PsychNology Journal, 2009
Volume 7, Number 2, 213 – 236

Evaluation of the Potential of Gaze Input for Game Interaction

Javier San Agustin[*][♦], Julio C. Mateo[♣], John Paulin Hansen[♦] and Arantxa Villanueva[♠]

[♦] University of Copenhagen (Denmark)
[♣] Wright State University (USA)
[♠] Public University of Navarra (Spain)

ABSTRACT

To evaluate the potential of gaze input for game interaction, we used two tasks commonly found in video game control, target acquisition and target tracking, in a set of two experiments. In the first experiment, we compared the target acquisition and target tracking performance of two eye trackers with four other input devices. Gaze input had a similar performance to the mouse for big targets, and better performance than a joystick, a device often used in gaming. In the second experiment, we compared target acquisition performance using either gaze or mouse for pointing, and either a mouse button or an EMG switch for clicking. The hands-free gaze-EMG input combination was faster than the mouse while maintaining a similar error rate. Our results suggest that there is a potential for gaze input in game interaction, given a sufficiently accurate and responsive eye tracker and a well-designed interface.

Keywords: *Gaze input, video games, electromyography, pointing devices, performance evaluation, Fitts' Law, human-computer interaction.*

Paper Received 14/11/2008; received in revised form 02/05/2009; accepted 05/05/2009.

1. Introduction

In recent years, the video game industry has introduced new and innovative ways of controlling games. In 2003, Sony presented the EyeToy, a camera that is connected to a PlayStation 2 console and tracks the body movements of the players, allowing them to control the on-screen characters by moving their bodies (Sony Computer Entertainment, Inc., 2008). In 2005, Nintendo presented the Wiimote, a novel gamepad for their console Wii (Nintendo of America, Inc., 2008). The Wiimote includes an

Cite as:
San Agustin, J., C. Mateo, J.,Hansen, J.P., & Villanueva, A. (2009). Evaluation of the Potential of Gaze Input for Game Interaction. *PsychNology Journal, 7*(2), 213 – 236. Retrieved [month] [day], [year], from www.psychnology.org.

[*] Corresponding Author:
Javier San Agustin
IT University of Copenhagen, Rued Langgaards Vej 7, 2300 – Copenhagen S, Denmark
E-mail: javier@itu.dk

accelerometer and optical sensor technology that allow games to be controlled by moving the pad in three-dimensional space. In 2007, Nintendo introduced a new peripheral for the Wii, the Wii Balance board, a board that measures the user's center of balance and body mass index.

Continuing with the trend of seeking alternative and more intuitive input devices for game interaction, gaze represents a fast and natural input method that can also be exploited. However, the potential of gaze input to increase the speed of interaction in gaming and possibly free the hands for other tasks has received little attention. Most past research on eye tracking technology has emphasized human-computer interaction for severely disabled people who cannot control traditional input devices (Majaranta & Räihä, 2002).

Interaction with a video game usually requires performing two main tasks: pointing at a target and selecting it (i.e., *target acquisition* tasks) and keeping the pointer on the target while this moves on the screen (i.e., *target tracking* tasks). Gaze interaction has been extensively evaluated in target acquisition tasks under the Fitts' Law framework (Sibert & Jacob, 2000; Zhang & MacKenzie, 2007). However, the performance in target tracking tasks using gaze input is yet to be investigated. These kinds of studies can provide an insight into the mechanics of smooth pursuit movements that would be fundamental in the development of gaze-controlled video games, such as first-person shooters.

Pointing using gaze-based systems has been shown to be both more intuitive and faster than mouse pointing (Sibert & Jacob, 2000). This may not be surprising given that humans naturally tend to direct their eyes toward the location to which they are moving and that eye movements are faster than hand movements (Zhai, Morimoto, & Ihde, 1999).

However, gaze-based systems are not as well suited for performing selections. Finding a method to perform selections reliably using only gaze is not a trivial problem. In gaze-based systems, the two most common selection methods are *dwelling* and *blinking*. When using dwelling as the selection method, the system issues an activation every time the user stares at a target for longer than a pre-specified threshold duration (i.e., *dwell time*). Common dwell times range from 0.5 to 1 s. When using blinking as the selection method, the system issues an activation every time the user closes his or her eyes. Although useful, these two selection methods have a range of usability problems due to the difficulty of inferring the user's intention and the fact that both prolonged fixations and blinks occur naturally and frequently when users do not intend to issue any activation. By relying exclusively on the duration of fixations for activation, dwelling sometimes leads to undesired activations when a user stares at an object to study it without the intention of giving any command. This is known as the *Midas Touch* problem (Jacob, 1991). Activation by blinking avoids this problem, but it is usually tiring for the user and, since blinking is a natural action, some natural blinks can be mistaken and taken for activations. Arguably, gaze-only selection techniques are unnatural and slow down the interaction.

Sibert and Jacob (2000) found that target acquisition performance was faster using gaze with short dwell times than using a mouse. They used a dwell time as low as 150 ms, which is too short if the task the user is performing causes a higher cognitive effort, such as typing on an on-screen keyboard (Majaranta & Räihä, 2002). The longer dwell times needed for these tasks can substantially slow down gaze interaction. As a consequence, for example, typing performance on an on-screen keyboard using gaze as the input tends to be slower than using the mouse (Hansen, Tørning, Johansen, Itoh, & Aoki, 2004). One way to solve the limitations of current selection methods is to combine gaze pointing with alternative modalities (e.g., facial-muscle signals) to perform the selection task. When using alternative modalities for selection, preservation of the hands-free advantage of gaze-based systems obviously depends on whether the chosen modality requires the use of hands (e.g., mouse button) or not (e.g., facial-muscle switch).

A complete evaluation of the use of gaze tracking in game interaction can provide an insight into how the limitations of eye movements might affect game performance and how design could help compensate for these limitations. In this study, we perform two experiments. In the first experiment, we compare the performance of six different input devices (i.e., two commercial eye tracking systems, a mouse, a touch screen, a joystick and a head tracker) on game-like target acquisition and target tracking tasks. The superior performance of the mouse over all other input devices in our first experiment suggests that the mouse is still the best device. In the second experiment, we explore the potential of combining gaze pointing with a facial-muscle *electromyographic* (*EMG*) signal for selection in order to compete with the speed of the mouse in target acquisition tasks. This particular hands-free gaze-EMG input combination showed the potential to match (and even outperform) the speed of mouse interaction. However, the limited accuracy of gaze tracking remains a challenging problem.

2. Previous Work

The use of gaze interaction for video game control has not been fully investigated yet. Smith and Graham (2006) compared the performance of gaze versus mouse in three different games by measuring the time participants required to complete a given task or by comparing the scores given by the game. Although participants felt more immersed in the game when using gaze, control by mouse was found to be more effective. Isokoski and Martin (2006) performed a similar study on a first-person shooter. They compared the score obtained when using gaze in combination with mouse and keyboard input, only mouse and keyboard input (without gaze), and an Xbox 360 controller. Using gaze input, participants obtained a performance similar to the Xbox controller, but worse than the performance using the keyboard and mouse combination. Dorr, Böhme, Martinetz and Barth (2007) compared the performance of gaze versus mouse in a modified version of the Breakout game, finding gaze to be superior to mouse.

Instead of focusing on specific games or game genres, in this paper we evaluate the performance of gaze interaction using Fitts' Law and the ISO 9241-9 standard. The results are applicable to video games as well as more generic gaze-based interfaces.

2.1. Target Acquisition Tasks: Fitts' Law and the ISO 9241-9 Standard

Many studies have been carried out to evaluate the performance of different input devices in target acquisition tasks. Most of them use Fitts' Law to calculate the index of performance (*IP*) of each input device in order to compare device performance. IP is measured in bits per second (bits/s) and is calculated with the following formula:

$$IP = \frac{ID}{MT} \tag{1}$$

where *ID* is the task's index of difficulty (*ID*), measured in bits, and *MT* is the average movement time required to complete the task, measured in seconds. The ID is usually given by the following expression:

$$ID = log_2\left(\frac{A}{W} + 1\right) \tag{2}$$

ID depends on the distance to the target (i.e., amplitude *A*) and the width of the target measured along the axis of movement (*W*). Equation 1 can be rewritten so that the predicted variable is MT, giving

$$MT = \frac{ID}{IP} \tag{3}$$

The IP can be determined as in Equation 1, or as a regression of *MT* on *ID*, which gives the following equation of a line

$$MT = a + b\,ID \tag{4}$$

where *a* and *b* (intercept and slope, respectively) are regression coefficients to be calculated empirically. The reciprocal of the slope, 1/*b*, corresponds to the IP in Equation 3.

Ware and Mikaelian (1987) conducted the first study of gaze interaction under the Fitts' Law framework. They evaluated the movement time and error rate of an eye tracker with three selection methods: dwell, a physical button, and an on-screen button to confirm a selection. Average movement times were below 1 s for the three techniques, with dwell and physical button being faster than the on-screen button.

In 2000, the ISO 9241-9 standard based on Fitts' law was introduced (ISO, 2000). It establishes the guidelines for evaluating computer input devices in terms of performance and comfort. The metric to measure performance is *throughput*, in bits/s. It combines both the speed and accuracy of the input device. The equation for throughput is based on the IP in Fitts' Law, but it uses an effective index of difficulty (ID_e) giving the expression:

$$Throughput = \frac{ID_e}{MT} \tag{5}$$

where ID_e is determined as follows:

$$ID_e = log_2\left(\frac{A}{W_e} + 1\right) \tag{6}$$

ID_e is calculated using the effective width (W_e) instead of the nominal width of the target. That is, ID_e is calculated from what the users actually did (i.e., distribution of movement endpoints) and not from what was expected (i.e., target width), therefore incorporating the variability in performance across participants. W_e is determined by

$$W_e = 4.133 \times SD \tag{7}$$

where *SD* is the standard deviation of the movement endpoints across participants, measured along the line from the origin of movement to the center of the target. Using W_e is necessary when an error rate different from 4% is observed. When the endpoints are not known, W_e can be calculated from the error rate (MacKenzie, 1992).

Douglas, Kirkpatrick and MacKenzie (1999) carried out the first evaluation of pointing devices using the ISO 9241-9 standard, when it was still a draft. The authors concluded that the scientific basis of the standard (the accepted Fitts' Law) was solid enough to be used for performance evaluations of input devices. Some of their considerations were taken into account in the final version of the standard.

Zhang and MacKenzie (2007) conducted the first evaluation of the performance of gaze interaction following the ISO 9241-9 standard. They studied the throughput of a mouse and an eye tracker with three different selection methods: short dwell (500 ms), long dwell (750 ms), and space bar. The throughput obtained when using gaze with the space bar was close to the throughput of the mouse, although the error rate was significantly higher.

2.2. Target Tracking Tasks: Time-On-Target Metric

There are few studies on the performance of input devices on target tracking tasks. The obvious metric to measure the accuracy of a device is *time on target* (*TOT*). For each sample during a trial, we check whether the pointer is on the target or not. The TOT for the trial is the number of samples "on" the target divided by the total number of samples (N):

$$TOT = \frac{\sum_{i=1}^{N} On(i)}{N} \quad (8)$$

On(i) returns '1' if the pointer is within the target's radius for sample *i*, and '0' otherwise.

Klochek and MacKenzie (2006) introduced several new metrics to measure the accuracy and smoothness of an input device and compared the performance of a mouse and a gamepad in a three-dimensional target tracking task in a game-like three-dimensional environment. Although the new metrics can help explain the differences in the performance of the two devices, TOT is the most relevant metric when the objective is to check whether two devices have a similar performance or not. The authors of this paper have not found any previous studies that evaluate gaze interaction in target tracking tasks.

2.3. Using Alternative Modalities for Selection: Gaze-EMG Input Combination

Facial-muscle activity can be measured through the electromyographic (EMG) signal and can be used to provide a fast and hands-free selection method (Junker & Hansen, 2006). Nelson et al. (1996) found indications that clicking by frowning could be up to 20% faster than clicking by using a mouse button. A combination of gaze pointing and

EMG clicking seems promising to compete with the speed of the mouse in target acquisition tasks.

Partala, Aula and Surakka (2001) studied the benefit of combining gaze pointing and facial-muscle EMG clicking compared to mouse input in target acquisition tasks. They found task completion times to be shorter for the new input technique for long distances (above 100 pixels) after removing the trials where selection occurred outside the target. However, a very high error rate (34%) was observed for the gaze-EMG combination. Throughput was not calculated.

Surakka, Illi and Isokoski (2004) extended the previous study with a more detailed Fitts' Law analysis. They compared the target acquisition performance of gaze pointing and EMG selection (i.e., frowning) to the mouse. The gaze-EMG input combination showed a higher index of performance than the mouse for error-free data, but for short distances the mouse was more effective. Surakka, Illi, & Isokoski (2004) suggested that gaze and EMG may be faster at longer distances, but their data did not show any speed advantage of gaze and EMG over the mouse.

3. Experiment 1: Performance Evaluation in Target Acquisition and Target Tracking Tasks

Experiment 1 compared the performance of six different input devices in target acquisition and target tracking tasks using the ISO 9241-9 standard. Specifically, the performance of two commercially available eye tracking systems (Tobii and Quick Glance 3) was compared to each other and to a mouse, a touch screen, a head tracker, and a joystick. This experiment extends the findings of Zhang and MacKenzie (2007) by using two different commercially available eye tracking systems. In addition to comparing gaze and mouse, this experiment compares gaze input with other input devices that are expected to perform worse than the mouse. Lastly, this experiment is possibly the first to explore target tracking performance using gaze input.

3.1 Method

Participants

A total of 6 participants, 5 males and 1 female, participated in the experiment. Ages ranged from 26 to 48 years old. All 6 participants were regular mouse users and had previous experience with joystick devices; 3 had previous experience with eye trackers, and 1 with head trackers.

Apparatus

The software used to present the targets was programmed in C# and ran at a constant frame rate of 30 Hz. The input devices tested were mouse (Logitech optical mouse), touch screen (Dell E157FPT), joystick (Logitech Attack 3), head tracker (NaturalPoint), and two remote eye trackers (Tobii 1750 and Quick Glance 3), both set with the minimum possible smoothing between images on estimated cursor position.

Design and Procedure

Participants performed two types of task during this experiment: target acquisition tasks and target tracking tasks. Target acquisition tasks required the participants to point at a target as quickly as possible and activate a button to select it. Participants always moved from the center to the single target present in the workspace at any time. The 16 targets were arranged in a circular layout (as proposed in ISO 9241-9) with a radius of 250 pixels, as shown in Figure 1. Targets could be 75 or 150 pixels in diameter (roughly 2 and 4 degrees of visual angle, respectively). Given that distance to the target was always constant (i.e., 250 pixels), the nominal indexes of difficulty were 2.1 and 1.4 bits. The performance metrics used in this task are *throughput* and *completion time*.

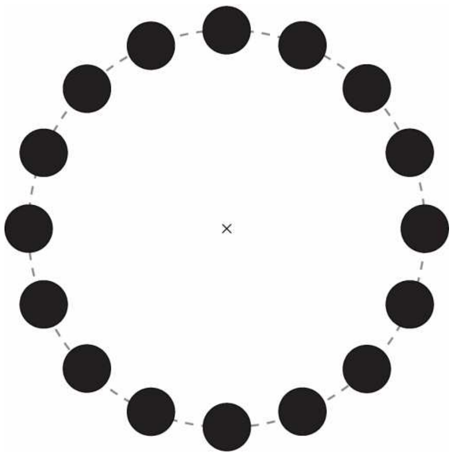

Figure 1. Layout of the 16 targets (only one target was shown at a time).

Target tracking tasks required the user to keep the pointer on the target while the target moved on the screen. In this study, targets moved at a constant velocity of 90 pixels/s and they always moved from one of the 16 target locations to the center of the screen. Two possible ways to alert the user when the pointer is not on target are auditory feedback, which alerts the user by emitting a sound, and movement feedback, which alerts the user by stopping the target. In our experiment, we tested two feedback conditions: one using only auditory feedback and the other using a combination of auditory and movement feedback. The metric used to evaluate the performance was *time on target* (TOT).

Each participant completed four blocks of 16 trials with each of the input devices, starting always with the mouse. The order of the other five devices was counter-balanced across participants using a balanced Latin square. The four blocks that participants completed with each device corresponded to different target-size and feedback conditions. The order of these four blocks was chosen to counterbalance the effects of order and practice across participants. Prior to starting the experiment, participants familiarized themselves with the task in a warm-up block using the mouse. All blocks were performed in one day, and the total experiment lasted about 2 hours with a short break after each device.

At the beginning of each block, the participant pointed at the X on the center of the screen to indicate he or she was ready to start, triggering the release of the first target. This procedure was repeated at the beginning of each trial to ensure that the starting position of the pointer was at the center of the screen for every trial. Targets appeared consecutively in random order in one of 16 locations on the circular layout shown in Figure 1. Participants were instructed to move the pointer to the target and select it as soon as possible after its appearance. Once the target was acquired, it started moving towards the center of the screen with a constant velocity of 90 pixels/s. Participants were instructed to keep the pointer on the target while the target was moving to the center. The target disappeared when reaching the center, and an X appeared in its place. The same sequence was repeated in each subsequent trial until the end of the block.

3.2 Results

Data analysis was performed using three 6×2×2 within-subjects ANOVAs, with *device* (mouse, touch screen, head tracker, joystick, Tobii or Quick Glance), *target size* (75 pixels or 150 pixels) and *feedback* (auditory or auditory plus movement) as the independent variables. Throughput, completion time, and time on target (TOT) were analyzed as the dependent variables. An average of the 16 trials conducted under each block was calculated for each subject. All data were included.

Throughput

An error rate of 4% is assumed in this experiment. Throughput is therefore calculated using Equations 1 and 2. Overall mean throughput was 1.85 bits/s. There was a significant effect of input device on throughput, $F(5, 25) = 5.61$, $p < 0.05$, with mean values ranging from 1.09 to 2.12 bits/s. Touch screen had the highest throughput (M = 2.12 bits/s, SD = 0.53 bits/s), and it was significantly different ($p < 0.05$, Scheffe post hoc test) from the head tracker (M = 1.35 bits/s, SD = 0.24 bits/s) and the joystick (M = 1.09 bits/s, SD = 0.18 bits/s). The throughput of mouse (M = 2.05 bits/s, SD = 0.39 bits/s) was significantly higher than the throughput of head tracker and joystick. The Tobii tracker (M = 1.92 bits/s, SD = 0.91 bits/s) showed a better performance ($p < 0.05$) than joystick. Quick Glance also had a higher throughput than the head tracker ($p <$ 0.05). The eye trackers did not differ significantly. Neither size, $F(1, 5) = 6.45$, $p > 0.05$, nor feedback, $F(1, 5) = 1.65$, $p > 0.05$, had a significant effect on throughput. Figure 2 shows the throughput of the different devices for each target size.

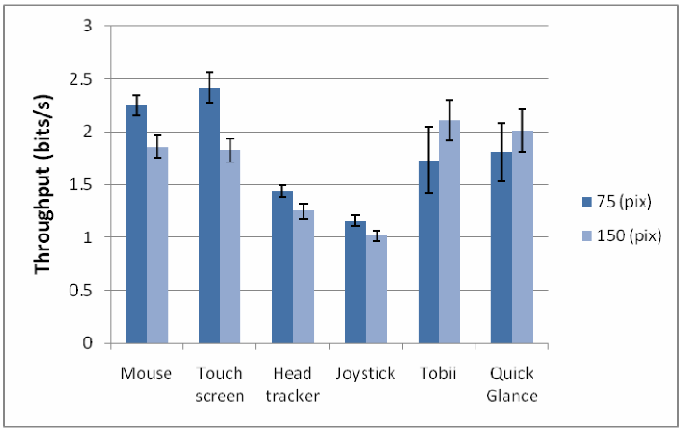

Figure 2. Mean throughput of each device for both target sizes. Error bars show standard errors of the mean.

Completion Time

Overall mean completion time was 1183 ms. There was a significant effect of input device on completion time, $F(5, 25) = 6.53$, $p < 0.05$. Touch screen had the lowest completion time (M = 859 ms, SD = 190 ms), and it was significantly different ($p < 0.05$, Scheffe post hoc test) from the head tracker (M = 1340 ms, SD = 308 ms) and the joystick (M = 1649 ms, SD = 341 ms). Mouse (M = 875 ms, SD = 177 ms) also had a significantly lower completion time than head tracking and joystick. Both of the eye trackers (Tobii M = 1159 ms, SD = 684 ms and Quick Glance M = 1219 ms, SD = 964 ms) had a lower completion time ($p < 0.05$) than joystick. Quick Glance had a significantly lower completion time than head tracker ($p < 0.05$). The eye trackers did not differ significantly. Size had a significant effect on completion time, $F(1, 5) = 26.88$, $p < 0.05$, but type of feedback did not, $F(1, 5) = 1.41$, $p > 0.05$. Figure 3 shows the completion time for the different devices and target sizes.

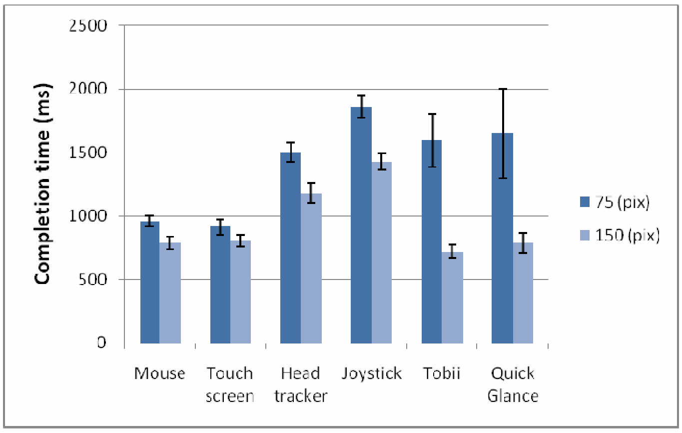

Figure 3. Mean completion time for each device and target size. Error bars show standard errors of the mean.

Time on Target

The overall mean time on target (TOT) was 0.90. There was a significant effect of input device on TOT, $F(5, 25) = 15.06$, $p < 0.05$. TOT was significantly lower on small targets ($M = 0.82$, $SD = 0.17$) than on big targets ($M = 0.97$, $SD = 0.04$), $F(1, 5) = 74.77$, $p < 0.05$. Feedback also had a significant effect on TOT, $F(1, 5) = 23.72$, $p < 0.05$, with TOT being higher when auditory and movement feedback were present ($M = 0.92$, $SD = 0.11$) than when only auditory feedback was used ($M = 0.88$, $SD = 0.18$). Figure 4 shows the mean TOT for each device and target size condition.

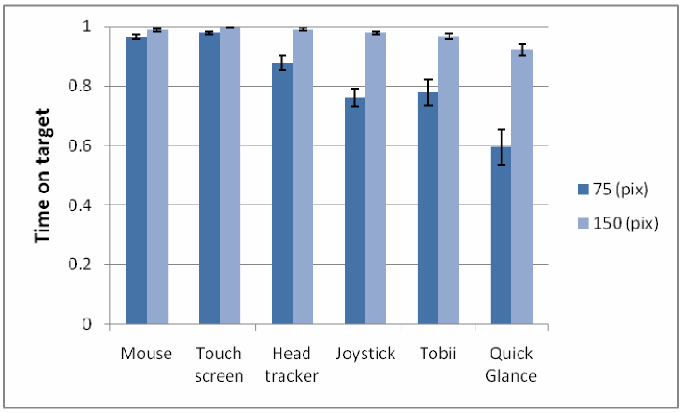

Figure 4. Mean time on target for each input device and target size condition. Error bars show standard errors of the mean.

The interaction between size and device on TOT was significant, $F(5, 25) = 10.68$, $p < 0.05$ (see Figure 5). The post hoc test showed that the difference between Quick Glance and the other 5 devices was significant for the small 75-pixel targets ($p < 0.05$).

The Tobii tracker had a lower TOT under that condition than mouse and touch screen ($p < 0.05$), but did not differ significantly from the joystick or head tracker. None of the devices differed under the large 150-pixel target condition.

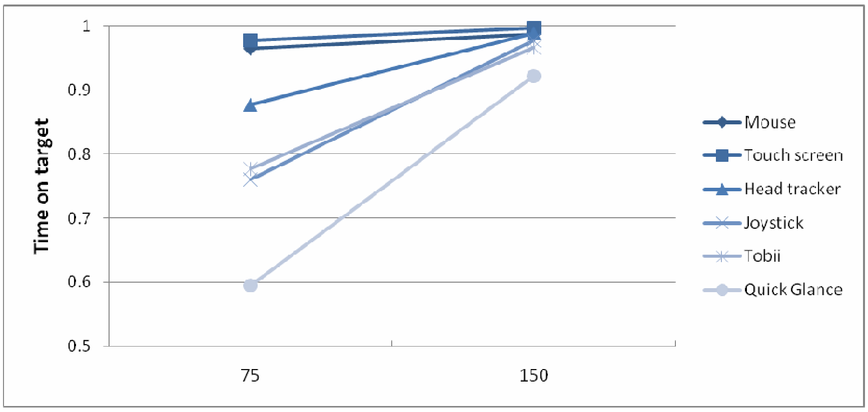

Figure 5. Mean time on target as a function of target size for all six input devices.

4. Experiment 2: Performance Evaluation of Gaze Pointing and EMG Clicking

When targets were big enough to compensate for inaccuracies of the gaze tracker, completion times for gaze pointing were found to be similar to mouse pointing. Therefore, our first experiment showed that, given a sufficiently accurate eye tracker, gaze pointing can be as fast as mouse pointing in target acquisition tasks. In order to compete with the speed of the mouse, we conducted a second experiment where we combined gaze pointing with EMG clicking. Specifically, we compared the performance of the combinations of mouse and gaze pointing with button and EMG clicking in a target acquisition task. The objective was to investigate whether the hands-free combination of gaze and EMG could outperform the mouse in target acquisition tasks. This experiment extends the experiments by Partala, Aula, & Surakka (2001) and by Surakka, Illi, & Isokoski (2004) by using the ISO 9241-9 standard. Furthermore, our study also evaluates the performance of mouse-EMG and gaze-button combinations.

4.1 Method

Participants

A total of 5 male volunteers participated in this study. They ranged in age from 25 to 30 years old. All 5 participants were regular mouse users, 4 had previous experience with gaze tracking, and 2 had tried an EMG system before.

Apparatus

Figure 6 shows all the equipment used in this experiment. Targets were presented by software programmed in C# that ran at a frame rate of 60 Hz on a Pentium IV. The display was a 17-inch monitor with a resolution of 1024×768 pixels. The sensitivity of the optical mouse (Acer) was set to an intermediate setting.

EMG activity was measured with a Cyberlink™ system (Nelson et al., 1996). Participants wore a headband that measured electrical signals from facial muscles on the forehead. The Cyberlink™ sent a click command to the computer via an RS-232 interface each time participants slightly frowned or tightened their jaw.

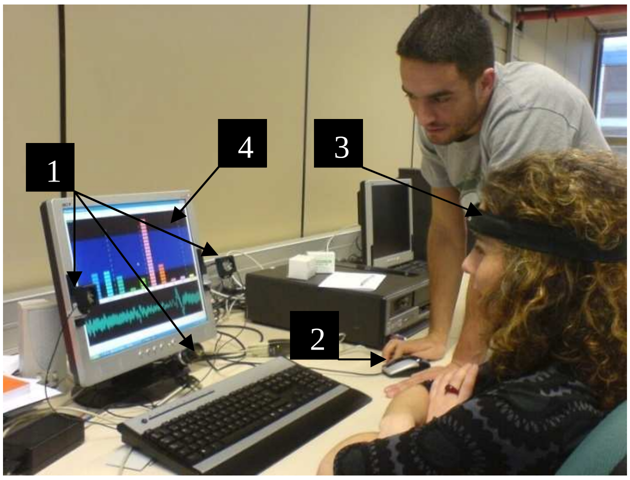

Figure 6. Experimental setup in Experiment 2: (1) Eye tracker. (2) Mouse. (3) Cyberlink™ headband. (4) 17-inch monitor. The display is showing the Cyberlink™ software.

We used an eye tracking system developed at the Public University of Navarra as the pointing device. It has an infrared light source on each side of the screen and uses a Pupil-Corneal-Reflection technique. The measured accuracy is better than 0.5º (around 16 pixels in our configuration), and the sampling rate is 30 Hz.

Design and Procedure

Participants performed a target acquisition task during this experiment. *Pointing method* (mouse or gaze) and *selection method* (mouse button or EMG switch) were manipulated across blocks, so that each participant used all four input combinations. There were 16 targets arranged in a circular layout, as shown in Figure 1. Targets could be 100, 125 and 150 pixels in diameter, and the distance to the center could be 200, 250 and 300 pixels. The nominal indexes of difficulty were between 1.2 and 2 bits. In each trial, we measured *completion time* and *unsuccessful activations* (i.e., clicks outside the target). Participants also completed a questionnaire rating the speed, accuracy, ease of use, and fatigue perceived in association with each input combination.

Each participant completed a block of trials for each input combination. The order of these four blocks was chosen to counterbalance the effects of order and practice across participants. The participants' task in this experiment was identical to the target acquisition task in Experiment 1 (see *Design and Procedure* in Section 3.1). However, no target tracking task was performed in this experiment.

In each block, 16 data points were collected for each width and distance combination, one for each of 16 possible directions of movement, as specified in ISO 9241-9 (ISO, 2000). The resulting 144 trials (16 directions × 3 widths × 3 distances) were presented in a random order in each block. Participants could take breaks at any time between trials by not moving the cursor back to the home position after the end of a trial. After each block, participants rated the input combination used during the block. At the end of the fourth block, they evaluated the four input combinations.

4.2 Results

Data analysis was performed using three 2×2×3×3 within-subjects ANOVAs, with *pointing method* (mouse or gaze), *selection method* (mouse button or EMG switch), *target size* (100, 125 or 150 pixels), and *distance to the target* (200, 250 or 300 pixels) as the independent variables. Completion time, throughput, and error rate were analyzed as the dependent variables. Our task required a successful activation to complete each trial. Unsuccessful activations resulted in longer completion times. To avoid the effect of unsuccessful activations on our speed measures, erroneous trials were removed from the data used for the ANOVAs of completion time and throughput. However, we also compared completion time data before and after removing erroneous trials in the Fitts' Law analysis described below. *Error rate* was defined as the proportion of erroneous trials (i.e., with one or multiple unsuccessful activations) in each condition.

Fitts' Law Analysis

The mean completion times for each combination of size and distance were used to analyze how well the data fitted Fitts' Law. As the index of difficulty (ID) increases, Fitts' Law predicts a linear increase in completion time. Following Equation 4, the regression lines for the four input combinations were calculated and plotted in Figure 7, together with their corresponding equations. The linear fits for all four input combinations show positive slopes, indicating that a positive correlation exists between ID and completion time, in accordance with Fitts' Law. The gaze-EMG combination had the shallowest slope of the four input combinations (slope = 0.14).

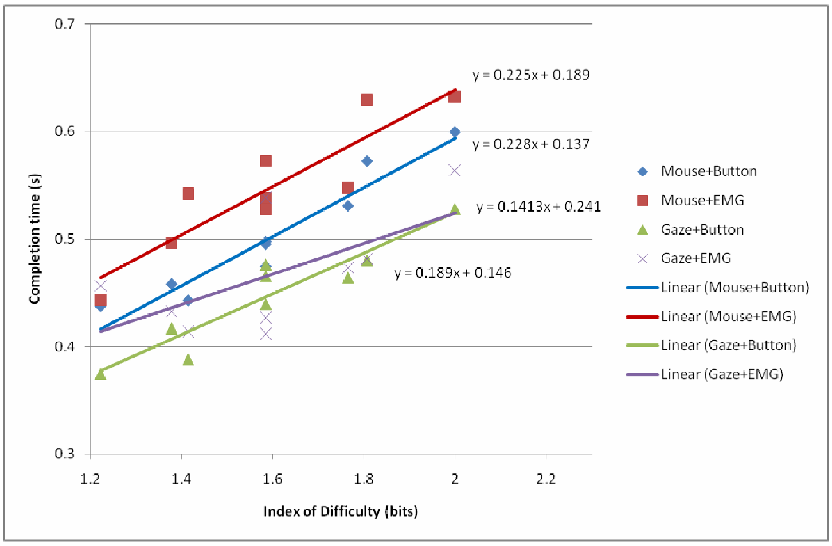

Figure 7. Completion time as a function of index of difficulty for all four input combinations.

A reanalysis of the data was performed after removing erroneous trials. The regression lines and corresponding equations are shown in Figure 8. When looking at these error-free data, input combinations in which the mouse was used for pointing present positive slopes (slope > 0.11), whereas combinations in which gaze was used for pointing present a virtually flat slope (slope < 0.01). This is in accordance with the findings by Partala, Aula, & Surakka (2001).

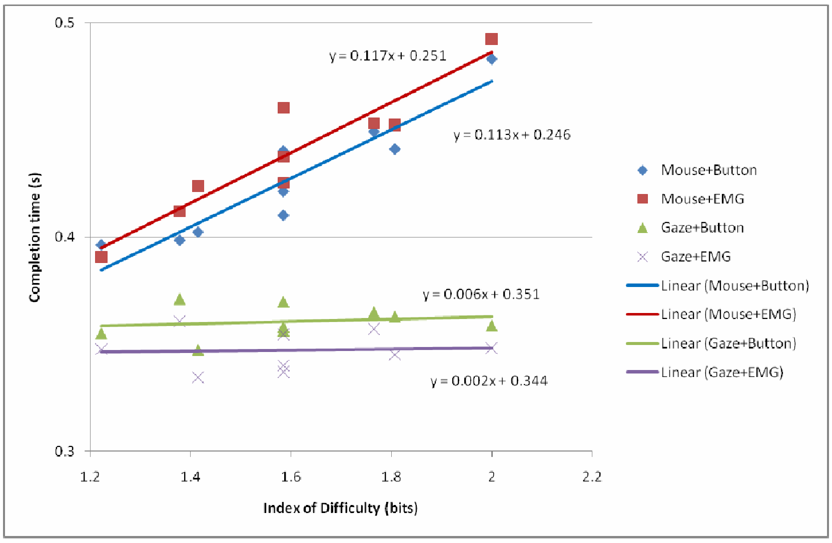

Figure 8. Completion time as a function of index of difficulty for all input combinations after removing erroneous trials.

Throughput

A high error rate was observed in this experiment. Therefore, a correction of the target width was performed by means of the error rate (MacKenzie, 1992). Overall mean throughput was 3.03 bits/s. Mean throughput was higher for gaze pointing (*M* = 3.31 bits/s, *SD* = 0.78 bits/s) than for mouse pointing (*M* = 2.76 bits/s, *SD* = 0.65 bits/s), $F(1, 4) = 7.98$, $p < 0.05$. Mean throughput was not significantly different between mouse selection (*M* = 3.10 bits/s, *SD* = 0.69 bits/s) and EMG selection (*M* = 2.97 bits/s, *SD* = 0.84 bits/s), $F(1, 4) = 1.52$, $p > 0.05$. Figure 9 shows the mean throughput obtained for each input combination. Target distance had a significant effect on throughput, $F(2, 8) = 5.12$, $p < 0.05$, but target size did not, $F(2, 8) = 0.58$, p > 0.05.

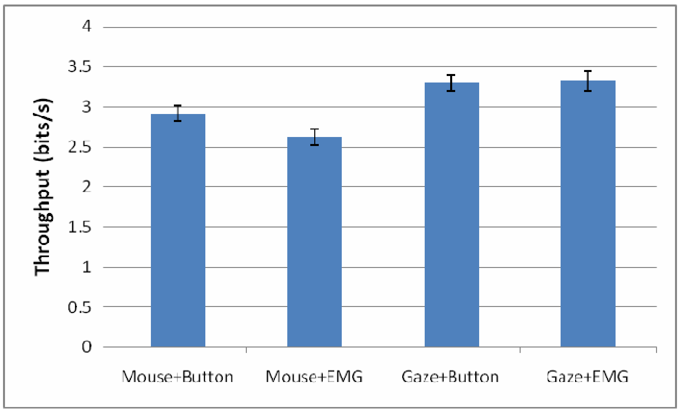

Figure 9. Mean throughput of each input combination. Error bars show standard errors of the mean.

Completion Time

Overall mean completion time was 393 ms. Mean completion time was lower for gaze pointing (*M* = 354 ms, *SD* = 46 ms) than for mouse pointing (*M* = 433 ms, *SD* = 43 ms), $F(1, 4) = 29.91$, $p < 0.05$. The mean completion times for mouse selection (*M* = 394 ms, *SD* = 57 ms) and EMG selection (*M* = 393 ms, *SD* = 62 ms) were not significantly different, $F(1, 4) = 0.004$, $p > 0.05$. Figure 10 shows the mean completion time for each input combination. Distance to the target, $F(2, 8) = 18.66$, $p < 0.05$, and target size, $F(2, 8) = 5.43$, $p < 0.05$, had an effect on completion time. Both longer distances and smaller sizes resulted in longer times.

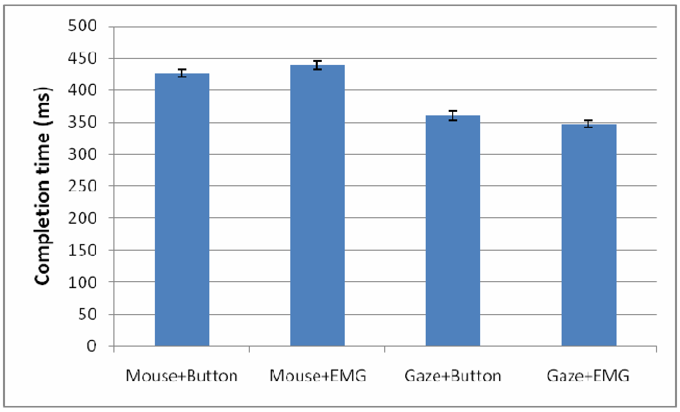

Figure 10. Mean completion time for each input combination. Error bars show standard errors of the mean.

Error Rate

Overall mean error rate was 22.25%. Neither pointing method, $F(1, 4) = 0.64$, $p > 0.05$, nor selection method, $F(1, 4) = 1.35$, $p > 0.05$, had a significant effect on error rate. Mean error rate was 21.45% (SD = 14.69%) for mouse pointing and 23.05% (SD = 13.27%) for gaze pointing. In the case of selection method, mean error rate was 20.69% (SD = 13.88%) for mouse selection and 23.82% (SD = 13.98%) for EMG selection. Figure 11 shows the mean error rate for each input combination. Target size affected error rate, $F(2, 8) = 15.63$, $p < 0.05$, while distance did not, $F(2, 8) = 3.32$, $p > 0.05$. Error rates were higher for distant and small targets than close, big ones.

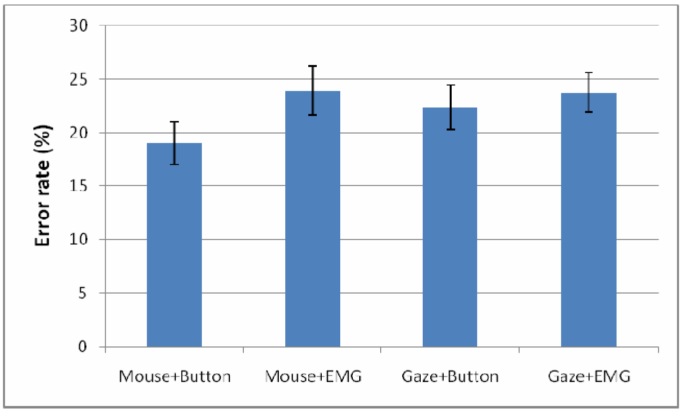

Figure 11. Mean error rate for each input combination. Error bars show standard errors of the mean.

Subjective Ratings

Participants rated gaze pointing as faster, but less accurate, than mouse pointing. Most of them reported that the gaze-EMG combination was natural to use, but they needed more practice to use it to its full potential. Gaze was also rated as fatiguing, in part because of the need to keep the head still for long periods of time. One participant even suggested using a chinrest.

5. Discussion

The results from the two experiments conducted in this study show a potential for gaze input to be used in videogames. Contrary to the findings of Sibert and Jacob (2000), our first experiment did not find the throughput of gaze to be higher than the throughput of the mouse. However, gaze throughput was higher than the throughput of a joystick, a device frequently used in games. Our second experiment did find gaze to have a higher throughput than the mouse (supporting Sibert and Jacob). Furthermore, it showed that the hands-free input gaze-EMG combination could perform at least as well as the mouse while allowing the user’s hands to be used to control other functions. Surakka, Illi, & Isokoski (2004) were not able to find a speed advantage of the gaze-EMG input combination over the mouse, and they suggested that such an advantage may become apparent if longer distances were used. However, we found such a speed advantage in our study even though the distances we used were, on average, shorter than those used by Surakka, Illi, & Isokoski (2004).

We attribute the different performance in our two experiments to the different eye trackers used in each. Although the Tobii tracker was set to the lowest possible smoothing between images, some smoothing was still performed on estimated gaze coordinates, which slowed down the cursor movement. Quick Glance did not apply any smoothing in our configuration, but the lower frame rate affected the responsiveness of the system, which again slowed down interaction. In comparison, the eye tracker used in our second experiment had no smoothing and a very low delay, allowing the participants to point at the targets much faster.

Unlike the other devices studied in Experiment 1, both eye trackers showed an improvement in throughput when target size increased. This finding can be attributed to the lower pointing accuracy of gaze pointing and the fact that bigger targets compensate for miscalibrations and possible offsets in the estimated cursor position. Interfaces designed specifically for gaze-based interaction should preferably present sufficiently large target areas to aid gaze input. However, it is important to note that the visual part of a target need not be as big as the target’s functional hit area. That is, a gaze-controlled game may well contain small targets that are difficult to discover – but easy to hit once they are detected.

In our first experiment, target tracking performance for small targets was relatively poor for both eye trackers, especially Quick Glance. Maintaining the pointer on the target can be challenging if the eye tracker is not accurate enough or if there is a lag between the eye movements and the cursor movement. In most of the popular shooting games, it is important not only to aim as quickly as possible, but also to accurately track a target while it is moving. Most commercial eye trackers are designed to detect user fixations and smooth the estimated gaze coordinates over a sequence of frames in order to make the cursor appear steady when the user fixates a point. Due to this smoothing, players using an eye tracker might experience the cursor as lagging behind when tracking a target. Eye trackers usually do not include algorithms for detection of smooth-pursuit movements. However, we believe that these kinds of algorithms would greatly benefit players using gaze input when performing target tracking tasks. In addition, it is possible that faster eye movements are especially useful under certain target tracking conditions (e.g., faster or less predictable moving targets). We did not study the effect of target speed or acceleration in our experiments, but it would be interesting to see, for instance, if gaze could outperform other input modes when following high-speed targets or when the speed of the target varies during its movement.

The participants in our study only tried each input device a few times, while real gamers will play over and over again before they master a new controller. In spite of this, participants with more than ten years of mouse experience were as good using gaze and EMG as they were using the mouse (or even better). We expect expert gaze-EMG users (e.g., gamers) to perform better and consistently outperform mouse users. A long-lasting learning experiment using more game-like stimuli may be more revealing of the true potential of gaze input for gaming. In addition, in order to obtain even more ecologically valid data on the value of gaze input for game interaction, it could be beneficial to develop a game that users can play from their home at their own pace. The game score could be calculated from the throughput and time-on-target performance metrics every time the user plays the game, providing feedback to them but also yielding data for statistical analysis. Data collected in this distributed and collaborative way could be used to obtain a better idea of the true potential of gaze-controlled game interaction.

EMG selections were as fast and accurate as mouse-button selections, but not faster (as Nelson et al., 1996, had found). This different result may be partially attributed to technical difficulties we encountered in our implementation. When interfacing EMG selection with our target presentation application, our program occasionally missed mouse clicks sent by the Cyberlink™ software, forcing the participant to issue another activation, and therefore increasing the completion time of the trial. However, differences between the pure reaction time task used by Nelson et al. (1996) and the target acquisition task we used may have played a role. Future studies should clarify this issue.

A Fitts' Law analysis of completion times for the different indexes of difficulty presented lines with positive slope, in accordance with the theoretical results. The gaze-EMG combination presented a shallower slope than the other input combinations, suggesting that this input combination may become more efficient as the ID of the task increases. A Fitts' Law regression analysis after removing erroneous trials presented a very flat response for gaze input. This is consistent with the study carried out by Partala, Aula, & Surakka (2001). The shallow (virtually flat) slope obtained for gaze pointing suggests that, in cases where the accuracy is high enough to acquire the target without errors, an increase in the index of difficulty (e.g., due to a higher distance to the target) does not affect the completion time. Since Fitts' Law implies that a positive correlation exists between ID and completion time, a reformulation of the law might be necessary for gaze interaction.

Subjective ratings suggest that discomfort associated with gaze input can be a serious drawback of this interaction technique, especially if the user needs to keep the head still for long periods of time. However, it is relevant to note that when the gaze tracking was particularly accurate, participants reported similar observations as those mentioned by Sibert and Jacob (2000). That is, pointing with gaze felt as if the system was "responding to their intentions, rather than to their explicit commands" (p. 282). In contrast, when there was an offset between actual and estimated point of regard (e.g., due to head movements), participants felt frustrated by their inability to correct the cursor position. Thus, given an eye tracker accurate enough and tolerant to naturally occurring head movements, participants may rate gaze pointing more positively.

In conclusion, we claim that, given a sufficiently accurate and responsive eye tracker and a well-designed interface, the use of gaze input holds interesting potential for game interaction. In our first experiment, we found that gaze had higher throughput than other input devices typically used in game interaction (e.g., joysticks). In our second experiment, we showed that a gaze-EMG input combination has the potential to perform at least as fast as the mouse while leaving the user's hands free to perform other functions. We obtained these results in spite of the fact that users received limited practice with a novel device and that we used very controlled tasks that do not fully reflect real-world gaming (and are less motivating to users). Future research should explore practice effects and use more ecologically valid tasks. For example, the idea of developing an online game with better graphics, sounds, and a motivating mission to accomplish may address the concerns about ecological validity. At the same time, it will also make the long-lasting study more feasible. One limitation of gaze input is its limited pointing accuracy. Using current technology, it is often necessary to use targets that are bigger than those found in most video games to obtain the results reported here. Future research should address some of these accuracy issues, both from the technological side (e.g., gaze estimation algorithms) and from the interface-design side. Given the demonstrated speed advantage of gaze over mouse pointing, the payoff of enabling reliable gaze input for game interaction could be invaluable.

6. Acknowledgments

This research was partly supported by the COGAIN Network of Excellence, IST IU 6, Contract Number 511598. We would like to thank Henrik Skovsgaard and Martin Tall from the IT University of Copenhagen for fruitful discussions and proofreading.

7. References

Dorr, M., Böhme, M., Martinetz, T., & Barth, E. (2007, September). Gaze beats mouse: a case study. Presented at 3rd *Annual Conference on Communication by Gaze Interaction, COGAIN 2007,* Leicester, UK.

Douglas, S. A., Kirkpatrick, A. E., & MacKenzie, I. S. (1999). Testing pointing device performance and user assessment with the ISO9241, Part 9 standard. *Proceedings of the ACM Conference on Human Factors in Computing Systems* (pp. 215-222), New York: ACM Press.

Hansen, J. P., Tørning, K., Johansen, A. S., Itoh, K., & Aoki, H. (2004). Gaze typing compared with input by head and hand. *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications, ETRA* (pp. 131-138) New York: ACM Press.

ISO (2000). *ISO/DIS 9241-9 Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9: Requirements for non-keyboard input devices*. International Standard, International Organization for Standardization.

Isokoski, P., & Martin, B. (2006, September). Eye tracker input in first person shooter games. Presented at the *2nd Conference on Communication by Gaze Interaction*. Torino, Italy.

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

Junker, A. M., & Hansen, J. P. (2006, September). Gaze pointing and facial EMG clicking. Presented at the *2nd Conference on Communication by Gaze Interaction*, Torino, Italy.

Klocheck, C., & MacKenzie, I. S. (2006). Performance Measures of Game Controllers in a Three-Dimensional Environment. *Proceedings of the 2006 conference on Graphics interface* (pp. 73–79). Toronto: Canadian Information Processing Society.

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

Majaranta, P., & Räihä, K. (2002, March). Twenty years of eye typing: Systems and design issues. Presented at the *2002 Symposium on Eye Tracking Research & Applications*, *ETRA*, New Orleans, Louisiana.

Nelson, W., Hettinger, L. J., Cunningham, J. A., Roe, M. M., Haas, M. W., Dennis, L. B., Pick, H. L., Junker, A., & Berg, C. (1996). Brain-body-actuated control: Assessment of an alternative control technology for virtual environments. *Proceedings of the 1996 IMAGE CONFERENCE* (pp. 225-232). Chandler, AZ: The IMAGE Society.

Nintendo of America, Inc. (2008). *Wii*. http://wii.com/

Partala, T., Aula, A., & Surakka, V. (2001). Combined voluntary gaze direction and facial muscle activity as a new pointing technique. In M. Hirose (Ed.). *INTERACT 2001* (pp. 100–107). Amsterdam: IOS Press.

Sibert, L. E., & Jacob, R. J. (2000). Evaluation of eye gaze interaction. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 281-288). New York: ACM Press.

Smith, J. D., & Graham, T. C. (2006, June). Use of eye movements for video game control. Presented *ACM SIGCHI International Conference on Advances in Computer Entertainment Technology (ACE '06).* Los Angeles, California, USA.

Sony Computer Entertainment, Inc. (2008). *EyeToy*. http://www.eyetoy.com.

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

Ware, C., & Mikaelian, H. H. (1987). An evaluation of an eye tracker as a device for computer input. *SIGCHI Bulletin, 17*, 183-188

Zhai, S., Morimoto, C., & Ihde, S. (1999). Manual and gaze input cascaded (MAGIC) pointing. *SIGCHI Conference on Human Factors in Computing Systems, CHI ’99* (pp. 246-253). New York: ACM Press.

Zhang, X., & MacKenzie, I. S. (2007). Evaluating eye tracking with ISO 9241 – Part 9. *Proceedings of HCI International 2007* (pp. 779-788). Berlin: Springer.