Interacting with Computers 24 (2012) 1–9

Contents lists available at SciVerse ScienceDirect

Interacting with Computers

journal homepage: www.elsevier.com/locate/intcom

Wireless Face Interface: Using voluntary gaze direction and facial muscle activations for human–computer interaction ☆

Outi Tuisku [a,*], Veikko Surakka [a], Toni Vanhala [a,c], Ville Rantanen [b], Jukka Lekkala [b]

[a] Research Group for Emotions, Sociality, and Computing, Tampere Unit for Computer–Human Interaction (TAUCHI), School of Information Sciences, University of Tampere, Kanslerinrinne 1, FI-33014 University of Tampere, Finland
[b] Sensor Technology and Biomeasurements, Department of Automation Science and Engineering, Tampere University of Technology, P.O. Box 692, FI-33101 Tampere, Finland
[c] VTT Technical Research Centre of Finland, Tekniikankatu 1, P.O. Box 1300, FI-33101 Tampere, Finland

ARTICLE INFO

Article history:
Received 27 January 2011
Received in revised form 10 October 2011
Accepted 20 October 2011
Available online 29 October 2011

Keywords:
Eye tracking
Facial muscle activations
Human–computer interaction
Fitts' law

ABSTRACT

The present aim was to investigate the functionality of a new wireless prototype called Face Interface. The prototype combines the use of voluntary gaze direction and facial muscle activations, for pointing and selecting objects on a computer screen, respectively. The subjective and objective functionality of the prototype was evaluated with a series of pointing tasks using either frowning (i.e., frowning technique) or raising the eyebrows (i.e., raising technique) as the selection technique. Pointing task times and accuracies were measured using three target diameters (i.e., 25, 30, 40 mm), seven pointing distances (i.e., 60, 120, 180, 240, 260, 450, and 520 mm), and eight pointing angles (0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°). The results showed that the raising technique was faster selection technique than the frowning technique for the objects that were presented in the pointing distances from 60 mm to 260 mm. For those pointing distances the overall pointing task times were 2.4 s for the frowning technique, and 1.6 s for the raising technique. Fitts' law computations showed that the correlations for the Fitts' law model were $r = 0.77$ for the frowning technique and $r = 0.51$ for the raising technique. Further, the index of performance (IP) value was 1.9 bits/s for the frowning technique and 5.4 bits/s for raising the eyebrows technique. Based on the results, the prototype functioned well and was adjustable so that two different facial activations can be used in combination with gaze direction for pointing and selecting objects on a computer screen.

1. Introduction

Several different techniques that utilize signals from the human head area have been developed and tested for human–computer interaction (HCI) (Barreto et al., 2000; Chin et al., 2008; Jacob, 1991; Majaranta and Räihä, 2002, 2007; Millán et al., 2010; Surakka et al., 2004, 2005). These methods include brain–computer interfaces (BCIs) which measure the electrical activity from the human brain (Millán et al., 2010; Wolpaw et al., 2002). Other methods include gaze based methods (Majaranta and Räihä, 2002, 2007; Sibert and Jacob, 2000), electro-oculography (EOG) based methods (Bulling et al., 2009), and facial electromyography (EMG) based methods (Barreto et al., 2000). In techniques that utilize the gaze direction as the computer controlling method, eye trackers are used to measure the user's point of gaze on a computer screen. Video-based eye trackers are normally used to image the user's eye and then software algorithms are used to transform the imaged eye movements to cursor movements on a computer screen. Methods that utilize facial EMG measure the level of the electrical activity of the facial muscles. A relatively recent approach is to combine the use of two modalities measured from the human face area. Surakka et al. (2004) developed a method in which voluntarily controlled eye movements and voluntarily generated changes in the level of electrical activity of the facial muscles were used for pointing and selecting graphical objects, respectively.

The performance of the above multimodal technique was evaluated using simple pointing and selection tasks with three target circle diameters (i.e., 25, 30, 40 mm) and three pointing distances (i.e., 60, 120, 180 mm) to evaluate the potential of the new interaction technique (Surakka et al., 2004, 2005). Their results showed that the technique worked well in comparison to the computer mouse. The pointing task time analysis revealed that the use of mouse was significantly faster than the new technique only at the shortest pointing distances (i.e., 60 mm). At medium and long pointing distances (i.e., 120, and 180 mm) there were no significant differences between the techniques. The overall mean pointing task time was

☆ This paper has been recommended for acceptance by Dianne Murray.

* Corresponding author. Tel.: +358 3 3551 8893; fax: +358 3 3551 6070.
E-mail addresses: Outi.Tuisku@uta.fi (O. Tuisku), Veikko.Surakka@uta.fi (V. Surakka), Toni.Vanhala@vtt.fi (T. Vanhala), Ville.Rantanen@tut.fi (V. Rantanen), Jukka.Lekkala@tut.fi (J. Lekkala).

0953-5438/$ - see front matter © 2011 British Informatics Society Limited. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.intcom.2011.10.002

0.7 s for the new technique and 0.6 s for the mouse. In a further study with gaze pointing, Surakka et al. (2005) compared the use of frowning and smiling (i.e., measured from above *zygomaticus major* muscle that is activated when smiling) as the selection technique and again, the gaze direction was used as the pointing method. The results showed that mean pointing task time was 0.8 s for the frowning technique and 0.5 s for the smiling technique.

These earlier studies (Surakka et al., 2004, 2005) used two separate technologies, remote eye tracker and EMG amplifier to implement the new pointing and selection method. In the present study we have developed a new device called wireless Face Interface (Face Interface project, 2011). The Face Interface device consists of an eyeglass-like frame that houses technologies for measuring both gaze and facial movement. Thus, the device is combination of a wearable video-based eye tracker and a capacitive sensor to detect the facial movement resulting either from the activation of the *corrugator supercilii* (i.e., activated when frowning) or the *frontalis* (i.e., activated when raising the eyebrows) facial muscles. The device includes also a scene camera that is used in compensating the user's head movements while the user interacts with a computer. The capacitive sensor provides an alternative to EMG measurement and it detects the distance to the facial skin. The distance changes by the activation of the facial muscles. For example, when a person frowns, the facial skin wrinkles and thus, the distance of the facial skin to the sensor is different from when the face is relaxed. One obvious advantage of this technique is that it is contact free and does not require any preparation of the skin (Rantanen et al., 2010).

We will start by briefly reviewing the related work. Then we will present the Face Interface device in detail. Finally, we will present the details of the current experiment.

2. Related work

Some studies have used setups which resemble the multichannel technique discussed above (Surakka et al., 2004, 2005). San Agustin et al. (2009) used two pointing devices, a remote eye tracker and mouse. As a selection method, they used a mouse button and facial EMG activity that was measured with commercial Cyberlink™ system (e.g., Nelson et al., 1997). They tested all the four possible combinations of the pointing and selection techniques. Three pointing distances (i.e., 200, 250, and 300 pixels) and three target diameters (i.e., 100, 125, and 150 pixels) were used. The results revealed an overall mean pointing task time of approximately 0.4 s. Gaze combined with facial EMG was the fastest one of the tested pointing and selection combinations.

Chin et al. (2008) combined the use of gaze direction and facial EMG in a slightly different manner than Surakka et al. (2004). They used facial EMG to correct the inaccuracy of the eye tracker in the following way. First, user gazed on the computer screen at the object to be selected. Second, if the cursor was not inside the object after the first step, user moved the cursor using facial movements while still gazing at the target. Left and right jaw clench resulted in the cursor to move left and right, respectively. Eyebrows up and eyebrows down movements resulted in the cursor to move up and down on a computer screen. Finally, user selected the target by clenching the whole jaw. The method was tested against gaze only method and regular computer mouse using pointing and selection tasks with three target diameters (i.e., 48, 66, and 96 pixels) and three pointing distances (i.e., 286, 578, and 778 pixels). The results revealed that the gaze direction combined with facial EMG was the slowest method. The overall mean pointing task time for it was 4.7 s. On the other hand, the error rate was lower when facial EMG was combined with gaze compared with the gaze only method. Thus, these results seem to suggest a trade-off between the accuracy and speed of their technique.

In earlier studies frowning action has been most frequently used as the selection technique (San Agustin et al., 2009; Surakka et al., 2004, 2005). Grauman et al. (2003) have been the only ones to use raising the eyebrows as the selection technique. Their *EyebrowClicker* system is based on computer vision which means that one camera captures video of the user from which the actions of the eyebrows are detected. The *EyebrowClicker* was tested using a simple game setup where the raising action made a frog to catch a fly with its tongue. Participants were instructed to raise their eyebrows when a circle appeared on the screen in order to "catch a fly". The results showed that participants were able to catch most of the flies. The success rate of 89% of detected eyebrow raises was reported. In other studies, raising the eyebrows action has been used also to move cursor upwards in the computer screen (Barreto et al., 2000; Chin et al., 2008).

Our own observations have indicated that some people are better (i.e., more accurate and quicker to learn) at performing the frowning action while other people are better at performing the raising the eyebrows action. Thus, it would be ideal if a person could freely choose which action they would use as the selection technique. The ability to choose the preferred selection technique means better adaptivity or adjustability of the user interface (Piwetz et al., 1995). This type of an approach is beneficial because it might be that, for example, not all people can frown, or they might find the raising the eyebrows action more pleasant to perform than the frowning action.

2.1. Fitts' law

The Fitts' law is often used to compare different pointing methods with each other (Card et al., 1978; Fitts, 1954). Fitts' law is a mathematical model which can be simplified to state that it is easier to point at targets that are big and near than targets that are small and far. According to the Fitts' law, the difficulty of a pointing task is called the index of difficulty (*ID*) which is calculated from the equation: $ID = log_2(A/W + 1)$, where A is the moved distance and W is the width of the target area. The *ID* is a linear relationship to pointing time and thus, can be described by linear regression equation of the movement time (*MT*). $MT = a + b\ ID$, where a and b are regression coefficients. The value of b can be used to compare different pointing devices. With low value of b, the ID of the task has less effect on the resulting movement times when using the technique in question. However, normally an index of performance (IP) value is calculated for the performance by dividing 1 with b, so that, $IP = 1/b$. The IP value is given in bits/s. The comparison of the IP values should reveal how well one input device is functioning as compared to the other input device, at least in theory. However, if one cannot be absolutely certain that the pointing and selection tasks have been conducted in similar manner and the IP values have been calculated with the same equation, then comparing IP values is not recommended (MacKenzie, 1992). Surakka et al. (2004) reported an IP value of 12.7 bits/s with a high correlation to a Fitts' law model (i.e., $r = 0.99$). San Agustin et al. (2009) reported an IP value of 3.03 bits/s but the correlation to Fitts' law model was not reported. Chin et al. (2008) did not calculate the Fitts' law values.

The analysis of the Fitts' law is based on the simple pointing and selection tasks (Douglas and Mithal, 1994; Surakka et al., 2004; Ware and Mikaelian, 1987). In those tasks, the pointing distances and the target sizes are varied to get as many ID values as possible. The variations in the pointing distances in different studies have been moderate, and thus, the increase in the ID values has also been moderate. This is partly due to the sizes of the computer displays, which have not allowed the use of longer distances in the pointing tasks. As the sizes of the computer displays have grown during the recent years, they offer now the possibility of adding more variability to the pointing distances.

2.2. Subjective ratings

In addition to Fitts' law, it is equally important to collect the subjective ratings of the used technique to see how the participants experience the used technique. The results of the ratings of the first study showed that the new technique was rated as significantly faster to use than the mouse (Surakka et al., 2004). Further, the new technique was rated also as significantly more difficult and less accurate to use as compared to mouse. In the second study, no statistically significant differences were found between the use of gazing and frowning and the use of gazing and smiling (Surakka et al., 2005). San Agustin et al. (2009) reported that the gaze pointing was rated as faster but less accurate than mouse pointing. Further, they reported that the gaze combined with facial EMG technique was rated as natural to use.

Surakka et al. (2004, 2005) collected the subjective ratings using six nine-point bipolar scales that varied from −4 to +4, and 0 represented the neutral evaluation. The scales were: general evaluation, difficulty, speed, accuracy, enjoyableness, and efficiency. The use of these types of scales has relatively long and well-studied background (Bradley and Lang, 1994; Osgood, 1952). Thus, the scales were adapted to the current study as well. However, four more bipolar rating scales were added (i.e., usefulness, naturalness, entertainment, and interestingness) to gain even deeper understanding of how the new wireless prototype was perceived.

3. Wireless Face Interface

Face Interface is a wireless wearable prototype device that combines the use of a wearable video-based eye tracker for pointing and a capacitive sensor to detect the facial movement resulting either from the activation of the *corrugator supercilii* (i.e., activated when frowning) or the *frontalis* (i.e., activated when raising the eyebrows) facial muscles for selecting the objects. Fig. 1 shows the Face Interface device that was built on the frames of protective glasses. The device includes two cameras, one for imaging the eye and the other for imaging the computer screen, an infrared (IR) light emitting diode for illumination of the eye and to provide the corneal reflection, a sensor device for detecting facial movements using a capacitive method, and a shoulder bag which contained radio frequency (RF) devices for wireless operation. The cameras used 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.

The shoulder bag contains a power supply unit for the prototype, two wireless analog video transmitters, a wireless (serial) transmitter for the capacitance measurement, and four AA batteries. A separate receiving station consists of two video receivers with 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 use the common free frequencies at 2.4 GHz.

For the eye tracking, the pupil detection and corneal reflection method was used (Duchowski, 2003). 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 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 and Kaehler, 2008) was utilized to extract features from the image streams of both eye and scene cameras. For the head movement compensation the location of six physical markers was extracted from the scene in order to track head orientation in relation to the computer display (see Fig. 2 for the placement of the markers). Movement of the eyebrow (i.e., the selection technique) was detected by finding increasing or decreasing slopes from the signal of the capacitive sensor.

The aim was to investigate the performance of the new wireless Face Interface device using either of the two available facial activations (i.e., frowning or raising) for object selection. Pointing task times and accuracies were measured using three target diameters (i.e., 25, 30, 40 mm), seven target distances (i.e., 60, 120, 180, 240, 260, 450, and 520 mm), and eight pointing angles (0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°) while performing pointing and selection tasks with the new prototype. The used computer screen was a 24″ widescreen display. We wanted to analyze if there are any radical differences in performing the tasks when two different facial muscle activations were used. Subjective ratings were also collected using 10 different nine point bipolar scales.

4. Methods

4.1. Participants

Twenty (seven male, 13 female) voluntary participants took part in the experiment. Their mean age was 27.8 years (range 19–43 years). All participants had normal or corrected-to-normal (i.e., by contact lenses) vision by their own report. Half of them preferred the object selection by frowning and the other half by raising the eyebrows.

Fig. 1. Left: the wireless Face Interface prototype. Right: a person wearing the prototype.

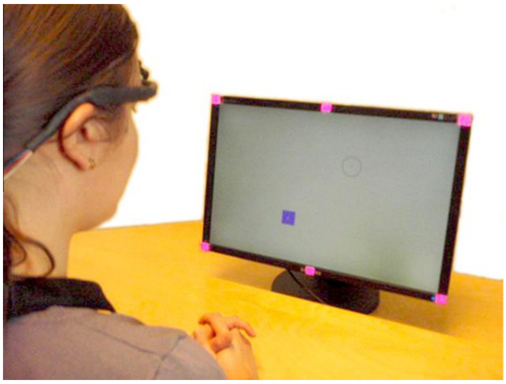

Fig. 2. Person performing an experimental task. The six markers for the head movement compensation are visible on the edges of the display.

4.2. Apparatus

The Face Interface device was used as the pointing and selection device. The display used in the experiment was a Samsung SyncMaster 24″ widescreen display with the resolution of 1920 × 1200. The viewing distance was approximately 60 cm. A desktop computer with Windows XP operating system was used to run the experiment.

4.3. Experimental task

The experimental task of this study was an extension of the task used by Surakka et al. (2004, 2005). Originally the task was created by Douglas and Mithal (1994). The task proceeded as follows. First, a home square and a target circle appeared simultaneously on the screen. The task of the participant was first to point and select the home square then to do the same for the target circle. The object became highlighted when participant's gaze was inside it and it disappeared after a successful click by frowning or raising the eyebrows. The target circle could not be selected before a successful click of the home square. After a successful click of the target circle there was a pause of 2000 ms and then the home square and target circle appeared again. The time between the two clicks was measured as the pointing task time. The width of the home square was kept constant and it was 30 mm. In order to make the longer pointing distances possible the place of the home square was varied so that it always appeared symmetrically on the opposite direction from the target circle (i.e., measured from the center of the screen). The target circle appeared in one of eight possible angles (four orthogonal and four diagonal directions, see Fig. 3 for the eight pointing angles) in relation to the home square. Three diameters of the target circle (i.e., 25, 30, and 40 mm) and seven pointing distances (i.e., 60, 120, 180, 240, 260, 450, and 520 mm) were used. For the distances 260, 450, and 520 mm the edges of the computer screen were used as the starting points. Thus, in total there were 120 tasks per participant.

4.4. Procedure

First, the laboratory and equipment were introduced to the participant. Then, the participant wore the prototype to see live video from the eye camera and from the scene camera. She or he was instructed to try different head orientations to see how large head movements were possible while still keeping the six markers in the scene camera image. Next, the participant was instructed to try and perform clicks by frowning or by raising the eyebrows. After several successful clicks were produced, the eye tracker was calibrated.

Before the actual experiment, there was a practice session of 20 trials which were not included in the actual experiment. The practice trial took approximately five minutes to complete. The participant was allowed to perform practice tasks either by using frowning or raising the eyebrows as the selection technique. The participant was told to perform the tasks as fast and as accurately as possible. The cursor for the current gaze location was shown as a cross. After practice, the participant chose the selection technique she or he preferred to use during the experiment. Then, there was a short relaxation period before the actual experiment. The eye tracker was calibrated and the actual experiment started. The eye tracker was re-calibrated when needed during the experiment (i.e., approximately once per participant).

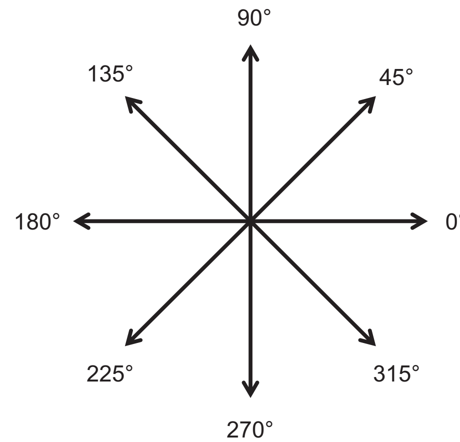

Fig. 3. Eight pointing angles used in the experiment.

At the end of the experiment, the participants rated the method with ten nine-point bipolar scales. The scales were: general evaluation (i.e., varying from bad to good), difficulty (i.e., from difficult to easy), speed (i.e., from slow to fast), accuracy (i.e., from inaccurate to accurate), enjoyableness (i.e., from unpleasant to pleasant), efficiency (i.e., from inefficient to efficient), usefulness (i.e., from unusable to usable), naturalness (i.e., from unnatural to natural), amusement (i.e., from boring to fun), and interestingness (i.e., from uninteresting to interesting). The scales varied from −4 (e.g., bad experience) to +4 (e.g., good experience), and 0 represented the neutral value (e.g., not bad nor good). Conducting the experiment took approximately 60 min in total.

4.5. Data analysis

Mixed-model analyses of variance (ANOVA) were first performed for seven pointing distances (i.e., 60, 120, 180, 240, 260, 450, and 520 mm) and three target circle diameters (i.e., 25, 30, and 40 mm) as within-subject factors and two selection techniques as between-subject factor. Then ANOVAs were performed for eight pointing angles (i.e., 0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°, and 360°) and three target circle diameters as within-subject factors and two selection techniques as between-subject factor. Bonferroni corrected *t*-tests were used for post hoc pairwise comparisons. The definition of an error was that if the first click on the target circle was not successful, the trial was marked as an erroneous one and was excluded from the data analysis. This happened in 27.3% of all of the trials. Pairwise Mann–Whitney *U*-test was used to compare the ratings of both techniques.

5. Results

5.1. Pointing task time analyses

5.1.1. All pointing distances

Fig. 4 shows mean pointing task times and standard error of the means (SEMs) for each of the stimuli averaged over all directions.

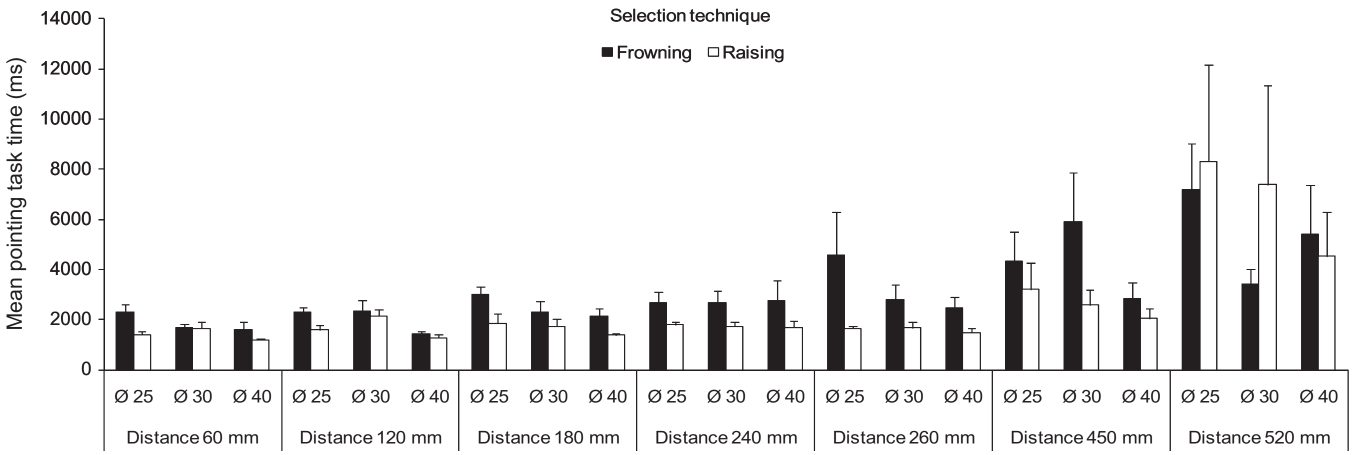

Fig. 4. Mean pointing task times and SEMs for each stimulus averaged over pointing angles.

The overall mean pointing task time (mean ± SEM) was 2823 ± 734 ms (i.e., 3149 ± 766 ms for the frowning technique, and 2497 ± 701 ms for the raising technique). A 7 × 3 × 2 (distance × diameter × selection technique) mixed-model ANOVA revealed a statistically significant main effect of the distance $F(6,90) = 6.8$, $p < 0.01$, and a statistically significant main effect of the diameter $F(2,30) = 5.4$, $p < 0.05$. The main effect of the selection technique and interactions between distance and diameter, between selection technique and distance, and between selection technique and diameter were not statistically significant.

Post hoc pairwise comparisons for the distance were not statistically significant. Post hoc pairwise comparisons for the diameter showed that the pointing task times were significantly faster for the 40 mm than the 25 mm targets, $MD = 999.2$, $p < 0.05$. Other pairwise comparisons were not statistically significant.

Fig. 4 shows that in comparison to task times using distances of from 60 to 260 mm there is relatively sharp increase in the pointing task times using distances of 450 and 520 mm. Thus, the most functional area of the current study (e.g., using 24″ widescreen display) settled to distances from 60 to 260 mm. In order to analyze the results over this most functional area of the display a mixed-model three-way ANOVA was computed using the area with five pointing distances (i.e., from 60 to 260 mm) and three target circle diameter (i.e., 25, 30, 40 mm) as the within-subjects factor and the selection technique (i.e., frowning, raising) as the between-subjects factor.

5.1.2. Pointing distances from 60 mm to 260 mm

The overall mean pointing task time (mean ± SEM) was 2048 ± 328 ms (i.e., 2473 ± 467 ms for the frowning technique and 1624 ± 188 ms for the raising technique). A 5 × 3 × 2 (distance × diameter × selection technique) mixed-model ANOVA revealed a statistically significant main effect of the distance $F(4,60) = 3.2$, $p < 0.05$, a statistically significant main effect of the diameter $F(2,30) = 18.2$, $p < 0.001$, a statistically significant main effect of the selection technique $F(1,15) = 5.3$, $p < 0.05$, and a statistically significant interaction effect between the selection technique and the diameter $F(2,30) = 8.7$, $p < 0.01$. The interactions between distance and diameter and between selection technique and distance were not statistically significant.

Post hoc pairwise comparisons for the distance were not statistically significant. Post hoc pairwise comparisons for the diameter showed that the pointing task times were significantly faster for the 40 mm diameter targets than for the 30 mm ($MD = 332.2$, $p < 0.01$), and for the 25 mm ($MD = 575.6$, $p < 0.001$) diameter targets. Post hoc pairwise comparisons for the selection technique showed that the pointing task times were significantly faster for the raising technique than for the frowning technique ($MD = 848.9$, $p < 0.05$). Other pairwise comparisons were not statistically significant.

Because of the significant interaction of the main effects of the selection technique and the diameter, one-way repeated measures ANOVAs for the diameter were performed for both interaction techniques separately. For the frowning technique, one-way ANOVA revealed a significant main effect of the diameter $F(2,14) = 17.9$, $p < 0.01$. Post hoc pairwise comparisons revealed that the pointing task times were significantly faster for the target circle diameter of 40 mm than for the target circle diameter 25 mm ($MD = 894.6$, $p < 0.01$). Also, pointing task times were significantly faster for the target circle diameter of 30 mm than for the target circle diameter of 25 mm ($MD = 610.4$, $p < 0.05$). Other post hoc pairwise comparisons were not statistically significant. For the raising technique one-way ANOVA revealed a significant main effect of the diameter $F(2,16) = 5.3$, $p < 0.05$. Post hoc pairwise comparisons were not statistically significant.

5.1.3. Pointing angle analysis

Fig. 5 shows mean pointing task times and SEMs for each of the stimuli averaged over all distances. An 8 × 3 × 2 (angle × diameter × selection technique) mixed-model ANOVA revealed a statistically significant main effect of the angle $F(7,105) = 4.5$, $p < 0.001$ and a statistically significant main effect of the diameter $F(2,30) = 7.5$, $p < 0.01$. The main effect of the selection technique and interactions were not statistically significant. Post hoc pairwise comparisons for the angle were not statistically significant. Post hoc pairwise comparisons for the diameter showed that the pointing task times were significantly faster for the 40 mm than the 25 mm targets, $MD = 770.0$, $p < 0.01$. Other pairwise comparisons were not statistically significant.

5.2. Error rate analyses

5.2.1. All pointing distances

Fig. 6 shows mean error rates and SEMs for each of the stimuli averaged over all directions. The overall mean error rate (mean ± SEM) was 27.3 ± 7.5% (i.e., 28.9 ± 7.9% for the frowning technique, and 25.7 ± 7.2% for the raising technique). A 7 × 3 × 2 (distance × diameter × selection technique) mixed-model ANOVA revealed a statistically significant main effect of the distance $F(6,90) = 5.8$, $p < 0.001$, and a statistically significant main effect of the diameter $F(2,30) = 5.4$, $p < 0.05$. The main effect of the selection technique and interactions were not statistically significant.

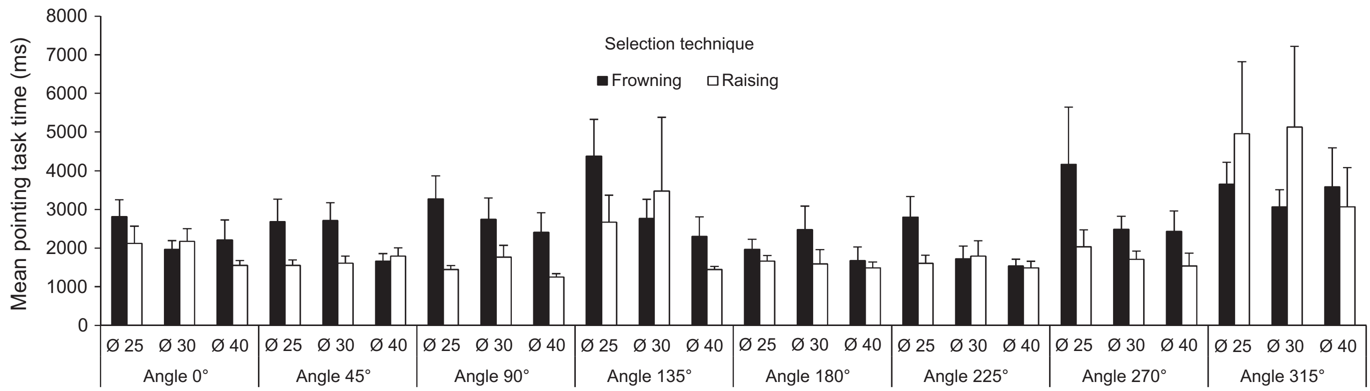

Fig. 5. Mean pointing task times and SEMs for each stimulus averaged over pointing distances.

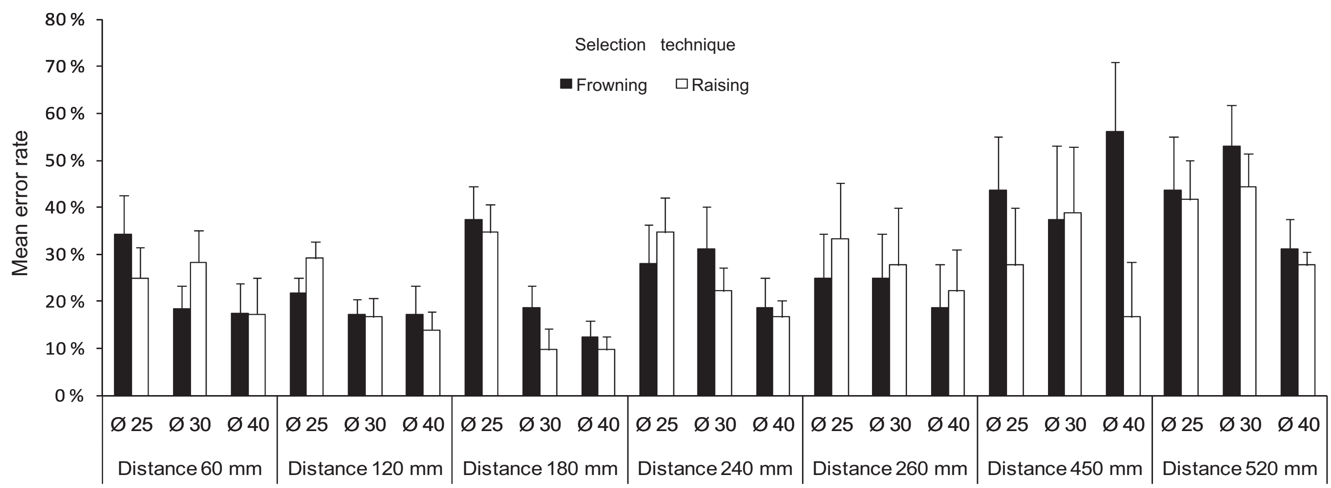

Fig. 6. Mean error rates and SEMs for each stimulus averaged over pointing angles.

Post hoc pairwise comparisons for the distance showed that the participants made significantly more errors when the distance was 520 mm than when the distance was 60 mm ($MD = 0.17$, $p < 0.05$), 120 mm ($MD = 0.21$, $p < 0.01$), 180 mm ($MD = 0.19$, $p < 0.05$), and 240 mm ($MD = 0.15$, $p < 0.05$). They also made significantly more errors when the distance was 450 mm than when the distance was 120 mm ($MD = 0.18$, $p < 0.05$), and 180 mm ($MD = 0.16$, $p < 0.05$). Other pairwise comparisons were not statistically significant.

Because it was found that the pointing distances of 450 and 520 mm had more errors as compared to other pointing distances and to imitate the analysis of the pointing task times, a separate analysis for the pointing distances from 60 mm to 260 mm were performed.

5.2.2. Pointing distances from 60 mm to 260 mm

The overall mean error rate (mean ± SEM) was 22.8 ± 6.4% (i.e., 22.8 ± 6.5% for the frowning technique and 22.8 ± 6.3% for the raising technique). A 5 × 3 × 2 (distance × diameter × selection technique) mixed-model ANOVA revealed a statistically significant main effect of the diameter $F(2,30) = 10.5$, $p < 0.001$. The main effect of the distance, the main effect of the selection technique, the interaction between distance and diameter, the interaction between distance and selection technique, and the interaction between selection technique and diameter were not statistically significant.

Post hoc pairwise comparisons for the diameter showed that participants made significantly more errors when the target diameter was 25 mm than when the diameter was 30 mm ($MD = 0.09$, $p < 0.05$) and when the diameter was 40 mm ($MD = 0.14$, $p < 0.05$).

5.2.3. Pointing angle analysis

Fig. 7 shows mean pointing task times and SEMs for each of the stimuli averaged over all distances. An 8 × 3 × 2 (angle × diameter × selection technique) mixed-model ANOVA revealed a statistically significant main effect of the diameter $F(2,30) = 8.0$, $p < 0.01$. The main effect of the angle, the main effect of the selection technique, the interaction between distance and diameter, the interaction between distance and selection technique, and the interaction between selection technique and diameter were not statistically significant. Post hoc pairwise comparisons for the diameter showed that participants made significantly more errors when the target diameter was 25 mm than when the diameter was 40 mm, $MD = 0.12$, $p < 0.05$.

5.3. Fitts' law analysis

Fitts' law analysis was performed for distances from 60 mm to 260 mm because in those distances the functioning of the Face Interface device was fastest and most accurate (see Fig. 4). The mean pointing task times were used to analyze how well the Fitts' law applied to eye tracking combined with both of the selection techniques (see Fig. 8).

The linear regression equation for the frowning technique was $MT = 913 + 526\ ID$, $r = 0.77$, $p < 0.01$. For the raising technique the linear regression equation was $MT = 1155 + 181\ ID$, $r = 0.51$, *ns*. Thus, the Fitts' law applied statistically significantly only for the gaze combined with the frowning selection technique.

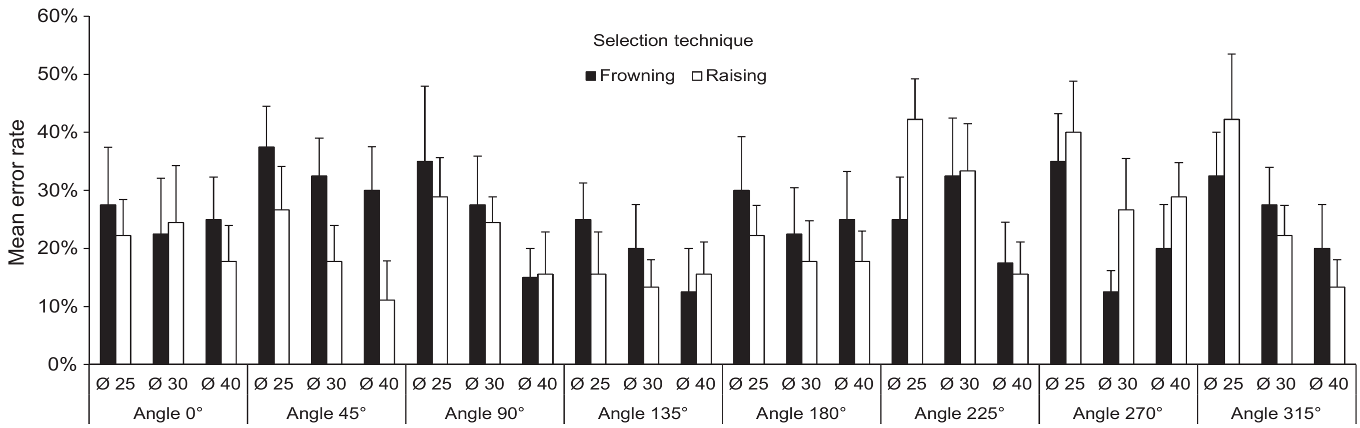

Fig. 7. Mean error rates and SEMs for each stimulus averaged over pointing distances.

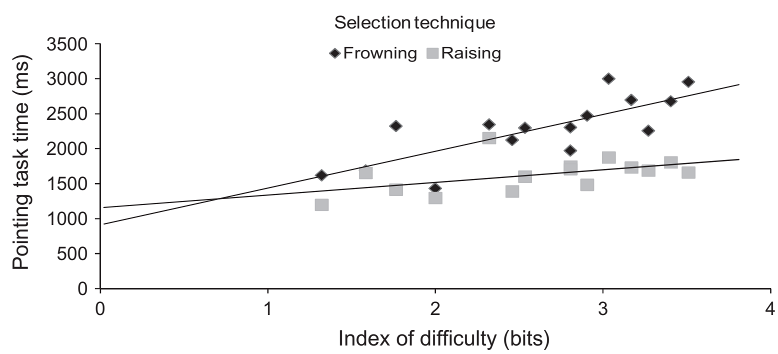

Fig. 8. Fitts' law regression lines for both selection techniques.

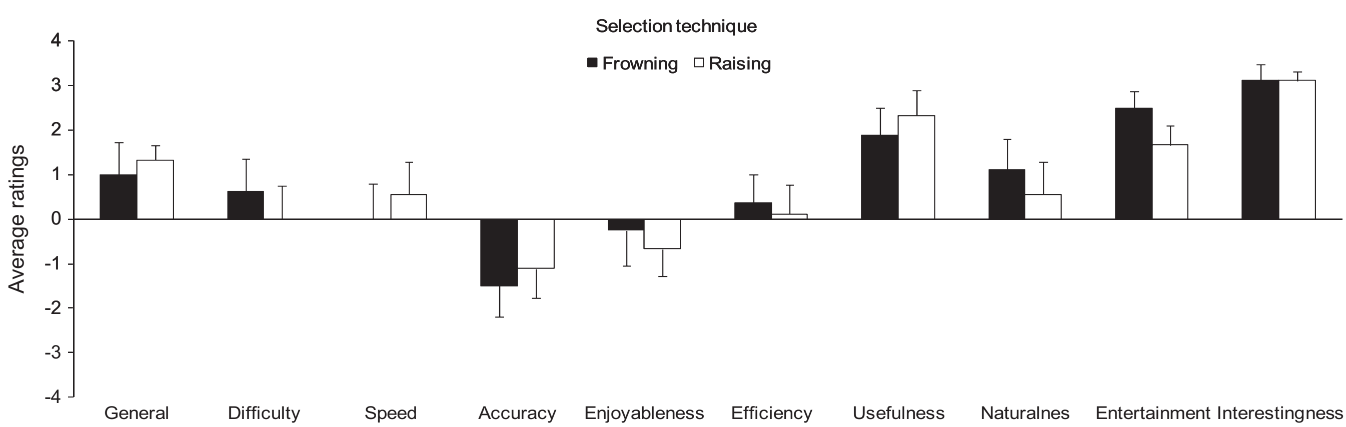

Fig. 9. Mean subjective ratings and SEMs for both selection techniques.

The IP values for both selection techniques were calculated through linear regression from the equation for IP reported in the Related Work section. The results showed that for the frowning technique the IP value was 1.9 bits/s and for the raising technique the IP value was 5.4 bits/s.

5.4. Ratings

Fig. 9 shows the mean ratings and SEMs for the ten different bipolar scales. Pairwise comparisons between selection techniques (Mann–Whitney *U*-test) were not statistically significant.

6. Discussion

The results showed that the wireless Face Interface prototype device was functional in object pointing and selecting tasks in a simple experimental setup. The pointing task times were fastest for the largest object sizes; a finding which has been typical in other studies as well (San Agustin et al., 2009; Surakka et al., 2004, 2005; Ware and Mikaelian, 1987). When taking all seven distances into account the overall mean pointing task times were 3.2 s for the frowning technique and 2.5 s for the raising technique. Although the raising technique was approximately one second faster than the frowning technique, there were no statistically significant differences between the two selection techniques as measured with pointing task times. For the pointing distance analysis, the ANOVA revealed a statistically significant main effect of the pointing distance. However, the post hoc pairwise comparisons were not statistically significant. This finding indicates that there is only a small increase in the pointing task times when the pointing distance increases. The overall error rates were 28.9% for the frowning technique and 25.7% for the raising technique. The pairwise comparisons showed that the participants made significantly more errors with both selection techniques when the pointing distances were 450 mm and 520 mm as compared to other pointing distances. Probably, part of the error rates of the Face Interface device are due to the fact that the eye camera is placed in front of the user's left eye. We note, however, that the current error rates are at about the same level than with other similar methods which do not place the camera in front of the eye (San Agustin et al., 2009; Surakka et al., 2004, 2005).

Specifically, pointing tasks with five pointing distances (i.e., 60, 120, 180, 240, 260 mm) proved to be the fastest and most accurate to point and select (see Figs. 4 and 6). When the raising action was used as the selection technique it resulted in significantly faster pointing task times than when the frowning action was used as the selection technique (i.e., 2.5 s for the frowning technique and 1.6 s for the raising technique). Hence, it seems that the raising technique was faster than the frowning selection technique. However, also the frowning technique was functional as the participants were able to perform all of the pointing tasks with it. The overall mean error rates for the pointing task distances from 60 mm to 260 mm were the same (i.e., 22.8%) for both selection techniques.

When comparing the current pointing task times (i.e. 2.5 s for the frowning and 1.6 s for the raising technique) with other similar type of existing methods we can see that they compare well. Earlier findings on pointing task times have varied from 0.4 s to 4.7 s (Chin et al., 2008; San Agustin et al., 2009; Surakka et al., 2004, 2005). Face Interface device can be compared to other types of interaction techniques. However, one needs to be cautious because they are profoundly different in respect to their operational principles and in their use of modalities. For example, in methods using gaze only for both object pointing and selection the dwell time protocol (i.e. the object is selected when the user's gaze has dwelled on it for a certain amount of time) is required. The pointing task times for the gaze only based methods obviously depends on the length of the chosen dwell time for the task at hand. For example, Ware and Mikaelian (1987) reported an overall pointing task time of one second when a dwell time of 400 ms was used. Vilimek and Zander (2009) reported an overall pointing task time of 4.5 s when a dwell time of 1000 ms was used. In the method that used only facial EMG, the object pointing and selection are both done by facial muscle activations. This can be time-consuming as, for example, Barreto et al. (2000) reported an overall pointing task time of 16.3 s.

The analysis of the pointing angle showed in the pointing task time analysis a significant main effect of the pointing angle. Further, the post hoc pairwise comparisons did not reveal statistically significant differences between task times in different pointing angles. In the error rate analysis there were no significant differences between the different pointing angles. In both analyses, the main effect of the target circle diameter was statistically significant. These are interesting findings and may indicate that this technique is not sensitive to variations in pointing angles. When using mouse as the pointing device, Whisenand and Emurian (1996) reported that there were differences in orthogonal and diagonal pointing angles, that is, participants were faster with orthogonal pointing angles. Still, they did not explain why this was the case. However, comparing mouse and this multimodal technique might not be that reasonable because the hand movements and eye movements are profoundly different.

Fitts' law analysis performed for the pointing tasks with the distances from 60 mm to 260 mm showed that Fitts' law applied statistically significantly only for the frowning technique. The correlation to the Fitts' law model was $r = 0.77$ ($p < 0.01$) for the frowning technique and $r = 0.51$ (*ns*) for the raising technique. These correlations indicate that the Fitts' law might not be perhaps the best possible method for evaluating the performance of these types of pointing devices. That is, for the gaze based techniques it might be more promising that correlation is not that high because it indicates that the increase in pointing task time is not that high with longer pointing distances. Perhaps a better approach would be to compare the IP values which were 1.9 bits/s for frowning technique and 5.4 bits/s for the raising technique. Earlier, Surakka et al. (2004) reported an IP value of 12.6 bits/s. San Agustin et al. (2009) reported an IP value of 3.03 bits/s. For the (different) regular computer mice, IP values between 4.6 and 5.7 bits/s have been reported (Isokoski and Raisamo, 2004; MacKenzie and Isokoski, 2008; Surakka et al., 2004). Thus, the IP values of the combination of gaze pointing and making selections with facial activity compare relatively well to those achieved with computer mice. This encourages the further development of these types of interaction methods.

The ratings of the use of the prototype gave interesting results. First, the ratings of the two selection techniques did not differ statistically significantly from each other. Thus, the ratings of the technique were not depended on the selection technique. Further, although the ratings of accuracy and the enjoyableness were on the negative side of scale the participants gave positive ratings of the prototype on the scales of usefulness, naturalness, entertainment, and interestingness. Also the rating of the general usability of the technique was on the positive side.

When the user uses the Face Interface prototype as a pointing and selection device for computers, some possible drawbacks may arise from the fact that not all people are capable of frowning or raising their eyebrows. While this interaction method may not be ideal for a particular segment of the population, such as those who suffer from paralyzed facial muscles, a large majority could still benefit from this method. People with disabilities should be able to use this technique, provided that normal eye movements and the ability to move their facial muscles still remain. Further, another problem with the frowning or raising related facial muscle activations might be that people have their own abilities and preferences in performing them (Levenson et al., 1990). Thus, this approach that will allow people to choose the facial muscle activation that they prefer as the selection technique was proved to be beneficial.

The current findings and comparisons show that the wireless Face Interface is a promising concept for future hands-free HCI. There are many advantages for the wireless Face Interface device. It compares especially well with other similar techniques in terms of comparable pointing task times and IP values. It is also very easy to learn to use the Face Interface technique as it takes only a couple of minutes. The method using capacitive sensing needs neither any skin contactor nor preparation of the skin in order to use it. Face Interface is wireless which offers more freedom when compared to techniques that use traditional electrode technology (e.g., Barreto et al., 2000; Sibert and Jacob, 2000; Surakka et al., 2004; Wolpaw et al., 2002). Since the Face Interface prototype device is wireless and wearable, it offers better possibilities for the user to move, for example, his or her head when interacting with computers than the existing methods. The results showed that the technique is not sensitive to the pointing distance, or to the pointing angle. The most important factor when interacting with the computer with wireless Face Interface prototype is the size of the object to be selected, that is, the object size needs to be large enough for it to be accurately selected. Finally, the results showed that people can use either the frowning or the raising technique for object selection. Thus, the wireless Face Interface technique offers support for selecting or adapting the channel which an individual user can handle better as a selection technique.

7. Conclusions

A novel prototype for HCI called the wireless Face Interface was presented. It was built on the frames of protective glasses housing both, the wearable eye tracker for pointing and the capacitive sensor for selecting the objects. The use of prototype was evaluated using a wider selection of tasks than earlier studies for similar techniques have done. The results showed that the Face Interface technique was functional in both cases, using either frowning technique or raising the eyebrows technique in object selection. Thus, it offers more flexibility for the user than previous similar techniques. The results showed that the Face Interface technique functions the best in the middle of the widescreen display, although, it was functional in the corners of the display as well. The results also showed that there were no differences in different pointing directions giving promise of an advantage over computer mouse. Thus, the current results suggest that the wireless Face Interface is a promising concept for future hands-free HCI and can be used in future in more advanced tasks, for example, writing with on-screen keyboard.

Acknowledgements

This research was funded by the Academy of Finland (project numbers 115997 and 116913) and the Finnish Doctoral Program in User-Centered Information Technology (UCIT). The authors thank Jarmo Verho for assistance in designing the electronic components used in the prototype. We would like to thank Dr. Pekka-Henrik Niemenlehto for the eye tracking and facial movement detection algorithms. Finally, the authors wish to express gratitude to Professor Poika Isokoski for providing the software that was used to run the pointing and selection tasks.

References

Barreto, A.B., Scargle, S.D., Adjouadi, M., 2000. A practical EMG-based human–computer interface for users with motor disabilities. J. Rehabil. Res. Dev. 37, 53–63.

Bradley, M., Lang, P.J., 1994. Measuring emotion: the self-assessment manikin and the semantic differential. J. Behav. Ther. Exp. Psychiatr. 25, 49–59.

Bradski, G., Kaehler, A., 2008. Learning OpenCV: Computer Vision with the OpenCV Library. O'Reilly Media, Sebastopol, California, USA.

Bulling, A., Roggen, D., Tröster, G., 2009. Wearable EOG goggles: eye-based interaction in everyday environments. Proceedings of CHI 2009. ACM Press, New York, pp. 3259–3264.

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

Chin, C.A., Barreto, A.B., Cremades, J.G., Adjouadi, M., 2008. Integrated electromyogram and eye-gaze tracking cursor control system for computer users with motor disabilities. J. Rehabil. Res. Dev. 45, 161–174.

Douglas, S.A., Mithal, A.K., 1994. The effect of reducing homing time in the speed of a finger-controlled isometric pointing device. Proceedings of CHI 1994. ACM Press, pp. 411–416.

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

Face Interface project. <https://www.cs.uta.fi/~wtpc/?q=node/7> (Checked 06.05.11).

Fitts, P.M., 1954. The information capacity of the human motor system in controlling the amplitude of movement. J. Exp. Psychol. 47, 381–391.

Grauman, K., Betke, M., Lombardi, J., Gips, J., Bradski, G.R., 2003. Communication via eye blinks and eyebrow raises: video-based human–computer interfaces. Univers. Access Inform. Soc., 359–373.

Isokoski, P., Raisamo, R., 2004. Speed and accuracy of six mice. Asian Inform.-Sci.-Life 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 Trans. Inform. Syst. 9, 152–169.

Levenson, R.W., Ekman, P., Friesen, W.V., 1990. Voluntary facial action generates emotion-specific autonomic nervous system activity. Psychophysiology 17, 363–384.

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

MacKenzie, I.S., 1992. Fitts' law as a research and design tool in human–computer interaction. Hum. Comput. Interact. 7, 91–139.

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

Majaranta, P., Räihä, K.-J., 2002. Twenty years of eye typing: systems and design issues. Proceedings of ETRA 2002. ACM Press, pp. 15–22.

Majaranta, P., Räihä, K.-J., 2007. Text entry by gaze: utilizing eye-tracking. In: MacKenzie, I.S., Tanaka-Ishii, K. (Eds.), Text Entry Systems: Mobility, Accessibility, Universality. Morgan Kaufmann, pp. 175–187.

Millán, J.d., Rupp, R., Müller-Putz, G.R., Murray-Smith, R., Giugliemma, C., Tangermann, M., Vidaurre, C., Cincotti, F., Kübler, A., Leeb, R., Neuper, C., Müller, K.R., Mattia, D., 2010. Combining brain–computer interfaces and assistive technologies: state-of-the-art and challenges. Front. Neurosci. 4, 161–193.

Nelson, W.T., Hettinger, L.J., Cunningham, J.A., Roe, M.M., Haas, M.W., Dennis, L.B., 1997. Navigating through flight environments using brain–body-actuated control. Proceedings of the 1997 Virtual Reality Annual International Symposium (VRAIS'1997). IEEE, pp. 30–37.

Osgood, C.E., 1952. Nature and measurement of meaning. Psychol. Bull. 49, 197–237.

Piwetz, C., Eiffert, F., Heck, H., Müller-Clostermann, B., 1995. An adjustable user interface providing transparent access to application programs for the physically disabled. SIGCAPH Comput. Phys. Handicap. 5, 11–16.

Rantanen, V., Niemenlehto, P.-H., Verho, J., Lekkala, J., 2010. Capacitive facial movement detection for human–computer interaction to click by frowning and lifting eyebrows. Med. Biol. Eng. Comput. 48, 39–47.

San Agustin, J., Mateo, J.C., Hansen, J.P., Villanueva, A., 2009. Evaluation of the potential of gaze input for game interaction. PsychNol. J. 7, 213–236.

Sibert, L.E., Jacob, R.J.K., 2000. Evaluation of eye gaze interaction. Proceedings of CHI 2000. ACM Press, pp. 281–288.

Surakka, V., Illi, M., Isokoski, P., 2004. Gazing and frowning as a new human–computer interaction technique. ACM Trans. Appl. Percept. 1, 40–56.

Surakka, V., Isokoski, P., Illi, M., Salminen, K., 2005. Is it better to gaze and frown or gaze and smile when controlling user interfaces? Proc. HCI Int..

Vilimek, R., Zander, T.O., 2009. BC(eye): combining eye-gaze input with brain-computer interaction. In: Stephanidis, C. (Ed.), Universal Access in HCI, Part II, HCII 2009, LNCS, vol. 5615. Springer-Verlag, Berlin Heidelberg, pp. 593–602.

Ware, C., Mikaelian, H.H., 1987. An evaluation of an eye tracker as a device for computer input. Proceedings of CHI 1987. ACM Press, pp. 183–188.

Whisenand, T.G., Emurian, H.H., 1996. Effects of angle of approach on cursor movement with a mouse: consideration of Fitts' law. Comput. Hum. Behav. 12, 481–495.

Wolpaw, R.J., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughn, T.M., 2002. Brain–computer interfaces for communication and control. Clin. Neurophysiol. 113, 767–791.