications (ETRA '04)*, 2004, 131-138.

[Hansen and Majaranta, 2012] D.W. Hansen and P. Majaranta, Basics of Camera-Based Gaze Tracking. In: Majaranta et al. (Ed.), *Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies,* IGI Global, 2012, 21-26.

[ISO 9241-9] Ergonomic requirements for office work with visual display terminals (VDTs). *Requirements for non-keyboard input devices,* ***9****, Suomen Standardisoimisliitto SFS, 2000.*

[Isokoski, 2000] P. Isokoski, Text input methods for eye trackers using off-screen targets. In: *Proc. of Eye Tracking Research & Applications (ETRA'02)*, 2000, 15-21.

[Isokoski and Linden, 2004] P. Isokoski and T. Linden, Effect of foreign language on text transcription performance: Finns writing English. In: *Proc. of NordiCHI* 2004, 105-108.

[Jacob, 1991] R.J.K. Jacob, The use of eye movements in human-computer interaction techniques: What you look at is what you get. In: *ACM Transactions on Information Systems,* **9**(2), 1991, 152-169.

[Jacob, 1995] Jacob, R.J.K., Eye Tracking in Advanced Interface Design, In: *W. Barfield and T.A. Furness (eds.),* Virtual Environments, 1995.

[Just and Carpenter, 1976] M.A. Just and P.A. Carpenter, Eye fixations and cognitive processes. In: *Cognitive Psychology,* **8**, 1976, 441-480.

[MacKenzie and Soukoreff, 2003] I.S. MacKenzie and R.W. Soukoreff, Phrase sets for evaluating text entry techniques. In: *Proc. of CHI 2003: Extended Abstracts on Human Factors in Computing Systems*, 2003, 754-755.

[MacKenzie and Zhang, 2008] I.S MacKenzi and X. Zhang, Eye typing using word and letter prediction and a fixation algorithm. In: *Proceedings of the Symposium on Eye Tracking Research & Applications* (ETRA'08), 2008, 55-58.

[Majaranta, 2009] P. Majaranta, *Text Entry by Eye Gaze*, University of Tampere, Dissertations in Interactive Technology, 2009.

[Majaranta et al., 2006] P. Majaranta, I.S. MacKenzie, A. Aula and K.-J. Räihä, Effects of feedback and dwell time on eye typing speed and accuracy, *Universal Access in the Information Society,* **5**(2), 2006, 199-208.

[Majaranta et al., 2009] P. Majaranta, U-K. Ahola and O. Špakov, Fast Gaze Typing with an Adjustable Dwell Time. In: *Proc. of CHI*, 2009, 357-360.

[Majaranta and Räihä, 2002] P. Majaranta and K.-J. Räihä, Twenty years of eye typing: Systems and design Issues. In: *Proc. of Eye Tracking Research & Applications (ETRA)*, 2002, 15-22.

[Majaranta and Räihä, 2007] P. Majaranta and K.-J. Räihä, Text entry by gaze: Utilizing eye-tracking. In: I. Scott MacKenzie and Kumiko Tanaka-Ishii (Ed.), *Text entry systems: Mobility, accessibility, universality.* San Francisco: Morgan Kaufmann, 2007, 175-187.

[Partala et al., 2001] t. Partala, A. Aula and V. Surakka, Combined Voluntary Gaze Direction and Facial Muscle Activity as a New Pointing Technique. In Hirose M. (Ed.) *Proceedings of INTERACT 2001*, Tokyo, Japan, IOS Press, 2001, 100-107.

[Partala and Surakka, 2004] T. Partala and V. Surakka, The effects of affective interventions in human-computer interaction. In: *Interacting with computers* **16** (2), 2004, 295-309.

[Rantanen et al., 2010] V. Rantanen, P. H. Niemenlehto, J. Verho, and J. Lekkala, Capacitive facial movement detection for human-computer interaction to click by frowning and lifting eyebrows, *Medical and Biological Engineering and Computing*, **48**(1), 2010, 39-47.

[Rantanen et al., 2011] V. Rantanen, T. Vanhala, O. Tuisku et al., A wearable, wireless gaze tracker with integrated selection command source for human-computer interaction, *IEEE Transactions on Information Technology in BioMedicine,* **15**(5), 2011, 795-801.

[Rantanen et al., 2012] V. Rantanen, J. Verho, J. Lekkala, O. Tuisku, V. Surakka, and T. Vanhala: The effect of clicking by smiling on the accuracy of head-mounted gaze tracking. In: *Proc. of the Symposium on Eye Tracking Research and Applications (ETRA '12)*, 2012, 345-348.

[San Agustin et al., 2009] J. San Agustin, J. C. Mateo, J. P. Hansen, and A. Villanueva, Evaluation of the potential of gaze input for game interaction, *PsychNology Journal*, **7**(2), 2009, 213-236.

[Sandnes and Aubert, 2007] F.E. Sandnes and A. Aubert, Bimanual Text Entry using Game Controllers: Relying on Users' Spatial Familiarity with QWERTY, *Interacting with Computers*, **19**(2), 2007, 140-150.

[Sibert and Jacob, 2000] L.E. Sibert and R.J.K. Jacob, Evaluation of eye gaze interaction. In: *Proc. of the SIGCHI Conference on Human Factors in Computing Systems (CHI ‘00),* 2000, 281-288.

[Špakov and Majaranta, 2008] O. Špakov and P. Majaranta, Scrollable keyboards for eye typing. In: *Proc. of the 4th Conference on Communication by Gaze Interaction* (COGAIN), 2008, 63-66.

[Špakov and Majaranta, 2009] O. Špakov and P. Majaranta, Scrollable keyboards for casual typing, *PsychNology Journal,* **7**(2), 2009, 159-173.

[Stampe and Reingold, 1995] D.M. Stampe and E.M. Reingold, Selection by looking: A novel computer interface and its application to psychological research. In: J.M. Findlay, R. Walker and R.W. Kentridge (Ed.) *Eye Movement Research: Mechanisms, Processes and Applications*, Amsterdam: Elsevier Science, 1995, 467-478.

[Surakka et al., 2003] V. Surakka, M. Illi, and P. Isokoski, Voluntary eye Movements in Human-Computer Interaction. In: J. Hyönä, R. Radach and H. Deubel (Ed.), *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, 2003, 473-491.

[Surakka et al., 2004] V. Surakka, M. Illi, and P. Isokoski, Gazing and frowning as a new human-computer interaction technique, *ACM Transactions on Applied Perceptions*, **1**(1), 2004, 40-56.

[Surakka et al., 2005] V. Surakka, P. Isokoski, M. Illi, and K. Salminen, Is it better to gaze and frown or gaze and smile when controlling user interfaces? In: *Proc. of the 11th International Conference on Human-Computer Interaction (HCI '05)*, CD-ROM, 2005, 7 pages.

[Surakka and Vanhala, 2011] V. Surakka and T. Vanhala, Emotions in human-computer interaction, In: A. Kappas Arvid and N. Krämer (Ed.) *Face-to-face communication over the Internet: emotions in a web of culture, language and technology: Cambridge University Press*, 2011, 213-236.

[Tobii Technology, 2013] Tobii Technology, homepage: http://www.tobii.com (checked on the 10th of December, 2013).

[Tuisku et al., 2008] O. Tuisku, P. Majaranta, P. Isokoski and K.-J. Räihä, Now Dasher! Dash Away! Longitudinal study of fast text entry by eye gaze. *Proc. of the Symposium on Eye Tracking Research & Applications (ETRA'08)*, 2008, 19-26.

[Tuisku et al., 2011] O. Tuisku, V. Surakka, Y. Gizatdinova et al., Gazing and frowning to computers can be enjoyable. In: *Proc. of the 3rd International Conference on Knowledge and Systems Engineering (KSE '11)*, 2011, 211-218.

[Tuisku et al., 2012] O. Tuisku, V. Surakka, T. Vanhala, V. Rantanen, and J. Lekkala, Wireless face interface: using voluntary gaze direction and facial muscle activations for human-computer interaction, *Interacting with Computers*, **24**(1), 2012, 1-9.

[Tuisku et al. 2013] O. Tuisku, V. Surakka, V. Rantanen, T. Vanhala, J. Lekkala, Text Entry by Gazing and Smiling. In: *Advances in Human-Computer Interaction,*2013.

[Villanueva et al., 2012]. A. Villanueva, R. Cabeza and J. S. Agustin, Gaze Estimation, In: *Gaze interaction and Applications of Eye Tracking: Advances in Assistive Technologies.* Hershey: IGI Global, 2012.

[Ward and McKay, 2002] D.J. Ward and D.J.C. MacKay, Fast hands-free writing by gaze direction. *Nature 414,* 2002, 838.

[Ware and Mikaelian, 1987] C. Ware and H. H. Mikaelian, An evaluation of an eye tracker as a device for computer input. In: *Proc. of CHI,* 1987, 183-188.

[Yamada and Fukuda, 1987] M. Yamada and T. Fukuda. Eye word processor (EWP) and peripheral controller for the ALS patient, *IEEE Proceedings Physical Science, Measurement and Instrumentation, Management and Education,* **134**(4), 328-330.

Appendix 1 Rating form

1. Laitteen yleinen toiminta:
1 .... 2 .... 3 .... 4 .... 5 .... 6...................... 7

Erittäin vaikea (käyttää)	Erittäin helppo (käyttää)

2. Sujuvuus toiminnan aikana:
1 .... 2 .... 3 .... 4 .... 5 .... 6...................... 7

Erittäin karkea	Erittäin sujuva

3. Toimintanopeus oli:
1 .... 2 .... 3 .... 4 .... 5 .... 6...................... 7

Mahdoton hyväksyä	Hyväksyttävä

4. Toimintaan vaadittu henkinen kuormitus:
1 .... 2 .... 3 .... 4 .... 5 .... 6...................... 7

Erittäin suuri	Erittäin pieni

5. Toimintaan vaadittu fyysinen kuormitus:
1 .... 2 .... 3 .... 4 .... 5 .... 6...................... 7

Erittäin suuri	Erittäin pieni

6. Osoittamisen tarkkuus oli:
1 .... 2 .... 3 .... 4 .... 5 .... 6...................... 7

Erittäin epätarkkaa	Erittäin tarkkaa

7. Kohteen valinta oli:
1 .... 2 .... 3 .... 4 .... 5 .... 6...................... 7

Erittäin epämukavaa	Erittäin mukavaa

8. Yleinen mukavuus:
1 .... 2 .... 3 .... 4 .... 5 .... 6...................... 7

Erittäin epämukava	Erittäin mukava

9. Silmien rasittuminen:
1 .... 2 .... 3 .... 4 .... 5 .... 6...................... 7

Erittäin rasittuneet	Ei lainkaan rasittuneet

10. Kasvolihasten rasittuminen:
1 .... 2 .... 3 .... 4 .... 5 .... 6 7

Erittäin rasittuneet	Ei lainkaan rasittuneet

11. Niskan rasittuminen:
1 .... 2 .... 3 .... 4 .... 5 .... 6...................... 7

Erittäin rasittunut	Ei lainkaan rasittunut

Appendix 2 Interview questions

1. Millainen yleisvaikutelma sinulle jäi tästä käyttöliittymästä?
2. Miltä tuntui katseen ja hymyilyn yhdistäminen?
3. Jos vertailet ensimmäistä ja viimeistä kertaa, tuntuiko erilaiselta?
4. Miten kommentoisit näppäimistöä ja sen asettelua? Opitko ne tässä tutkimuksen aikana?
5. Millaisissa tilanteissa voisit käyttää tätä menetelmää (katse-hymyily)?
6. Tuleeko mieleesi parannusehdotuksia tähän menetelmään?