Outi Tuisku

Face Interface

ACADEMIC DISSERTATION
To be presented with the permission of the School of Information Sciences of the University of Tampere, for public discussion in the Pinni auditorium B1100 on May 23rd, 2014, at noon.

School of Information Sciences
University of Tampere

Dissertations in Interactive Technology, Number 16
Tampere 2014

ACADEMIC DISSERTATION IN INTERACTIVE TECHNOLOGY

Supervisor:	Professor Veikko Surakka, Ph.D. School of Information Sciences, University of Tampere, Finland
Opponent:	Professor Jukka Hyönä, Ph.D. Department of Psychology, University of Turku, Finland
Reviewers:	Associate Professor John Paulin Hansen, Ph.D. Innovative Communication Group, IT University of Copenhagen, Denmark Professor Markku Tukiainen, Ph.D. School of Computing, University of Eastern Finland, Finland

Acta Electronica Universitatis Tamperensis 1428
ISBN 978-951-44-9473-4 (pdf)
ISSN 1456-954X
http://tampub.uta.fi

The originality of this thesis has been checked using the Turnitin OriginalityCheck service in accordance with the quality management system of the University of Tampere.

Dissertations in Interactive Technology, Number 16

School of Information Sciences
FIN-33014 University of Tampere
FINLAND

ISBN 978-951-44-9463-5
ISSN 1795-9489

Juvenes Print - Suomen Yliopistopaino Oy
Tampere 2014

Abstract

The aim of the thesis was to iteratively develop and experimentally test a new kind of Face Interface prototype for human-computer interaction (HCI). Face Interface combined the use of two modalities: voluntarily-controlled gaze direction and voluntarily-controlled facial muscle activations for pointing and selecting objects in a graphical user interface (GUI), respectively. The measurement technologies were embedded in wearable, eyeglass-like frames that housed both an eye tracker to measure the gaze direction and capacitive sensor(s) to measure the level(s) of facial activations.

The work for this doctoral thesis consisted of two closely connected tasks as follows: First, Face Interface was rigorously tested. In these studies, simple point-and-select tasks were used in which the pointing distances and object sizes were varied. Especially the speed and accuracy of the Face Interface prototype was tested in a series of experimental studies. Second, Face Interface was used for entering text on an on-screen keyboard. For that, three on-screen keyboard layouts were designed. Then, they were experimentally tested so that the places of the characters were randomized after every typed word. This was done in order to exclude the effect of any previously learned layouts. The use of Face Interface was then compared against the use of a computer mouse.

In this thesis, three different versions of Face Interface have been used. The first one was wired and had one capacitive sensor placed in the bridge of the nose of the prototype so that it was able to monitor only the frowning related movements. Also, a chin rest was used in order to prevent head movements. The second version was wireless and it was able to monitor either frowning or eyebrow-related movements, depending on a person. The eye tracker was also improved so that the pupil detection algorithm was improved and the corneal reflection detection was added. Moreover, a scene camera was added so that head movements could be compensated using a head-movement-compensation algorithm. The third version was further improved by using five capacitive sensors to detect different facial activations: frowning, raising the eyebrows, and smiling.

The results showed that Face Interface functioned promisingly as a pointing and selection technique. From the iterations, significant improvements have been achieved in the pointing task times (i.e., from 2.5 seconds with the first prototype to 1.3 seconds with the third prototype). The subjective ratings showed that users felt positive about using the Face

Interface. The text entry rates for first-time users were encouraging (i.e., four words per minute on average).

To conclude, this thesis introduced a novel, multimodal, and wearable Face Interface device for pointing and selecting objects on a computer screen. It seems that the use of facial behaviors to interact with technology has great potential. The research has shown, for example, that it is easy to learn the use of these two different modalities together, and the use of it does not require much practice. These are clear indications for the use of facial information in human-computer interaction.

Acknowledgements

This thesis process has been mentally demanding. At times, it has taken an overwhelming hold of my life. However, as the end result finally approaches, it has been worth of every sleepless night—and, of course, the times of joy and success. There are many of whom I would like to express my gratitude for helping me get through this long process. First, I sincerely thank my supervisor, Professor Veikko Surakka, who has generously supported me throughout this thesis work. He has unstintingly provided his time and advice.

This thesis has been funded by the Finnish Doctoral Program in User-Centered Information Technology (UCIT) and the Academy of Finland. I thank the reviewers Associate Professor John Paulin Hansen and Professor Markku Tukiainen for their time and efforts in reviewing this thesis.

This thesis would not exist without the efforts of the members of Wireless User Interface (WUI) consortium. Thus, I owe my greatest appreciation for the past and present members of WUI consortium. More specifically, I wish to thank all my co-authors who have directly contributed to this thesis. I wish to especially thank Ville Rantanen with whom the collaboration has been fluent. I extend my thanks also to Toni Vanhala.

Tampere Unit for Computer-Human Interaction (TAUCHI) Research Center has been a great place to work, and I wish to thank former and current heads of TAUCHI, Professor Kari-Jouko Räihä and Professor Roope Raisamo, for doing such a good job of providing excellent research facilities. I appreciated all of the administrative personnel. The members of research group for Emotions, Sociality, and Computing (ESC) have been supportive throughout this process. Thus, I want to thank all the members of ESC with whom I have had the pleasure of working with. I wish to thank Mirja Ilves for the mental support and engaging discussions. I thank Päivi Majaranta for introducing me to research work.

I wish to express gratitude toward my friend, Outi, who has shared this journey with me and has understood me when nobody else did. I want to express my deepest gratitude to my family, mother, father, Arto, and Kaisa, for being there for me. My loving thanks to my husband, Mika, for supporting me every step of the way, for tolerating me at times when I was being difficult, and for being by my side. *Nothing Else Matters*.

Tampere, 1st of April, 2014, *Outi Tuisku*

Contents

List of Publications

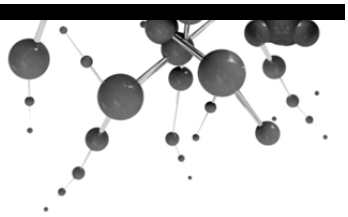
This thesis consists of a summary and the following original publications, reproduced here by permission of their publishers.

Author's Contributions to the Publications

Each publication included in this thesis was coauthored, indicating that all of them originated from collaborative research between the authors. The present author was the main author of Publications I, III, IV, and V. The empirical work for Publication II was designed and implemented by the present author. Publication II was first drafted by Ville Rantanen and then revised by the present author. The present author also wrote the descriptions regarding the empirical work for Publication II.

List of Abbreviations

ASL	Applied Science Laboratories	p. 49
BCI	Brain-computer interface	p. 23
CMOS	Complementary metal oxide semiconductor	p. 36
CPM	Characters per minute	p. 11
CRT	Cathode ray tube	p. 28
EMG	Electromyography	p. 2
EOG	Electro-oculography	p. 9
GUI	Graphical User Interface	p. 1
HCI	Human-Computer Interaction	p. 1
ID	Index of difficulty	p. 27
IR	Infrared	p. 33
KSPC	Keystrokes per character	p. 11
MSD	Minimum string distance	p. 11
MT	Movement time	p. 28
SAK	Ambiguous scanning method	p. 24
SMI	SensoMotoric Instruments	p. 8
WPM	Words per minute	p. 11



1 Introduction

The computer mouse has been the most common pointing and selecting technique in graphical user interfaces (GUIs) since it was developed about 50 years ago (English et al., 1967). Almost for an equally long time, the search for alternative interaction techniques has been going on in human-computer interaction (HCI) research. In HCI, it has been an important goal to try to take into account natural human behavior when creating new interaction techniques. It is envisioned that this eventually leads to a HCI that would be intuitive and versatile. One special area of development has been the utilization of human eye movements when interacting with computers. The eyes move naturally according to one's visual attention, so pointing and selecting objects with eye movements should be convenient. Further, it can be argued that eye movements serve important functions in human to human interaction. In addition to the direction of visual attention while working, eye behavior serves for communicative purposes—which is another argument for the use of gaze in HCI.

While eyes are centrally a perceptual organ and as such are intended for perceiving visual information, it is known that they can be voluntarily-controlled (Ware & Mikaelian, 1987; Surakka et al., 2003; Zhai, 2003). People can, for example, gaze at their interaction partner or any object of interest. Using eye trackers, eye movements can be converted to computer cursor movements in order to be able to control computers. Gaze-based interaction uses only one modality (i.e., unimodal interaction). Simple functions—such as pointing and selecting objects—require special arrangements in order to differentiate these two different functions from one modality. The solution for this has been the use of a dwell time. Dwell time means that in order to select an object, the gaze needs to be held above the object for a certain predefined time period in order for the object to be selected. Without this solution, or with short dwell times, it becomes

difficult to make a distinction between glances when the user is just looking around and fixations with the intention to start a selection function. This leads to a so-called Midas touch problem in which everything that user gazes at becomes selected (Jacob, 1991). Another disadvantage could be that video-based eye tracking requires expensive equipment, and not everyone that needs an eye tracker is able to afford one. The development of HCI, however, has taken such an approach that low-cost eye trackers do exist (Rantanen et al., 2012b; San Agustin et al., 2009a).

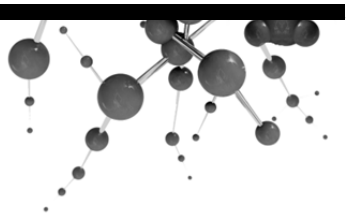
Another (behavioral) modality that centrally is used for human communicative purposes—and is under both spontaneous and voluntary control—is human facial movements. It is known that many facial actions and expressions are activated spontaneously, but they can also be activated voluntarily in human to human communication (Dimberg, 1990; Fridlund, 1991; Surakka & Hietanen, 1998). Although the suitability of facial behaviors for pointing and selecting objects in a GUI has been studied, there is evidence that using facial behavior alone may result in relatively slow interaction (e.g., Barreto et al., 2000). As a unimodal interaction technique, the use of facial expression is arguably a promising technique. However, pointing to objects can be quite cumbersome in contrast to eye movements because there is no direct route in applying the facial expression in controlling computers. This is because people need to twist and turn their faces in order to be able to move the cursor on a computer screen. Although the cursor movement might be challenging, the object selection could easily be done by, for example, frowning or raising one's eyebrows. Thus, combining the use of eye movements and facial behaviors would offer a potentially new means for interaction with computers. There are several arguments in support of this. To mention some, both modalities serve communicative purposes; they are well under voluntary control; and both function relatively fast.

The idea of combining voluntarily-directed eye movements and voluntarily-controlled facial muscles as a new multimodal HCI technique has been introduced quite recently (Chin et al., 2008, 2009; San Agustin et al., 2009a, 2009b; Surakka et al., 2004, 2005). In these techniques, two different measurement techniques have been used: an eye tracker for measuring the gaze direction and an electromyography (EMG) device for measuring the facial activations. The simple starting point of these studies has been to model the functionalities that the computer mouse has (i.e., pointing and selecting objects on a computer display). This multimodal technique has proved to be functional, although more research is needed in order to find out which facial muscles would be most usable in the case of selecting objects on a computer screen.

This thesis introduces a series of studies investigating the potential of combining gaze and face behaviors for multimodal HCI. A central technological innovation used for these studies has been a prototype called Face Interface. Thus, the thesis at hand also deals centrally with an iterative development of the prototype technology. Face Interface combines the use of two above-mentioned modalities: voluntarily-controlled gaze direction and voluntarily-controlled facial muscle activations for pointing and selecting objects in a GUI, respectively. The measurement technologies were embedded in wearable, eyeglass-like frames that house both an eye tracker and capacitive sensor(s) to measure the levels of facial activations. In the course of this thesis work, three different facial actions were used as the selection technique: frowning, raising the eyebrows, and smiling. The development of Face Interface has been iterative so that its limitations as well as its potential functionality could be understood. This thesis introduces five original publications in which different versions of Face Interface for pointing and selecting has been used.

In the course of the thesis work, the functionality of Face Interface was improved iteratively. The number channels for measuring the facial activity was increased from one to five. Also the eye tracker was improved, first by improving the pupil detection algorithm and then by adding a scene camera in order to compensate the head movements. In each state, the functionality of the multimodal interaction was experimentally tested. The results were used in order to find out the requirements for developmental changes for the functionality of the prototype from the technological point-of-view. In each state, the functionality of the Face Interface prototype was experimentally tested in order to find out the feasibility of the changes. This was done by using simple pointing and selecting tasks where the pointing distances and target sizes were varied. The new interaction method was used for entering text with an on-screen keyboard.

It seems that combining the use of facial behaviors to interact with technology has great potential. The research has shown, for example, that it is easy to learn the use of these two different modalities together—and the use of it does not require much practice. These are clear indications for the use of facial information in human technology interaction.



2 Facial Information for HCI

This chapter provides an overview about the functioning of two different modalities—the gaze and facial system—and their use in HCI. Both of these systems can be used independently or complementary to each other in order to create multimodal HCI. Thus, they are first introduced separately, and then their functioning in combination is discussed.

2.1 Gaze-Based Interaction

Background Information

Gaze can be used for different purposes (e.g., as a perceptual organ and in social interaction). In social interaction, people naturally look at the person that they are interacting with (Jacob, 1991; Vertegaal, 1999). It is known that gaze direction can reveal the direction of one's attention whether it is another person or object on a computer screen. For these reasons, researchers have been interested in studying eye movements since 1950s-1960s (Gibson, 1950; Klein & Ettinger, 2008; Stark et al., 1962; Wade & Tatler, 2009). By studying the eye movements, information on the cognitive processes such as reading behavior can be produced (e.g., Hautala et al., 2010; Hyönä, 2009; Hyönä & Niemi, 1990; Sharmin et al., 2012). A newer application area for eye movement research is to use gaze as the input modality for controlling computers (Ware & Mikaelian, 1987; Jacob, 1991; Sibert & Jacob, 2000; Duchowski, 2002; 2003; Majaranta & Räihä, 2002; 2007). Before going into details on how the gaze direction can be tracked, some general background information on the eye is provided. As can be seen from the Figure 1, the eye is a complex organ.

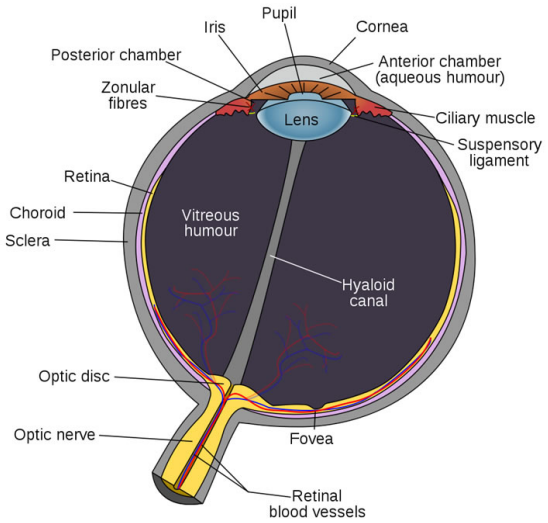


Figure 1. Structure of the eye. Picture adapted from public domain: http://www.sciencekids.co.nz/pictures/humanbody/eyediagram.html

From the perspective of eye tracking, it is important to understand how the eyes move. Eye movements can be divided in three different categories: *fixations*, *saccades*, and *smooth pursuits*. People have the ability to hold the eyes on some object of interest for a short time period, which is called a *fixation*. All of the visual information is gained during fixations. They last for a brief duration, approximately 100-200 ms (Jacob & Karn, 2003). Of course, the length of the fixation depends on the task at hand. For example, while reading, the fixation duration could be as long as 1000 ms (Just & Carpenter, 1980). Generally, it could be argued that the more cognitively demanding the task, the longer the fixation is. Because of the fact that the visual perception occurs in the brain, the fixations need to be long enough so that there is enough time to formulate the perception.

Eyes move from one fixation to another with ballistic movements called *saccades*. Once a saccade is started, it cannot be stopped nor can its direction of movement be changed. The length of the saccade varies, but usually it lasts approximately 30-120 ms (Jacob, 1995). During saccades, people do not gain any visual information. Thus, eyes move so that they are a combination of fixations and saccades and the eye movements can be described as quite 'jumpy'. The eyes move smoothly only when they are following a moving target, such as a moving car in a distance. This movement is known as a *smooth pursuit*.

In order to create an accurate view of the world, eyes need to move actively. This is because the accurate field of vision is approximately one to two degrees. An often used example to describe the accurate field of vision is that a thumbnail at an arm length is approximately 1.5-2° of field of vision (e.g., Duchowski, 2003; Holmqvist et al., 2011) which

corresponds to the accurate field of vision. This narrow field of accurate vision is due to the fact that only a small part of the eye—the *fovea*—is responsible for accurate viewing (see Figure 1). Because of this, eyes need to actively move in order to gain a broader sense of the world.

For obtaining and understanding (visual) sensations, cognitive processes are imperative. It has been suggested that (visual) attention acts as a spotlight or a zoom-able lens that is directed to any object of interest (Posner et al., 1980; Eriksen & St. James, 1986), which means that attention should be directed at the object of interest in order to create the experience of perception from the visual sensations. If attention is not directed to the object of interest (e.g., attention can be directed at thoughts), the visual information is not perceived—and, thus, it is not remembered.

There are many pros for using gaze direction as the input modality for controlling computers. For one, gaze functions fast when compared to other modalities (Ware & Mikaelian, 1987). Other pros include the fact that the gaze is natural, and it can be directed at will. Because people may control gaze, the possibility to interact with computers by gaze has emerged.

Eye Movements in HCI

In order to be able to use gaze direction as an input method for controlling computers, the gaze direction needs to be transformed to cursor movements. For that, eye trackers are used. Early eye trackers used a lens that was placed in the eye—similar to contact lenses. This made the eye tracking invasive because the contact lens was placed directly in the eye. Since then, eye trackers have evolved and are non-invasive. Modern eye trackers are based on the technique that was developed in 1960s, known as video-oculography (i.e., a video-based eye tracker is used). It usually consists of a video camera that images the user's eye(s), and, with different algorithms, finds the pupil and/or corneal reflection from that video. From this information, the eye movements can be calculated and the gaze direction can be transformed to cursor movements on a computer screen.

Two types of techniques for detecting the pupil from video exist: light pupil method and dark pupil method. In the light pupil technique, the eye is illuminated with a light source (e.g., infrared light) that is placed close to the optimal axis of the imaging device. Thus, the light goes through the lens and pupil to the retina and reflects back. This in turn causes the pupil to appear brighter in the image than the iris that surrounds it. In the dark pupil method, the light source is placed so that the pupil appears to be darker than the iris, and the darkest part of the image is then searched and recognized as the pupil.

In addition to finding the pupil, corneal reflection is used for eye tracking as a reference. It is achieved by using an infrared light source to create an

illumination on the eye (i.e., inside an iris area) which is called a Purkinje image. The corneal reflection can be used as a reference point because it stays static while the eye moves (i.e., the gaze direction can be calculated in relation to the corneal reflection). The corneal reflection stays still in the video image of the eye and the position of the pupil changes in relation to the corneal reflection when eye(s) move (e.g., Duchowski, 2003).

In order to use the eye tracker, a calibration procedure is needed so that correct position of the gaze on the computer screen can be identified for every user. It is necessary for finding the correct point of the gaze on a computer screen (or in environment). Usually calibration is done using a 3 × 3 grid so that user follows a moving dot with his or her gaze. The dot starts to move and stops in nine places on the screen. While stopped, the gaze data is collected for calibration. Then, after calibration, users can begin to interact with a computer. While calibration is needed in order to be able to use eye trackers, users see it as a tedious process (Villanueva et al., 2004). Another weak feature is the calibration drift. This means that after a while, the calibration weakens (i.e., drifts from the correct position) and therefore there is a need for re-calibration (Ashmore et al., 2005).

Eye trackers can be roughly divided into two different categories: remote and wearable (e.g., head-mounted). For remote eye trackers, the camera can be placed in a remote location like a computer screen. This makes it necessary for the user to stay in front of the computer in a quite static position in order for the eye tracker to find the eye(s). It is often stated that remote eye trackers are so-called state-of-the-art eye trackers. In the wearable eye trackers, the eye camera(s) is placed in front of the user's eye(s) using, for example, special eye glasses. Interestingly, the head-mounted eye trackers have been mainly research prototypes with low-cost parts (e.g., Babcock & Pelz, 2004; Franchak et al., 2012; Li et al., 2006; Noris et al., 2011; Rantanen et al., 2012b; Ryan et al., 2008). Quite recently, however, the large eye tracking manufacturers—like Tobii Technology or SensoMotoric Instruments (SMI)—have developed their own head-mounted eye trackers because they are seen as promising solutions for the future of eye tracking research. The advantage of the head-mounted eye tracker over the remote one is that user is able to move more freely with the device because the eye(s) are visible to the camera regardless of the user's (head) position.

While eye tracking might sound easy to use and develop, there are many challenges to overcome before eye trackers can be widely implemented. For example, the accuracy of the eye tracker can still be problematic. In general, the accuracy of the eye tracking is approximately 0.5-1° (Ashmore et al., 2005; Duchowski, 2003). This means that the accuracy of the eye pointing on a computer screen is approximately 16-33 pixels, if the monitor is a 17″ display with a resolution of 1280 × 1024 and the viewing

distance is 50 cm. This means that the objects on a computer screen need to be large enough so that users are able to easily point to them.

Of course, eyes are mainly a perceptual organ and are not intended for cursor control (Zhai, 2003). This knowledge offer challenges for the eye tracking technology because—while eye movements can be controlled at will—they also move compulsively. There are several reasons for involuntary eye movements (Ashmore et al., 2005). First, fixation jitter means that eyes never stay still; there are always small (involuntary) movements. Second, peripheral vision (i.e., the vision outside the accurate field of vision) is sensitive to the changes in the environment that if something happens in the background, it "catches the eye;" eyes move towards that distraction.

Another possibility to measure gaze direction (i.e., the point of gaze) is to use an electro-oculography (EOG)-based technique (Bulling et al., 2012). The EOG technique measures the resting potential of the retina. When the eyes move, that causes changes to the resting potential. The EOG sensors are attached around the eye—usually two in both sides and/or two above and below the eye to detect the changes in the resting potential. The calibration for EOG signal detection is done so that the baseline signal is calculated for each user. And, from the baseline, it is possible to detect the changes in the resting potential (Bulling et al., 2012). With EOG, it is not possible to detect the accurate point of gaze; and, for that reason, it is a more suitable technique for detecting gaze gestures (e.g., Bulling & Gellersen, 2010). As an example, Bulling et al. (2009) developed wearable EOG glasses so that they placed EOG sensors to the frames of the eye glasses, with the sensors attached to the skin around eyes. They tested the use of their EOG eye tracker with a simple experimental setting where the task of the participants was to produce different gaze gestures as fast and as accurate as possible. Their results showed that EOG glasses suited them well for recognizing gaze gestures but there might be restrictions in using them in tasks that need more accurate eye tracking (e.g., a certain button needs to be hit). The advantage that the EOG-based eye tracker has over the video-based techniques is that it requires much less computing power—making it easier to use it with mobile devices.

If we take into account the involuntary eye movements (i.e., jitter), we can conclude that eye tracking might never be as accurate to use as a computer mouse (e.g., Zhai, 2003). This is the case especially when eye trackers are self-built from low-cost parts. The low-cost parts will make the eye trackers affordable, but there might be trade-off in accuracy as compared to commercial eye trackers (Johansen et al., 2011). For example, if the object to be selected in user interface is too small, it might not be possible to select by gaze—and, for that reason, the design of the user interface becomes an important factor for the functionality of the eye trackers.

Selection Techniques for Gaze-Based HCI

The most commonly used selection technique in the case of gaze pointing has been the use of dwell time, which means that the user needs to fixate his or her gaze on the object for a certain predefined time period to select it. Different dwell times have been used, quite often they vary somewhere from 400 ms to 1000 ms (Majaranta & Räihä, 2002; Ware & Mikaelian, 1987). The use of a longer dwell time may slow down the interaction, causing difficulty or frustration for some people. On the other hand, with shorter dwell times, it may become difficult to differentiate whether the user is looking around or indicating a selection. This introduces a so-called Midas touch problem (Jacob, 1991)—meaning everything that the user gazes at becomes selected, even though the user might only be looking around.

Alternatives for the dwell time have been developed. One possibility is to use gaze gestures, which can be defined as patterns of eye movements. Gaze gestures can be issued as commands similar to mouse clicks (e.g., by first gazing at the object to be selected and then performing the corresponding gaze gestures) (Heikkilä & Räihä, 2012). Different types of sets for gaze gestures have been created from simple one directional eye movement (Heikkilä & Räihä, 2012; Møllenbach et al., 2010) to more complex sets of eye movements (Heikkilä & Räihä, 2009; Porta & Turina, 2008; Wobbrock et al., 2008). The use of complex gaze gestures may require that they need to be memorized before they can be used, which may make the use of gaze gestures unnatural. Other possibilities for selection techniques with eye pointing include winking, blinking, and eye closure (Ashtiani & MacKenzie, 2010; Heikkilä & Räihä, 2012; Królak & Strumiłło, 2011). Blinking and gaze gestures can be measured by an eye tracker but they can also be measured using EOG measurements too (e.g., Vehkaoja et al., 2005).

Gaze in Pointing and Selecting

The most direct route in using the gaze for HCI is to use it as a pointing and selection technique, similar to the computer mouse. The experimental studies on pure gaze pointing are rare. One of the earliest studies of using the gaze for HCI is a study by Ware and Mikaelian (1987). They used the gaze for pointing at objects. For selection, a dwell time of 400 ms, a screen button (i.e., a large area of the screen was designated as a button), or a physical hardware button was used. The task of the participants ($N = 4$) was to point to an object by gaze and to make the selection with one of the three aforementioned selection techniques. The results showed that, overall, the mean task time for the dwell time technique was approximately 0.8 seconds, and was approximately the same for the hardware button technique. For the screen button, however, the task time was slightly slower at approximately 0.9 seconds. The error percentages were 12% for the dwell time technique, 22% for the screen button

technique, and 8.5% for the hardware button technique. Thus, it seems that adding another modality for object selection decreases the error percentage—although, differences between the error percentages were not statistically significant.

Sibert and Jacob (2000) performed a point-and-select experiment where they compared the use of gaze to the computer mouse as an input method. The task of the participants ($N = 16$) was to select a circle from a 3 × 4 grid so that the target circle was highlighted, indicating that it was to be selected. After the selection of the highlighted circle, another circle was highlighted and participants pointed and selected that. For the gaze interface, a dwell time of 150 ms was used. Each circle had a diameter of 1.12", and the distance from the neighbor circles was 2.3". The results showed that the overall task completion time was 0.5 seconds for the gaze pointing, and 0.9 seconds for the mouse pointing. On the other hand, they did not report error percentages. The error percentages would have given more detailed information on the difference between the eye tracker and the mouse. However, they reported momentary equipment problems that happened for 11% of all eye tracking trials and only 3% for mouse trials. These percentages indicate some problems that eye trackers have. Mainly these issues are due to the fact that they do not find the pupil all the time, or they might find the pupil from a place where there is no pupil.

Text Entry

Today, gaze as an input method has been used for entering text for over 30 years (Majaranta & Räihä, 2002; 2007). The most direct route to apply eye tracking for text entry is to use on-screen keyboards, which can be modeled after the physical keyboards or after alternative keyboard solutions (Majaranta et al., 2006, 2009; Räihä & Ovaska, 2012). The characters have mainly been selected using a predefined dwell time and the length of that dwell time differs from one study to another. The text entry speed (in every text entry experiment, not just gaze-based) is measured as characters per minute (cpm) or as words per minute (wpm). Wpm is a reproduction from cpm, and they can be measured with the same quantity. In wpm, one word is defined to be 5 characters, including space and punctuation (Wobbrock, 2007). Thus, in the case of wpm, the time to write a sentence is divided with 5. To measure the errors in text entry tasks, usually two quantifications are used: minimum string distance (MSD) error rate, and keystrokes per character (KSPC). The MSD error rate compares the transcribed text (i.e., the text that was written by the participant) with the presented text, using a minimum string distance (Soukoreff & MacKenzie, 2003). The MSD error rate does not take into account how the text was produced—just the main result. KSPC, on the other hand, is used to give descriptive measures of the writing process itself, which means that the KSPC value indicates how often the participants corrected already typed characters (Soukoreff & MacKenzie,

2003). Ideally, KSPC value is 1.00, which indicates that each individual key press has produced a correct character. However, if a participant makes a correction during text entry process (i.e., presses the delete key and chooses another letter), the value of KSPC is larger than one.

Helmert et al. (2008) compared the use of three different dwell times (i.e., 350, 500, and 700 ms) while typing text on an on-screen keyboard. The task of the participants was to enter 12 words with each of the dwell times. Each participant started with 700 ms dwell time, then moved to 500 ms dwell time, and finally used the 350 ms dwell time. The results showed that the pointing task time was fastest with the shortest dwell time (59.5 cpm) and slowest for the longest dwell time (40.1 cpm). The pointing task time for the medium dwell time was 49.2 cpm.

Majaranta et al. (2006) studied the effect of feedback to text entry. They compared four different types of feedback for indicating that a key had been pressed on an on-screen keyboard. The used feedbacks were as follows: visual only, visual and auditory, speech and visual, and speech only. In the visual feedback, the key that was focused on was highlighted. It started shrinking, and when the key was selected (i.e., pressed down), the letter was colored as red. In the auditory feedback, a 'click' sound was played when the key was pressed down. For the speech feedback, the letter was spoken out loud when the key was pressed down. In a combination feedback, both of the mentioned feedbacks were used (e.g., in visual and auditory feedback both were used simultaneously). Thirteen participants took part in the experiment where the task was to enter five short phrases of text utilizing four feedback modes in four blocks, using a predefined dwell time of 700 ms. The results revealed that the feedback mode influenced the text entry rate. Typing with visual-auditory feedback was the fastest one. To conclude, by adding a simple 'click' sound when the key is pressed, a typing speed can be significantly improved when dwell time is used as the selection technique. On a longitudinal eye typing study, where participants were allowed to adjust themselves the length of the dwell time, results showed that it is possible to be quite fast with eye typing (Majaranta et al., 2009).

When an on-screen keyboard is used, some examples of different layouts of the places of the characters exist. For example, Špakov and Majaranta (2009) designed an alternative character layout to QWERTY. They used scrollable keyboards so that one, two, or three lines were visible of the keyboard. They designed an optimized keyboard arrangement so that they placed the most frequently used (in Finnish language) letters in the top row, the less frequently used letters in the second row, and the least used letters in the third row. The participants were able to scroll the rows using the buttons on the left or right-hand side of the keyboard. The designed letter placement was compared against the traditional QWERTY

layout. The results were encouraging, as the participants wrote slightly faster with the optimized layout than with the QWERTY layout. Their results showed that the mean writing speeds were 11.1 wpm for QWERTY and 12.18 wpm for the optimized letter placement. Similar results on keyboard design have been shown in other text-entry studies, where the QWERTY layout had been replaced (Bi et al., 2010; MacKenzie & Zhang, 1999). For gaze-based text entry, the QWERTY layout might not be the most convenient alternative because the accuracy of eye tracking varies depending on gaze direction. Gazing with the eye closer to the extremities of its rotational range makes the tracking less accurate because the eyelid(s) may cover the eye(s) and, thus, the pupil would not be visible to the camera. In the QWERTY layout, for example, the most frequently used characters are placed on the edge of the keyboard (e.g., the character 'a'), which may result in difficulties to the selection of the character when using gaze tracking (Räihä & Ovaska, 2012).

In most eye typing studies, the layout design (e.g., key size and placement) of the keyboard was not explicitly considered. One example of a different layout is called GazeTalk (Aoki et al., 2008; Hansen et al., 2003; 2004). GazeTalk consists of a 3 × 4 table that is divided in 11 cells that include a (1 × 2) text field and 10 (1 × 1) buttons. The size of the buttons was approximately 8 × 8 cm and the size of the text field was approximately 16 × 8 cm. Out of the 10 buttons, six were reserved for single characters that changed dynamically based on the written text; one button was reserved for selecting characters from an alphabetic listing; one button was for the eight most likely used words based on what the user had typed; and the last two was for the spacebar and backspace. The buttons were selected by dwelling on them. The results of a longitudinal study showed that the maximum text entry speed after one thousand typed sentences was approximately 9.4 wpm for Danish text and 29.9 cpm for Japanese text. The results are reported in two different metrics because Japanese text is different in its style as compared to Western text. And, thus, it is comparative to cpm value.

Dasher is another example of a type of text-entry software. It is a dynamic keyboard that adapts itself according to the entered text. Dasher uses one modality (i.e., mouse, gaze) for entering text (Ward & MacKay, 2002). It is a zooming interface in which a user operates with continuous pointing gestures. In its initial state, the letters are placed on the right-hand side of a computer screen. When the user enters text, the characters start to zoom out in the direction of where the cursor is (i.e., the area surrounding the cursor grows in size to display the most probable characters). The character is selected once it crosses a vertical line in the middle of the screen. The user navigates through the characters simply by looking at them. At first glance, the characters may seem to be unorganized and may cause initial difficulties to a novice. However, after about one hour of

practice, most users learn the logic of Dasher and are able to use it quite fluently. In a longitudinal study where 12 participants used gaze-controlled Dasher for ten fifteen-minute long sessions, the overall mean text-entry rate was approximately 17 wpm after the last session (Tuisku et al., 2008). However, after the first session, the mean text-entry rate was only approximately 2.5 wpm.

It is noteworthy to mention that, for the most part, these GUIs are rarely, if ever, modeled to compensate for the technical weaknesses of pointing techniques. It is important to take into account the challenges of the new pointing techniques into the design of the GUI to improve functionality. As a general example of this type of adaption is a keyboard layout that Oulasvirta et al. (2013) have designed to be used with touchscreen devices (e.g., a tablet computer). The software—called KALQ—consists of two, rectangular 4 × 4 key grids placed in the regions that are within reach of a user's thumbs. Oulasvirta et al. (2013) tested the KALQ layout against the traditional QWERTY layout. KALQ led to a faster text-entry rate than QWERTY (i.e., 37.1 wpm for KALQ and 27.7 wpm for QWERTY). This is once again proof that QWERTY might not be the best solution for entering text with alternative pointing techniques, despite its familiarity for users. Based on these findings, it could be concluded that it is important to design the keyboard layout according to the features of the used pointing device.

2.2 Face-Based Interaction

Background Information

In contrast to vision as a perceptual system, facial behavior system is mainly an expressive system. Facial expressions result from the contraction of facial muscles, which in turn, causes the facial skin to move accordingly (Rinn, 1984). The human facial muscle system is well advanced (as Figure 2 demonstrates). There are over 40 muscles that are used in generating facial expressions by contracting one or more of them (Rinn, 1984). Thus, faces are capable of producing versatile expressions (Mehrabian, 1981).The face area is well represented in the primary motor cortex. Facial muscles are, in this way, under good control.

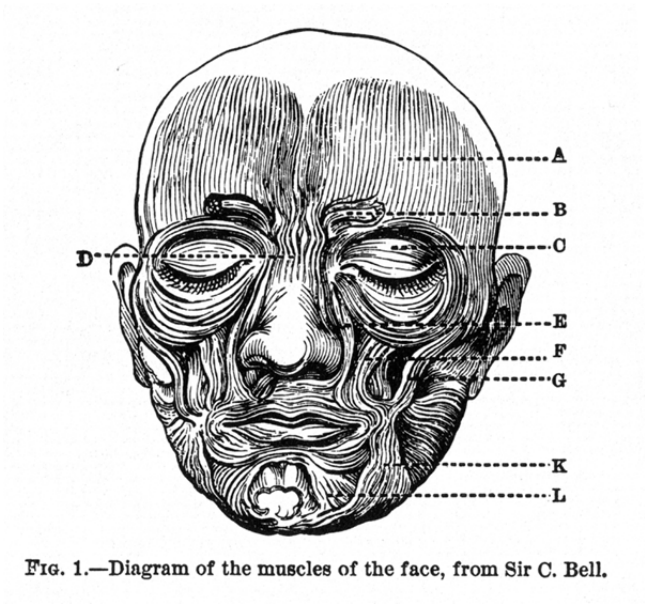


FIG. 1.—Diagram of the muscles of the face, from Sir C. Bell.

Figure 2.Representation of facial muscles. (The important ones for the scope of this thesis are: *frontalis* (A), *corrugator supercilii* (B), *zygomaticus major* (G) (Picture adapted from Wikimedia Commons, public domain).

In addition to spontaneous facial behavior (e.g., spontaneous emotional behavior), people are able to control their facial muscles voluntarily. It is known that people easily and frequently use their facial behavior on a voluntary basis in social interaction (Ekman, 1992; Ekman & Davidson, 1993; Hietanen et al., 1998; Surakka & Hietanen, 1998). The knowledge that the facial muscles can be controlled at will has made it possible to utilize the facial system in controlled tasks, such as pointing and selecting objects on a computer screen. The facial information can be used in simple pointing and selecting tasks that are modeled after the use of a mouse or perhaps in more advanced tasks like entering text (that can still involve pointing and selecting).

Measurement Techniques

For utilizing facial behavior (i.e., facial expressions) for interacting with computers, different measurement techniques can be used to track facial behavior. The activity of facial muscles can be measured with EMG, which is one method for transferring facial signals for the HCI purposes. EMG measures the levels of electrical activity in the facial muscles (Davidson et al., 2000; Fridlund & Cacioppo, 1987). EMG measurements can be so accurate that it measures the activity that is not visible in the face. With facial EMG most often electrical activity of *corrugator supercilii* (i.e., activated when frowning) and/or *zygomaticus major* (i.e., activated when smiling) muscles have been measured (Fridlund & Cacioppo, 1986). With frowning and smiling actions, for example, objects can be selected on a computer screen (Barreto et al., 2000; Surakka et al., 2004, 2005). In addition to the face area, EMG has been used for measuring the activity of other muscles in the human body such as muscles in the hand have been

used for controlling computers (Chen et al., 2007; Kim et al., 2004; Xion et al., 2011).

EMG has the downside that electrodes need to be attached to the skin. Plus, the skin needs to be prepared for the electrodes by being cleansed with ethanol, scrubbed with cotton sticks, and applied with abrasive paste need to remove the dead skin cells. All of these measures ensure a lower impedance of the EMG electrode. It is easy to realize that it might be quite cumbersome to use EMG on a daily basis. Further, there might be artifacts in EMG signals (e.g., because of body movement, teeth grinding, or extensive blinking) which has caused the signal to be unreliable (Rymarczyk et al., 2011).

Another possibility to measure facial activations is to use a capacitive sensing method (Rantanen et al., 2010; 2012a)—first introduced by Russian Léon Theremin in 1919 as a music player named after him. The theremin consisted of two metal antennas that sensed the position of the hands of the musician. One hand controlled the frequency of the sound, and the other controlled the volume. By moving the hands closer and farther away from the theremin, sound was created. Since then, applications of the capacitive sensors vary from sensitive clothes (Holleis et al., 2008) and posture recognition (Valtonen et al., 2011) to guitar strings (Wimmer & Baudisch, 2011), and much more. The capacitive measurement has the same principle as capacitive push buttons (e.g., traffic light buttons) and touchpads (e.g., a touchpad on a laptop) have. The principle for the capacitive measurement is simple. Only a single electrode that produces an electric field is needed for one measurement channel. Thus, the capacitive method is based on the proximity of the object to the electrodes. When an object nears the device, the electric field alters. Then, this change can be interpreted by signal processing algorithms (e.g., to create a mouse click using proper signal processing algorithms). In short, the capacitive measurement uses the distance between the electrode and the target (see Figure 3 for an illustration of the capacitive measurement). To use a capacitive sensing method for measuring facial behavior is a recent application area in HCI (Rantanen et al., 2010).

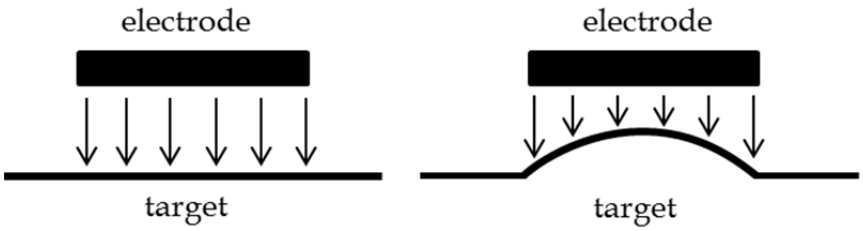


Figure 3. Left: Target is further away from the measurement electrode, and thus, the capacitance is larger. Right: The target is closer to the electrode, thus, decreasing the capacitance. Arrows represents the electric field between the electrode and the target.

Rantanen et al. (2010) studied the feasibility of the capacitive sensing method for HCI. They placed the sensor on a bridge of the nose of eye glasses so that it was able to detect both: raising the eyebrows and lowering the eyebrows (i.e., frowning). A short test was run to find out the feasibility of the capacitive measurements. The task of the participants (N = 10) was to move their eyebrows (i.e., to frown or to raise them) according to a corresponding sound clip. The signal collected from the capacitive sensor was recorded and analyzed offline with an algorithm that was designed for detecting the eyebrow movements from the signal. Even though the capacitive sensor was not used for real-time interaction tasks, the results showed that the capacitive sensing method detected facial movements.

To continue the study of a capacitive sensing technique, Rantanen et al. (2012a) investigated the use of it for more complex facial activity than basic frowning and raising of the eyebrows. For that, they built a wearable measurement prototype device in which the capacitive sensors were attached to a headset with six whiskers-like extensions (see Figure 4). There were three extensions in each side of the prototype so that the top extensions were placed above the eyebrows; the middle extensions were placed on the top of each cheek; and the bottom extensions were placed in the mouth and the jaw area. The task of the participants ($N = 10$) was to produce six facial actions that were: lowering the eyebrows, raising the eyebrows, closing the eyes, opening the mouth, raising the mouth corners, and lowering the mouth corners. They were told to perform these actions so that other parts of the faces—that were not involved on the current action—would stay still during the activations. It was found that even with these predefined facial actions, some facial movements activated parts of the face that were not meant to be activated. This indicates that the addition of measurement channels might introduce a potential problem when different facial muscles could be used at the same time. Further, results revealed that, with capacitive sensors, it might be possible to detect more complex facial activity than simple frowning and raising the eyebrows such as a combination of them.

Figure 4. The measurement prototype device (Figure printed by permission of Ville Rantanen).

Rantanen et al. (2013) continued their work with the above measurement device and found out that, with the capacitive sensing method, it is possible to detect the intensity of facial movement.

The Use of Facial Information in HCI

In HCI, studies that have monitored signals of the human neuromuscular system, as an alternative interaction method, have emerged. In psychophysiological research, human physiological signals have been used for quite a long time. However, the idea of using the signals measured from human body as a HCI method is more recent. Both spontaneous and voluntarily produced changes in the electrical activity of the human body have been utilized both for controlling computers (Kübler et al., 1999; Surakka et al., 2004; Wolpaw, 2007) and also for social-emotional HCI purposes (Baxter & Sommerville, 2011; Picard, 1997; Surakka & Vanhala, 2011; Vanhala & Surakka, 2008).

Most studies involving facial EMG have been conducted so that facial EMG has been used to record the facial muscle activity to find out the reactions that participants have on the phenomena that is under investigation. Mostly the activation of *zygomaticus major* (activated when smiling) and *corrugator supercilii* (activated when frowning) is measured to find out the reactions to different stimulations (Partala et al., 2006; Rymarczyk et al., 2011; Surakka & Hietanen, 1998; Vanhala et al., 2010; 2012).

Barreto et al. (2000) were among the first ones who used facial EMG for controlling computers. They measured the activity of *frontalis* (activated when raising the eyebrows) and left and right *temporalis* (activated when

moving the jaws) using three EMG electrodes. The facial muscle activations controlled the cursor on a computer screen. The action of raising the eyebrows resulted in moving the cursor upwards; the lowering of the eyebrows action moved the cursor downwards; left jaw movement moved the cursor left; right jaw movement moved the cursor right; and, finally, full jaw movement resulted in a mouse click (left mouse click). They tested their EMG system using simple pointing and selecting tasks. The tasks of the participants were to first point and select the start button and then point and select the stop button. The start button's diameter stayed static throughout the experiment and was 8.5 mm. For the target buttons, three diameters were used (8.5, 12.5, 17.0, and 22.0 mm). The start button was placed in the middle of the display, and the location of the stop button was varied so that it was placed in every corner of the display. The overall mean task time was 16.4 seconds.

Chin and Barreto (2006) continued the research on using facial EMG as a pointing and selecting technique. They measured the activity of right *frontalis*, left and right *temporalis,* and *procerus* (which is activated when lowering the eyebrows). Otherwise, the procedure was the same as earlier (Barreto et al., 2000). They reported an overall mean pointing task time of 13.2 seconds. In a follow up study, Chin et al. (2006) measured the activations of right *frontalis*, and left and right *temporalis* similar to Barreto et al. (2000)—and the activations of *frontalis*, left and right *temporalis,* and *procerus* similar to Chin and Barreto (2006)—for measuring the facial activity. Again, the procedure was the same as before (Barreto et al., 2000). The results showed an overall mean task time of 16.4 seconds for the first system and 13.2 seconds for the second system.

The problem with facial EMG method as a unimodal interaction technique is the fact that it might be difficult and even slow to point to objects using only facial EMG (Barreto et al., 2000; Chin & Barreto, 2006; Chin et al., 2006). This probably results from the fact that there was no direct route to move the cursor diagonally. This means that, for diagonal movement, two different facial muscles needed to be activated one after the other.

Text Entry

Text-entry studies are rare for techniques that measure information from human face. One example was presented by Gizatdinova et al. (2012). They used a computer vision technique, in which the cursor was moved by moving the head and characters were selected either by opening the mouth or raising the eyebrows. These actions were detected by using a simple web camera. A regular QWERTY on-screen keyboard was used for entering the text. The results showed an overall mean text-entry rate of 3 wpm. Based on these results, it seems that—while using computer vision for entering text is a promising approach—there is still room for improvement.

Dasher has been used for computer vision-based text entry (De Silva et al., 2003). In this case, Dasher was controlled by head movements (i.e., moving the head to the right caused Dasher to move right), which was detected using a web camera. The average text-entry rate was reported being 38 cpm (i.e., 7.3 wpm) for two users.

2.3 Multimodal Interaction

When using synchronously two or more different modalities—in this case the facial system and visual system—it is called multimodality. Multimodal interaction can be divided into two parts: input, and input and output research. This thesis introduces only the multimodal input as measured from the face area. Multimodal input in HCI can be defined as "*a system that responds to inputs in more than one communication channel*" (Jaimes & Sebe, 2007). In the current thesis, multimodality refers to the use of facial input (i.e., face-based multimodality) by the means of combining eye movements (i.e., gaze direction) and facial behavior (i.e., behavior or signals measured from face area) input in HCI. Research on face-based multimodality has emerged quite recently in HCI (Chin et al., 2008; D'Mello & Kory, 2012; San Agustin et al., 2009b; Surakka et al., 2004).

The advantage of the multimodal interaction, as compared to unimodal interaction, comes from the fact that the most functional or most convenient parts of both modalities can be utilized. That is, with gaze it is easy to look at any place, but making the selection does not come naturally. Further, with the facial muscles pointing, it might be slow and somewhat unnatural. The object selection by activating facial muscles, however, comes quite naturally for people.

Background Information

People naturally use these two rather complex systems (i.e., gaze and face) so that they do not need to actively think about the use of them. They are used to generating facial expressions and directing their gaze to any object of interest—without giving it much thought. This type of behavior happens every day and is natural for people—even automatic. However, when discussing about the voluntary use of these two systems in combination in HCI, it becomes more complex.

If a task of the user is to point and select some predefined objects using gaze direction and facial muscle activations for point and select objects, however, there are many processes needed. First, the user needs to gaze at the object to be selected and then actively keep the eyes focused on the target. Next, the user needs to create a conscious perception and understanding about the fact that the gaze is on the object. Only after that can the facial system be activated for object selection. It is virtually impossible to make the decision to activate facial muscles properly in

respect to the task before the visual information is received (and understood). When the user understands that the gaze is on the object she/he can activate the facial muscle(s) in order to select the object. After a successful activation of facial muscles, the user needs to understand that the object was selected (e.g., to see that the object disappeared after a successful selection). Following this, a new task may begin.

Based on the above example, it is easy to realize that multimodality can be seen as a cognitive process in a sense that one needs to always actively process and decide when to activate the first modality and when to activate the second modality. Thus, face-based interaction requires perception, memory, and thinking (Atkinson & Shiffrin, 1968; Baddeley & Hitch, 1974; Baddeley, 2000; Matlin, 2009; Whitman, 2011). These processes are not in the scope of the current thesis and, therefore, are not further discussed.

Pointing and Selecting

The most direct route in using gaze and facial information for multimodal interaction is to imitate the same functions that the mouse has (i.e., pointing and selecting). The work on this area is quite recent, and mainly eye trackers and EMG have been used.

Partala et al. (2001) tested an idea that used gaze direction for pointing and facial muscle activations for selecting objects. The idea was that a remote eye tracker could be used for measuring gaze direction for pointing and facial EMG from above the *corrugator supercilii* (i.e., activated when frowning) facial muscles for object selection. The system was an offline one, so that the data from these two systems were combined and analyzed offline. The new technique was compared to a regular computer mouse. The task was to first point and select a home square and then to point and select a target circle. They used three pointing distances (50, 100, and 150 pixels) and one target size (32 pixels). The target circle appeared in each of the eight angles in relation to the home square. Seven people participated in the experiment. The results showed an overall mean pointing task time of approximately 0.6 seconds. For the mouse, the overall mean task time was approximately 0.8 seconds.

Later, Surakka et al. (2004) introduced a real time system where gaze was used for pointing, and frowning was used for object selection. They used three pointing distances (60, 120, and 180 mm) and three diameters for target circle (25, 30, and 40 mm). Again, the target circle appeared in each of the eight angles in relation to the home square. The new technique was again compared to the computer mouse. Fourteen people participated in the experiment. The results showed an overall mean task time of 0.7 seconds for the new technique and 0.6 seconds for the mouse. In a follow-up study, Surakka et al. (2005) compared the use of frowning and smiling as the selection technique with gaze. They used the same task, as in the

earlier study, with eight participants. The results showed that smiling outperformed frowning as the selection technique, because the overall mean task times were 0.9 seconds for the frowning technique and 0.5 seconds for the smiling technique.

San Agustin et al. (2009a) used a self-built eye tracker for pointing and a commercial CyberLink™ headband for EMG measurements for object selection. They ran the experiment (*N = 6*) in a static condition (i.e., sitting in front of a desktop computer screen) and in a mobile condition (i.e., walking in a treadmill wearing a head-mounted display) for comparing the use four pointing and selection techniques. The pointing techniques were gaze and mouse, and the selection techniques were EMG and a mouse. These two pointing techniques—and two selection techniques—were then combined in to total of four pointing and selection techniques. The results showed that the overall mean task time for the static condition was 0.8 seconds, and—for the mobile condition—it was approximately one second.

Mateo et al. (2008) and San Agustin et al. (2009b) tested two pointing techniques (i.e., gaze and mouse) and two selection techniques (i.e., EMG selection and mouse click selection) in an experiment where each of the four combined pointing and selection techniques. With the EMG selection, the selection was indicated either by frowning or by tightening the jaws. The task of the participants (*N = 5*) was to point and select targets. Three target sizes (100, 125, and 150 pixels) and three pointing distances (200, 250, and 300 pixels) were used. The results showed that the overall mean task time was 0.4 seconds when all the pointing and selection techniques were taken into account. The fastest technique was gaze-pointing combined with EMG-selection, with a mean task time of 0.35 seconds.

Navallas et al. (2011) used the activation of *frontalis* facial muscle for object selection when pointing was done by gaze. They had three different groups of eight people performing the pointing and selecting tasks. One group tested the system with no communication protocols between the EMG and eye tracker (i.e., offline analysis); the second group tested the system with communication between the EMG and eye tracker (i.e., real time interaction); and the third group tested the system with communication between the EMG, eye tracker, and fixation delay (i.e., the participant needed to fixate on the target long enough before the selection could be made). They used three different noise levels for the signals. The used experimental setup was the same as in San Agustin et al. (2009b). They reported an overall mean task times of approximately 0.7 to 1.4 seconds, depending on the setup.

Lyons et al. (2001) used gaze direction and facial muscle activations differently than the above studies. That is, they used facial EMG for correcting the inaccuracies of the eye tracker and for selecting objects. The

method functioned so that participants first pointed to an object by gaze. If—after a gaze pointing—the cursor was not inside the object, the cursor was moved inside the object using facial muscle activations. Left and right yaw clenches caused the cursor to move left and right, respectively. Similarly, up and down movements of the eyebrows caused the cursor to move up and down, respectively. The object was selected with a full jaw clench. They also compared the use of the combined method to a computer mouse and unimodal EMG method (Barreto et al., 2000). They used the same task as Barreto et al. (2000) and compared the use of the combined system of the EMG-only system and the use of a mouse. The overall mean task time for the combined system was 6.8 seconds; for EMG, only it was 16.4 seconds; and for the mouse, it was one second.

Chin et al. (2008) improved the combined system introduced by Lyons et al. (2001) and reported an overall mean task time was 4.7 seconds. Chin et al. (2009) further improved the combined system. They compared the use of it to the use of a regular computer mouse and to the use of a purely gaze-based system (with a dwell time of 100 ms). They had 10 participants for each of the input methods. The results showed that the overall mean completion time was 0.98 s for the mouse, 4.68 s for the combined system, and 3.07 s for the eye tracking system. In these studies (Chin et al., 2008, 2009; Lyons et al., 2001), the object sizes were quite small—which can certainly affect the pointing task times when eye tracking is used for pointing because there could be a calibration error that makes it difficult (or even impossible) to hit the smallest targets.

Although, in this thesis, brain-computer interfaces (BCIs) are not in the focus, one interesting multimodal example of BCIs are introduced. Vilimek and Zander (2009) and Zander et al. (2010) have used BCI for indicating the selections when gaze direction was used for pointing. They called their multimodal system "BC (eye)." The task of the participants (*N = 10*) was to perform a search-and-select task using BC (eye). Participants were presented with reference stimuli in the middle of the screen. Around it—in a circular arrangement—were 12 stimuli, of which one was the target stimulus. The task of the participant was to find the target stimulus that was the same as the reference stimulus. The stimuli consisted of either four characters in an easy condition or seven characters in a difficult condition. The time that it took to find and select the target stimulus was measured as the task time. Three different selection techniques were used with eye tracker: short dwell time (1000 ms), long dwell time (which was either 1300 ms in Zander et al., 2010 or 1500 ms in Vilimek & Zander, 2009), and BCI (i.e., by thinking the activation). The overall mean task time for short dwell time was 4.68 seconds, for long dwell time, 6.08 seconds, and for BCI, 7.37 seconds.

Zhai et al. (1999) proposed an alternative pointing technique for gaze called Manual and Gaze Input Cascade (MAGIC) pointing. MAGIC pointing combined the use of gaze pointing to manual selection in two different methods. In conservative MAGIC pointing, the cursor appeared where the user was gazing at after the user moved the pointing device slightly, and the selection was made using a manual pointing device (e.g., by mouse button). In liberal MAGIC pointing, the cursor appears next to every object that the user is gazing at, without the need to physically move the cursor. The use of the two MAGIC techniques was compared to mouse pointing. The task of the participants ($N = 9$) was to point and select targets using two target sizes (20 and 60 pixels) and three pointing distances (200, 500, and 800 pixels) were used. The completion time for the mouse technique was 1.4 seconds, 1.52 seconds for conservative MAGIC technique, and 1.33 seconds for the liberal MAGIC pointing technique. This MAGIC technique shows that the multimodal technique where the hand is used for indicating the selection as compared to facial muscle activation, the speed of the operation is comparable to the face-based techniques.

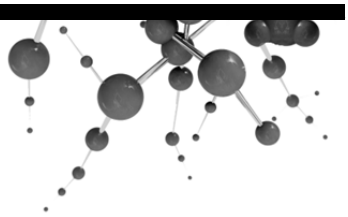
Text Entry

The above described multimodal techniques have been proven functional in simple pointing and selecting tasks. They show such a promise that they could very well be used in more advanced tasks that still involve pointing and selecting. Even though text entry would be a valuable application, the studies on text entry with multimodal (face-based) techniques are virtually non-existing. Examples can be found only on gaze-based systems and from such a system where BCI is used for selecting objects.

Ashtiani and MacKenzie (2010) introduced text entry software, named as BlinkWrite2, which uses an ambiguous scanning method (SAK). BlinkWrite2 consists of two regions: one is for selecting letters, and the other is for the word selection region. The letter selection consists of four keys so that letters (26, according to English alphabet) are divided in alphabetical order to three keys, and the fourth key includes the spacebar. The letters are scanned, and the letters are selected by blinking once the scanning is on the desired key. After each character selection, the word list was updated according to the selections that were made. They ran an experiment where the task of the participants was to enter five phrases, using three different scanning intervals. The results revealed an overall mean text entry rate of 4.71 wpm.

Yong et al. (2011) ran an experiment ($N = 7$) where the eye tracker was used as a pointing device, and BCI was used as a selection device. Their study was offline once, so that the data from eye tracker and BCI was combined offline and not in the real-time. The task of the participants was

to enter text so that the character was pointed by gaze and when the gaze had been on the object for an amount of dwell time (i.e., 0.75 and 1.00 seconds), the user could select the character by thinking of a hand extension movement. The results showed that the text entry rate was 9.1 cpm on average.



3 Evaluation of Pointing Devices

Evaluation of different pointing devices is important in order to gain knowledge on how they are functioning and how they can be improved. Pointing devices need to be developed further—and, in a sense, they are never ready. The mouse, as an example, has changed in its design and functionalities throughout the years so that it would be easier and perhaps more natural to use.

Different methods exist on how the pointing devices can be objectively and subjectively evaluated and compared with each other. The most commonly used objective method is called as Fitts' law (Fitts, 1954). While objective metrics are important, equally important is the measurement of subjective metrics. That is because—even if the pointing device would be very effective to use—if the potential users do not appreciate it, they will not use it.

3.1 Fitts' Law

The underlying model for Fitts' law comes from human behavioral research—mainly from studies concerning the functioning of the motor system. Fitts' law can describe the information capacity of the human motor system by measuring the consistent movement from among several movement types. Fitts' law is a mathematical model, which measures the relationship between the difficulty of the pointing task and the pointing task time (Fitts, 1954; MacKenzie, 1992). The difficulty of the pointing task can be varied by using different pointing distances and the width of the targets.

The task difficulty is defined as an index of difficulty (ID), which is calculated through the equation $ID = log_2(A/W + 1)$—where A is the

moved distance (*amplitude*) and W is the width of the target area. The tasks that involve Fitts' law analysis are usually relatively simple pointing and selection tasks, where the pointing distance and size of the target are varied in order to gain different ID values. The ID has a linear relationship to pointing time, and, thus, it can be described by the linear regression equation of the movement time (MT). $MT = a + b\ ID$, where a and b are the regression coefficients. Different pointing techniques can be compared with each other using an index of performance value (later called as throughput), that can be retrieved from the equation of the MT. Throughput is calculated by taking the inverse of b; that is, $throughput = 1/b$ (Zhai, 2004).

Originally, Fitts (1954) studied only one-dimensional movement with a stylus tapping. He used two physical discs with a height of 6 inches that the participants tapped with a stylus. The pointing distances (2, 4, 8, and 16 inches) and the width of the targets (0.25, 0.5, 1, and 2 inches) were varied, so that the tasks were at a different level as measured with task difficulty. The task of the participants ($N = 16$) was to hit these discs with a stylus as many time as possible in 15 seconds. Two styluses were used: one was lighter (1 ounce) than the other (1 pound). The results showed that the movement time increased when the distance grew. The error percentages grew when the width of the target was small as compared to larger target widths. The average error rate with a lighter stylus was 1.2% and 1.3% with the heavier stylus. The throughput varied from 10.3 bits/s to 11.5 bits/s with the lighter stylus and from 7.5 bits/s to 10.4 bits/s with the heavier stylus. Fitts (1954) concluded that the measured performance describes quite well the capacity of the human motor system because the throughput values stayed consistent throughout the experiment.

Since the 1950s, when the first article about the Fitts' law was published, it has widely been used by other researchers in evaluating different types of pointing devices (e.g., Card et al., 1978; MacKenzie & Buxton, 1992; MacKenzie & Isokoski, 2008; Surakka et al., 2004; Whisenand & Emurian, 1996; Zhai, 2004). In HCI, the first study that has exploited the Fitts' law in evaluating a performance of pointing device was a study by Card et al. (1978). They compared the use of a mouse, joystick, and two sets of keys on a keyboard (i.e., step keys, or arrows, and text keys) for object selection. The keys were selected so that they were standard ones in the 1970s to be used with cathode ray tube (CRT) displays. In the experiment, the participants ($N = 5$) were shown a page of text. The task was to point and select a target word or phrase that was marked by highlighting it. Five distances from starting position (1, 2, 4, 8, and 16 cm) and four target sizes (1, 2, 4, 10 characters) were used. The results showed that the mouse was the fastest and most efficient on the tested pointing devices, whereas the key-based techniques were much slower. Card et al. (1978) did not report

the throughput values in such a way that is almost a standard today, but their study has been an important starting point on using Fitts' law in HCI.

The Fitts' law was first introduced only for one-dimensional pointing (Fitts, 1954). It has later been extended for taking into account the angle of the movement. For example, Whisenand and Emurian (1996) studied the angle of the movement for the mouse. They used four target widths (0.25, 0.5., 1.0, and 1.5 cm), five pointing distances (1, 2, 4, 8, and 10 cm), and eight pointing angles (0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°). The results showed the following: First, the movement time increased when the target size decreased; and, secondly, it was shown that horizontal and vertical angles resulted in faster pointing task times than diagonal angles. The throughput values were not calculated.

The growth in the display sizes has allowed the use of more variety in ID values because longer pointing distances are possible to be used in analyzing the pointing device employing Fitts' law. The longer pointing distances could give more reliable results on pointing task times, since nowadays people tend to use increasingly bigger displays when interacting with computers (Tan et al., 2006). Of course, the limits of the pointing method need to be taken into account. Finger tapping, for example, is not possible in too large displays if a person is not able to reach everywhere on the screen.

The Fitts' law has been seen to be theoretical, and it has been criticized for the fact that it does not take into account the actual movement that the users do while performing these tasks. Thus, Fitts' law has been extended to take into account the movement that user performs by calculating the effective target width, or W_e (MacKenzie & Soukoreff, 2003). W_e is calculated so that $W_e = 4.1333 s_x$, where s_x is the standard deviation of the selection coordinates in the place that user has selected the object. With this method, more reliable values for Fitts' law can be calculated because they are based on the actual movement that users perform. Thus, in the function for ID, the value of W is replaced with the value of W_e.

Quite naturally, the throughput values depend on the experimental setup (i.e., on the used target sizes and pointing distances). For that reason, one needs to be cautious when comparing different studies with each other. When taking this point into account, the following comparison can be made: The mouse is the most commonly used pointing device in HCI and has thus been often evaluated using Fitts' law. The throughput value has been reported in many studies of being approximately 5 bits/s (Isokoski & Raisamo, 2004; Surakka et al., 2004). It is a good value for reference and to compare other alternative pointing devices to it.

For the techniques that utilize voluntary gaze direction and facial muscle activations, only a few publications have calculated the throughput values.

Surakka et al. (2004) calculated the Fitts' law and found that the new technique was more efficient than the computer mouse in terms of throughput. The throughput value for the new technique was 12.7 bits/s. San Agustin et al. (2009b) reported a throughput value of 3.03 bits/s. Chin et al. (2009) did not report a throughput value; they only mentioned a poor match to the Fitts' law model.

3.2 Subjective Ratings

While Fitts' law analysis compares the performance of the pointing devices objectively, it is equally important to collect the ratings of the subjective experience as well. However, there are almost as many possibilities to measure the subjective ratings as there are researchers. It is necessary to take a look on what kind of scales there are that are intended for evaluating the subjective experience.

A profound theory and method that is presented for collecting subjective rating is called semantic differential (Osgood, 1952). The semantic differential method uses the combination of associational and scaling procedures. When this method was originally used, the subjective experience was rated along seven-point continuous scales. Both ends of the scale had opposite adjective pairs (e.g., good and bad) and task of the evaluator was to rate his or her experience using the scale. The method is independent on what the object is evaluated as. That is, it can be used to evaluate a wide variety of objects, including user interfaces, buildings, persons, ad infinitum.

Based on the semantic differential method, Bradley and Lang (1994) introduced three nine-point bipolar scales for measuring emotion-related experiences. The scales were: valence, arousal, and dominance. The bipolar scales varied from -4 to +4, so that 0 represented the neutral evaluation. These scales have been used in many HCI studies (Anttonen & Surakka, 2005; Ilves & Surakka, 2013; Salminen et al., 2008). However, these scales are not well suitable for evaluating pointing devices because they focus more on evaluating stimulus that the participant is affected to, rather than on the used device.

That is why Surakka et al. (2004) created a set of six bipolar rating scales that adopted the use of both above theories. The scales were general evaluation (i.e., varies from bad to good), difficulty (from difficult to easy), speed (from slow to fast), accuracy (from inaccurate to accurate), enjoyableness (from not enjoyable to enjoyable), and efficiency (from inefficient to efficient). These scales have been successfully used in studies that new pointing devices have been evaluated (Surakka et al. 2004; 2005). For example, Surakka et al. (2004) reported that the new technique was

rated as more difficult and less accurate to use than the mouse. On the other hand, the new technique was rated as faster than the mouse.

Another possibility is to use independent rating scales that ISO 9241-9 standard (ISO 9241-9 standard, 2000) provides. It is a 7-point Likert scale and is meant for evaluating non-keyboard input devices. The scales include an evaluation of the speed of the operation and mental and physical effort that is required for operation (ISO 9241-9 standard, 2000). The idea behind these scales is to gain some more information on the used pointing device, other than merely objective metrics.

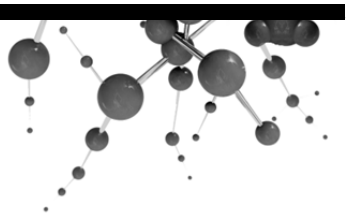
3.3 Interviews

Interviews can provide more in-depth answers than subjective rating scales. In interviews, participants are able to explain their visions in more detail than is possible when using simple rating scales (Vilkko-Riihelä, 1999).

For interviews, there are different possibilities on how to conduct the interview. The questions, of course, need to be created separately for every topic, but the method for conducting the actual interview might change. For a really strict interview, a structured method may be used (DiCicco-Bloom & Crabtree, 2006). A participant is presented only the predefined set of questions, and other questions are not asked.

A freer interview is called a semi-structured interview (DiCicco-Bloom & Crabtree, 2006). This is a type of interview where a predefined set of questions is created similarly as in structured interview. However, the interviewer is allowed to ask clarifying questions or continuations to the questions as she/he thinks are needed in order for the interviewee to explain his/her thoughts on a deeper level. The freest interview is called as unstructured or a freeform interview, in which there are only themes that the participant is presented with.

There are pros and cons for each of these methods. For example, in a structured interview, the answers can easily be analyzed against each other. On the other hand, the information gained from a freeform interview may be more diversified than in a structured interview. On the analysis side, it is of course easier to analyze the answers from a structured interview than from the freeform interview.



4 Introduction to Face Interface and publications

Developing the Face Interface prototype has been iterative. During the process of this thesis work, altogether three versions of the Face Interface prototype have been iteratively developed and experimentally tested. This chapter introduces the development of Face Interface and its publications. The thesis consists of five original publications where the Face Interface prototype has been used as the pointing and selecting device for HCI. Different versions of the Face Interface prototype are introduced, and the five different publications are summarized below:

4.1 PROTOTYPE 1

The first version of Face Interface prototype is shown in Figure 5. The prototype device was built on the frames of protective glasses. The device consisted of a web camera for imaging the eye, an infrared (IR) light source for illumination of the eye, and a capacitive sensor for detecting facial movements resulting from the activation of the *corrugator supercilii*. The eye tracker was especially build for the Face Interface prototype using a commercial, low-cost USB web camera (Creative Live! Cam Notebook) at approximately 30 frames per second. It was slightly modified to operate in the IR wavelengths.

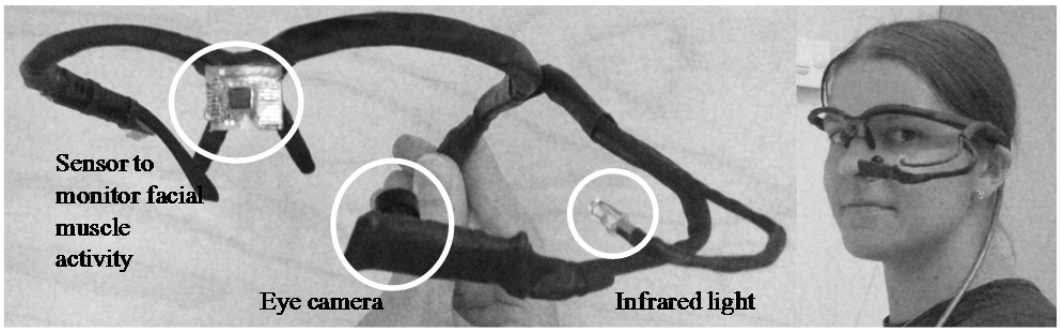


Figure 5. A close-up of the Face Interface prototype (a) and person wearing the prototype (b) (Tuisku et al., 2011 © 2011 IEEE).

The sensor unit (shown in Figure 5) consisted of a programmable controller for capacitive sensors and the electrodes that were on the bottom side of the sensor circuit board. The capacitance was measured at a sampling rate of approximately 90 Hz. Electronics for sending the measurement data to a computer were hidden in the frames of the device on top of the user's right ear. The electronics had a microcontroller that gathered the data from the sensor unit and an RS232 transceiver that provided the connection to a serial port of a computer (Rantanen et al., 2010).

The pupil detection algorithm was implemented to extract the location of the center of the pupil from the camera image in order to calculate the point of gaze on the computer screen from the measured center of the pupil. The use of infrared light ensured that the pupil was represented as the darkest and biggest region in the image. A simple, yet effective, procedure of intensity thresholding was used to find the position of the pupil in the image. First, the image of the eye was thresholded using a pixel grid of size 4 × 4, which was found to be optimal for a given resolution of the video frame (320 × 240 pixels). The output of the thresholding procedure was a binary image, containing areas in which the pixel values were above the threshold. The biggest region—which was located close to the center of the image and corresponded to a number of rules, defining an elliptic shape of the pupil—was selected as a pupil candidate.

The calibration of the eye tracker was done similar to what was proposed by Li et al. (2005) in the OpenEyes project. In particular, the program code from the OpenEyes project was utilized to calculate the approximation coefficients, which defined the matrix of correspondence between the found pupil center and the point of gaze on the screen.

Movement of the eyebrow was detected by finding increasing or decreasing slopes from the capacitive signal. First, the absolute value of the signal derivative was calculated. Then, if the current value of the signal exceeded an adaptive threshold (i.e., 5.0 times the mean of 500

previous samples), a mouse click or, in other words, an object selection was generated.

Publication I: Gazing and Frowning to Computers Can Be Enjoyable

In Publication I, the first version of the Face Interface prototype was introduced and experimentally tested. Because the prototype did not take into account the head movements, a chin-rest was used for preventing involuntary head movements in order to improve the reliability of eye tracking. The chin rest was used to evaluate the possible potential future of the device. Two experiments were ran to investigate the functionality of the Face Interface prototype. In Experiment 1, the participants ($N = 10$) used only the Face Interface. In Experiment 2 ($N = 10$), the eye tracker of the prototype was replaced with a commercial Tobii 1750 eye tracker for pointing, and the same sensor as in the Experiment 1 was used as the selection device. This was done to find out the full potential of the Face Interface technique.

The task of the participants was to perform a simple Fitts' law style pointing and selection task. First, the participant pointed a home square by gaze, and selected it by frowning—and then did the same for target circle. The targets were highlighted when the gaze was inside it; and, after a successful selection of the target, it disappeared. When both targets were successfully selected, there was a pause of 2 seconds, and then the home square and the target circle appeared again in different locations. The target circle appeared in eight directions in relation to the home square. Three pointing distances (60, 120, and 180 mm) and three target circle diameters (25, 30, and 40 mm) were used. Thus, in total, one participant performed 72 (8 × 3 × 3) trials.

The results showed a mean task completion time of 2.5 seconds in Experiment 1 and 1.2 seconds in Experiment 2. The overall mean error rate was 28.5% in Experiment 1 and 9.6% in Experiment 2. Results from Fitts' law showed that the throughput value was 1.4 bits/s for Experiment 1 and 6.3 bits/s for Experiment 2.

These results clearly indicated that, by improving the eye tracker of the prototype significantly, better results could be achieved; because the device that was used to detect the frowning action was the same in both of the experiments. Interestingly, however, the subjective ratings were in a much better level in Experiment 1 than in Experiment 2. Participants of Experiment 1 seemed to clearly like the use of the prototype and rated it on a positive level. There could be many reasons for this, but it was suggested that people might prefer using only one device for interaction. Further, it seems that when the new interaction method functions on a good level, then the ratings might be on a lower level because people start to expect more.

The first experiment on using this type of prototype included both a wearable eye tracker for pointing to objects and a capacitive sensor for selecting objects.

A proof of concept in developing an interaction device that combines eye tracking and capacitive measurement of facial behavior was achieved. The results were encouraging in respect to further development of the device.

4.2 PROTOTYPE 2

A new prototype version was developed on the basis of the results from Publication I. Figure 6 shows the second generation of the Face Interface device. The pupil detection algorithm was improved so that it took into account the corneal reflection as well. A scene camera was added in order to be able to compensate for the head movements so that the chin rest would not be needed as with Prototype 1. The prototype was also made wireless, so that it would allow the freedom of movement to the participants.

Figure 6. Face Interface prototype and person wearing it. Figure adapted from Publication III.

At this stage, the device included two cameras: one for imaging the eye and the other for imaging the computer screen—an infrared light emitting diode for illumination of the eye and to provide the corneal reflection. It also included a sensor device for detecting facial movements using a capacitive method and a shoulder bag, which contained radio frequency devices for wireless operation. The used cameras were commercial low-cost complementary metal oxide semiconductor (CMOS) cameras. The eye camera was a greyscale camera that was modified to image IR wavelengths, and the resolution was 352 × 288 pixels. The scene camera was a color camera with a resolution of 597 × 537 pixels. The frame rate for both of the cameras was 25 frames per second. The eye camera was placed near the user's left eye and the IR light source was placed next to it. The sensor that was used in the capacitance measurement was a programmable capacitance touch sensor (AD7142 by Analog Devices). The sampling frequency for the capacitive sensor was approximately 90 Hz.

The capacitive sensor in the glasses was placed on the bridge of the nose and the scene camera was placed above it. It was able to detect the facial movement resulting either from the activation of the *corrugator supercilii* (activated when frowning) or the *frontalis* (activated when raising the eyebrows) facial muscles for selecting the objects.

The shoulder bag contained a power supply unit for the prototype, two wireless analogue video transmitters, a wireless (serial) transmitter for the capacitance measurement, and four AA batteries. A separate receiving station consisted of two video receivers with a power supply, a radio receiver for the capacitive sensor signal, and two frame-grabbers for the video signals. The radios for both the capacitive measurement and wireless video transmission used the common free frequencies at 2.4 GHz.

For eye tracking, the pupil detection and corneal reflection method was used (Duchowski, 2003). The pupil was detected using the dark pupil method which, in short, detects the darkest ellipse inside the iris as the pupil (Li et al., 2005). Calibration of the eye tracker was again done in a similar manner as in the OpenEyes project (Li et al., 2005).

The scene camera was used to compensate for head movements throughout the experiment. Computer vision library, OpenCV version 2.0 (Bradski & Kaehler, 2008), was utilized to extract features from the image streams of both eye and scene cameras. For the head movement, compensation of the location of six physical markers attached to the computer screen was extracted from the scene in order to track the head orientation in relation to the computer display (see Figure 7 for the placement of the markers). Movement of the eyebrow (i.e., as the selection technique) was again detected by finding increasing or decreasing slopes from the signal of the capacitive sensor (Rantanen et al., 2010).

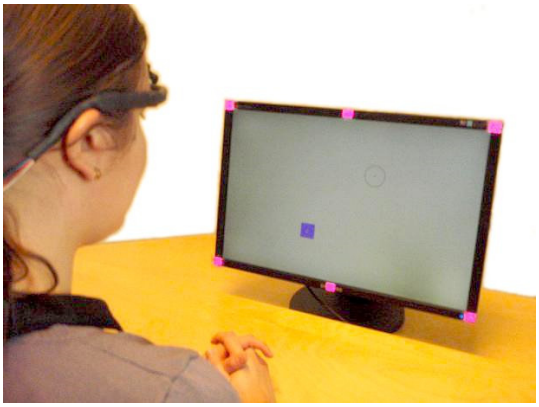


Figure 7. Markers on the screen. Figure adapted from Publication III, reprinted with permission.

Publication II: A Wearable, Wireless Gaze Tracker with Integrated Selection Command Source for Human-Computer Interaction

In this study, the aim was to perform a technical analysis on functioning of this prototype in two conditions: in an electromagnetically shielded laboratory and in a regular office environment. In the laboratory, the lighting was kept constant—and, in the office conditions, the lighting was kept as constant as possible. In office, there was a window with Venetian blinds that were closed if needed—depending on a sun light. In the laboratory, a 15 inch display with a resolution of 1024 × 768 was used and, in the office, a 24 inch widescreen display with a resolution of 1920 × 1200 was used.

Ten participants took part in this study. They performed a same set of pointing and selecting tasks as in Publication I in the office and the laboratory. During the experiments, such eye camera frames—where the corneal reflection was not found—were saved as well as the same frame from the scene camera in order to analyze the percentage on the found corneal reflections. The frame pairs from the eye camera and scene camera were also saved periodically.

The results showed that the Face Interface functioned about equally well in the office conditions as in the laboratory conditions as measured with the accuracy of the eye tracker. The accuracies were 0.67° in the laboratory and 0.79° in the office conditions. Further, it was found that the corneal reflection was found on average with an accuracy of 92% in the office and with an accuracy of 98% in the laboratory. It was discovered that the prototype itself was functional even in the office condition—but, of course, functioned even better when the environment was more stable than the office was.

Publication III: Wireless Face Interface: Using Voluntary Gaze Direction and Facial Muscle Activations for Human-Computer Interaction

The task of the participants was to perform simple pointing and selecting tasks similar to Publication I. In this study, however, seven pointing distances (60, 120, 180, 240, 260, 450, and 520 mm) and three target sizes (25, 30, and 40 mm) were used. These distances were chosen because of the radical growth in display sizes in recent years. For that reason, the distances 260, 450, and 520 were selected so that all eight pointing directions could be used in the edges of the display as well. The participants used either frowning or raising of the eyebrows as the selection technique, according to their own preference.

The results revealed that the most functional area for pointing and selecting with Face Interface was from 60 mm to 260 mm. The results showed that the task completion time was 2.4 seconds on average for the frowning technique; and, in the case of the raising technique, the task completion time was 1.6 seconds on average. Throughput values were 1.9

bits/s for the frowning technique and 5.4 bits/s for the raising of the eyebrows technique.

The error rate was the same 22.8% for both selection techniques. These were clearly improved from the values achieved in Publication I. Thus, an improved pupil detector and the use of a scene camera for head-movement compensation proved to be valuable. The subjective ratings did not reveal any statistically significant differences in between the two selection techniques. The ratings were on a positive level for both techniques, which indicated that the participants were positive about the Face Interface technique. Especially, they rated the use of the Face Interface as interesting and fun. As mentioned, these ratings did not depend on the used selection technique.

4.3 PROTOTYPE 3

The third version of the Face Interface prototype (see Figure 8) was created based on the results from Publications II and III. The amount of sensors to monitor the facial activity was added. This was done in order so that it was possible to investigate the use of different facial activations as the selection technique (i.e., frowning, raising the eyebrows, and smiling). Further, because Publications II and III showed that participants were not able to use both of the facial activations (i.e., frowning and raising the eyebrows), it was clear that the places of the capacitive sensors should be changed. As a result, the places of the capacitive sensors were designed according to the guidelines for EMG studies that were introduced by Fridlund and Cacioppo (1986). That is, the sensors were placed so that they were supported in front of the corresponding facial tissue.

Figure 8. Face Interface prototype: On the left the wearable part and on the right the unit for the wireless functionality. Figure adapted from Publication IV.

The head-worn device included two cameras: one for imaging the eye and the other for imaging the computer screen—an IR light emitting diode for illuminating the eye and providing a corneal reflection, sensors and electronics for detecting facial movements using a capacitive method, and a Class 2 Bluetooth radio (RN-42 by Roving Networks) for serial transmission of the measured capacitance signal. The used cameras were the same ones that were used in Prototype 2. The facial movement sensors were based on capacitance measurement with a programmable controller for capacitance touch sensors (AD7147 by Analog Devices). The capacitive sensors in the frames were placed in front of both eyebrows and cheeks, and one was placed in front of the forehead.

In addition to the head-worn device, a separate carry-on unit to house some components responsible for the wireless operation was included. The unit included a power supply, four AA batteries, and two wireless analogue video transmitters that used the common free frequencies at 2.4 GHz. The PC computer was connected to a receiving station that consisted of two video receivers with a power supply and two frame-grabbers for the video signals. The capacitive signal was received with computer's Bluetooth functionality.

Computer vision library OpenCV version 2.1 (Bradski & Kaehler, 2008) was utilized to extract features from the image streams of both eyes and scene cameras. Pupil detection was based on the corneal reflection method. The algorithm that was used for pupil detection and corneal reflection detection was the same that was thoroughly introduced in Publication II. Calibration of the eye tracker was again done in a similar manner as in the OpenEyes project (Li et al., 2005).

The screen detection algorithm was further improved so that there would be no need to use the physical markers anymore. Head movements in relation to the computer screen were compensated using this screen detection algorithm. The screen detection algorithm aimed to find the frames of a dark rimmed computer display from the scene camera video. The algorithm was based on three observations: First, there were one or two highly contrasted edges that separated the display surface from the surrounding background. The screen is typically brightly illuminated and, thus, lighter than the surroundings. Many monitors have a black frame that surrounds the display surface. Thus, there is a sharp contrast between the illumination of the display surface and the surrounding space (e.g., the monitor frame or background), and there may also be another edge with high contrast between the dark monitor edge and the background. Second, both the display surface and the monitor frame are typically rectangular, which means that they have four straight corners. Third, the corners of the outer border of the monitor frame are relatively close to the corners of the display surface. These three features were used to rank potential screen

candidates to select a best one. For example, a candidate with a dark rimmed border was preferred to one without.

Publication IV: Pointing and Selecting with Facial Activity

The aim of this study was to compare three different facial actions (i.e., frowning, raising the eyebrows, and smiling) as the selection technique when the gaze was used for pointing. In addition, because the dwell time is the most commonly used selection technique when the gaze is used for pointing, a dwell time of 400 ms was used as a reference selection technique. Interestingly, while these selection techniques have been used in pointing and selection tasks, their use has not been compared with each other. While dwell time differs from the facial selection techniques, it is as important to compare the use of it to the facial selection techniques to gain an even deeper insight on how users evaluate it.

The task of the participants was again to perform simple pointing and selecting tasks using three pointing distances (60, 120, and 240 mm) and three target circle diameters (25, 30, and 40 mm). Participants completed two different subjective ratings forms of each of the selection techniques. The first one was the same bipolar rating used in Publications I, III, and V. The second scale was introduced in ISO 9421-9 standard (ISO 9241-9 standard, 2000), and it aimed at giving more perspective on the used prototype.

The results revealed an overall mean pointing task time of 1.4 seconds in the case of frowning, raising the eyebrows, and smiling. In case of the dwell time, the overall mean task time was 1.3 seconds. The error percentages revealed for frowning technique was 22%; for raising technique, 21%; for smiling technique, 16%; and for dwell time technique, 16%.

The overall mean error distances for the frowning technique was 14.9 mm; for the raising technique, 13.8 mm; for the smiling technique, 12.9 mm; and for the dwell time technique, 19.5 mm. The statistical analysis of error distance revealed that the facial selection techniques were significantly more accurate to use than the dwell time selection technique.

Subjective ratings showed that the use of dwell time was rated as significantly easier to use than the frowning technique and the raising technique. Further, the use of the smiling technique was rated as more accurate than the use of the frowning technique and the use of the raising technique. The ISO 9421-9 (ISO 9241-9 standard, 2000) rating scales showed that the ratings were on average above the medium value so that participants generally liked the use of Face Interface. Ratings revealed interesting points, such as the experienced eye fatigue was higher with the dwell time technique than it was with the face-based techniques.

Publication V: Text Entry by Gazing and Smiling

The Publication V aimed at extending the use of Face Interface. For that, text entry with an on-screen keyboard was one reasonable option, because it involves pointing and selecting objects on a computer screen. Most gaze-based text-entry studies are mainly done though on-screen keyboards modeled after the regular QWERTY keyboard layout. Based on the findings on the functionality of Face Interface, there were several considerations in respect to why QWERTY might not be a good solution for text entry with Face Interface. As Publication III revealed, there are some parts in a computer screen—mainly near the edges of the screen—that are difficult to point and select objects with Face Interface. For that reason, a regular on-screen keyboard may not be functional with Face Interface because, for example, it has frequently used characters near the edges such as character "a." This is a problem also with high-end commercial eye trackers (Räihä & Ovaska, 2012).

The aim of this study was two-fold. First, the aim was to design and experimentally test different on-screen keyboard layouts that would be most functional with Face Interface. Another important factor was also the user acceptance of the layouts so that potential users would evaluate the use of them. For that, three different on-screen keyboard layouts were designed. The designed layouts are presented in Figure 9. In all layouts, it was taken into account that it is easier to select objects in the middle than in the edges of the screen. Second, the aim was to compare the use of the Face Interface to the use of a regular computer mouse in entering text on a computer screen. A special feature that was used in both of the experiments was a randomization of the characters on keyboard. That is, the places of characters were randomized after every typed word. The randomization was chosen because it allows the possibility that the participants would be forced to select characters on every part of the keyboard—so that they are able to get a profound opinion on the used layouts. Further, it was used to at least partly cancel out previous experience of the places of the characters. By using QWERTY layout, it would be likely that mouse would outperform any new interaction techniques because off its familiarity.

Figure 9. Three designed on-screen keyboard layouts: Layouts 1, 2, and 3 from left to right, respectively.

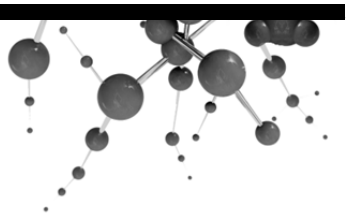
In the first experiment (i.e., the Layout Selection Experiment), the designed layouts were tested with ten participants so that each of the participant entered a word "aurinko" (i.e., "sun" in English) ten times with each of the layout. The order of the layouts was counterbalanced so that every participant did not start with Layout 1, for example. The word "aurinko" was chosen because it is a quite common word and also because every character is different from each other. After the participant had completed the task with each of the layouts, participants rated the used layouts and a short interview was conducted.

The text entry speed was approximately the same: 15 cpm for each layout. Further, the participants rated Layout 2 as clearly the most enjoyable, clearest, and most functional out of the three designed layouts. Also, the results on KSPC metrics supported the subjective ratings, as they showed that the participants were significantly more accurate with Layout 2 than with Layout 1 or Layout 3 (i.e., they needed the least amount of key strokes in order to write the word).

In the second experiment (i.e., the Text Entry Experiment), Layout 2 was used based on the results of the first experiment. The task of the participants (N = 12) was to enter the word "aurinko" 20 times with both Face Interface and mouse. The order of the input devices was counterbalanced. After the participant had finished the task with the first input device, subjective ratings were collected. Finally, the participant was shortly interviewed at the end of the experiment.

The results showed that the mouse was significantly faster in terms of text entry rate (i.e., 19.4 cpm for Face Interface and 27.1 cpm for the mouse). Participants were significantly more accurate with the mouse than with Face Interface in terms of KSPC, because the KSPC value was 1.1 for Face Interface and 1.0 for in the case of the mouse. The MSD error rate revealed a similar result: it was 0.12 for Face Interface and 0.0 for the mouse.

Publication V was the first experiment using gaze direction in conjunction with facial actions for text entry.



5 Discussion

The aim of this thesis has been to iteratively develop and experimentally test the use of the new Face Interface prototype and also to investigate the use of two different modalities for interacting with computers. The concept of Face Interface was investigated in Publication I. The results revealed that the Face Interface prototype was functional. The mean task completion time was 2.5 seconds. And, when the wearable eye tracker was replaced with Tobii 1750 eye tracker, the task completion time was reduced to 1.2 seconds. The error rates were reduced when Tobii 1750 eye tracker was used. The subjective ratings showed that participants clearly liked to use the prototype regardless of its slowness and inaccuracies, as they rated it as easy and fast to use. Thus, it was concluded that the prototype was worth developing further.

For Publications II and III, a new version of the prototype was created in order to further study the Face Interface technique. The results achieved from Publication I was taken into account, and the prototype was improved accordingly. That is, it was imperative that the accuracy of eye tracking needed improvements. In order to do that, a solution for head-movement compensation was needed. A scene camera was chosen for the task, because it had previously been used in wearable eye trackers (Ryan et al., 2008)—although, it had not been used earlier for head-movement compensation. Earlier objects had been identified using a scene camera, but—to identify the screen that user is looking at and for calculating the head-movement compensation—was new idea. The use of the scene camera proved to be functional, as the results in Publications II and III were clearly better than in Publication I. The need to recalibrate the eye tracker was reduced as compared to Publication I.

To add more features to the Face Interface, a third version of the Face Interface prototype introduced the use of smiling as an additional selection technique. Thus, Publication IV aimed at comparing the three different facial selection techniques with each other. Publication V took a more application-oriented approach and, in that, participants entered text by gazing and smiling. All these iterations have led to more functional Face Interface prototypes than the first stage of this thesis.

If we look at the results of Publications I-V, clear improvements in the functioning of the prototype were achieved. The most obvious result is that the overall mean pointing task times reduced gradually from Publication I to Publication IV. The overall completion time reported in Publication I was 2.5 seconds. This was admittedly quite slow, although the participants were all able to accomplish the trials. In Publication III, the reported overall mean completion times were 2.4 seconds for frowning technique and 1.6 seconds for the raising technique. The improvements in the pointing task times resulted from a new pupil detection algorithm that used corneal reflection, and the addition of the head movement compensation algorithm. It was identified that the placement of the capacitive sensor, however, was not the most convenient. Thus, third generation of the prototype able to detect frowning, raising the eyebrows, and smiling related facial movements was created. The task completion times were again improved, as they were 1.4 seconds for each of the facial selection techniques. Thus, the replacement of the capacitive sensors was efficient. Of course, the pointing distances have somewhat varied in between these studies—but, regardless of that, the reduction in pointing task times has been an important achievement in the prototype development.

The achieved task completion times from Publications I, II, and IV compares well with the pointing task times of other similar techniques. For example, Surakka et al. (2004) reported a task completion time of 0.7 seconds. In a follow up study, Surakka et al. (2005) reported a task completion time of 0.5 seconds for the smiling technique and 0.9 seconds for the frowning technique. San Agustin et al. (2009b) reported a task completion time of 0.3 seconds. On the other hand, Chin et al. (2008) reported a task completion time of 4.7 seconds when they used EMG for correcting the inaccuracies that the eye tracker might have. Thus, it seems that the pointing task times achieved from current studies are comparable to other similar studies. Further, when BCI was used as the indication for selection with eye tracker, the task time was reported on being 7.37 seconds on average (Vilimek & Zander, 2009). Based on these pointing task times, it is clear that Face Interface prototype is not yet as fast as those techniques that use EMG, but it is—at this stage—much faster than BCI techniques.

The error percentages got gradually smaller throughout this thesis work. That is, Publication I resulted in an error percentage of 28.5%; Publication III resulted in 22.8%; and, finally, Publication IV resulted in 18.8%, on average. Similar error rates have been reported in studies with the same approach as in the current thesis. The error rates in studies, in which the gaze pointing was combined with EMG selection, has varied between 12% and 27.1% (Mateo et al., 2008; Navallas et al., 2011; San Agustin et al., 2009; Surakka et al., 2004, 2005). While these error rates might seem to be rather high, it must be taken into account that these types of experimental studies are bound to have a quite high error rate. Mainly this is because of the strict definition of an error (i.e., if the first click on the target was not successful, a trial was marked as an erroneous one). In these studies, however, all participants were able to perform all tasks successfully because the trials would not have proceeded before a successful click on the second target. These pointing and selecting studies have been carefully controlled experimental studies; and, as such, there might be a speed-accuracy trade-off. This means that, while these studies investigate the functionality of the pointing and selecting techniques, they are not the types of tasks that people usually/naturally perform while interacting with a computer. Participants might try to perform the tasks fast and not think about the errors. And, of course, from the participants' perspective, they do not see that they are making errors. Thus, it is important to keep in mind that the seemingly high error rates in strictly controlled experimental studies might not reflect the actual use of the pointing device. For example, when Face Interface was used for entering text in Publication V, participants did not make as many errors as measured with MSD error rate and KSPC values as compared to other gaze-based text entry studies (Majaranta et al., 2009; Tuisku et al., 2008).

The collected subjective ratings have been on a same level throughout the course of this thesis work—although, they were on the highest level in Publication I. In Publication I, it was shown that participant rated the use of prototype in all six scales (i.e., general evaluation, difficulty, speed, accuracy, enjoyableness, and efficiency) to the positive end of the scale. In Publications III, IV, and V, however, the bipolar ratings were in a quite neutral level. This is an interesting finding and might indicate that, when the pointing device improves, more is expected from it—and, thus, they are rated on a more neutral level. Differences in ratings were found when four different selection techniques were used in Publication IV, which showed that the smiling was rated as more accurate and faster than the use of frowning or raising the eyebrows. In Publication V, where the use of Face Interface was compared against computer mouse, the mouse was rated as significantly easier, faster, and more accurate than the use of the Face Interface. Quite similar results have been achieved on other similar studies. For example, Surakka et al. (2004) reported that their gazing and frowning technique was rated as faster than the use of the mouse. They

also reported that the mouse was rated as significantly easier and more accurate to use than the facial technique. San Agustin et al. (2009b) reported similarly that the gaze pointing was evaluated as less accurate than mouse pointing. Thus, there is a clear coherence in the subjective ratings in studies where gaze has been used for pointing and facial muscle activation for object selection (regardless of the used activation). This suggests that all the facial activations as a selection technique could be considered to function equally well from participants' standing point.

Similarly as the pointing task times have improved, the Fitts' law based throughput values have improved as well. Publication I showed a throughput value of 1.4 bits/s. Publication III showed throughput values of 1.9 bits/s for the frowning technique and 5.4 bits/s for raising the eyebrows technique. Finally, in Publication IV, the throughput values of 9.13 bits/s for the frowning technique, 8.38 bits/s for the raising technique, 15.33 bits/s for the smiling technique, and 10.24 bits/s for the dwell time technique were found. To compare these to other results, Surakka et al. (2004) reported a throughput value of 12.7 bits/s and San Agustin et al. (2009b) reported an overall value of 3.03 bits/s. However, it is important to realize that the correlation to Fitts' law is an important factor to take into account. This means, that in some gaze-based studies, the correlation to the Fitts' law is found to be quite low. When the correlation to the Fitts' law model is high, the throughput value might be low (and the other way around). This indicates that the gaze-based interaction techniques function the other way around as compared to the traditional pointing devices. It is known that mouse (i.e., hand movements) is faster with short pointing distances than with longer pointing distances. In the case of gaze, for example, Heikkilä and Räihä (2012) have recently shown that longer eye movements are faster than shorter eye movements. A similar effect was also found in Publication I, when the wearable eye tracker was replaced with the commercial one. The computer mouse is a traditional pointing device in a sense that it is faster to point with it at shorter distances than is at longer distances. For these reasons, there have been discussions as to whether the Fitts' law suits (at all) the gaze-based systems and for face-based multimodal systems (Chin et al., 2009). Based on Publications I, III, and IV, it could be argued that the better the gaze-based pointing and selection device (or technique) is, the worse is the correlation to the Fitts' law model.

Table 1 presents each of the key studies that has used gaze for pointing objects and facial activation for object selection together with all three versions of Face Interface.

Table 1. The used devices and mean pointing task times and throughput values for main face-based multimodal studies, including Face Interface. Values that were not reported are marked as not applicable (n/a).

Authors	**Eye tracker**	**Facial measurement device**	**Selection technique**	**Mean task time (s)**	**Through-put (bits/s)**
Surakka et al. 2004	Applied Science Laboratories (ASL) 4000	EMG (electrodes)	frowning	0.7	12.7
Surakka et al. 2005	Tobii 1750	EMG (electrodes)	frowning	0.9	n/a
			smiling	0.5	n/a
San Agustin et al. 2009a	self-built by Public University of Navalla	EMG (Cyberlink™ Headband)	frowning/ jaw tightening	0.3	3.3
Chin et al. 2008*	ASL R6-HS	EMG (electrodes)	jaw clench	4.7	n/a
Tuisku et al. 2011	Face Interface 1	Capacitive sensor	frowning	2.5	1.4
	Tobii 1750			1.2	6.3
Tuisku et al. 2012	Face Interface 2	Capacitive sensor	frowning	2.5	1.9
			raising the eyebrows	1.6	5.4
Tuisku et al. Submitted	Face Interface 3	Capacitive sensor	frowning	1.4	9.13
			raising the eyebrows	1.4	8.38
			smiling	1.4	15.33
			dwell time of 400 ms	1.3	10.24

* Used facial movements to correct inaccuracies of the eye tracker

In this thesis work, the use of three different facial actions as the selection technique was investigated. At first, modeled after the experiment of Surakka et al. (2004), only frowning technique was used as the selection technique. Then, as the Face Interface prototype technique was found to be functional, it was improved according to the results that were attained. The second version of Face Interface made it possible to use either frowning or raising the eyebrows as the selection technique. The raising of the eyebrows technique especially proved to be a well-chosen technique because it was faster than frowning technique in terms of task completion time (i.e., 1.6 seconds vs. 2.4 seconds). In the third prototype version, the amount of capacitive sensors was added so that it was possible to detect the facial movements related to frowning, raising the eyebrows, and smiling.

The facial activations that were used with the Face Interface prototype were short-term and strict movements which are therefore easy to perform. The used movements (i.e., frowning, raising the eyebrows, and smiling) are natural for people to perform involuntarily because they are closely connected to the human emotion system. It is also known that they are easy to perform at will (Surakka & Hietanen, 1998). The present result shows that even for novice users, it is quite fast to learn to produce these simple facial actions that are needed for the Face Interface. Further, Publication IV and Surakka et al. (2005) showed that smiling fits especially well for this type of multimodal task, where only a short activation of the *zygomaticus major* facial muscle is needed. In Publication IV, the participants rated the use of smiling as more accurate to use than the frowning or the raising techniques. Interestingly, it seems that the use of the smiling movement depends on the length of the required movement. That is, there is evidence that keeping a voluntary smile on for longer time period becomes tedious and might be difficult to hold on. For example, Vanhala and Surakka (2007) investigated the intensity of facial muscle activations. The task of the participants was to activate *corrugator supercilii* and *zygomaticus major* facial muscles separately from each other for 30 seconds, using three different intensity levels (low, medium, and high). The results showed that the higher the intensity, the less the participants liked to perform the facial activation. The subjective rating showed that the more intensive the smiling movement, the less enjoyable it was to perform. Thus, in comparison to the tasks where the required activation is short-lasting and less intensive, it seems that the smiling is easier to perform than when the required activation is longer-lasting (Vanhala & Surakka, 2007; Rantanen et al., 2013).

For eye tracking studies, dwell time is and has been the most used selection technique (Jacob, 1991; Ware & Mikaelian, 1987). Mainly the reason might be that it is simple to measure dwell time with an eye tracker, and that there is no other natural selection technique to use when only

gaze is used for pointing. The use of dwell time requires that one needs to hold his or her gaze still on the object to be selected for a certain time period, which might be tedious for some users. Evidence for this was revealed in Publication IV because it was shown that the dwell time was not as accurate to use as were the facial selection techniques in terms of error distance. The error distance analysis showed that the dwell time was the least accurate selection technique as compared to the facial selection technique. This means that with dwell time selection there was significantly more variation in the point where the selection was indicated than with the facial selection techniques. Further, Publication IV showed that the experienced eye fatigue was not as high when using facial activations as the selection technique as compared to the use of dwell time. It was found by Zhang and MacKenzie (2007) that when using only an eye tracker for interacting with computers, eye fatigue becomes a disturbing factor. Thus, for longer use of the gaze-based techniques, the use of the facial selection technique might offer a potential solution against the fatigue.

During this thesis work, it has been shown that the accuracy of the wearable eye tracker is not compromised when using the facial muscle activations for selection technique as compared to the use of dwell time (Publication II, IV). The angular accuracy of the wearable/head-mounted eye tracking was studied on Publication II. The study showed that there could be limitations on the accuracy on the eye tracking when the eye is directed to the extremities on the gaze direction. That is, when the gaze is at its extremities, the pupil is covered by eye lids which makes it virtually impossible to find the pupil and thus, eye tracking is not possible. This was also confirmed in Publication III where the targets at the edges of the display were more difficult to point and select than the targets in the middle of the display. In addition to Publications II and IV, Rantanen et al. (2012b) studied the effect that the selection made by smiling has on the accuracy of the head-mounted eye tracker. Their results showed that by adding an additional modality for the selection does not compromise the accuracy of the eye tracker.

There are several interesting features in combining gaze and face behavior into a multimodal interaction technique. First, eye is primarily a perceptual organ (Zhai, 2003), which means that people are not used to use gaze as an interaction technique. Second, while people actively use their facial muscles when discussing with other people, it is still a rather strange function to use facial muscle activation as a selection technique when interacting with computers (e.g., Barreto et al., 2000; Surakka et al., 2004). Because of these reasons, when using these modalities together for controlling computers, it would intuitively indicate slow functioning. However, as Publications I-V has shown, participants needed only five minutes of practice prior the experiment to learn to use it. This indicates

that learning to use two different modalities together was easy. After a longer practice, it is expected that the use of these two modalities becomes automatic. Each part of the face has its own representation in the motor cortex. For example, lips have a quite large representation on the motor cortex (Penfield & Boldrey, 1937; Penfield & Rasmussen, 1950). Penfield and Boldrey (1937) suggested that the larger the representation, the easier it is to control that part of the face voluntarily. It is likely that, by practice, the muscle representation in brain will gradually evolve—and people are able to more easily to control their muscles.

Two different techniques have been used in the face-based multimodal studies to detect the facial activity: the capacitive sensing (Rantanen et al., 2012b; Publications I-V) and EMG (San Agustin et al., 2009a, 2009b; Surakka et al., 2004, 2005). While these two measurement techniques can be used for the same task, they are profoundly different from each other. EMG measures the electrical activation of facial muscles and can detect even slight changes in the muscles. The capacitive sensor detects the movement of the facial skin that results from the facial muscle activation (Rantanen et al., 2010). Thus, as compared to EMG, there could be a small delay as in indicating the selection because the skin movement that is resulting from the activation of facial muscles is detected as contrast to the electrical activity of the facial muscles. In this dissertation, sampling frequencies of 70-90 Hz has been used for the capacitive method. When this is compared to EMG, in which the sampling frequency can be up to at least 400 Hz (Fridlund & Cacioppo, 1986), it is easy to realize that to same accuracy and speed than with EMG might be difficult to achieve.

The design for the user interface for new interaction techniques needs to take into account the possible difficulties that the pointing technique might have. For example, Publication III showed the most functional area in the computer screen when Face Interface was used as the pointing device. This area was in the center of the screen, for pointing distances from 60 mm to 260 mm. It was shown that the participants had difficulties in hitting the objects in the edges of the display. This means that the most functioning area of the computer display needs to be taken into account when designing UIs for Face Interface. For these reasons, different types of keyboard layouts where designed in order to adapt the functioning of the prototype and UI.

The results of Publication V showed that the three designed layouts proved to be all functional. In a sense, this was not surprising because they all took into account the functionality and limitations that the Face Interface technique has (Publication III). A most functional layout was found among the three layouts. The layout that had larger keys around the edges and smaller keys in the middle of the keyboard (Layout 2 in Figure 9) was evaluated as most enjoyable, clearest, and most functional among

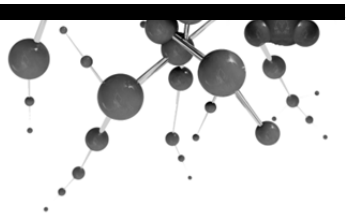
the three layouts. Further, participants needed to use significantly smaller amount of keystrokes when they used this layout, as compared to other layouts. For these reasons, it was selected to be used with Face Interface in the second study. The results of the text-entry experiment showed an overall mean text entry rate of 20 cpm (approximately 4 wpm) for Face Interface and 27 cpm (approximately 5 wpm) for the mouse. This is a relatively nice result when considering the participants had years of experience in using the mouse. The text-entry rate of Face Interface compares well to other gaze-based text-entry tasks, where the layout has been different from the regular QWERTY layout. For example, with Dasher, the first-time users have achieved a text-entry rate of 2.5 wpm. With GazeTalk, the mean text entry rate was reported being 6.22 wpm for Danish text and 11.71 cpm for Japanese text (Hansen et al., 2004). For face-based text entry, Gizatdinova et al. (2012) reported a mean text-entry rate of 3 wpm. For the multimodal text entry, where gaze for pointing and selection was made utilizing BCIs, Yong et al. (2011) reported a text-entry rate of 9.1 cpm. Thus, it can be seen that the results on Publication V compare well with other similar text entry techniques.

To summarize, this thesis has introduced a novel, multimodal pointing and selecting technique for HCI. The achieved results suggest that face-based HCI methods can be competitive future technologies, because they rely on actions that humans use naturally when interacting with other people. In the course of this thesis work, it has been shown that the use of Face Interface is easy to learn—it takes practically only five minutes of practice before it is possible to use Face Interface. As compared to other face-based techniques, Face Interface functions well. Even when compared to techniques where selection is made by hand, Face Interface functions slightly faster (Zhai et al., 1999).

There are many possibilities to use Face Interface in the future. For example, the use of Google Glass (Google Glass, 2013) has become more popular. The design of Google Glass is similar to Face Interface, as it is worn like eyeglasses. By wearing Google Glasses, the user is able, for example, to take a picture of an object she/he is looking at or to find information on an object. Thus, it is easy to realize that Face Interface could add more functionality to the Google Glass so that with Google Glass could detect facial expressions.

Facial muscle actions for selection (i.e., frowning, raising the eyebrows, and smiling)—which were introduced in this thesis—have been short lasting, and thus, relatively easy to perform. These could add value, for example, to videogames. That is, they could be used in games with a different meaning (e.g., raising the eyebrows could mean that information is needed, and frowning could indicate that assistance is needed). With the Face Interface, however, even more functionalities could be offered in

gaming—as the gaze direction could add much functionality in gaming. Thus, it is expected that Face Interface could replace a joystick as a controlling device when playing computer games. It is expected that face-based interaction technique offers many possibilities for future HCI.



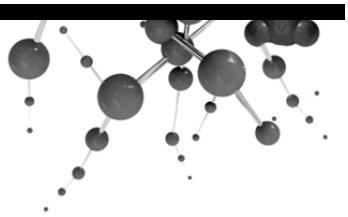
6 Conclusions

The current thesis has iteratively studied a new wearable prototype called Face Interface. It is an eyeglass like device that houses both wearable video-based eye tracker for eye pointing and capacitive sensor(s) to measure facial activity for selecting objects. The facial activations that has been used for selecting objects were: frowning (Publications I, II, III, and IV), raising the eyebrows (Publications II, III, and IV), and smiling (Publications IV and V).

The results of the five original publications have shown that the Face Interface prototype was functional. Three iterations on the Face Interface prototype were introduced and experimentally tested using simple pointing and selecting tasks. Improvements were achieved to the speed and accuracy of the prototype in terms of pointing task time (i.e., from 2.5 seconds in Publication I to 1.3 seconds in Publication V) and in terms of error rates (i.e., from 28.5% in Publication I to 19.7% in Publication IV).

In Publication V, the Face Interface was used for entering text using a specially designed on-screen keyboard. Entering text with Face Interface was compared to entering text with a regular computer mouse. The results showed that, even with a randomized keyboard, first time Face Interface users achieved a text-entry speed of 4 wpm. For the mouse, the text-entry speed was 5 wpm.

The results from five original publications suggest that face-based human-technology interaction methods can be competitive future technologies, because they rely on actions that humans use naturally when interacting with other people. The results indicate that Face Interface is a promising real multimodal technique for future HCI.



7 References

Anttonen, J., and Surakka, V. (2005). Emotions and heart rate while sitting on a chair. In *Proceedings of the CHI 2005*, ACM Press, 491-499.

Aoki, H., Hansen, J. P., and Itoh, K. (2008). Learning to interact with a computer by gaze. *Behaviour & Information Technology*, 27(4), 339-344.

Ashmore, M., Duchowski, A. T., and Shoemaker, G. (2005). Efficient eye pointing with a fisheye lens. In *Proceedings of the Graphics Interface 2005*, ACM Press, 203-210.

Ashtiani, B., and MacKenzie, I. S. (2010). BlinkWrite2: An improved text entry method using eye blinks. In *Proceedings of the ETRA 2010*, ACM Press, 339-345.

Atkinson, R. C., and Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Ed.): *The psychology of learning and motivation: Advances in research and theory*, New York, Academic Press, 89-195.

Babcock, J. S., and Pelz, J. B. (2004). Building a lightweight eyetracking headgear. In *Proceedings of the ETRA 2004*, ACM Press, 109-114.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417-423.

Baddeley, A. D., and Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.): *The psychology of learning and motivation*. New York: Academic Press, 47-89.

Barreto, A. B., Scargle, S. D., and Adjouadi, M. (2000). A practical EMG-based human-computer interface for users with motor disabilities. *Journal of Rehabilitation Research & Development*, 37(1), 53-63.

Baxter, G., and Sommerville, I. (2011). Socio-technical systems: For design methods to systems engineering. *Interacting with Computers*, 23(1), 4-17.

Bi, X., Smith, B. A., and Zhai, S. (2010). Quasi-Qwerty soft keyboard optimization. In *Proceedings of the CHI 2010*, ACM Press, 283-286.

Bradley, M., and Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49-59.

Bradski, G., and Kaehler, A. (2008). *Learning opencv: Computer vision with the opencv library*.O'Reilly Media, Sebastopol, California, USA.

Bulling, A., and Gellersen, H. (2010). Toward mobile eye-based human-computer interaction.*Pervasive Computing*, 9(4), 8-12.

Bulling, A., Roggen, D., and Tröster, G. (2009). Wearable EOG goggles: Eye-based interaction in everyday environments. In *Proceedings of the CHI 2009: Extended Abstracts*, ACM Press, 3259-3264.

Bulling, A., Ward, J. A., and Gellersen, H. (2012). Multimodal recognition of reading activity in transit using bodyworn sensors. *ACM Transactions on Applied Perceptions*, 9(1), Article 2, 21 pages.

Card, S. K., English, W. K., and Burr, B. J. (1978). Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. *Ergonomics*, 21(8), 601-613.

Chen, X., Zhang, X., Zhao, Z.-Y., Yang, J.-H., Lantz, V., and Wang, K.-Q. (2007). Multiple hand gesture recognition based in surface EMG signal. In *Proceedings of the 1st International Conference on Bioinformatics and Biomedical Engineering*, IEEE, 506-509.

Chin, C. A., and Barreto A. (2006). Hands-free manipulation of the computer cursor based on the electromyogram. In *Proceedings of the Florida Conference on Recent Advances in Robotics (FCRAR 2006)*, 6 pages.

Chin, C. A., Barreto A., and Adjouadi M. (2006). Enhanced real-time cursor control algorithm, based on the spectral analysis of electromyograms. *Biomedical Science Instrumentation*, 42, 249-254.

Chin, C. A., Barreto A., and Adjouadi M. (2009). Integration of EMG and EGT modalities for the development of an enhanced cursor control system. *International Journal on Artificial Intelligence Tools*, 18(3), 399-414.

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

Davidson, R. J., Jackson, D. C., and Larson, C. L. (2000). Human electroencephalography. In J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson (Eds.): *Handbook of Psychophysiology*. Cambridge University Press, 27-52.

De Silva, G. C., Lyons, M. J., Kawato, S., and Tetsutani, N. (2003). Human factors evaluation of a vision-based facial gesture interface. In *Proceeding of the 2003 Conference on Computer Vision and Pattern Recognition Workshop*, IEEE, 52.

DiCicco-Bloom, B., and Crabtree, B. F. (2006). The qualitative research interview. *Medical Education*, 40(4), 314-321.

Dimberg, U. (1990). Facial electromyography and emotional reactions. *Psychophysiology*, 27(5), 481-494.

D'Mello, S., and Kory, J. (2012). Consistent but modest: A meta-analysis on unimodal and multimodal affect detection accuracies from 30 studies. In *Proceedings of the ICMI 2012*, ACM Press, 31-38.

Duchowski, A. T. (2002). A breadth-first survey of eye tracking applications. *Behavior Research Methods, Instruments, & Computers*, 34(4), 455-470.

Duchowski, A. T. (2003). *Eye Tracking Methodology: Theory and Practice*. Springer-Verlag.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6(3/4), 169-200.

Ekman, P., and Davidson, D. J. (1993). Voluntary smiling changes regional brain activity. *Psychological Science*, 4(5), 342-345.

English, W. K., Engelbart, D. C., and Berman, M. L. (1967). Display-selection technique for text manipulation. *IEEE Transactions in Electronics*, HFE-8(1), 5-15.

Eriksen, C. W., and St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, 40(4), 225-240.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6), 381-391.

Franchak, J. M., Kretch, K. S., Soska, K. C., Babcock, J. S., and Adolph, K. E. (2010). Head-mounted eye-tracking of infants' natural interactions: a new method. In *Proceedings of the ETRA 2010*, ACM Press, 21-27.

Fridlund, A. J. (1991). Evolution and facial action in reflex, social motive, and paralanguage. *Biological Psychology*, 32(1), 3-100.

Fridlund, A. J., and Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology*, 23(5), 567-589.

Gibson, J. J. (1950). *The perception of the visual world*. Houghton Mifflin, Boston, MA.

Gizatdinova, Y., Špakov, O., and Surakka, V. (2012). Comparison of video-based pointing and selection techniques for hands-free text entry. In *Proceedings of the AVI 2012*, ACM Press, 132-139.

Google Glass. (2013). http://www.google.com/glass/start/ (Checked 19.11.2013)

Hansen, J. P., Johansen, A. S., Hansen, D. W., Itoh, K., and Mashino, S. (2003). Command without a click: Dwell time typing by mouse and gaze selections. In M. Rauterberg, M. Menozzi, and J. Wesson (Eds.): *Human – Computer Interaction, INTERACT '03*, IOS Press, 121-128.

Hansen, J. P., Tørning, K., Johansen, A. S., Itoh, K., and Aoki, H. (2004). Gaze typing compared with input by head and hand. In *Proceedings of the ETRA 2004*, ACM Press, 131-138.

Hautala, J., Hyönä, J., Aro, M., and Lyytinen, H. (2011). Sublexical effects on eye movements during repeated reading of words and pseudowords in Finnish. *Psychology of Language and Communication*, 15(2), 129-149.

Heikkilä, H., and Räihä, K.-J. (2009). Speed and accuracy of gaze gestures. *Journal of Eye Movement Research*, 3(2), 1-14.

Heikkilä, H., and Räihä, K.-J. (2012). Simple gaze gestures and the closure of the eyes as an interaction technique. In *Proceedings of the ETRA 2012*, ACM Press, 147-154.

Helmert, J. R., Pannasch, S., and Velichkovsky, B. M. (2008). Influences of dwell time and cursor control on the performance in gaze driven typing. *Journal of Eye Movement Research*, 2(4), 1-8.

Hietanen, J. K., Surakka, V., and Linnankoski, I. (1998). Facial electromyographic responses to vocal affect expressions. *Psychophysiology*, 35(5), 530-536.

Holleis, P., Schmidt, A., Paasovaara, S., Puikkonen, A., and Häkkilä, J. (2008). Evaluating capacitive touch input on clothes. In *Proceedings of the MobileHCI*, ACM Press, 81-90.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., and van de Weijer, H. (2011). *Eye Tracking - A Comprehensive Guide to Methods and Measures*, Oxford University Press.

Hyönä, J. (2009). Foveal and parafoveal processing during reading. In S. P. Liversedge, I. D. Gilchrist, and S. Everling (Eds.): *The Oxford Handbook of Eye Movements*, Oxford University Press, 819-838.

Hyönä, J., and Niemi, P. (1990). Eye movements during repeated reading of a text. *ActaPsychologica*, 73(3), 259-280.

ISO 9241-9:2000 (2000). Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9: Requirements for non-keyboard input devices, CEN.

Ilves, M., and Surakka, V. (2013). Subjective responses to synthesised speech with lexical emotional content: The effect of the naturalness of the synthetic voice. *Behaviour & Information Technology*, 32(2), 117-131.

Isokoski, P., and Raisamo, R. (2004). Speed and accuracy of six mice. *Asian Information-Science-Life*, 2(2), 131-140.

Jacob, R. J. K. (1991). The use of eye movements in human-computer interaction techniques: What you look is what you get. *ACM Transactions on Information Systems*, 9(3), 152-169.

Jacob, R. J. K. (1995). Eye tracking in advanced interface design. In W. Barfield, and T. A. Furness (Eds.): *Virtual Environments and Advanced Interface Design*. New York: Oxford University Press, 258-288.

Jacob, R. J. K., and Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In Hyönä, J., Radach, R., and Deubel, H. (Eds.): *The Mind's Eyes: Cognitive and Applied Aspects of Oculomotor Research*. Elsevier Science, Oxford, 573-605.

Jaimes, A., and Sebe, N. (2007). Multimodal human-computer interaction: A survey. *Computer Vision and Image Understanding*, 108(1-2), 116-134.

Johansen, S. A., San Agustin, J., Skovsgaard, H., Hansen, J. P., and Tall, M.(2011). Low cost vs. high-end eye tracking for usability testing. In *Proceedings of the CHI 2011: Extended Abstracts*, ACM Press, 1177-1182.

Just, M. A., and Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329-354.

Kim, J.-S., Jeong, H., and Son, W. (2004). A new means of HCI: EMG-mouse. In *Proceedings of the International Conference on Systems, Man and Cybernetics*, IEEE, 100-104.

Klein, C., and Ettinger, U. (2008). A hundred years of eye movement research in psychiatry. *Brain and Cognition*, 68(3), 215-218.

Królak, A., and Strumiłło, P. (2012). Eye-blink detection system for human-computer interaction. *Universal Access in the Information Society*, 11(4), 409-219.

Kübler, A., Kotchoubey, B., Hinterberger, T., Ghanayim, N., Perelmouter, J., Schauer, M., Fritsch, C., Taub, E., and Birbaumer, N. (1999). The thought translation device: A neurophysiological approach to communication in total motor paralysis. *Experimental Brain Research*, 124(2), 223-232.

Li, D., Babcock, J., and Parkhurst, D. J. (2006). Open eyes: A low-cost head-mounted eye-tracking solution. In *Proceedings of the ETRA 2006*, ACM Press, 95-100.

Li, D., Winfield, D., and Parkhurst, D. J. (2005). Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches. In *Proceedings of the IEEE Vision for Human-Computer Interaction Workshop at CVPR*, 1-8.

Lyons, E. C., Barreto, A. B., and Adjouadi, M. (2001). Development of a hybrid hands-off human computer interface based on electromyogram signals and eye-gaze tracking. In *Proceedings of 23rd Annual EMBS International Conference*, IEEE, 1423-1426.

MacKenzie, I. S. (1992). Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction*, 7(1), 91-139.

MacKenzie, I. S., and Buxton, W. (1992). Extending Fitts' law to two-dimensional tasks. In *Proceedings of the CHI 1992*, 219-226.

MacKenzie, I. S., and Isokoski, P. (2008). Fitts' throughput and the speed-accuracy tradeoff. In *Proceedings of the CHI 2008*, ACM Press, 1633-1636.

MacKenzie, I. S., and Soukoreff, R. W. (2003). Card, English, and Burr (1978) - 25 years later. In *Proceedings of the CHI 2003*, ACM Press, 760-761.

MacKenzie, I. S., and Zhang, S. X. (1999). The design and evaluation of a high-performance soft keyboard. In *Proceedings of the CHI 1999*, ACM Press, 25-31.

Majaranta, P., Ahola, U.-K., and Špakov, O. (2009). Fast gaze typing with an adjustable dwell time. In *Proceedings of the CHI 2009*, ACM Press, 357-360.

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

Majaranta, P., and Räihä, K.-J. (2002). Twenty years of eye typing: Systems and design issues. In *Proceedings of the ETRA 2002*, ACM Press, 15-22.

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

Mateo, J. C., San Agustin, J., and Hansen, J. P. (2008). Gaze beats mouse: hands-free selection by combining gaze and EMG. In *Proceedings of the CHI 2008: Extended Abstracts*, ACM press, 3039-3044.

Matlin, M. W. (2009). *Cognitive Psychology*, 7th Edition. Asia: John Wiley & Sons, Inc.

Mehrabian, A. (1981). *Silent messages: Implicit communication of emotions and attitudes*. Wadsworth, Belmont, California.

Møllenbach, E., Lillholm, M., Gail, A., and Hansen, J. P. (2010). Single gaze gestures. In *Proceedings of the ETRA 2010*, ACM Press, 177-180.

Navallas, J., Ariz, M., Villanueva, A., San Agustin, J., and Cabeza, R. (2011). Optimizing interoperability between video-oculographic and electromyographic systems. *Journal of Rehabilitation Research & Development*, 48(3), 254-266.

Noris, B., Keller, J.-B., and Billard, A. (2011). A wearable gaze tracking system for children in unconstrained environments. *Computer Vision and Image Understanding*, 115(4), 476-486.

Osgood, C. E. (1952). Nature and measurement of meaning. *Psychological Bulletin*, 49(3), 197-237.

Oulasvirta, A., Reichel, A., Li, W., Zhang, Y., Bachynskyi, M., Vertanen, K., and Kristensson, P. O. (2013). Improving two-thumb text entry on touchscreen devices. In *Proceedings of the CHI 2013*, ACM Press, 2765-2774.

Partala, T., Aula, A., and Surakka, V. (2001). Combined voluntary gaze direction and facial muscle activity as a new pointing technique. In *Proceedings of the INTERACT 2001*, IOS Press, 100-107.

Partala, T., Surakka, V., and Vanhala, T. (2006). Real-time estimation of emotional experiences from facial expressions. *Interacting with Computers*, 18(2), 208-226.

Penfield, W., and Boldrey, E. (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain*, 60(4), 389-443.

Penfield W., and Rasmussen, T. (1950). *The cerebral cortex of man: A clinical study of localization of function*. New York: Macmillan.

Picard, R. (1997). *Affective Computing*. Massachusetts: The MIT Press.

Porta, M., and Turina, M. (2008). Eye-S: A full-screen input modality for pure eye-based communication. In *Proceedings of the ETRA 2008*, ACM Press, 27-34.

Posner, M. I., Snyder, C. R. R., and Davidson, B. J. (1980). Attention and detection of signals. *Journal of Experimental Psychology: General®*, 109, 160-174.

Rantanen, V., Kumpulainen, P., Venesvirta, H., Verho, J., Špakov, O., Lylykangas, J., Vetek, A., Surakka, V., and Lekkala, J. (2012a). Capacitive facial activity measurement. In *Proceedings of XX IMEKO World Congress*, Busan, South Korea, 6 pages.

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

Rantanen, V., Venesvirta, H., Špakov, O., Verho, J., Vetek, A., Surakka, V., and Lekkala, J. (2013). Capacitive measurement of facial activity intensity. *IEEE Sensors Journal*, 13(11), 4329-4338.

Rantanen, V., Verho, J., Lekkala, J., Tuisku, O., Surakka, V., and Vanhala, T. (2012b). The effect of clicking by smiling on the accuracy of head-mounted gaze tracking. In *Proceedings of the ETRA 2012*, ACM Press, 345-348.

Rinn, W. E. (1984). The neuropsychology of facial expression: A review of the neurological and psychological mechanisms for producing facial expressions. *Psychological Bulleting*, 95(1), 52-77.

Ryan, W. J., Duchowski, A. T., and Birchfield, S. T. (2008). Limbus/pupil switching for wearable eye tracking under variable lighting conditions.In *Proceedings of the ETRA 2008*, ACM Press, 61-64.

Rymarczyk, K., Biele, C., Grabowska, A., and Majczynski, H. (2011). EMG activity in response to static and dynamic facial expressions.*International Journal of Psychophysiology*, 79(2), 330-333.

Räihä, K.-J., and Ovaska, S. (2012). An exploratory study of eye typing fundamentals: dwell time, text entry rate, errors, and workload. In *Proceedings of the CHI 2012,* ACM Press, 3001-3010.

Salminen, K., Surakka, V., Lylykangas, J., Raisamo, J., Saarinen, R., Raisamo, R., Rantala, J., and Evreinov, G. (2008). Emotional and behavioral responses to haptic stimulation. In *Proceedings of the CHI 2008*, ACM Press, 1555-1562.

San Agustin, J., Hansen, J. P., Hansen, D. W., and Skovsgaard, H. (2009a). Low-cost gaze pointing and EMG clicking. In *Proceedings of the CHI 2009*, ACM Press, 3247-3252.

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

Sharmin, S., Špakov, O., and Räihä, K.-J. (2012). The effect of different text presentation formats on eye movement metrics in reading. *Journal of Eye Movement Research*, 5(3), 1-9.

Sibert, L. E., and Jacob, R. J. K. (2000). Evaluation of eye gaze interaction. In *Proceedings of the CHI 2000*, ACM Press, 281-288.

Soukoreff, R. W., and MacKenzie, I. S. (2003). Metrics for text entry research: an evaluation of MSD and KSPC, and a new unified error metric. In *Proceedings of the CHI 2003*, ACM Press, 113-120.

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

Stark, L., Vossius, G., and Young, L. (1962).Predictive control of eye tracking movements. *IRE Transactions on Human Factors in Electronics*, HFE-3, 52-57.

Surakka, V., and Hietanen, J. K. (1998). Facial and emotional reactions to Duchenne and non-Duchenne smiles. *International Journal of Psychophysiology*, 29(1), 23-33.

Surakka, V., Illi, M., and Isokoski, P. (2003). Voluntary eye movements in human-computer interaction. In J. Hyönä, R. Radach, and H. Deubel (Eds.): *The Mind's Eyes: Cognitive and Applied Aspects of Oculomotor Research*. Elsevier Science: Oxford, 473-491.

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

Surakka, V., Isokoski, P., Illi, M., and Salminen, K. (2005). Is it better to gaze and frown or gaze and smile when controlling user interfaces? In *Proceedings of the HCI International 2005*, CD-Rom, 7 pages.

Surakka, V., and Vanhala, T. (2011). Emotions in human-computer interaction. In A. Kappas, and N. C. Krämer (Eds.): *Face-to-Face Communication over the Internet*. Cambridge, UK: Cambridge University Press, 213-236.

Tan, D. S., Gergle, D., Scupelli, P., and Pausch, R. (2006). Physically large displays improve performance on spatial tasks. *ACM Transactions on Computer-Human Interaction*, 13(1), 71-99.

Tuisku, O., Majaranta, P., Isokoski, P., and Räihä, K.-J. (2008). Now Dasher! Dash away! Longitudinal study of fast text entry by eye gaze. In *Proceedings of the ETRA 2008*, ACM Press, 19-26.

Valtonen, M., Kaila, L., Mäentausta, J., and Vanhala, J. (2011). Unobtrusive human height and posture recognition with a capacitive sensor. *Journal of Ambient Intelligence and Smart Environments*, 3(4), IOS Press, 305-332.

Vanhala, T., and Surakka, V. (2007). Facial activation control effect (FACE). In *Proceedings of the ACII 2007*, Lecture Notes in Computer Science, 4738, Springer, 278-289.

Vanhala, T., and Surakka, V. (2008). Computer-assisted regulation of emotional and social processes. In J. Or (Ed.): *Affective Computing: Focus on Emotion Expression, Synthesis, and Recognition*. I-Tech Education and Publishing: Vienna, Austria, 405-420.

Vanhala, T., Surakka, V., Courgeon, M., and Martin, J.-C. (2012). Voluntary facial activations regulate physiological arousal and subjective experiences during virtual social stimulation. *ACM Transactions on Applied Perception*, 9(1), Article 1, 21 pages.

Vanhala, T., Surakka, V., Siirtola, H., Räihä, K.-J., Morel, B., and Ach, L. (2010). Virtual proximity and facial expressions of computer agents regulate human emotions and attention. *Computer Animation and Virtual Worlds*, 21(3-4), 215-224.

Vehkaoja, A., Verho, J., Puurtinen, M., Nöjd, N., Lekkala, J., and Hyttinen, J. (2005). Wireless head cap for EOG and facial EMG measurements. In *Proceedings of the 2005 IEEE Engineering in Medicine and Biology*, IEEE, 5865-5868.

Vertegaal, R. (1999). The GAZE groupware system: Mediating joint attention in multiparty communication and collaboration. In *Proceedings of the CHI 1999*, ACM Press, 294-301.

Vilimek, R., and Zander, T. O. (2009). BC(eye): Combining eye-gaze input with brain-computer interaction. In C. Stephanidis (Ed.): *Universal Access in HCI, Part II, HCII 2009*. LNCS 5615, 593-602.

Vilkko-Riihelä, A. (1999). *PSYYKE – Psykologian käsikirja*.WSOY.

Villanueva, A., Cabeza, R., and Porta, S. (2004). Eye tracking system with easy calibration. In *Proceedings of the ETRA 2004*, ACM Press, 55.

Wade, N. J., and Tatler, B. W. (2009). Origins and applications of eye movement research. In S. P. Liversedge, I. D. Gilchrist, and S. Everling (Eds.): *The Oxford Handbook of Eye Movements*. Oxford University Press, 17-43.

Ward, D. J., and MacKay, D. J. C. (2002). Fast hands-free writing by gaze direction. *Nature*, 418(6900), 838.

Ware, C., and Mikaelian, H. H. (1987). An evaluation of an eye tracker as a device for computer input. In *Proceedings of the CHI 1987*, ACM Press, 183-188.

Whisenand, T. G., and Emurian, H. H. (1996). Effects of angle approach on cursor movement with a mouse: Consideration of Fitts' law. *Computers in Human Behavior*, 12(3), 481-495.

Whitman, D. (2011). *Cognition*. United States of America: John Wiley & Sons, Inc.

Wimmer, R., and Baudisch, P. (2011). Modular and deformable touch-sensitive surfaces based on time domain reflectometry. In *Proceedings of the UIST 2011*, ACM Press, 517-526.

Wobbrock, J. O. (2007). Measures of text entry performance. In I. S. MacKenzie, and K. Tanaka-Ishii (Eds.): *Text Entry Systems: Mobility, Accessibility, Universality*. Morgan Kaufmann, 47-74.

Wobbrock, J. O., Rubinstein, J., Sawyer, M. W., and Duchowski, A. T. (2008). Longitudinal evaluation of discrete consecutive gaze gestures for text entry. In *Proceedings of the ETRA 2008*, ACM Press, 11-18.

Wolpaw, J. R. (2007). Brain-computer interfaces as new brain output pathways. *Journal of Physiology*, 579(3), 613-619.

Yong, X., Fatourechi, M., Ward, R. K., and Birch, G. E. (2011). The design of a point-and-click system by integrating a self-paced brain-computer interface with an eye-tracker. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 1(4), 590-602.

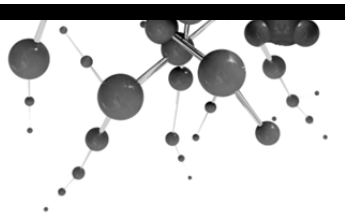
Zander, T. O., Gaertner, M., Kothe, C., and Vilimek, R. (2010). Combining eye gaze input with a brain-computer interface for touchless human-computer interaction. *International Journal of Human-Computer Interaction*, 27(1), 38-51.

Zhai, S. (2003). What's in the eyes for attentive input. *Communications of the ACM*, 46(3), 34-39.

Zhai, S. (2004). Characterizing computer input with Fitts' law parameters - The information and non-information aspects of pointing. *International Journal of Human-Computer Studies*, 61(6), 791-809.

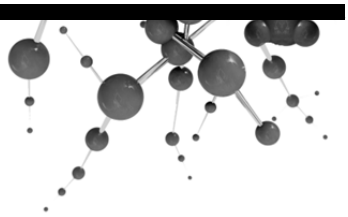
Zhai, S., Morimoto, C., and Ihde, S. (1999). Manual and gaze input cascaded (MAGIC) pointing. In *Proceedings of the CHI 1999*, ACM Press, 246-253.

Zhang, X., and MacKenzie, I. S. (2007).Evaluating eye tracking with ISO 9241 - Part 9. In J. Jacko (Ed.): *Human-Computer Interaction, Part III, HCII 2007*. LNCS 4552, 779–788.



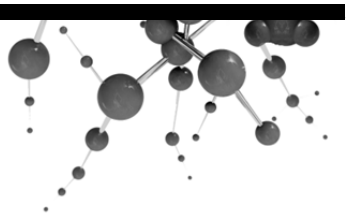
Publication I

Tuisku, O., Surakka, V., Gizatdinova, Y., Vanhala, T., Rantanen, V., Verho, J., and Lekkala, J. (2011). Gazing and Frowning to Computers Can Be Enjoyable. In *Proceedings of the Third International Conference on Knowledge and Systems Engineering, KSE 2011 (Hanoi, Vietnam)*, October 2011, IEEE Computer Society, pages 211-218.



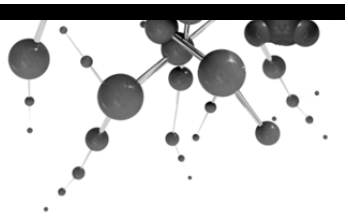
Publication II

Rantanen, V., Vanhala, T., Tuisku, O., Niemenlehto, P.-H., Verho, J., Surakka, V., Juhola, M., and Lekkala, J. (2011). A Wearable, Wireless Gaze Tracker with Integrated Selection Command Source for Human-Computer Interaction. *IEEE Transactions on Information Technology in BioMedicine*, 15(5), 795-801.



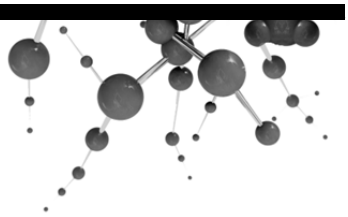
Publication III

Tuisku, O., Surakka, V., Vanhala, T., Rantanen, V., and Lekkala, J. (2012). Wireless Face Wireless: Using Voluntary Gaze Direction and Facial Muscle Activations for Human-Computer Interaction. *Interacting with Computers*, 24(1), 1-9.



Publication IV

Tuisku, O., Rantanen, V., Špakov, O., Surakka, V., and Lekkala, J. (Submitted). Pointing and Selecting with Facial Activity. Revised version submitted to *Interacting with Computers*.



Publication V

Tuisku, O., Surakka, V., Rantanen, V., Vanhala, T., and Lekkala, J. (2013). Text Entry by Gazing and Smiling. *Advances in Human-Computer Interaction*, Article ID 218084, 13 pages.

1. **Timo Partala:** Affective Information in Human-Computer Interaction
2. **Mika Käki:** Enhancing Web Search Result Access with Automatic Categorization
3. **Anne Aula:** Studying User Strategies and Characteristics for Developing Web Search Interfaces
4. **Aulikki Hyrskykari:** Eyes in Attentive Interfaces: Experiences from Creating iDict, a Gaze-Aware Reading Aid
5. **Johanna Höysniemi:** Design and Evaluation of Physically Interactive Games
6. **Jaakko Hakulinen:** Software Tutoring in Speech User Interfaces
7. **Harri Siirtola:** Interactive Visualization of Multidimensional Data
8. **Erno Mäkinen:** Face Analysis Techniques for Human-Computer Interaction
9. **Oleg Špakov:** iComponent - Device-Independent Platform for Analyzing Eye Movement Data and Developing Eye-Based Applications
10. **Yulia Gizatdinova:** Automatic Detection of Face and Facial Features from Images of Neutral and Expressive Faces
11. **Päivi Majaranta:** Text Entry by Eye Gaze
12. **Ying Liu:** Chinese Text Entry with Mobile Devices
13. **Toni Vanhala:** Towards Computer-Assisted Regulation of Emotions
14. **Tomi Heimonen:** Design and Evaluation of User Interfaces for Mobile Web Search
15. **Mirja Ilves:** Human Responses to Machine-Generated Speech with Emotional Content
16. **Outi Tuisku:** Face Interface