

# PsychNology Journal



Vol. 7 N.2

ISSN 1720 - 7525

IN COOPERATION WITH  
**COGAIN**

Edited by  
Martin Boehme, John-Paulin Hansen, Fiona Mulvey

## Communication by Gaze Interaction



---

# PSYCHOLOGY JOURNAL

## The Other Side of Technology

---

### EDITORS-IN-CHIEF

**Luciano Gamberini**

Department of General Psychology, Padova University, Italy.

**Giuseppe Riva**

Catholic University of Milan, Italy.

**Anna Spagnoli**

Department of General Psychology, Padova University, Italy.

---

### EDITORIAL BOARD

**Mariano Alcañiz Raya:** Universidad Politecnica de Valencia. Valencia, Spain.

**Cristian Berrío Zapata:** Pontificia Universidad Javeriana. Bogotá, Colombia.

**Rosa Baños:** Universidad de Valencia, Valencia, Spain.

**David Benyon:** Napier University, Edinburgh, UK.

**Cristina Botella:** Univeritat Jaume I. Castellón, Spain.

**Antonella de Angeli:** University of Manchester. United Kingdom

**Jonathan Freeman:** Goldsmiths College, University of London. United Kingdom.

**Christine Hine:** University of Surrey. Guildford, United Kingdom.

**Christian Heath:** King's College. London, United Kingdom.

**Wijnand Ijsselsteijn:** Eindhoven University of Technology. The Netherlands.

**Giulio Jacucci:** Helsinki Institute for Information Technology. Helsinki, Finland.

**David Kirsh:** University of California. San Diego (CA), USA.

**Matthew Lombard:** Temple University. Philadelphia (PA), USA.

**Albert "Skip" Rizzo:** University of Southern California. Los Angeles (CA), USA.

**Ramakoti Sadananda:** Rangsit University. Bangkok, Thailand.

**Angela Schorr:** Universität Siegen. Siegen, Germany.

**Paul F.M.J. Verschure:** Universitat Pompeu Fabra. Barcelona, Spain.

**Alexander Voiskounsky:** Moscow State University. Moscow, Russia.

**John A Waterworth:** Umeå University. Umea, Sweden.

**Brenda K. Wiederhold:** Interactive Media Institute-Europe. Brussels, Belgium.

### CONSULTING EDITORS

**Hans Christian Arnseth:** University of Oslo. Oslo, Norway.

**Marco Casarotti:** University of Padova. Padova, Italy.

**Roy Davies:** Lund University. Lund, Sweden.

**Andrea Gaggioli:** Catholic University of Milan. Milan, Italy.

**Pietro Guardini:** Padova University. Padova, Italy.

**Frode Guribye:** University of Bergen. Bergen, Norway.

**Raquel Navarro-Prieto:** Universitat Oberta de Catalunya. Castelldefels, Spain.

**Stephan Roy:** Hospital Sainte Anne. Paris, France.

**Carlos Ruggeroni:** National University of Rosario. Rosario, Argentina.

### EDITORIAL ASSISTANTS

**Bruno Seraglia, Concetta Alberti:** University of Padova. Padova, Italy.

---

PSYCHOLOGY JOURNAL, PNJ  
PUBLISHED ON-LINE SINCE SUMMER 2002  
WEB SITE: [HTTP://WWW.PSYCHOLOGY.ORG](http://www.psychology.org)  
SUBMISSIONS: [ARTICLES@PSYCHOLOGY.ORG](mailto:ARTICLES@PSYCHOLOGY.ORG)



## TABLE OF CONTENTS

Editorial Preface..... p. 139

### **SPECIAL ISSUE: Gaze control for work and play**

Predicting preference from fixations..... p. 141  
Mackenzie G. Glaholt, Mei-Chun Wu, Eyal M. Reingold

Scrollable Keyboards for Casual Eye Typing..... p. 159  
Oleg Špakov, Päivi Majaranta

Hands Free Interaction with Virtual Information in a Real Environment:  
Eye Gaze as an Interaction Tool in an Augmented Reality System..... p. 175  
Susanna Nilsson, Torbjörn Gustafsson, Per Carleberg

Gaze beats mouse: A case study on a gaze-controlled breakout..... p. 197  
Michael Dorr, Laura Pomarjanschi, Erhardt Barth

Evaluation of the Potential of Gaze Input for Game Interaction..... p. 213  
Javier San Agustin, Julio C. Mateo, John Paulin Hansen, Arantxa Villanueva



## Editorial Preface

In many motor disorders, such as Amyotrophic Lateral Sclerosis (ALS), the eyes are among the last parts of the body to retain voluntary motor control. Eye tracking technology, which measures the direction of gaze, is one way of assisting people with these kinds of motor impairments to communicate and to access modern information technology.

COGAIN (“Communication by Gaze Interaction”) is a European Network of Excellence (NoE) aimed at developing new eye tracking technologies and applications for people with disabilities, as well as establishing standards and disseminating information about existing systems to those who may need or want them. COGAIN also aims to research and develop the potential of gaze as a viable input modality for non disabled users as well as a rich source of information on the attention and intention of users in human computer interaction in general.

The COGAIN project hosts a yearly conference on eye gaze interaction, and this special issue of *PsychNology Journal* (7.2) contains extended versions of selected papers from the 2007 and 2008 conferences. COGAIN 2007 was held on September 2 and 3, 2007, at De Montfort University, Leicester, UK; the theme of the conference was “Gaze-based Creativity, Interacting with Games and On-line Communities”. For COGAIN 2008, the theme was “Communication, Environment and Mobility Control by Gaze”, and the conference was held in Prague, Czech Republic, on September 2 and 3, 2008.

A call for extended papers based on the material presented at the conferences was published. From this call, we present five papers in this special issue. The first paper, “Predicting preference from fixations” by Mackenzie G.

Glaholt, Mei-Chun Wu and Eyal M. Reingold, examines the relationship between personal preference for certain items (faces and mock company logos) and gaze fixation on those items. The authors show that there is a strong positive correlation between total fixation time and preference. Moreover, they show that fixation time can be used to infer preference for individual features of an item (for example, the font, shape or texture of a logo) and demonstrate how this information can be used to predict preference for new, unseen items. The authors suggest that these insights could be used to assist users in selecting items from a large array and to create smart applications that automatically adapt their appearance to the user's inferred preferences.

The second paper, “Scrollable Keyboards for Casual Eye Typing” by Oleg Špakov and Päivi Majaranta, presents a novel interface for eye typing, i.e. for inputting text using an eye tracker. Instead of presenting a full keyboard on screen, from which characters are typically selected using dwell time, the idea of a scrollable keyboard is to show only one or two rows of the keyboard and allow the user to access rows that are currently not visible using special scroll keys. Compared to a full keyboard, the scrollable keyboard saves space or, alternatively, allows larger keys to be displayed in the same space. Špakov and Majaranta compare typing speeds on a full keyboard with 1- and 2-row scrollable keyboards. They show that, for a 2-row keyboard using an optimized layout, the reduction in typing speed is less than 20%, which, they argue, is entirely tolerable for casual typing, such as filling in web forms.

The next paper, “Hands Free Interaction with Virtual Information in a

Real Environment – Eye Gaze as an Interaction Tool in an Augmented Reality System” by Susanna Nilsson, Torbjörn Gustafsson and Per Carleberg, examines eye tracking as a medium for interaction in medical applications, where traditional user interfaces may not be practicable because of the need to keep the hands sterile. The authors describe two applications where an eye tracker is integrated into a head-mounted augmented-reality display. In the first application, gaze interaction is used to set up an “electrical knife” for surgery. The second application displays instructions on how to assemble a surgical tool, and gaze interaction is used to move from one step of the procedure to the next. The paper presents the results of user trials on the two systems and closes with a discussion of possible modifications and improvements.

Gaze interaction in gaming applications is the topic of the two remaining papers. The first of these, “Gaze beats mouse: A case study on a gaze-controlled Breakout” by Michael Dorr, Laura Pomarjanschi and Erhardt Barth, compares the performance of gaze and mouse input for the computer game Breakout, where players have to move a bat to hit one or several balls against a wall of bricks. The authors compare the performance of the two input options by pitting a gaze player against a mouse player; in two thirds of the rounds that were played, the gaze player won, demonstrating that gaze can be a superior input option even for able-bodied users who might otherwise use a mouse.

The final paper, “Evaluation of the Potential of Gaze Input for Game Interaction” by Javier San Agustin, Julio C. Mateo, John Paulin Hansen and Arantxa Villanueva, also compares gaze to other input modalities, but on the level of individual tasks (target acquisition and target tracking). In the

first of two experiments, two eye trackers are compared against a mouse, a touch screen, a head tracker and a joystick. Gaze is shown to be superior to the head tracker and the joystick; it is also superior to mouse and touch screen for large target sizes, though not for small ones. The second experiment combines gaze for pointing with an electromyographic (EMG) facial signal for selection and shows that this combination outperforms the mouse. Interestingly, unlike for traditional pointing devices, the completion time for the gaze-EMG combination does not seem to increase significantly with distance to the target.

These papers were chosen based on anonymous review and with the aim of presenting a broad spectrum of topics from the COGAIN conferences. We would like to thank our reviewers and authors as well as the editors-in-chief of PsychNology Journal for their time and work, and encourage interested readers to follow the research results of the COGAIN network through the COGAIN Association (see [www.cogain.org](http://www.cogain.org)).

**Martin Böhme**

University of Lübeck

**John Paulin Hansen**

IT University of Copenhagen

**Fiona Mulvey**

Technical University of Dresden

Guest Editors

## Predicting preference from fixations

Mackenzie G. Glaholt<sup>\*♦</sup>, Mei-Chun Wu<sup>\*</sup>, and Eyal M. Reingold<sup>♦</sup>

<sup>♦</sup>University of Toronto  
Mississauga  
(Canada)

<sup>\*</sup>National Sun Yat-sen University  
(Taiwan)

---

### ABSTRACT

We measured the strength of the association between looking behaviour and preference. Participants selected the most preferred face out of a grid of 8 faces. Fixation times were correlated with selection on a trial-by-trial basis, as well as with explicit preference ratings. Furthermore, by ranking features based on fixation times, we were able to successfully predict participants' preferences for novel feature combinations in a two-alternative forced choice task. In addition, we obtained a similar pattern of findings in a very different stimulus domain: mock company logos. Our results indicated that fixation times can be used to predict selection in large arrays and they might also be employed to estimate preferences for whole stimuli as well as their constituent features.

---

Keywords: *Eye movements, preference, gaze bias, decision making.*

Paper Received 14/11/2008; received in revised form 24/02/2009; accepted 29/04/2009.

### 1. Introduction

Measures of fixation duration and location have proven to be invaluable in the context of human factors, usability engineering, and marketing and advertising research (see Duchowski, 2002 for a recent review). In many real world and computer based applications the user is confronted with a cluttered array of options, in which objects and locations are serially and often repeatedly selected for detailed or attentive processing (e.g. looking through a gallery of image thumbnails). Given that attention moves among the options in a

---

Cite as:

Glaholt, M.G., Wu, M., & Reingold, E.M. (2009). Predicting preference from fixations. *PsychNology Journal*, 7(2), 141 – 158. Retrieved [month] [day], [year], from [www.psychology.org](http://www.psychology.org).

\* Corresponding Author:

Mackenzie G. Glaholt  
University of Toronto Mississauga  
Department of Psychology  
3359 Mississauga Road N. RM 2037B  
Mississauga, Ontario, Canada L5L 1C6  
e-mail: mackenzie.glaholt@gmail.com

largely serial fashion, such a process can be slow and effortful. This is especially the case for tasks that require detailed visual comparison between non-adjacent objects (e.g., online shopping, e-learning). Consequently, in the context of visual choice and comparison tasks, monitoring the distribution and duration of eye fixations has the potential to provide an excellent measure of an observer's interests and preferences. This, in turn, would allow the development of smart applications that facilitate and customize information retrieval and inspection based on the users' manifest preferences (e.g., an art image database that learns the user preferences and biases image retrieval accordingly).

The goal of the present research was to examine the usefulness of looking behavior as an indirect measure of the observer's preferences. If shown to be a reliable and sensitive measure of preference, looking behavior would potentially have several advantages over traditional self-report measures of preference (e.g., ratings, questionnaires, interviews). First, current eye movement monitoring systems allow for relatively unobtrusive measurement of looking behavior while the observer is interacting naturally with their visual environment. Thus, unlike overt preference ratings, the observer is not required to produce additional responses to indicate his/her preferences. Second, compared to preference ratings, looking behavior is likely to provide better measurement of unconscious preferences. Third, looking behavior is likely to be less susceptible to attempts on the part of the user to only report socially desirable, appropriate, or justifiable preferences. Finally, measurement of preferences by looking behavior can be obtained quickly and efficiently across multi-element arrays of items.

Given these possible advantages, it is surprising that relatively few empirical studies have examined the relation between preference and looking behavior in adults (but see Isaacowitz, Wadlinger, Goren, & Wilson, 2006; and for a review see Murphy & Isaacowitz, 2008). However, several recent studies suggest that such an effort may be feasible and promising. Specifically, Shimojo, Simion, Shimojo and Scheier (2003) and Simion and Shimojo (2006, 2007) reported a gaze bias that exists during selection between two visually presented items. On each trial, two faces were presented and participants had to select the more attractive face. Gaze was shown to be biased towards the face that was later selected. This gaze bias became evident between 1 and 1.5 seconds prior to the response that marked the overt decision. Building on this finding, Glaholt and Reingold (in press) demonstrated that the bias in looking behavior was particularly robust in eight-alternative forced-choice (8-AFC) decision tasks. These findings indicate that by monitoring eye movements it may be possible to predict the observers' choice or preference prior to the

overt response and possibly prior to the point at which the choice is consciously made. Bee, Prendinger, André and Ishizuka (2006) demonstrated the feasibility of using eye movements to predict the visual preference decisions of users in real-time, for the purpose of designing applications that would automatically detect users' visual preferences solely based on eye movements in a two-alternative forced choice (2-AFC) setting. These authors reported that in a pilot study involving the selection of neckties, their system correctly classified participants' choices with an average accuracy of 81% (with 50% constituting chance performance).

Our study aims to extend these prior findings in several ways. Specifically, the present study was designed to provide quantitative estimates for the strength of the association between fixation patterns and observers' preferences. We examined gaze behavior during preference decisions in multi-element arrays. Multi-element arrays mimic the kinds of displays that are present in a variety of applied settings, such as web-based image catalogues where the decision maker searches through a large set of decision alternatives. In addition, we examined the potential for using fixation times extracted during the viewing of multi-element arrays to circumvent the need for overt selection between pairs of items and thereby boost the efficiency of search through large sets of potential alternatives.

We applied the analysis of the within-trial gaze bias reported by Shimojo, Simion, Shimojo and Scheier (2003) to our 8-AFC task and measured the accuracy of the within-trial prediction of selection from gaze data. In order to estimate the strength of the gaze-selection and gaze-preference relationships, measures of fixation duration (see Rayner, 1998 for a review of eye movement measures) were then correlated with both overt choice behavior and explicit preference ratings. For such correlations to be useful they must be reliable and robust at the level of an individual and not just across a group of participants, and consequently we computed correlations separately for each participant. We also wanted to see if, once the individual's preferences for stimulus features are determined, it would be possible to construct novel combinations of these features and accurately predict the pattern of choices among these stimuli. If so, this would demonstrate how fixation times can narrow search for preferred items in a large feature space.

Accordingly, in the first part of the experiment, each participant was asked to select the most attractive stimulus in visual arrays of 8 faces (group 1) or arrays of 8 company logos (group 2) (see the 8-AFC task, Figure 1). Each item in each array represented a unique combination of 3 stimulus dimensions with 8 possible features in each dimension (see Figures 3 & 4). In the second component, participants explicitly rated their preference for each of the stimuli from part 1. These data allowed us to examine the relationship of fixation

time to selection (within-trial) and fixation time to overt preference rating. In the final part of the experiment, pairs of novel stimuli were constructed that were expected to constitute easy or difficult preference decisions based on the participant's fixation patterns from part 1. For each trial of this 2-AFC task, we had a prediction about which of the two stimuli the participant would prefer.

## **2. Method**

### **2.1 Participants**

One group of eight participants took part in the face version of eight-alternative task and another group of eight participants took part in the logo version. In the second part of the experiment, five participants from each group completed the two-alternative task. All participants were students at the University of Toronto at Mississauga, and each received \$10 compensation for their time. All participants provided informed consent, and the reported research was conducted in strict compliance with APA ethical principles.

### **2.2 Apparatus**

The eyetracker employed in this research was the SR Research Ltd. EyeLink 1000 system. This system has high spatial resolution ( $0.005^\circ$ ) and a sampling rate of 1000 Hz (1-msec temporal resolution). By default, only the participant's dominant eye was tracked in our study. In the present study, the configurable acceleration and velocity thresholds were set to detect saccades of  $0.5^\circ$  or greater. Stimulus displays were presented on two monitors, one for the participant (a 19-in. Viewsonic) and one for the experimenter. The experimenter monitor was used to give feedback in real time about the participant's computed gaze position. This feedback was given in the form of a cursor measuring  $1^\circ$  in diameter that was overlaid on the same image being viewed by the participant. This allowed the experimenter to evaluate system accuracy and to initiate a recalibration if necessary. In general, the average error in the computation of gaze position was less than  $0.5^\circ$  of visual angle. The participant used a chinrest with a head support to minimize head movement.

### **2.3 Stimuli**

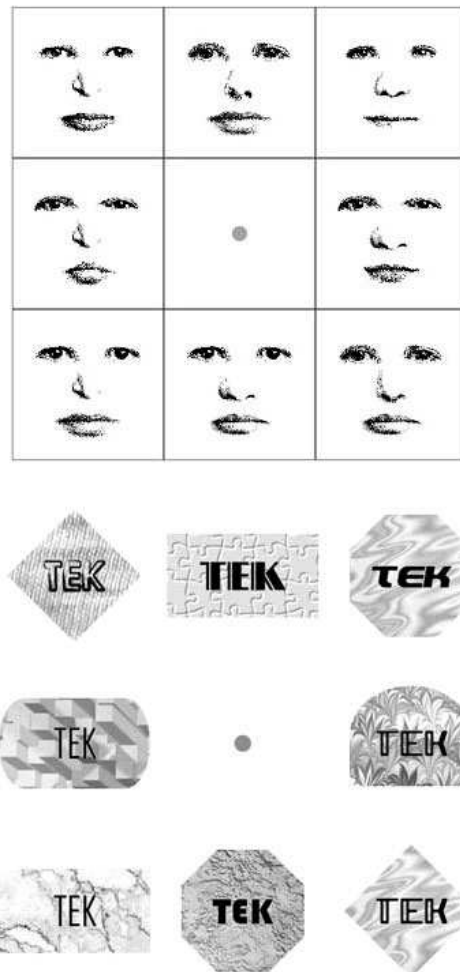
Faces were constructed as unique combinations of 3 stimulus dimensions (eyes, nose, and mouth) with 8 possible features (i.e., exemplars) in each dimension. The features were

stored as bitmaps and assembled into faces on each trial (see Figure 3). Of the 512 possible faces in our feature space, we selected a set of 64 in which all pairs of features occurred once, and all individual features occurred eight times. These faces were used in the 8-AFC task and the Preference Rating task. Of the remaining 448 faces, a subset (determined separately for each participant) was used in the two-alternative forced choice task. The logo version of the experiment was analogous to the face version. We constructed logos by combining features from each of three stimulus dimensions: font, shape, and texture (see Figure 4). All the logos were portrayed as possible logos for a fictional company with the initials 'TEK'.

## 2.4 Procedure

The experiment consisted of three components that were completed by the participant in a fixed order: eight-alternative forced-choice, preference rating, and two-alternative forced choice. Only five of the eight participants in each group were available to complete the 2-AFC task. One group of participants was given the face version of the experiment, and another group was given the logo version. Each participant was given instructions prior to each component of the experiment. In the 8-AFC task, following a 9-point calibration, eye movements were recorded while the participant selected, from each display presented, the most attractive face. All 64 faces (see Stimuli) were presented 8 times, across 64 stimulus displays, where each display contained a unique combination of 8 stimuli (see Figure 1). At the beginning of each trial the display appeared, and the participant decided which of the eight stimuli (faces or logos) was most attractive. To terminate trials, the participant first had to look at the grey dot located in the center of the display and fixate it for 500 ms (the beginning of the saccade preceding this fixation was considered the end of the trial for the purpose of analysis). Following this fixation, the grey center dot turned green indicating to the participant to move his/her gaze to the preferred stimulus in order to select it and terminate the trial.

In the second component, Preference Rating, the participant viewed each of the 64 stimuli from the 8-AFC task, one at a time, and rated the attractiveness of the face or logo on a 300-pixel wide sliding scale anchored by 'Unattractive' and 'Attractive' on the left and right ends, respectively. On each trial the stimulus appeared in the center of the display and the rating scale appeared horizontally below. The participant responded using the mouse, and the preference rating was recorded as the position of the participant's mark along the axis of the rating scale.



**Figure 1.** Stimulus displays from the 8-AFC task in the face version (top) and the logo version (bottom).

Following the Preference Rating task, there was a short break during which the experimenter analyzed the data from the 8-AFC task. Average total fixation time, across presentations, was computed for each of the 24 features (see Results for feature analysis). For each of the 448 stimuli (logos or faces) not yet seen by the participant, we computed an average total fixation time (i.e., the average of the mean fixation times for its three component features). We then ranked the new stimuli according to total fixation time and selected 'high preference' items from the top quartile 'low preference' items from the bottom quartile. In the 2-AFC component, the participant completed 64 trials, half of which were high-low preference pairs (Easy decision), and the other half were high-high preference pairs (Difficult decision). Importantly, the pairs were selected such that the two stimuli did not

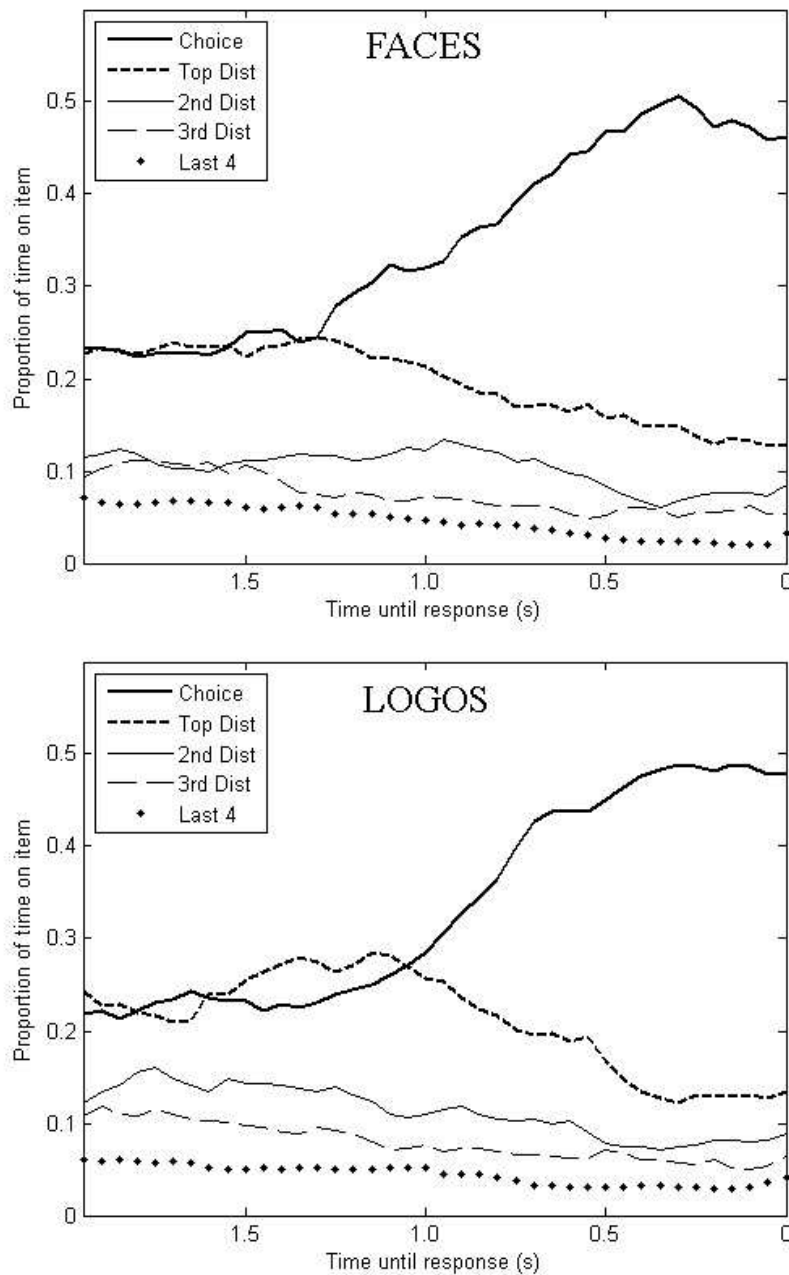
share features. The stimulus display was similar to the display in the 8-AFC task, but containing only the middle row of boxes (see Figure 1). Participants were asked to select the stimulus they preferred. For both Easy and Difficult pairs, we recorded both the correspondence between the choice made by the participant and the predicted choice, as well as the response latency for the decision. We expected to find higher prediction accuracy and shorter response latencies for the Easy trials compared to the Difficult trials.

### 3. Results

To analyze the gaze behavior in the 8-AFC task, we defined 8 non-overlapping regions of interest (corresponding to the boxes in Figure 1, top), each of which contained one stimulus. The summed duration of all fixations on a stimulus throughout the trial (i.e., from stimulus onset to the participant's re-fixating the center prior to selection) is referred to as total fixation time. Our analyses were aimed at exploring two different issues: 1) in order to compare the present results to previous investigations, we looked at the association between gaze behavior and preference decisions within a trial, and 2) we examined how total fixation times for individual stimuli and features, averaged across trials, could be used for predicting overt preference decisions, subjective preference ratings, and preference decisions for novel stimuli.

#### 3.1 Gaze and Preference Within-trial

Similar to previous research (Glaholt & Reingold, in press, 2009; Shimojo, Simion, Shimojo and Scheier, 2003; Simion & Shimojo, 2006, 2007), we explored the gaze bias toward the chosen item on each trial of the 8-AFC by plotting the proportion of time spent on the chosen item prior to the response. In addition, we ranked each of the items that were not chosen (referred to here as 'distractors') according to their total fixation time for the trial. We then labeled them as the 'top distractor', '2<sup>nd</sup> distractor', '3<sup>rd</sup> distractor', and the four lowest-rank distractors were averaged together and labeled 'Last 4'. In Figure 2, for each of forty 50-millisecond time bins (i.e., a 2 second interval) prior to the response, we plot the proportion of time the eye spent on the chosen item, and on each distractor category. Data for the face (Figure 2, top) and logo (Figure 2, bottom) versions of the task were plotted separately, collapsing across all participants and trials. In total, 3% of the trials were excluded because they lasted less than 2 seconds.



**Figure 2.** Plots of the proportion of time that gaze was directed to the chosen item or distractor items (ranked by total fixation time) over the 2 seconds prior to the response in the 8-AFC task for faces (top) and logos (bottom).

As can be seen in Figure 2, the time course plots for the Face and Logo versions of the experiment are very similar. Replicating previous findings (Glaholt & Reingold, in press, 2009; Shimojo, Simion, Shimojo and Scheier, 2003; Simion & Shimojo, 2006, 2007), gaze

was biased towards the chosen item during the last second prior to the decision. The first distractor item seems distinguishable from the other distractors; it enjoyed a similar proportion of the gaze to the chosen item up until approximately a second prior to the decision, after which the chosen item dominated. This may indicate that towards the end of the trial, the participant was choosing primarily between the top two options. The other distractor items received a smaller amount of gaze throughout the interval. Interestingly, the ranking of the other distractors seems stable over the interval, suggesting that gaze duration may provide a stable estimate of the preference ranking of each option.

To quantitatively assess the apparent gaze bias towards the chosen item, we computed the percentage of trials in which the chosen item had the highest total fixation time (chance = 12.5%). We also computed the percentage of trials in which total fixation time for the chosen item was one of the two largest (chance = 25%), or four largest fixation times (chance = 50%). For all of these measures, total fixation time was substantially longer on the selected item than predicted by chance (for all comparisons,  $t(7) > 11.85$ ,  $p < 0.001$ ). Specifically, for faces, the chosen item had the highest total fixation time on 65.2% of the trials, was within the top 2 total fixation times on 85.5% of the trials, and was in the top four total fixation times on 95.3% of the trials. For logos, the chosen item had the highest total fixation time on 67.8% of the trials, was within the top 2 total fixation times on 90.0% of the trials, and was in the top four total fixation times on 97.7% of the trials. Thus, total fixation time within a trial is a powerful predictor of the item chosen, and an even stronger predictor of the active subset of the options from which the participant is choosing.

The time course plots clearly depict a pattern of increasing gaze selectivity throughout the trial. To quantify this narrowing of actively considered items over the course of the trial, we divided the trial into dwells (where a dwell is a run of one or more consecutive fixations on an item). We contrasted the first four dwells (beginning of trial) and last four dwells (end of trial) in each trial (13% of trials had less than 8 dwells and were excluded). Total fixation time on the chosen item and on other items was computed separately for the beginning and the end of the trial, for either the face or logo versions of the task. The result of this analysis is shown in Table 1 and clearly depicts a dramatically stronger gaze bias in the end than in the beginning of the trial (for all comparisons,  $t(7) > 4.89$ , all  $p < 0.01$ ). The chosen item had the longest total fixation time, and was in the top two total fixation times, more often in the end than in the beginning of the trial. Interestingly, even in the beginning of the trial, the chosen item had the longest total fixation time and was one of the top two total fixation times, more often than chance. This increase in gaze bias toward the end of the trial is also reflected in

the fact that the chosen item was more frequently visited and fixated on for a longer duration during the end than in the beginning of the trial.

	Dwell set	# of visits to chosen item	Chosen item total fixation time (ms)	Chosen item is top fixation time (prop'n)	Chosen item within top 2 fix. times (prop'n)
Faces	First 4	0.55	244	0.24	0.48
	Last 4	1.12	763	0.56	0.87
Logos	First 4	0.6	301	0.27	0.53
	Last 4	1.07	691	0.49	0.84

**Table 1.** Analysis of the first 4 and last 4 dwells in each trial in the 8-AFC task.

### 3.2 Total Fixation Time and Preference Across-trials

Having illustrated the relationship between gaze and preference decisions in the 8-AFC task within-trial, we asked what relationships hold across trials. We examined the correlation between total fixation time and preference, at two levels: 1) at the level of whole stimuli (faces or logos), and 2) at the level of individual features. These analyses are discussed separately below.

*Correlations across stimuli.* Each stimulus (face or logo) was part of the stimulus array in eight trials during the 8-AFC task. Across these eight presentations, we counted the number of times the stimulus was chosen (Selection) and computed the average total fixation time on this item. In addition, for each stimulus we obtained the subjective preference rating (Preference) from the rating task. For each participant, the correlations between total fixation time and Selection, and between total fixation time and Preference, were computed across stimuli and are displayed in Table 2 (faces) and Table 3 (logos). In addition, to measure the strength of the relationship between the three variables we computed the multiple correlation between total fixation time and both the number of times selected and mean preference rating (Selection + Preference). As can be seen in the Tables, for all participants, total fixation time was strongly and positively correlated with Selection (i.e., number of times an item is chosen) and Preference (i.e., subjective preference ratings). Furthermore, taken

together both Selection and Preference were highly predictive of total fixation time, reflected in a high multiple correlation.

*Correlations across features.* We derived measures of total fixation time, number of times selected, and preference rating for individual features. Each feature appeared in different stimuli, which allowed us to derive the average contribution of each feature to total fixation time, number of times selected, and preference rating. We assigned the values for these variables from each whole stimulus to its component features. For example, for a given face, the total fixation time, number of times selected, and preference rating were assigned to the eyes, nose, and mouth features that composed that face (font, shape, and texture for logos). For each feature, these values were averaged across 64 occurrences (in 8 different items, each presented 8 times). The correlations, as computed across features, show a particularly strong positive relationship between total fixation time and selection and also between total fixation time and overt preference rating (see Tables 2 and 3). Taken together both Selection and Preference were highly predictive of total fixation time, as reflected in a high multiple correlation.




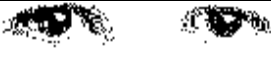


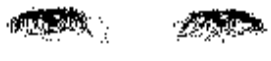


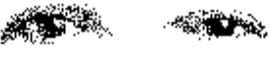

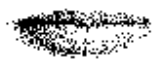
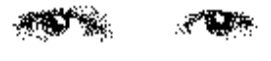




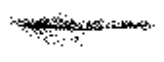
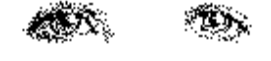





Participant	Whole Stimuli (Faces) (Pearson's r)			Single Features (Pearson's r)		
	Selection	Preference	Selection + Preference	Selection	Preference	Selection + Preference
1	0.65†	0.62†	0.69	0.89†	0.75†	0.89
2	0.74†	0.58†	0.76	0.84†	0.70†	0.85
3	0.70†	0.66†	0.74	0.86†	0.83†	0.89
4	0.80†	0.48†	0.78	0.89†	0.56§	0.89
5	0.66†	0.39§	0.69	0.81†	0.56§	0.83
6	0.75†	0.74†	0.83	0.88†	0.86†	0.92
7	0.65†	0.39§	0.65	0.83†	0.48§	0.84
8	0.70†	0.39§	0.70	0.83†	0.70†	0.84
Mean	0.71	0.53	0.73	0.85	0.68	0.87

**Table 2.** 8-AFC with faces. Correlations between total fixation time and number of times selected (Selection), between total fixation time and mean preference rating (Preference), and the multiple-correlation between total fixation time and both the number of times selected and mean preference rating (Selection + Preference), computed across whole stimuli and across single features, for each participant. Note: †:  $p < 0.001$ , §:  $p < 0.01$ , \*:  $p < 0.05$ .

Participant	Whole Stimuli (Logos) (Pearson's r)			Single Features (Pearson's r)		
	Selection	Preference	Selection + Preference	Selection	Preference	Selection + Preference
1	0.75†	0.57†	0.75	0.92†	0.80†	0.92
2	0.84†	0.56†	0.84	0.91†	0.66†	0.91
3	0.65†	0.34§	0.66	0.83†	0.44*	0.84
4	0.74†	0.57†	0.76	0.86†	0.65†	0.86
5	0.87†	0.60†	0.88	0.95†	0.71†	0.96
6	0.81†	0.67†	0.81	0.91†	0.76†	0.91
7	0.74†	0.88†	0.91	0.96†	0.88†	0.97
8	0.71†	0.65†	0.76	0.89†	0.81†	0.90
Mean	0.76	0.61	0.80	0.90	0.71	0.91

**Table 3.** 8-AFC with logos. Correlations between total fixation time and number of times selected (Selection), between total fixation time and mean preference rating (Preference), and the multiple-correlation between total fixation time and both the number of times selected and mean preference rating (Selection + Preference) computed across whole stimuli and across single features, for each participant. Note: †:  $p < 0.001$ , §:  $p < 0.01$ , \*:  $p < 0.05$ .

*Correlations across participants.* In order to evaluate the consistency of preferences across participants, we computed between-participant correlations across features for average total fixation time, number of times selected (Selection), and preference rating (Preference). For both faces and logos, these correlations were quite variable, often low, and sometimes negative. For faces, total fixation time: range = -0.11 to 0.84, mean = 0.43; Selection: range = -0.02 to 0.89, mean = 0.51; Preference: range = -0.08 to 0.78, mean = 0.48. For logos, total fixation time: range = -0.27 to 0.64, mean = 0.11; Selection: range = -0.37 to 0.57, mean = 0.07; Preference: range = -0.46 to 0.54, mean = 0.13. This indicates that, at least for the stimuli used in this experiment, individual differences in preference are substantial, and consequently preference predictors should be derived separately for each participant. Note that the correlations across participants tended to be higher for faces than logos (for each measure), and hence the consistency of preference across participants may depend strongly on the stimulus domain. Nevertheless, it is of interest to portray the central tendency of preference for a group of participants. To demonstrate this, we normalized the average total fixation time for each feature for each participant by converting it to a z-score, and we then ranked each feature according to its mean z-score across participants (see Figures 3 and 4).

Rank	Eyes	Noses	Mouths
1			
2			
3			
4			
5			
6			
7			
8			

**Figure 3.** Face features in rank order (from highest-1 to lowest-8) according to normalized average total fixation time across participants.

### 3.3 Predicting Preference Decisions for Novel Stimuli

To further establish the usefulness of fixation information across trials, we used average total fixation time on individual features to rank the expected preference for stimuli other than the 64 that were used in the 8-AFC task (i.e., the 448 combinations of features that were not used in the 8-AFC task). High-high preference pairs (Difficult trials) and High-Low preference pairs (Easy trials) were then used in a 2-AFC preference decision task. For each pair of stimuli, we predicted which of the two items would be preferred based on the average total fixation time for individual features obtained from the 8-AFC task.

Rank	Font	Shape	Texture
1			
2			
3			
4			
5			
6			
7			
8			

**Figure 4.** Logo features in rank order (from highest-1 to lowest-8) according to normalized average total fixation time across participants.

Generally, these predicted choices corresponded quite well to the actual choices in the 2-AFC task. For faces, on 93.8% of the Easy trials and 65.6% of the Difficult trials, the predicted item was chosen by the participant. Both of these values differ significantly from chance (Easy:  $t(4) = 43.95$ ,  $p < 0.001$ ; Difficult:  $t(4) = 4.58$ ,  $p < 0.05$ ). Logos showed a similar pattern of results, with the predicted item being chosen on 98.8% of the Easy trials and 63.8% of the Difficult trials, again both values differing from chance (Easy:  $t(4) = 64.21$ ,  $p <$

0.001; Difficult:  $t(4) = 3.43, p < 0.05$ ). In addition, participants took longer to make Difficult decisions than Easy decisions (2387 ms vs. 1758 ms for faces; 2409 ms vs. 1748 ms for logos), though the difference was only significant for logos ( $t(4) = 3.51, p < 0.05$ ).

#### 4. Discussion

We examined the relationship between eye movements and selection during preference decisions. Consistent with previous findings (Ford, Schmitt, Schechtman, Hults, & Doherty, 1989; Glaholt & Reingold, in press, 2009; Lohse & Johnson, 1996; Pieters & Warlop, 1999; Rosen & Rosenkoetter, 1976; Russo, 1978; Russo & Doshier, 1983; Russo & Leclerc, 1994; Russo & Rosen, 1975; Shimojo, Simion, Shimojo and Scheier, 2003; Simion & Shimojo, 2006, 2007), the present results clearly demonstrate the usefulness of eye movement measurements for the study of visual decision making. In the present study, we found that during 8-AFC preference decisions, the amount of time the eye spends on a particular stimulus is positively related to the likelihood of that stimulus being selected and preferred. These results confirm and extend previous findings of a bias in looking behavior towards the item-to-be-chosen prior to the response in preference decisions (Glaholt & Reingold, in press, 2009; Shimojo, Simion, Shimojo and Scheier, 2003; Simion & Shimojo, 2006, 2007). Specifically, the selected item was substantially more likely than chance to be the item with the longest total fixation time, to be among the top two total fixation times, and to be among the top four total fixation times. In addition, the 8-AFC task revealed a clear differentiation among the items that were not chosen. Consequently, gaze behavior prior to decision can potentially provide a sensitive moment-by-moment indication of the relative strength (i.e., activation) of the items and as such may be very useful in the investigation of the time course of preference decisions in large arrays of items. For example, we speculated that there was a narrowing of the 'active' options considered by participants over the course of the trial, and we explored this by contrasting gaze bias in the beginning versus the end of the trial. Consistent with this idea, gaze bias was substantially stronger in the end than the beginning of the trial. Interestingly, and consistent with previous findings (Glaholt & Reingold, in press, 2009), gaze bias was also evident early in the trial.

From an applied perspective, our results suggest a possible extension to the ideas of Bee, Prendinger, André and Ishizuka (2006), who implemented an auto-selection mechanism by monitoring the emerging gaze bias. Specifically, in large arrays such as the ones used here,

although gaze bias is a reasonable predictor of the chosen item, it is probably not accurate enough to substitute for overt selection. However, gaze bias clearly contains information concerning the subset of items that are being actively considered by the participant at a particular time period during the trial. This suggests that gaze bias may be useful in identifying an active subset of the items competing for choice, and providing a graded measure of the preference ranking of items. Future studies are required to explore the use of eye movement recordings in real-time to aid selection among alternatives in large arrays of items such as web-pages and image arrays. Monitoring of cumulative fixation times could be applied in order to gradually and dynamically reduce the number of alternatives in a large array, potentially assisting the decision process. However, note that in pursuing such an application, one should consider the potential obstacles that could arise due to known effects of display changes on low-level visual processing (Pannasch, Dornhoefer, Unema, & Velichkovsky, 2001; Reingold & Stampe, 2002, 2004).

Furthermore, the present results indicate that fixation times are also able to convey stable preference information about stimuli and individual features. This is shown by the strong positive correlations, shown in each participant, between fixation time and selection, and between fixation time and overt preference rating. To demonstrate the power of this technique, we created new combinations of features with high and low expected preference. The results of the 2-AFC task showed that these predictions were very accurate for coarse differences between stimuli (Easy trials). Fine differentiation (Difficult trials) was less robust, but yet consistently above chance. This demonstrates that the fixation times collected 'passively' during preference decisions in large arrays can provide a particularly accurate appraisal of a person's preference along elementary feature dimensions, and this could be used to constrain a preference-search through a large set of stimuli generated from these features. We also documented large individual differences in preference. Further research is required to explore whether with large samples, it is possible to derive stable estimates of central tendency or identify clusters of individuals who share preferences.

Together with prior work (Bee, Prendinger, André and Ishizuka, 2006; Glaholt & Reingold, in press, 2009; Shimojo, Simion, Shimojo and Scheier, 2003; Simion & Shimojo, 2006, 2007), the present results convincingly demonstrate the usefulness of monitoring gaze selectivity during preference decisions. Theoretically, gaze bias may be a valuable tool for exploring the time course of preference decisions and informing the development of qualitative and quantitative models in this area. From an applied perspective, gaze bias may be exploited in applications that attempt to facilitate users' selection among items in a

complex visual display, and in devising smart applications which are able to extract information about users' preferences and customize their visual environment accordingly.

## 5. Acknowledgments

We thank Jiye Shen for essential technical assistance in these experiments. This research was funded by an NSERC research grant to Eyal M. Reingold.

## 6. References

- Bee, N., Prendinger, H., André, E., & Ishizuka, M. (2006, September). Automatic Preference Detection by Analyzing the Gaze 'Cascade Effect'. Presented at *the 2<sup>nd</sup> Annual Conference on Communication by Gaze Interaction, COGAIN 2006*, Turin, Italy.
- Duchowski, A. T. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments & Computers*, 34(4), 455-470.
- Ford, J. K., Schmitt, N., Schechtman, S. L., Hults, B. M., & Doherty, M. L. (1989). Process tracing methods: Contributions, problems, and neglected research questions. *Organizational Behavior and Human Decision Processes*, 43, 75-117.
- Glaholt, M. G., & Reingold, E. M. (in press). The time course of gaze bias in visual decision tasks. *Visual Cognition*.
- Glaholt, M. G., & Reingold, E. M. (2009). Stimulus exposure and gaze bias: a further test of the gaze cascade model. *Attention, Perception & Psychophysics*, 71, 445-450.
- Isaacowitz, D. M., Wadlinger, H. A., Goren, D., & Wilson, H. R. (2006). Selective preference in visual fixation away from negative images in old age? An eye-tracking study. *Psychology and Aging*, 21(1), 40-48.
- Lohse, G. L., & Johnson, E. J. (1996). A comparison of two process tracing methods for choice tasks. *Organizational Behavior and Human Decision Processes*, 68, 28-43.
- Murphy, N. A., & Isaacowitz, D. M. (2008). Preferences for emotional information in older and younger adults: A meta-analysis of memory and attention tasks. *Psychology and Aging*, 23(2), 263-286.

- Pannasch, S., Dornhoefer, S. M., Unema, P. J. A., & Velichkovsky, B. M. (2001). The omnipresent prolongation of visual fixations: saccades are inhibited by changes in situation and in subject's activity. *Vision Research*, 41, 3345-3351.
- Pieters, R., & Warlop, L. (1999). Visual attention during brand choice: The impact of time pressure and task motivation. *International Journal of Research in Marketing*, 16, 1-16.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 102, 21-38.
- Reingold, E. M., & Stampe, D. M. (2002). Saccadic inhibition in voluntary and reflexive saccades. *Journal of Cognitive Neuroscience*, 14(3), 371-388.
- Reingold, E. M., & Stampe, D. M. (2004). Saccadic inhibition in reading. *Journal of Experimental Psychology: Human Perception and Performance*, 30(1), 194-211.
- Rosen, L. D., & Rosenkoetter, P. (1976). An eye fixation analysis of choice and judgment with multiattribute stimuli. *Memory & Cognition*, 4(6), 747-752.
- Russo, J. E. (1978). Eye fixations can save the world: A critical evaluation and a comparison between eye fixations and other information processing methodologies. *Advances in Consumer Research*, 5(1), 561-570.
- Russo, J. E., & Doshier, B. A. (1983). Strategies for multiattribute binary choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(4), 676-696.
- Russo, J. E., & Leclerc, F. (1994). An eye-fixation analysis of choice processes for consumer nondurables. *Journal of Consumer Research*, 21, 274-290.
- Russo, J. E., & Rosen, L. D. (1975). An eye fixation analysis of multialternative choice. *Memory & Cognition*, 3(3), 267-276.
- Simion, C., & Shimojo, S. (2007). Interrupting the cascade: Orienting contributes to decision making even in the absence of visual stimulation. *Perception & Psychophysics*, 69(4), 591-595.
- Simion, C., & Shimojo, S. (2006). Early interactions between orienting, visual sampling and decision making in facial preference. *Vision Research*, 46, 3331-3335.
- Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience*, 6(12), 1317-1322.

# Scrollable Keyboards for Casual Eye Typing

Oleg Špakov\*♦ and Päivi Majaranta♦

♦University of Tampere  
(Finland)

---

## ABSTRACT

In eye typing, a full on-screen keyboard often takes a lot of space because the inaccuracy in eye tracking requires big keys. We propose “scrollable keyboards” where one or more rows are hidden to save space. Results from an experiment with 8 expert participants show that the typing speed reduced by 51.4% for a 1-row keyboard and 25.3% for a 2-row keyboard compared to a full (3-row) QWERTY. By optimizing the keyboard layout according to letter-to-letter probabilities we were able to reduce the scroll button usage, which further increased the typing speed from 7.26 wpm (QWERTY) to 8.86 wpm (optimized layout) on the 1-row keyboard, and from 11.17 wpm to 12.18 wpm on the 2-row keyboard, respectively.

---

Keywords: *Eye typing, text entry, eye tracking, gaze input.*

Paper Received 12/11/2008; received in revised form 23/03/2009; accepted 17/04/2009.

## 1. Introduction

Text entry is one of the main interaction tasks in gaze-controlled interfaces. The primary method of eye typing consists of selection of keys from an on-screen virtual keyboard (for a review of gaze-based text entry methods, see Majaranta & Rähä, 2007). The user types by pointing at each character by gaze and dwelling on it for a certain amount of time, using dwell time as an activation command. Typically, only one keystroke per character (KSPC) is needed since most letters can be directly pointed at and selected.

Having all characters visible at the same time requires space. The keys on the virtual keyboard must be big enough because of the accuracy limitations of eye tracking devices. This is true especially with “low-cost” systems that are based on off-the-shelf video or web cameras and have limited spatial resolution. Obviously, if the keyboard

---

Cite as:

Špakov, O. & Majaranta, P. (2009). Scrollable Keyboards for Casual Eye Typing. <i>PsychNology Journal</i> , 7(2), 159 – 173. Retrieved [month] [day], [year], from <a href="http://www.psychology.org">www.psychology.org</a> .
---

\* Corresponding Author:

Oleg Špakov

Department of Computer Sciences / TAUCHI, FIN-33014 University of Tampere, Finland

E-mail: [oleg.spakov@cs.uta.fi](mailto:oleg.spakov@cs.uta.fi)

occupies most of the screen estate, it significantly limits the space available for other applications.

Several attempts have been made to solve the problem of coping with the inaccuracy of the measured point of gaze and still preserving maximum screen space. Decreasing the number of keys can serve to save screen space (Miniotas, Špakov, & Evreinov, 2003). However, bigger keys are more often needed to enable the use of an eye tracker with low spatial resolution (Hansen, Hansen, & Johansen, 2001), or to enable an end-user with eye tremor or involuntary movements to point at items on screen comfortably enough (Donegan et al., 2006). Thus, in some cases, having fewer keys is a requirement for any tracking at all and would therefore not save screen space.

If only a few big keys (e.g.  $3 \times 3 = 9$  keys) can be shown at any time due to inaccuracy problems, it means the whole alphabet cannot fit on the reduced keyboard. This requires a menu structure in which reaching a certain letter can take two or more steps. Similarly to text messaging with mobile phones, several letters can be placed on one key, for example, 'abc', 'def', etc. In mobile phones with fixed (physical) keys the 'c' can be reached with three strokes on the 'abc' key. In virtual (soft) keyboards, it is possible to re-draw the keyboard and fill the cells with subsequent (groups of) keys. In this way, it is possible to type with keyboard layouts with very few keys, such as the 2x2 layout illustrated in Figure 1 (Donegan et al., 2006).

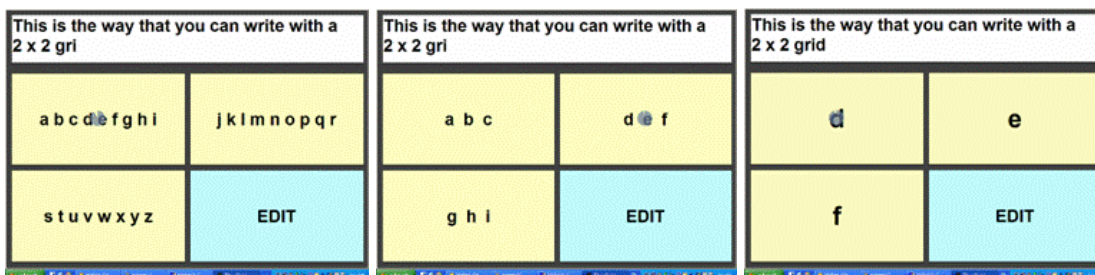


Figure 1. 3-step typing with a 2x2 keyboard layout.

If the order and layout of the keys remain constant in the sub-menus, the keyboard is easy to learn. However, typing can be very slow since several keystrokes are needed, depending on the dwell time threshold. For example, in a 3-step 2x2 layout it may take over three seconds to type one letter using three strokes with a 1-second dwell time. The typing speed can be increased by using an optimized layout in which the letters are organized according to the probabilities of the letters (Frey, White, & Hutchinson, 1990; Hansen, Hansen, & Johansen, 2001). The optimized layout reduces the number

of strokes needed to reach a certain letter; however, learning such a non-standard layout takes time and can be confusing to novice users.

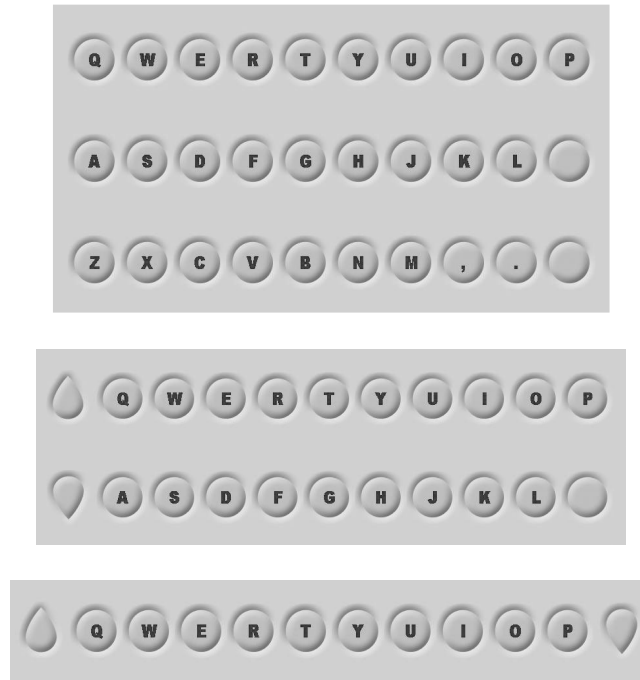
Isokoski (2000) used off-screen targets to preserve maximum screen space. To type a character, the user fixates at the off-screen targets in a certain sequence. The resulting gaze gesture is mapped to a character or command. Some recent gaze gesture systems use parts of the screen itself as active areas for the gesture recognition (Drewes & Schmidt, 2007; Porta & Turina, 2008) or show a small special area where entering of the gaze gestures happens (Bee & Andre, 2008; Wobbrock, Rubinstein, Sawyer, & Duchowski, 2008). All these systems save screen space but learning the gesture-based alphabet takes time. They also require several strokes per character (typically 2–4). In experiments, users have achieved an average typing speed of 5–8 words per minute (Porta & Turina, 2008; Wobbrock, Rubinstein, Sawyer, & Duchowski, 2008).

Miniotas, Špakov, & Evreinov (2003) developed Symbol Creator. A character is created by combining two (or more) symbols. Hence, two keystrokes produce one character (with few exceptions). The symbol parts and their combinations resemble hand written characters or their parts (for instance, 'o' and 'l' put together form 'd'), which helps in learning the symbols. Symbol Creator has eight keys in a 1-row virtual keyboard. Showing only one row of keys leaves most of the screen estate free for other purposes. Authors reported an average typing speed of 8.5 wpm in the last session.

Our goal was to develop a keyboard that saves screen space but will still be immediately usable and not require any special training. Our idea is to use a keyboard layout that is already familiar to the user (such as QWERTY) and to save screen space by only showing part of the keyboard. In the following sections, we first describe the design of the reduced keyboards, which we call scrollable keyboards. We will then report results from an experiment where the keyboard was tested. Initial results from the first experiment with the standard QWERTY layout were presented in (Špakov & Majaranta, 2008). This paper extends the research by presenting a re-design of the scrollable keyboard with an optimized layout. Full results from both experiments are reported below. We close with a discussion and conclusions.

## 2. Scrollable Keyboards

For the “full” keyboard, we used a common keyboard layout, QWERTY, shown in Figure 2. For the experiment, we decided to leave out special characters and punctuation (other than the comma and period keys). Two space keys were placed at the end of the second and the third row.



**Figure 2.** Full (3-row) keyboard, 2-row and 1-row scrollable keyboards.

The 2-row keyboard (Figure 2, middle) has only two rows of keys visible at any time. To reach the third row, the user needs to select one of the special scroll keys, “up” or “down”, on the left. The 1-row keyboard (Figure 2, bottom) only shows one row. The scroll keys are located on the sides of the keyboard. In both, the scrolling is cyclic; an invisible row can be reached using either one of the scroll buttons. The scrolling produces animated feedback, which takes 150 ms. Obviously, the KSPC measure is more than one for the scrollable keyboard, since at least one extra keystroke (scroll key) is required to reach a hidden row.

The visible distance between rows was extended because the drifting of the measured gaze position is higher in vertical direction than in horizontal direction with the tracker we used (see the method section below). Even though the visible buttons are circles, the gaze reactive area for each button is a rectangle (approximately 1.5\*3.0 degrees if the distance between the user and the monitor is 45 cm). The

buttons were selected using dwell time of 500 ms, which remained constant throughout the experiment. Animated feedback indicated the progression of the dwell time, and the key became “pressed” (shown as pressed “down” for 150 ms) when selected. The dwell progress was animated on the letter itself: the blue letter (or a progress bar for non-letter buttons) “filled up” with the red color. The end of the dwell time (button selection) was accompanied by a short “tap” sound.

### 3. Method and Procedure

Eight volunteers (aged 23–47 years, 5 male, 3 female) took part in the test. They were students or staff at the University of Tampere, and all had previously participated in other related eye typing experiments. Experienced participants were used to minimize the learning period. All were fluent in English and familiar with the QWERTY layout. Prior to the experiment, participants were informed about the experiment, participants’ rights and anonymity of the data in the experiment.

The experiment was conducted in the usability laboratory at the University of Tampere. The head-mounted EyeLink 1 eye tracking system was used to measure participants’ eye movements. The iComponent<sup>1</sup> software, which has a plug-in for EyeLink, was used to implement the experimental keyboard and to record data. The setup consisted of operator and subject monitors, adjustable chairs and tables. The chair was set so that the participant’s eyes were at approximately 45 cm from the 17-inch monitor.

For the experiment, 30 easy to memorize phrases were chosen from a set of 500 phrases by MacKenzie and Soukoreff (2003). Punctuation was removed and the phrases were case-insensitive. Participants were instructed to eye type the phrases as fast and accurately as possible, and press a key on the ordinary keyboard when they were done with each phrase. They were instructed to ignore mistakes and to carry on with a phrase when a mistake was made (the experimental keyboard did not have a backspace key).

Each session started with a short training period on the 2-row keyboard. To provide a basic level of familiarity with the experimental software, participants were given one practice phrase (about 25 characters) prior to data collection.

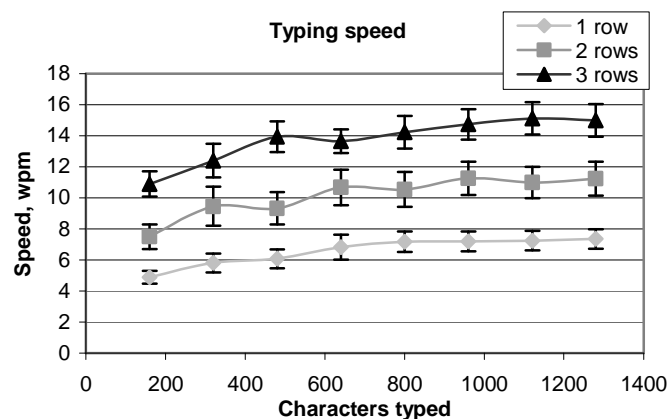
---

<sup>1</sup> iComponent is available for download at <http://www.cs.uta.fi/~oleg/downloads.html#iComp>.

The experiment used a within-subjects design with 3 conditions: 3-row (full), 2-row, and 1-row keyboard. There were 8 sessions, each including all three testing conditions (1 session per day). The order of conditions within the same session was counterbalanced between participants. Each session included 6 phrases (average length of 26.3 characters) for each condition, shown one at a time. Thus, the number of entered characters was approximately  $8 \cdot 8 \cdot 3 \cdot 6 \cdot 26.3 \approx 30300$  (1152 phrases). A session lasted approximately 10–15 minutes.

#### 4. Results

The typing speed was measured in words per minute (wpm). The typing speed results are presented in Figure 3. The increasing typing speed values during the first five sessions of each condition clearly indicate a learning process, thus we report here the average typing speed of the last three sessions. The average typing speed was 7.26 wpm (STD = 0.95), 11.17 wpm (STD = 1.43) and 14.95 wpm (STD = 1.16) for the 1-row, 2-row and 3-row keyboard, respectively. The worst (5.77, 8.93 and 12.72 wpm) and the best (8.73, 12.03 and 16.77 wpm) session speeds differed by 3–4 wpm. Maximum typing speeds registered during a single trial (typing a single phrase) were 15.3, 18.4 and 24.0 wpm, respectively.



**Figure 3.** Average typing speed in words per minute and error bars for the eight sessions.

The average error rates varied between 1–5%, with large variance between participants during the whole experiment. In the last session, the average error rates were below 2% for all conditions (see Figure 4).

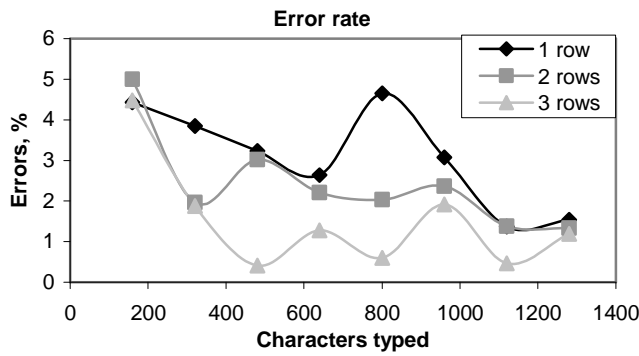


Figure 4. Error rate (%).

The selection time for the scroll buttons, letter keys and space was measured. Monitoring the usage of the scroll buttons is especially interesting because it shows how the participants learned to use the scrollable keyboards with only partially visible layout. Figure 5 shows the selection times for the 1-row (on the left) and 2-row (on the right) keyboards.

The decreasing values for the scroll buttons' selection time on both graphs during the first five sessions show the approximate amount of text required to type (~1000 characters) to learn this input technique. The average selection times of the scroll buttons in the last (8<sup>th</sup>) session were 1107 and 1268 ms for the 1-row and 2-row keyboard, respectively. These values are still higher than the letter buttons' selection times (1016 and 961 ms), especially in the case of the 2-row keyboard.

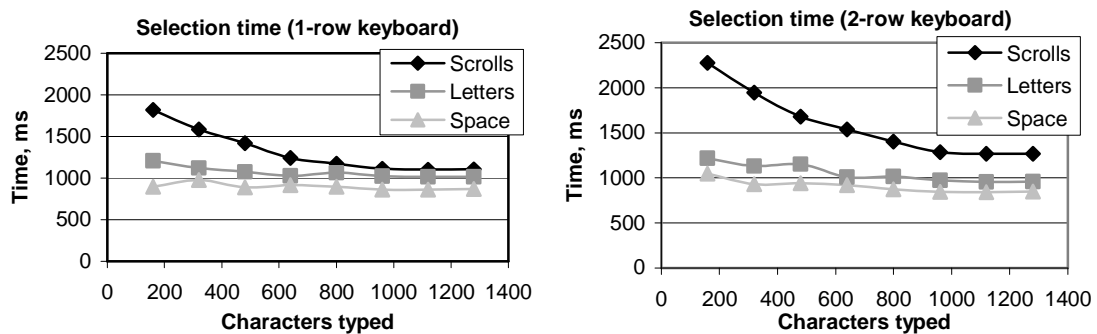


Figure 5. Selection times for the 1-row (left) and 2-row (right) scrollable keyboards.

Analysis of the scroll button usage shows that it slightly decreased in time, and the average percentage of the scroll button clicks among all clicks were 39% (1.64 KSPC) and 16.5% (1.2 KSPC) for the 1-row and 2-row keyboards, respectively. Participants used different strategies with the scrolling keyboards. Half of them memorized the location of the letters and rows so that they could choose the shortest route to the invisible row and thus minimize the scroll button usage. For example, after 'e' (located

on the top row) the user can reach 'n' (on the bottom row) by one scroll up instead of two scrolls down in the 1-row keyboard. Some participants never scrolled the layout from top line up (to the bottom) or vice versa, because they did not want to lose orientation in scrolling. In this case, more scrolling was required but the participants still did not spend time in searching for the target letter. Finally, one participant did not memorize the distribution of letters across rows but always visually scanned the rows to find the desired letter, and used only one direction of scrolling (up). This strategy resulted in the slowest typing speed. The typical difference between the fastest and slowest participant was approximately 3 wpm within each condition.

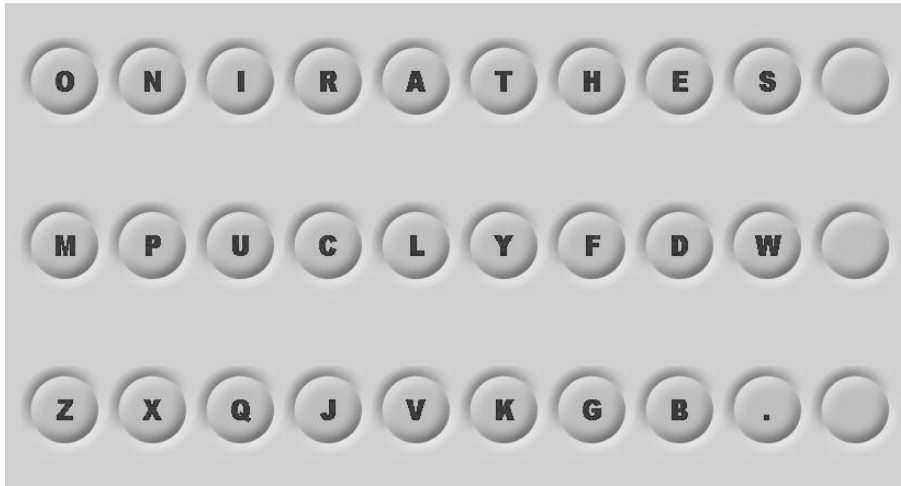
The typing speed and KSPC can be further improved using an optimized layout organized according to letter-to-letter probabilities. Even though the optimized layout requires longer learning time, it might be useful for expert users. In the following sections we present a re-design of the scrollable keyboard and results from an experiment where the efficiency of the optimized layout was tested.

## 5. Re-design – Layout Optimization

The analysis of the usage of the scrolling buttons revealed that the keyboard could benefit from optimization of the layout. The analysis is based on the calculation of each row's "weight" (or "value")  $RW$ , which is the sum of the relative frequencies of the letters in the row, and the calculation of the amount of (normalized) scrolling functions required for the input of two consecutive letters. We used the single-letter and digram (two-letter pair) frequency calculations for letter-position combinations by Mayzner and Tresselt (1965) to estimate the letter probabilities of the QWERTY layout (for more information, see Soukoreff & MacKenzie, 1995). The first row of the QWERTY layout has a row weight of  $RW_1 = 0.52$ , the second row has  $RW_2 = 0.33$ , and the third has  $RW_3 = 0.15$ . The proportion of digrams with letters on different rows is  $DG_{diff} = 61\%$ .

Our optimized layout was created based on the assumption that the usage of the scroll buttons would be reduced by grouping the most frequent letters on the same row. The most frequent letters were placed in the first row, the least frequent letters in the last row, and the space button (which is the most frequently used of all) in each row (we removed the comma button). The  $RW$  and  $DG$  values for this layout are as follows:  $RW_1 = 0.71$ ,  $RW_2 = 0.24$ ,  $RW_3 = 0.05$ ,  $DG_{diff} = 29.4\%$ . Thus, we expect that the optimized layout should reduce the usage of the scrolling function by half compared to

the QWERTY layout. Typing with the 2-row keyboard will be affected (improved) most by this optimization. The spatial distribution of the letters in the same row is based on the digram analysis: it is optimized so that the length of the gaze path is minimized within the row (keeping fixed the position of space button, which is always the right-most key). The optimized layout is shown in Figure 6.



**Figure 6.** Optimized layout.

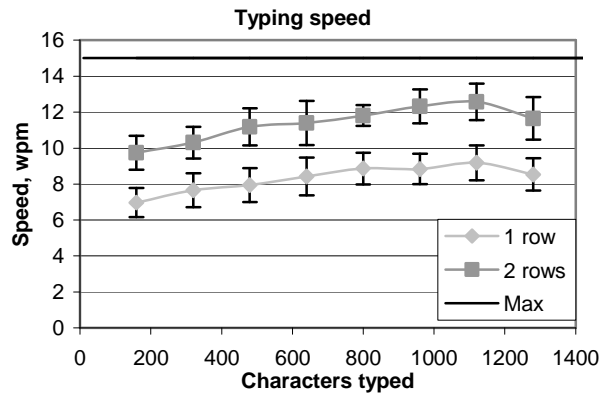
We tested the optimized layout in an experiment that followed the method and procedure of the first experiment. The only difference was that the condition with full-sized keyboard was omitted since we assumed that the typing speed would be the same after participants learn the layout. The number of entered characters was approximately  $8 \cdot 8 \cdot 2 \cdot 6 \cdot 26.3 \approx 20200$ .

## 6. Layout Optimization – Results of the Experiment

Eight participants were involved in the second experiment. Four of them were not involved in the first experiment; however, the analysis of the personal typing speed shows that there was no significant difference in the results and strategies between experienced (in this typing technique) participants and novices.

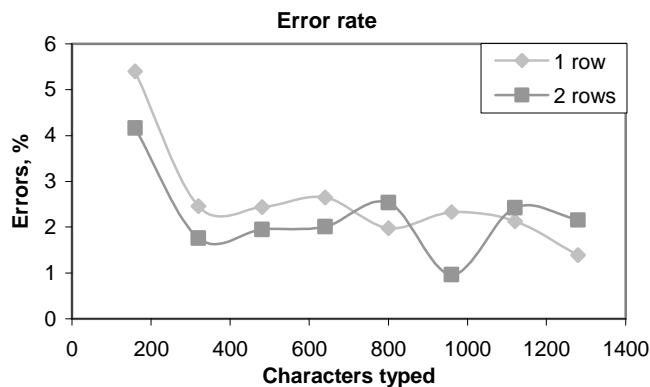
The typing speed results are presented in Figure 7. The increasing typing speed values of first five sessions of each condition clearly indicate a learning process, thus we report here the average values of the last three sessions. The average typing speed was 8.86 wpm (STD = 1.70) and 12.18 wpm (STD = 1.99) for the 1-row and 2-row keyboard, respectively. The worst (6.16 and 11.57 wpm) and the best (8.26 and

14.85 wpm) session speeds differed by 5–6 wpm. Maximum typing speeds registered during a single trial were 17.4 and 21.7 wpm, respectively. For comparison, the horizontal line (“Max”) in Figure 7 illustrates the average typing speed (14.95 wpm) on the full (3-row) keyboard with the QWERTY layout.



**Figure 7.** Typing speed (in words per minute) with the optimized layout.

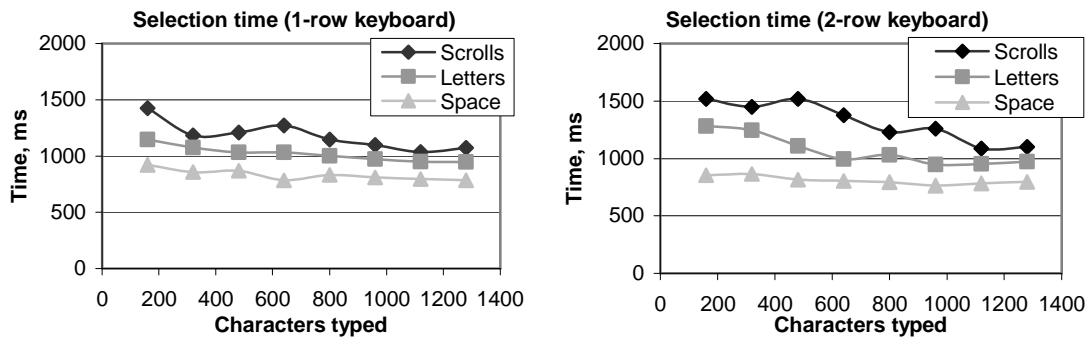
Analysis of errors shows that the users made relatively more errors only during the first session; however, the error rate decreased substantially for all other sessions, where the rate remained within approximately 2% independently of the keyboard (see Figure 8). There was a large variation in the error rate between participants. Most of them usually did not make more than 1% errors but for some, the error rate was as high as 4–7%. The main source of errors was the drifting of the calibration (causing inaccuracy in the eye tracking and thus false selections) and, in some cases, poor spelling (English was not the native language of the participants even though their language skills were considered good). Some errors occurred when the participants held their gaze too long over a (wrong) key while making the decision for the next action(s).



**Figure 8.** Error rate (%) with the optimized layout.

Typing with the optimized layout required less scroll button usage compared to the reduced QWERTY layout. The scroll button selections produced 33% (1.49 KSPC) of all clicks on the 1-row keyboard, and 10% (1.11 KSPC) on the 2-row keyboard. The usage of the scroll buttons remained approximately on the same level within all sessions.

The selection of a scroll button took 1074 ms on average, the selection of a letter button took 949 ms, and the selection of the space button took 784 ms while typing on the 1-row keyboard (see Figure 9, left). The selection times for the 2-row keyboard were 1102, 973 and 795 ms, respectively.



**Figure 9.** Selection times for the 1-row (left) and 2-row (right) scrollable keyboards with the optimized layout.

The typing strategies applied by the participants were similar to the strategies with the QWERTY layout. Again, the strategy where the participant searched all rows for the desired letter produced the slowest typing speed as well as the largest error rate.

## 7. Discussion

As expected, the optimized layout was initially harder because of the unfamiliar distributions of letters. However, the results show that the optimized layout did indeed improve the efficiency of typing by decreasing the usage of the scroll buttons: 33% versus 39% using the 1-row keyboard (reduced by 18%), and 10% versus 16.5% using 2-row keyboard (reduced by 40%). The reduction in the frequency of the scroll button usage helped to increase the typing speed from 7.26 to 8.86 wpm (increased by 22%) on the 1-row keyboard, and from 11.17 to 12.18 wpm (increased by 9%) on the 2-row keyboard.

Since every third click is produced by the selection of a scroll button in the optimized 1-row condition, the over production rate caused by the scrolling is 1.49 KSPC. When

typing using the optimized 2-row keyboard, every tenth click is produced over a scroll button, with a rate of 1.11 KSPC. These keystroke rates are quite reasonable compared to direct pointing with a fully visible keyboard with the optimum of 1 KSPC.

At the end of the experiments, the selection times were approximately the same for the letter and scroll buttons. The selection times of the space buttons were slightly shorter when typing with the keyboards with optimized layout. This was expected, since the optimized layout contains a space button on every row at the same position, therefore the users were able to find it easily. However, the selection times of the scroll buttons were always slightly longer than the selection times of other buttons in all conditions. A summary of the comparison between the two layouts (QWERTY and optimized) is presented in Table 1.

Rows	Speed wpm		Error rate %		KSPC		Selection time, ms					
	QWE	OPT	QWE	OPT	QWE	OPT	Scrolls		Letters		Spaces	
							QWE	OPT	QWE	OPT	QWE	OPT
1	7.26	8.86	1.45	1.74	1.64	1.49	1107	1074	1016	949	866	784
2	11.17	12.18	1.35	2.28	1.2	1.11	1268	1102	961	973	846	795
3	14.95		0.79		1				796		764	

**Table 1.** Comparison of the QWERTY and optimized layout.

With both keyboards, the scrolling was cyclic so that the users could scroll the keyboard around both ways. Even though this is considered efficient especially for the one-row keyboard, since the user can always select the shortest route (one scroll) to the desired key, it may be confusing for some users who want to maintain the orientation of the layout. Thus, for some users it might be useful to provide an option to prevent scrolling from the first (topmost) row to the third (bottommost) row. Furthermore, if the feedback on the scroll button reflected this constraint (e.g. by turning into a “disabled” mode), it might help the user to maintain orientation within the partly shown (partly hidden) keyboard.

Variations in experimental setup and duration of this study and the studies presented in the introduction do not allow direct comparison in performance between the proposed method and the existing eye typing methods designed to save screen space. However, the analysis of the typing speed in the last session reveals an advantage of the scrolling keyboards (9-13 wpm) over gesture-based and other reduced keyboards (5-9 wpm).

Further improvement of the scrolling keyboards might be achieved by introducing a method to enter two or more characters per selection. For example, the application

could provide a list of predicted words (or word completions) based on the text written so far (Hansen, Johansen, Hansen, Itoh, & Mashino, 2003). Another interesting direction for improvement is the implementation of keyboards with dynamic layouts. The algorithm for dynamic layout construction uses a language model and word prediction to organize the characters so that the most probable ones are always located in the visible row(s) at any given moment. However, the dynamic nature of such a layout may introduce additional cognitive and perceptual load (Koester & Levine, 1994) and reduce the ease of learning (MacKenzie & Zhang, 2008). Thus, implementation of word prediction or dynamic layout requires careful analysis and testing.

## **8. Conclusion**

We have shown that scrollable keyboards, which reduce the space taken by the full (3-row) keyboard by 1/3 or 2/3, can be efficiently used to enter text by gaze. The typing speed reduced by only 51.4% for the 1-row and 25.3% for the 2-row keyboard compared with the conventional QWERTY layout. Furthermore, the increase in the rate of keystrokes was quite reasonable, from 1 KSPC to 1.64 KSPC and 1.2 KSPC with the 1-row and 2-row keyboard, respectively.

By optimizing the keyboard layout according to the letter-to-letter probabilities we were able to reduce the frequency of the scroll button usage, which enabled a further increase in the typing speed, from 7.26 in the first experiment (with the QWERTY layout) to 8.85 wpm (with the optimized layout) on the 1-row keyboard, and from 11.17 (QWERTY) to 12.18 wpm (optimized) on the 2-row keyboard. The keystroke rates were 1.49 and 1.11 for the optimized 1-row and 2-row keyboards, respectively. The results are encouraging compared, for example, to gesture-based interfaces, which always require at least 2 KSPC.

Scrolling keyboards may be especially useful in casual typing situations, for example, filling in web forms where the overview of the full web page is important. Scrolling could also be useful in accessing the key rows that are not needed as often as letters, such as number, punctuation and function keys. Finally, the user should be able to easily adjust the number of visible rows to support the optimal layout in each situation.

## 9. Acknowledgments

We would like to thank all participants who volunteered for the study and Tatjana Evreinova for comments.

## 10. References

- Bee, N. & Andre, E. (2008). Writing with your eye: A dwell time free writing system adapted to the nature of human eye gaze. In E. André, L. Dybkjær, W. Minker, H. Neumann, R. Pieraccini, M. Weber (Eds.) *Perception in Multimodal Dialogue Systems* (pp. 111-122). Berlin: Springer.
- Donegan, M., Oosthuizen, L., Bates, R., Istance, H., Holmqvist, E., Lundälv, M., Buchholz, M., & Signorile, I. (2006). *D3.3 Report of user trials and usability studies*. Communication by Gaze Interaction, COGAIN. Retrieved on-line August, 10, 2009 from: <http://www.cogain.org/results/reports/COGAIN-D3.3.pdf>
- Drewes, H., & Schmidt, A. (2007). Interacting with the Computer Using Gaze Gestures. In *Proceedings of INTERACT 2007* (pp. 475-488). Berlin: Springer.
- Frey, L.A., White, K. P. JR., & Hutchinson, T. E. (1990). Eye-Gaze Word Processing. *IEEE Transactions on Systems, Man, and Cybernetics*, 20(4), 944-950.
- Hansen J.P., Hansen D.W., & Johansen A.S. (2001). Bringing Gaze-based Interaction Back to Basics. In C. Stephanidis (Ed.), *Universal Access in HCI* (pp. 325-328). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hansen, J.P., Johansen, A.S., Hansen, D.W., Itoh, K., & Mashino, S. (2003, April). Language technology in a predictive, restricted on-screen keyboard with ambiguous layout for severely disabled people. Presented at the *Workshop on Language Modeling for Text Entry Methods, EAACL 2003*, Budapest, Hungary.
- Isokoski, P. (2000). Text Input Methods for Eye Trackers Using Off-Screen targets. In *Proceedings of the Eye Tracking Research & Applications Symposium, ETRA 2000* (pp. 15-21). New York: ACM Press.
- Koester, H.H. & Levine, S.P. (1994). Modeling the speed of text entry with a word prediction interface. *IEEE Transactions on Rehabilitation Engineering*, 2(3), 177-187.

- MacKenzie, I.S., & Soukoreff, R.W. (2003). Phrase sets for evaluating text entry techniques. In *Proceedings of CHI 2003: Extended Abstracts on Human Factors in Computing Systems* (pp. 754-755). New York: ACM Press.
- MacKenzie, I.S. & Zhang, X. (2008). Eye typing using word and letter prediction and a fixation algorithm. In *Proceedings of the Symposium on Eye Tracking Research & Applications, ETRA 2008* (pp. 55-58). New York: ACM Press.
- Majaranta, P., & Rähkä, K.-J. (2007). Text Entry by Gaze: Utilizing Eye-Tracking. In I.S. MacKenzie & K. Tanaka-Ishii (Eds.), *Text entry systems: Mobility, accessibility, universality* (pp. 175-187). San Francisco: Morgan Kaufmann.
- Mayzner, M. S., & Tresselt, M. E. (1965). Table of single-letter and digram frequency counts for various word-length and letter-position combinations. *Psychonomic Monograph Supplements* 1, 13-32.
- Miniotas, D., Špakov, O., & Evreinov, G.E. (2003). Symbol Creator: An Alternative Eye-based Text Entry Technique with Low Demand for Screen Space. In M. Rauterberg, M. Menozzi, & J. Wesson (Eds.). *Proceedings of INTERACT 2003* (pp. 137-143). Amsterdam: IOS Press.
- Porta, M., & Turina, M. (2008). Eye-S: a full-screen input modality for pure eye-based communication. In *Proceedings of the Eye Tracking Research & Applications Symposium, ETRA 2008* (pp. 27-34). New York: ACM Press.
- Soukoreff, R. W. & MacKenzie, I. S. (1995). Theoretical upper and lower bounds on typing speed using a stylus and soft keyboard. *Behaviour & Information Technology*, 14, 370-379.
- Špakov, O. & Majaranta, P. (2008) Scrollable Keyboards for Eye Typing. In *Proceedings of the 4<sup>th</sup> Annual Conference on Communication by Gaze Interaction, COGAIN 2008* (pp. 63-66). Prague: Czech Technical University Publishing House.
- Wobbrock, J. O., Rubinstein, J., Sawyer, M. W., & Duchowski, A. T. (2008). Longitudinal evaluation of discrete consecutive gaze gestures for text entry. In *Proceedings of the Eye Tracking Research & Applications Symposium, ETRA 2008* (pp. 11-18). New York: ACM Press.



# Hands Free Interaction with Virtual Information in a Real Environment: Eye Gaze as an Interaction Tool in an Augmented Reality System

Susanna Nilsson<sup>\*♦</sup>, Torbjörn Gustafsson<sup>\*</sup> and Per Carleberg<sup>\*</sup>

<sup>♦</sup>Linköping University  
(Sweden)

<sup>\*</sup>XMReality, Linköping  
(Sweden)

---

## ABSTRACT

Eye contact in human conversations is a natural source of information about the visual attention of people talking, and also indicates who is speaking to whom. Eye gaze can be used as an interaction method, but gaze tracking can also be used to monitor a user's eye movements and visual interest. This paper describes how gaze-based interaction can be used and implemented in an Augmented Reality (AR) system. The results of the preliminary tests of the gaze-controlled AR system show that the system does work, but that it needs considerable development and further user studies before it can be a realistic option in real end user settings.

---

Keywords: *Augmented Reality, gaze-controlled augmented reality, mixed reality, gaze-based interaction*

Paper Received 14/11/2008; received in revised form 04/04/2009; accepted 28/04/2009.

## 1. Introduction

Using eye gaze as input to technical systems is not a new concept. However, to a large extent, the methods of interacting with systems through eye gaze or eye gestures have been too strenuous, complicated or expensive for widespread use in the domain of human-computer interaction. Instead, eye gaze research has mainly developed as a tool for people with various limitations that make it difficult or impossible to interact with computers through traditional means such as keyboards, mouse or voice control.

---

Cite as:

Nilsson, S., Gustafsson, T., & Carleberg, P. (2009). Hands Free Interaction with Virtual Information in a Real Environment: Eye Gaze as an Interaction Tool in an Augmented Reality System. *PsychNology Journal*, 7(2), 175 – 196. Retrieved [month] [day], [year], from [www.psychology.org](http://www.psychology.org).

\* Corresponding Author:

Susanna Nilsson  
Linköping University, SE-58183 Linköping, Sweden  
e-mail: susni@ida.liu.se

Eye contact in human conversations is a natural source of information about the visual attention of people talking, and also indicates who is speaking to whom.

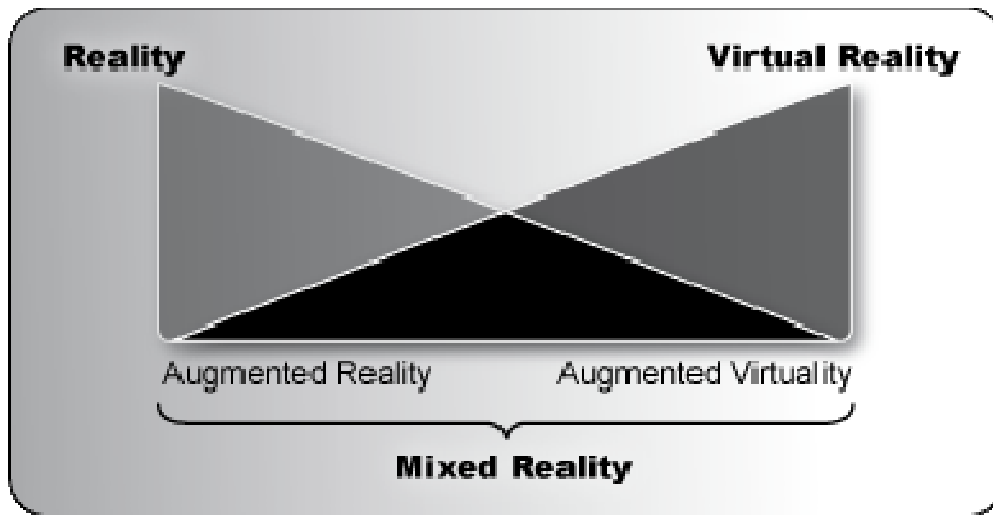
This source of information can also be of use when communicating with a technological system. Eye gaze tracking can be used as part of an interaction method, but gaze tracking can also be used to monitor a user's eye movements and visual interest. This allows the system to recognize the possible intentions of the user quickly and thereby improve the user interface.

In Augmented Reality (AR) systems, real and virtual objects are merged and aligned in relation to a real environment, and presented in the field of view of a user. AR applications that give hierarchical instructions to users often require some feedback or acknowledgement from the user in order to move to the next step in the instructions. It should be possible to give this feedback quickly and without interrupting the ongoing task. Many different types of interaction techniques have been used in the domain of AR; there are numerous examples of systems that use manual input, gestures and/or speech interfaces (Billingham, Kato, & Poupayev, 2001; Gandy et al., 2005; Henrysson, Ollila, & Billingham, 2007). However, there are situations where speech and gesture may not be appropriate. For instance, during surgical procedures in an operating room, the surgeon may have difficulties manually interacting with technical devices because of the need to keep the hands sterile. Voice interaction with a system may also not be appropriate due to surrounding noise or filtering problems. There is one modality that can overcome the issues of noisy environments, keeping hands sterile and the need to work with both hands while at the same time trying to interact with a computer or an AR system, and that is the visual modality. The aim of this paper is to present an AR system with an integrated gaze tracker, allowing quick feedback from the user to the system, as well as an analysis of the user's gaze behaviour.

## **2. Mixed and Augmented Reality**

Mixed Reality is a field of research as well as the collective name for the techniques aiming at combining and fusing virtual elements into the perceived real world. The term "mixed reality" aims to capture the conceptual ideas behind the technologies used – the blending, merging or mixing of different realities. Milgram and Kishino (1994) described the field of Mixed Reality (MR) as a continuum in which the different types of systems and applications can be placed (see Figure 1). In the far left of the continuum, systems

that aim to add only minimal virtual elements to the otherwise un-manipulated perceived real world can be found. These systems are known as Augmented Reality systems, where the aim is not to create an immersive virtual world (as opposed to Virtual Reality (VR) systems to the far right in the continuum), but rather augment the real world with added information or experiences. This paper focuses on an Augmented Reality (AR) system.



**Figure 1.** The Mixed Reality Continuum (after Milgram and Kishino, 1994).

Even though it may be an interesting question, this paper will not discuss the notion of "reality" or go into a philosophical debate about what is real and what is not. For the purpose of this text, "reality" is simply what humans perceive as their surrounding in their normal life. Given this definition of reality, "merging realities" could simply mean merging two images of the world together. Azuma (1997), however, mentions three criteria that have to be fulfilled for a system to be classified as an Augmented Reality system: It combines the real and the virtual, it is supposed to be interactive in real time (meaning that the user can interact with the system and get a response from it without delay), and it is registered and aligned in three dimensions. AR applications can be found in diverse domains, such as medicine, military applications, entertainment, technical support and industry applications, distance operation and geographic applications.

Designing usable and user-friendly interfaces is crucial when developing new systems. One area of research that has a close connection to usability and the end user community is research on interaction modalities. The user of an AR system can either be a passive recipient of information or an active part of an interaction process.

AR systems that use gestures and speech for interaction have been developed (Billinghurst, Kato, & Poupyrev, 2001; Gandy et al., 2005; Henrysson, Ollila, & Billinghurst, 2007). However, there are examples of situations where speech and gesture may not be appropriate. For instance, during surgical procedures in an operating room, the surgeon may have difficulties interacting with technical devices since both hands are usually occupied. Voice interaction with a system may not be appropriate due to surrounding noise or filtering problems. Another example can be found in a modern-day auto repair workshop, where a mechanic is doing repair work on an engine. Having greasy hands or being occupied with tools in both hands certainly would make interacting with a manual or technical support system difficult. There is one modality that can overcome both the issues of noisy environments and the need to work with both hands while at the same time trying to interact with a computer or an AR system, and that is the visual modality.

Many advances in direct manipulation, gaze-based interaction, pattern matching and recognition, gesture and natural language interaction have been made during the past decades. Apart from being used as an interaction method, gaze can also be used to monitor a user's eye movements and visual interest in usability studies and design evaluation. Gaze tracking allows the system to recognize the possible intentions of the user quickly, and this has been used to improve the interaction in multimodal user interfaces (Qvarfordt, 2004). Several of these results have also been implemented in the AR research area, but there are few examples of gaze-based interaction in this domain (Gustafsson & Carleberg, 2003; Gustafsson et al., 2005a; Sareika, 2005).

### **3. Technology Used in Augmented Reality Systems**

Technically, there are two different solutions for merging reality and virtuality in real time today: video see-through and optical see-through. At first glance, the latter of these is the preferable solution, but it has some technical and practical difficulties (Azuma, 1997, 2001; Kiyokawa, 2007). For one, the virtual projection cannot completely obscure the real world image – the see-through display does not have the ability to block off incoming light to an extent that would allow for a non-transparent virtual object. This means that real objects will shine through the virtual objects, making them difficult to see clearly. The problem can be solved in theory, but the result is a system with a complex configuration. There are also some issues with the placement of

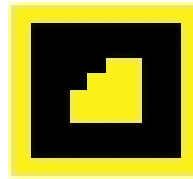
virtual images in relation to the surroundings in optical see-through displays. Since the virtual objects presented to the user are semi-transparent, they give no depth clues to the user. Instead, the virtual objects seem to lie in the same focal plane, whereas in natural vision, objects are perceived in different focal planes (Gustafsson et al., 2004; Haller, Billinghamurst & Thomas, 2007).

A way to overcome the problems with optical see-through is by using a camera in front of the user's eyes, and then projecting the camera image onto a small display in front of the user's eyes (video see-through) (Azuma, 1997; Gustafsson et al., 2004; Kiyokawa, 2007). The virtual images are added to the real image before it is projected, which solves the optic see-through problem of surrounding light, as well as gives control over where the virtual objects are placed. This method, however, has other problems, such as the limited field of view, lack of peripheral displays and the slight offset caused by the fact that the camera's position can never be exactly where the eyes are located. This gives the user a somewhat distorted experience, since the visual viewpoint is perceived to be where the camera is (Azuma, 1997). The difference between the bodily perceived movement and the visual movement as seen through the display can have an effect on the user's experience of the system, in some cases even causing motion sickness (Stanney, 1995). Despite these problems, there are important advantages with a video see-through solution. One has already been pointed out – the ability to occlude real objects – and another is that the application designer has complete control over the presented image in real time since it is run through the computer before it is presented to the user. In the optical see-through design, only the user will see the final augmented image. To conclude: there is a trade-off between the optic see-through systems and the camera based video see-through systems, and the available resources often determine the choice of solution.

Regardless of what display solution has been chosen for an AR application, the most important issue to solve is how and where to place the virtually generated image. In order to place the virtual information correctly, the AR system needs to know where the user and user viewpoint is. This means that the system has to use some kind of tracking or registration of the surrounding environment. There are different techniques to do this, and several of them can be combined to ensure more reliable tracking of the environment (Haller, Billinghamurst & Thomas, 2007). Tracking is normally done by using different sensors to register the surrounding environment. This sensor information is then used as a basis for placing the virtual information (Azuma et al., 2001). When using a video see-through technique, the AR system is already equipped with a visual

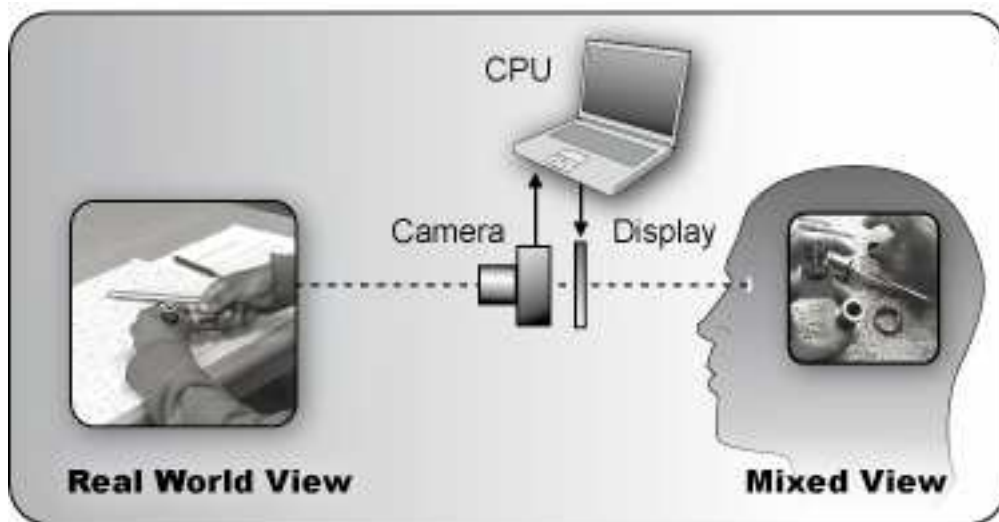
sensor – the camera – which allows vision based tracking. Although the development of marker-less techniques may dominate the future, the marker tracking technique has been one of the most commonly used techniques; it makes use of visual markers that can be recognized by feature tracking software (Kato & Billinghurst, 1999). Figure 2 below shows an example of a marker that can be used for this purpose.

The marker tracking technique used in the studies presented in this chapter is based on ARToolkit, which is an open software library for building AR applications (Kato & Billinghurst, 1999).



**Figure 2.** An example of a fiducial marker for vision based tracking.

By tracking and identifying markers placed in the environment, the algorithm calculates the position of the camera in relation to the markers, and hence the virtual information can be placed in the display relative to the marker position. Figure 3 illustrates how a marker (seen on the user's finger) is used to place the virtual 3D object in the user's field of view.



**Figure 3.** A schematic view of an AR system using marker tracking. Note the marker on the user's index finger.

One problem that can occur when using AR systems with marker based tracking is that when two markers are in the user's field of view, information is displayed at different locations at the same time. If the user only needs to see the information on one marker at a time, the system should be able to recognise this and switch off the

“unnecessary” information until it becomes significant for the user to see it. (In the case of the image above, the information related to the top marker is not relevant for the current task). This problem could perhaps be solved in a variety of ways, through hard coding, tracking technology and different input and interaction modalities. An alternative is to solve the problem by adding some sense of “user awareness” to the system, i.e. letting the AR system know what area and information the user is interested in at the moment. One possible way to do this is by making the AR system aware of the user’s visual interest with the aid of gaze tracking devices.

Adding user awareness to the system is an even more interesting approach in the promising emerging technique called markerless or feature based tracking. Markerless tracking will create opportunities to build more flexible systems where the user can view the world freely and still see virtual elements without being constrained to predefined areas in the world (Klein & Murray, 2007).

#### **4. Gaze-based Interaction**

For over twenty years, eye gaze interaction has been used effectively for eye typing, i.e. interacting with word processors through eye gaze instead of the traditional keyboard and mouse input (Majaranta & Rähkä, 2002). For a gaze-based interaction system to be useful, the gaze detection process should be implemented in a way that does not interfere with the user’s behaviour (Ohno & Mukawa, 2004). A gaze tracker should be able to work in different display configurations and in varying lighting conditions and, therefore, a gaze tracker is often a complex system. Many technologies are involved in a gaze tracker: displays, optic devices, image sensors and real-time signal processing. Each technology, with its characteristics, can affect the use of the tracker. For example, the gaze tracker’s accuracy and especially the system’s time delay are important in real-time interaction applications.

##### **4.1 Eye Gaze and Usability**

Usability testing methods presently involve a lot of trial and error when developing new interfaces. By using eye tracking, much time can be saved by directly giving answers to questions of layout, control functionality and visibility of different objects (Karn, Ellis & Juliano, 1999). Eye gaze tracking has been used to evaluate system usability as well as in psychological studies of human visual behaviour for a long time

(see for example Hyrskykari, 1997; Oviatt & Cohen, 2000; Maglio & Campbell, 2003; Qvarfordt, 2004).

A number of studies have shown that eye gaze patterns differ between expert and novice users of a system or method (Law, Atkins, Kirkpatrick, & Lomax, 2004). This implies that it is possible to use gaze patterns to see how a user interacts with a system, and the changes in gaze patterns may also reflect how the user's skills improve during training. This would make it possible to use changes in gaze patterns to evaluate the effectiveness of a training method or the design of an interface. Gaze recognition and logging can be a useful method for evaluating and developing AR applications for training and technical maintenance.

As noted previously, gaze behaviour can provide a system with implicit input, and this way of tracking the intentions of a user has been used successfully in multi-modal applications (Qvarfordt, 2004; Maglio & Campbell, 2003). In these studies, eye gaze direction was used to solve ambiguity problems in, for example, speech interaction: By registering on which visual area of the interface the user focuses, speech input can be interpreted faster and with more accuracy. In the example of a map, the visual area put some constraints on the possible input, which in this case would be names on the map. This is a way of adding an environmental context, the area of visual interest, to improve speech recognition (Shell, Selker & Vertegaal, 2003).

#### **4.2 Problems with Gaze-controlled Interaction**

Several studies discuss the potential problems of eye-gaze-based interaction (Vertegaal, 2002; Ohno, 1998; Zhai, Morimoto & Ihde, 1999; Zhai, 2003). One example of these problems is the "Midas touch" problem, which occurs when the user looks at an object without meaning to choose it but activates it anyway just by the first glance (Jakob, 1991; Hyrskykari, 1997; Ohno, 1998). Another problem occurs when using dwell time to determine whether a user has "clicked" on an object or not, since it is hard to define what the exact time should be before the object is activated (Majaranta, Aula, & R ih a, 2004). If the time chosen is too short, the "Midas touch" problem occurs; if the time is too long, the user may become annoyed. In both cases, the user will probably consider the interface to be quite user-unfriendly and will be reluctant to use it again (Ohno, 1998; Jakob, 1991). If dwell time is used as a means of activating objects on the screen or in typing with eye gaze, it is important that the feedback to the user is sharp and unambiguous (Majaranta, Aula, & R ih a, 2004).

An alternative way to interact visually with the system is to use eye gestures such as winks, but this is very difficult since the system has to be able to differentiate between

voluntary eye gestures and natural, involuntary eye gestures (Ohno, 1998). An additional important limitation of using gaze movements to interact with a system is that gaze will probably indicate visual interest, but not necessarily the cognitive interest of the user. It is one thing to determine if the user has had the information in the visual field of attention but a completely different issue to determine whether that information actually has been acknowledged and understood (Vertegaal, 2002; Bates & Istance, 2002). This could possibly be solved to some extent by requesting feedback from the user.

## **5. A Gaze-controlled Augmented Reality System**

As stated above, gaze control as a method for both usability studies as well as interaction is not a new research domain. However, gaze-controlled interaction is new to the Mixed Reality domain. The main focus of implementing gaze control in a head-mounted mixed reality display is to make the interaction between the system and the user easier and more efficient. Integrating gaze detection into the AR system could be a way to predict the user's intention and to anticipate what actions will be requested of the system. The use of gaze control in the AR system will also be useful for developing the system's interaction methods and designing the displays. The user's eye movements can be used to interpret what the problem might be when the user does not handle the interface in the way it was intended (Karn, Ellis & Juliano, 1999). By analyzing the gaze patterns, one might be able to see if the user has observed objects or if the gaze is distracted by other objects.

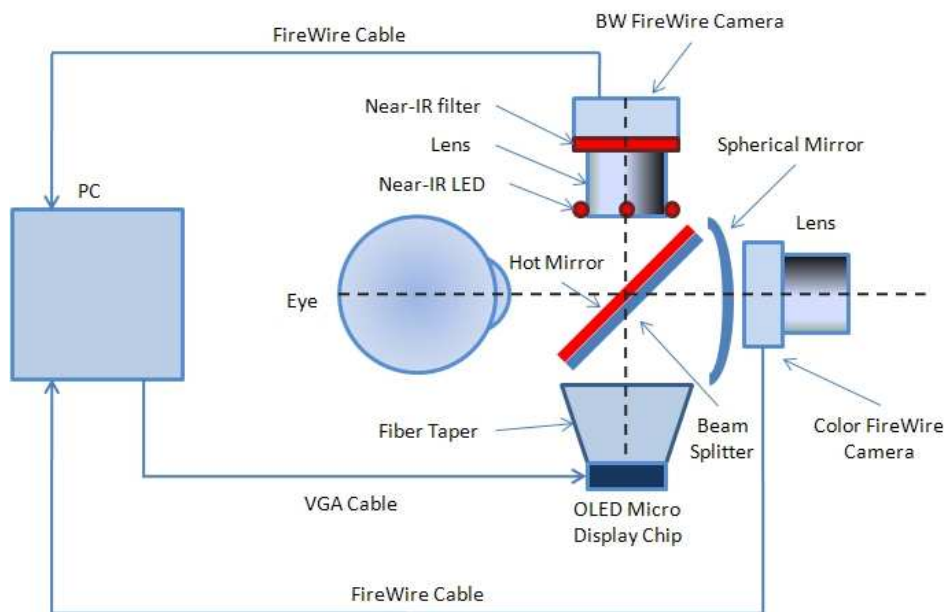
In interactive AR applications that require responses from the user, there must be an efficient and non-interruptive way to deliver responses to the AR system. Gaze control in the AR system could make interaction between the system and the user easier and more efficient.

For an AR system using gaze-based interaction to be useful, the gaze detection process should be implemented in a way that does not interfere with the user's normal behavior (Oviatt & Cohen, 2000). A gaze tracker for AR applications should be able to integrate with micro-displays and must function in varying conditions of illumination. The following sections present a helmet-mounted AR system with an integrated gaze tracker, which can be used both to monitor the user's gaze behaviour as well as for interaction. The system has been tested to reach stage 2 in the testing process

described by Sommerville (2004). The first stage is component testing, and the third stage of the process is the acceptance testing stage, where a more elaborate usability study would take place.

### 5.1 HMD and Integrated Gaze Tracker

We have developed a head-mounted video see-through AR system with an integrated gaze tracker (see Figures 4 and 5). The integrated head-mounted display, black/white gaze tracker camera (640 × 480 pixels resolution) and scene camera is an in-house construction, and the different components used are illustrated in Figure 4.



**Figure 4.** A schematic view of the gaze-controlled AR system.

The micro-displays have a resolution of 800 × 600 pixels and a field of view of 37 × 28 degrees. The gaze tracker camera and the micro-display are integrated and have co-aligned optic axes to facilitate future studies of vergence-movement-controlled systems (Gustafsson, Carleberg, Sohlberg, Persson, & Aldrin, 2005b).

The system utilizes the dark pupil tracking principle, which means that the NIR (near infrared) illumination is placed off-axis in relation to the gaze tracker (BW) camera (see Figure 5). Gaze tracking techniques generally use an illumination source, which may be placed either on- or off-axis with the gaze tracker camera. When the source is on-axis, the camera sees the pupil illuminated similar to the red-eye effect in flash photography. Thus on-axis tracking is also referred to as bright-pupil tracking. Similarly, off-axis tracking is referred to as dark-pupil tracking (Duchowski, 2007; Weigle & Banks, 2008; Young & Sheena 1975). The NIR illumination source is fully integrated

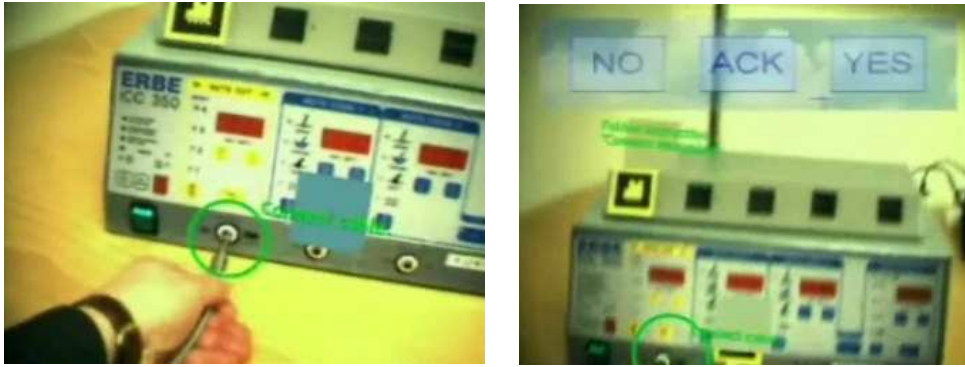
into the system and is not a separate device. The gaze tracker camera detects the pupil and its reflections by filtering and thresholding the image information. The position of the pupil and the positions of the reflections on the cornea caused by the NIR illumination are calculated. These positions are used as input parameters to a rather straightforward geometric algorithm that computes the gaze direction. Four illumination sources are used; however, only one reflection is needed to calculate gaze. The system can choose between the four different reflections, which increases the robustness of the system. Gaze is calculated using information on display geometry, camera placement and lens characteristics as well as initial assumptions on eye geometry. Interactive tuning/calibration is necessary in this prototype system, but this is normally a fast procedure (taking a few seconds) and could be automated.

## **5.2 Augmented Reality Software**

The AR system described here uses hybrid tracking technology, basically a marker detection system based on ARToolKit (HITLAB webpage), ARToolKit Plus (Schmalstieg, 2005) and ARTag (Fiala, 2005) but with the addition of a 3DOF (3-degrees-of-freedom) inertial tracker (isense.com and xsens.com).

The software allows applications to be built and defined in a scenario file in XML syntax. For gaze controlled interaction, the application designer can define the layout of the gaze control dialog areas as well as gaze action specifications. With this tool, the application developer can experiment with, compare and verify the functionality of different gaze-controlled interaction schemes.

In the system that was developed, eye gaze interaction can be restricted both temporally and spatially – only certain parts of the display will have a function, and only when there is a need for gaze interaction. These interaction areas are defined in the application scenario XML file, which is also where eye gaze dwell times and command actions are configured. The gaze dialog area positions can either be fixed or dynamic, i.e. relative to a detected marker position, which allows flexible design of the application. The areas in which the gaze interaction is active are represented by transparent images, as can be seen in Figures 5 and 6. Transparency, color, appearance and placement of the interaction areas are defined in XML syntax in the scenario file.



**Figure 5.** Gaze interaction areas can be concealed until the user needs them, thereby avoiding unnecessary clutter in the user's field of view.

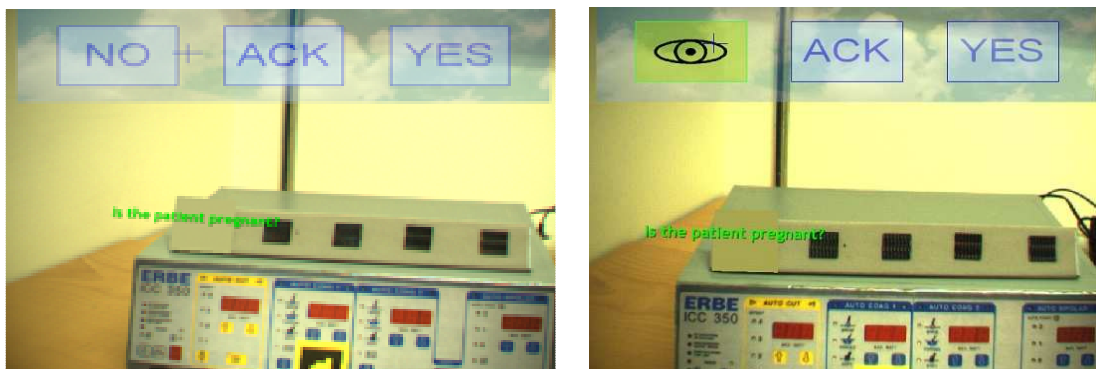
The interactive area in the display can, as noted, be defined in XML syntax. The implementation allows several different ways of interacting with the system through gaze control. As an example, in one implementation, the central and peripheral areas of the display are defined as active and non-active interaction areas, respectively. This means that when the user's gaze is in the central part of the display, no interaction areas are displayed, as seen in the left-hand image in Figure 5. When the user wants to interact with the system, she/he gazes at the peripheral part of the display, which activates the interaction area in the display, as seen in the right image in the same figure. The software solution is described in detail in Gustafsson, Carleberg, Sohlberg, Persson and Aldrin (2005b).

## 6. Preliminary Test of the System

The gaze-controlled AR system was tested in a laboratory setting using two different instructional applications. Both applications were designed and evaluated previously, but without gaze control as an interaction device (see for instance Nilsson & Johansson, 2006; Nilsson & Johansson, 2008). Both applications were developed in cooperation with the target user group, which was medical staff at a hospital, and the task for the user was to complete a set of instructions. Three different settings of the gaze interaction dialogue were tested: a static interaction area in the upper part of the display, a static interaction area in the lower part of the display and, finally, a dynamic interaction area in which the interaction area appears relative to the placement of a marker. Limiting the interaction area to the upper part of the user's field of view solves part of the "Midas touch" problem, since users prefer not to use this part of the display in natural interaction. For example, Mon-Williams, Plooy, Burgess-Limerick, and Wann,

(1998) discuss problems with the visual angle of displays in VR). However, since looking at this area requires an uncomfortable gaze angle, users may not appreciate this variant, even though it solves the Midas touch problem. Placing the interaction dialogue in the lower part of the display instead is more gaze-friendly, but the problem of accidentally activating the gaze interaction dialogue is more prominent. To address this issue, the area that activates the dialogue was reduced in size to only two smaller boxes, one in each corner, as can be seen in Figure 7. When the user looks at one of the areas, this is indicated by a change in appearance (colour and image), so that the user receives feedback, acknowledging that the system “knows” that s/he is looking at the area.

The dynamic layout of the interaction dialog allows the application designer to link the gaze interaction directly to a marker, for instance, thus allowing interaction in the central field of view of the user when necessary. In the applications described, the point of gaze fixation is only visible when it is positioned in the interaction area in the upper/lower part of the display. However, it is also possible to show the gaze fixation point at all times if desirable. Interaction feedback is given to the user by changing the image and colour of the button (see Figure 6). This change indicates that the “press” has been acknowledged by the AR system.



**Figure 6.** An example of the gaze interaction dialog. The user can respond to the question asked by looking at the different regions/“buttons” in the upper part of the field of view.

### 6.1 The First Application Example

This task has previously been used in a user study investigating the usability of AR instructions in the medical domain (Nilsson & Johansson, 2006, 2008). The main goal of the instructions is to activate an electro-surgical generator (ESG) and prepare it for operation. An ESG, often referred to as an “electrical knife”, is used for deep heating of tissue and is employed during surgery to minimise the amount of bleeding in the patient.

The instructions were given as statements and questions that had to be confirmed or denied via the input device, in this case the gaze interaction dialog, where the user can choose to look at “yes”, “no” and “ack” (short for “acknowledged”) (see Figure 6). The dwell time used in the test series was set to 1 second.

## 6.2 The Second Application

The task in the second application example has also been used previously in user studies of AR instructions in the medical domain (Nilsson & Johansson, 2007, 2008). In this application, the goal was to follow instructions on how to put a trocar together. A trocar is used as a “gateway” into a patient during minimal invasive surgeries. The trocar is relatively small and consists of seven separate parts, which have to be assembled correctly for it to function properly as a lock preventing blood and gas from leaking out of the patient’s body. The trocar was too small to have several different markers attached to each part. Markers attached to the object (such as the ones in study 1) would also not be realistic considering the type of object and its usage – it needs to be kept sterile and clean of other materials. Instead, the marker was mounted on a small ring with adjustable size, which the participants wore on their index finger (see Figure 7).



**Figure 7.** The gaze-controlled AR system in the second application. The images to the right show what the user sees. The “no” represents repeating an instruction (going back in the sequence), and the “yes” means going to the next step. Note that, in this setting, the interactive area is placed in the lower part of the display rather than the upper part as in the examples above.

The AR system used in this application was previously equipped with speech recognition and was controlled by verbal interaction from the user. The application,

which was originally designed for voice interaction, was redesigned and used in the gaze-controlled AR system. In contrast to the other application, in this task the users did not have to answer any questions but only needed to follow the instructions given, and the only input to the system was to acknowledge when an instruction was completed. After seeing the visual instruction and completing it, the user used gaze to let the system know if he/she wanted to repeat the instruction or go to the next step (see Figure 7).

The gaze-controlled AR application was tested in a laboratory setting and was not a full user study, but rather a functionality test. Trials that were performed with the system indicated that the system functions as intended.

### **6.3 Results of the Preliminary Tests**

During the trials, it was found that the users tended to turn and tilt their head so that the focus of attention was always in the centre of the field of view. Looking at things in the upper section of the display was therefore a conscious effort and probably not part of casual gazing. Users reported that it was strenuous to interact with this gaze angle and preferred the lower static interaction dialogue. There were, however, some problematic issues with the lower interaction dialogue as well: When the users tried to activate the dialogue, they tended to tilt their head, thus often losing camera contact with the marker, which in turn led to the loss of the visual instructions. In conclusion, the static dialogues are not ideal since they are not adapted to normal human behaviour: When humans want to focus their gaze on an object, virtual or real, they tend to place it in the central field of view. This is of course not possible when the instruction dialogue is static.

The dynamic dialogue does not have the problem of the users tilting and moving their heads. However, one important problem was found. Although the interaction dialogue was only visible when the system required an input from the user, it still sometimes covered too much of the user's field of view. This caused the participants to experience it as being cluttered. This could be addressed with a redesign of the interaction dialogue in future development of the system.

In general, the participants reacted positively to the concept of gaze control in these types of applications, but the system was experienced as clumsy and not entirely stable, since they sometimes lost the virtual information when the marker was not detected by the camera. The latter problem can be addressed by further refining the AR/MR system and the software. The clumsiness of the system is harder to address in the technical solution presented here. Video-see-through AR and gaze control requires

cameras (two at a minimum), and these are too heavy to be comfortably placed on a regular head-mounted setup – the helmet helps to balance the weight of the system. For gaze-controlled video see-through AR, a helmet-mounted solution is currently the best option.

## **7. Discussion and Future Research**

As mentioned previously, this paper presents a system test, which means that the focus of the study was to test the functionality of the gaze-controlled application, as opposed to a more elaborate end user study, where more aspects than the technical functionality are of interest. Nevertheless, the test resulted in some preliminary insights into the user aspects of the system. The gaze interaction in the applications tested was basically the same as with ordinary keyboard buttons. The experience so far is that gaze controlled interaction is as fast and distinct as pressing an ordinary keyboard button. This is in accordance with earlier research and the results of Ware and Mikaelian (1986), who showed that gaze interaction may even be faster because the time it takes to shift the position of the cursor manually causes a delay. Nevertheless, there are still many areas in which the system can be improved. The main factor is the size of the system – the cameras and displays still require a helmet-mounted set-up in order for calibration and weight distribution to work properly. The system would also require some more extensive user studies to ensure the effectiveness of the gaze control, and also to evaluate the system not only in technical terms but also more importantly in terms of end user satisfaction.

The use of dwell time as an interaction method was a choice made during the development process; to decrease the risk of the Midas touch phenomenon, the system also gives the user visual indicators of the gaze response. That is, when the user has dwelled on a gaze button area for about a half second, the area changes its appearance to make the user aware of the interaction activity. This is one way to give clear feedback to the user. The dwell times used to determine selection in the interaction have not yet been fully investigated but rely on previous studies in the domain of eye gaze interaction.

There may also be alternatives to using dwell time (as in the described system) and eye gestures such as winks, gaze gestures (Drewes & Schmidt, 2007) and “eye graffiti” (Porta & Turina, 2008) that have not been investigated for use in this application as of

today. For instance, eye gestures other than winks could be used. Using eye gaze patterns that can easily be identified as voluntary input instead of dwell time or winks could improve the usability. An example of such voluntary eye gestures could be moving the gaze quickly in one direction and then back again. This is similar to the gestures used with other pointing devices, such as the pointer gestures used for typing on small screens such as those of mobile phones.

Gaze interaction allows the user to work freely with her/his hands while stepping through an instruction hierarchy. This freedom of movement is of value in many situations, not only in the application described above but also in other applications involving maintenance and repair tasks. The conclusion therefore is that the system can be used as an alternative to traditional manual interaction. This is of particular interest for applications where manual input or speech input is not appropriate or possible. The experiences from the limited test runs are important for the further development of the system and have clearly indicated that the system is functional. Future tests, including a larger user group, will investigate the robustness of the system as well as give more insight into the speed and accuracy in other applications than the hierarchical instructions used here.

The gaze recognition in the system is not restricted to use for direct interaction but can also be used for indirect communication with the AR system. Gaze recognition can add a “user awareness” dimension to the system, which can monitor the user’s visual interest and act upon this. Gaze awareness can also allow the system to infer (via the user’s gaze direction) when and if the user wants to interact with the system. If the user has two or more markers in the field of view, the gaze direction can be used to indicate which marker’s virtual information should be displayed and is thus a relatively easy way to de-clutter the user’s field of view.

Combining the concepts from AR with gaze recognition and input enables quick and easy interaction in an AR system that allows for natural human communication, such as communicating intention by the use of eye gaze. Cognitive interest may not always be the same as visual interest, but, in many cases, visual interest can be an indicator of what the user is focusing on both visually and cognitively and can therefore allow the system to respond to this, for instance by presenting requested virtual information.

The proposed AR system with added gaze awareness and gaze-controlled interaction will be further tested in a user application. Another aspect of the gaze-aware AR system is its ability to log and analyze the gaze patterns of AR users, possibly allowing further usability studies and evaluations of the AR system. Gaze control in the AR-

system may also be a useful tool for developing the system's interaction methods and the designing the displays.

## 8. Acknowledgments

This research was a collaboration project between the department of Computer and Information Science (IDA) at Linköping University and the Swedish Defence Research Agency, funded by the Swedish Defence Materiel Administration (FMV).

## 9. References

- Azuma, R. (1997). A survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments*, 6(4), 355 – 385.
- Azuma, R., Bailot, Y., Behringer, R., Feiner, S., Simon, J. & MacIntyre, B. (2001). Recent Advances in Augmented Reality. *IEEE Computer Graphics and Applications*, 21(6), 34-47.
- Bates, R. & Istance, H. (2002). Zooming interfaces! Enhancing the performance of eye controlled pointing devices. In *Proceedings of the Fifth International ACM Conference on Assistive Technologies* (pp. 119-126). New York: ACM Press.
- Billinghurst, M., Kato, H. & Poupyrev, I. (2001). The MagicBook: Moving Seamlessly between Reality and Virtuality. *IEEE Computer Graphics and applications*, 21(3), 2-4.
- Drewes, H. & Schmidt, A. (2007, September). Interacting with the Computer using Gaze Gestures. Presented at *Interact 2007*, Rio De Janeiro, Brazil.
- Duchowski, A. T. (2007). *Eye Tracking Methodology: Theory and practice*. London: Springer.
- Fiala, M. (2005, October). ARTAG Rev2 Fiducial Marker System: Vision based Tracking for AR. Presented at the *Workshop of Industrial Augmented Reality*, Vienna, Austria.
- Gandy, M., MacIntyre, P., Presti, P., Dpw, J. Bolter, B. Yarbrough, & O'Rear, N. (2005). AR Karaoke: Acting in your favourite scenes. In *Proceedings of the fourth IEEE and ACM International conference on Mixed and Augmented Reality, ISMAR* (pp. 114-117). California: IEEE Computer Society.

- Gustafsson, T. & Carleberg, C. (2003). *Mixed Reality for Technical Support*. Technical report FOI--0857--SE, Swedish Defence Research Agency.
- Gustafsson, T., Carleberg, P., Nilsson, S., Svensson, P., Sivertun, Å. & LeDuc, M. (2004). *Mixed Reality for technical support*. Technical report ISSN 1650-1942, Swedish Defence Research Agency.
- Gustafsson, T., Carleberg, P., Svensson, P., Nilsson, S., & Sivertun, Å. (2005a). *Mixed Reality Systems for Technical Maintenance and Gaze-controlled interaction*. Technical report ISSN 1650-1942, Swedish Defence Research Agency.
- Gustafsson, T., Carleberg, P., Sohlberg, P., Persson, B. & Aldrin, C. (2005b). *Interactive Method of Presenting Information in an Image*. No WO/2005/124429, patent granted 2005-05-10.
- Haller, M., Billinghamurst, M., & Thomas, B. (Eds) (2007). *Emerging Technologies of Augmented Reality: Interfaces and Design*. London, UK: Idea Group Publishing.
- Henrysson, A., Ollila, M. & Billinghamurst, M. (2007). Mobile Phone Based Augmented Reality. In Haller, M., Billinghamurst, M. & Thomas, B. (Eds.), *Emerging technologies of Augmented Reality. Interfaces and design* (pp. 90-109). London: Idea Group Publishing.
- HITLAB, <http://www.hitl.washington.edu/artoolkit/> (as of 2008-06-10).
- Hyrskykari, A., Majaranta, P. & R  ih  , K.J. (2005, July). From Gaze Control to Attentive Interfaces. Presented at the *3rd International Conference on Universal Access in Human-Computer Interaction*, Las Vegas, NV, USA.
- Hyrskykari, A. (1997). *Gaze Control as an Input Device*. Report B-1997-4. University of Tampere, Department of Computer Science. Retrieved on-line August, 15, 2009 from: <http://www.cs.uta.fi/hci/gaze/publications.php>
- Jakob, R.K. (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(3), 152-169.
- Karn, K.S., Ellis, S. & Juliano, C. (1999). The Hunt for Usability: Tracking Eye Movements. In *Proceedings of CHI'99* (pp. 173-173). New York: ACM Press.
- Kato, H. & Billinghamurst, M. (1999, October). Marker Tracking and HMD Calibration for a video-based Augmented Reality Conferencing System. Presented at the *2nd International Workshop on Augmented Reality, IWAR 99*, San Francisco, USA.
- Kiyokawa, K. (2007). An Introduction to Head Mounted Displays for Augmented Reality. In Haller, M., Billinghamurst, M. & Thomas, B. (Eds.). *Emerging Technologies*

- of Augmented Reality: Interfaces and Design* (pp. 43-63). Idea Group Publishing, London, UK.
- Klein, G. & Murray D. (2007). Parallel Tracking and Mapping for Small AR Workspaces. In *Proceedings of the sixth IEEE and ACM International conference on Mixed and Augmented Reality* (pp. 225--234). Washington DC: IEEE Computer Society.
- Law, B., Atkins, M. S., Kirkpatrick, A. E. & Lomax, A. J. (2004). Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment. In *Proceedings of the Eye Tracking Research and Application Symposium* (pp. 41-48). New York: ACM Press.
- Maglio, P. & Campbell, C. (2003). Attentive Agents. *Communications of the ACM*, 46(3), 47-51.
- Majaranta, P. & R ih a, K-J. (2002). Twenty years of eye typing. In *Proceedings of Eye Tracking Research and Applications, ETRA'02* (pp. 15-22). New York: ACM Press.
- Majaranta, P., Aula, A. & R ih a, K-J. (2004). Effects of Feedback on Eye Typing with Short Dwell Time. In *Proceedings of Eye Tracking Research and Applications, ETRA'04* (pp.139-146). New York: ACM Press.
- Milgram, P. & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions on Information Systems, E77-D(12)*, 1321-1329.
- Mon-Williams, M., Plooy, A., Burgess-Limerick, R & Wann, J. (1998). Gaze angle: A possible mechanism of visual stress in virtual reality headsets. *Ergonomics*, 48(3), 280-285.
- Nilsson, S. & Johansson, B.J.E. (2006). A cognitive Systems Engineering Perspective on the Design of Mixed Reality Systems. In *Proceedings of the 13th European Conference on Cognitive Ergonomics* (pp. 154-161). New York: ACM
- Nilsson, S. & Johansson, B. (2007). Fun and usable: Augmented reality instructions in a hospital setting. In *Proceedings of the 19th Australasian Conference on Computer-Human interaction: Entertaining User interfaces* (pp. 123-130). New York: ACM Press.
- Nilsson, S. & Johansson, B.J.E. (2008). Acceptance of Augmented Reality Instructions in a Real Work Setting. In *Proceedings of Conference on Human Factors in Computing Systems* (pp 2025-2032). New York: ACM Press.
- Ohno, T. (1998). Features of Eye Gaze Interface for Selection Tasks. In *Proceedings of the 3rd Asia Pacific Computer-Human Interaction, APCHI'98* (pp.176-182). Washington, DC: IEEE Computer Society.

- Ohno, T. & Mukawa, N. (2004). A Free-head, Simple Calibration, Gaze Tracking System That Enables Gaze-Based Interaction. In *Proceedings of Eye Tracking Research and Applications, ETRA'04* (pp. 115 – 122). New York: ACM Press.
- Oviatt, S. & Cohen, P. (2000). Multimodal Interfaces that Process What Comes Naturally. *Communications of the ACM*, 43(39), 45-53.
- Porta, M. & Turing, M. (2008, March). Eye-S: a Full-Screen Input Modality for Pure Eye-based Communication. Presented at *the 2008 Symposium on Eye Tracking Research and Applications*. Savannah, GA, USA.
- Qvarfordt, P. (2004). *Eyes on Multimodal Interaction*. Linköping Studies in Science and Technology, Dissertation no. 893. Linköping: Unitryck.
- Sareika, M. (2005). *Einsatz von Eye-Tracking zur Interaktion in Mixed Reality Umgebungen*. Diplomarbeit. Fraunhofer Institut Angewandte Informationstechnik.
- Schmalstieg, D. (2005, October). Rapid Prototyping of Augmented Reality Applications with the STUDIERSTUBE Framework. Presented at the *Workshop of Industrial Augmented Reality*, Vienna, Austria.
- Shell, J. S., Selker, T., & Vertegaal, R. (2003). Interacting with groups of computers. *Communications of the ACM*, 46(3),40-46.
- Sommerville, I. (2004). *Software Engineering*. Edinburgh Gate: Pearson Education.
- Stanney, K. (1995, March). Realizing the full potential of virtual reality: human factors issues that could stand in the way. Presented at the *Virtual Reality Annual International Symposium, VRAIS'95*, Research Triangle Park, NC, USA.
- XSens: <http://www.xsens.com/> (as of 2008-06-10)
- Vertegaal, R. (2002). Designing Attentive Interfaces. In *Proceedings of the 2002 Symposium on Eye Tracking Research and Applications, ETRA'02* (pp. 23-30). New York: ACM Press.
- Ware, C. & Mikaelian, H.H. (1986). An evaluation of an eye tracker as a device for computer input. In *Proceedings of the SIGCHI/GI conference on Human Factors in computing systems and graphic interfaces* (pp. 183-188). New York: ACM Press.
- Weigle, C. & Banks, D.C. (2008, January). Analysis of eye-tracking experiments performed on a Tobii T60. Presented at the *Conference on Visualization and Data Analysis*, San José, California, USA
- Young, L., R. & Sheena, D. (1975) Survey of eye movement recording methods. *Behavior Research methods & Instrumentation*, 7(5), 397-429.
- Zhai, S. (2003). What's in the eyes for attentive input? *Communications of the ACM*, 46(3), 34 – 39.

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

# Gaze beats mouse: A case study on a gaze-controlled breakout

Michael Dorr<sup>\*♦</sup>, Laura Pomarjanschi<sup>♦</sup>, and Erhardt Barth<sup>♦</sup>

<sup>♦</sup> Institute for Neuro- and Bioinformatics  
University of Lübeck  
(Germany)

---

## ABSTRACT

We present an open-source, gaze-controlled adaptation of the well-known Breakout computer game. In a tournament where 20 subjects took turns playing this game against each other, one using gaze and one using a mouse, we demonstrate that gaze can be a superior input modality. In another experiment, we collected eye movement data from 9 subjects playing this game and find that expert and novice players differ in their employed eye movement strategies.

---

Keywords: *Gaming with gaze, human-computer interaction, alternative input devices*

Paper Received 14/11/2008; received in revised form 24/02/2009; accepted 07/05/2009.

## 1. Introduction

Eye tracking has become cheaper and more robust over the last years because of the rapid progress in digital camera technology and the steady advance in computing power (Li & Parkhurst, 2006). With the advent of remote trackers and miniaturized eye cameras that can be integrated into comfortably worn glasses, it may soon become technologically feasible to deploy gaze-based systems in the mass market. However, there is still a lack of a “killer application” and existing, keyboard- or mouse-controlled applications often cannot easily be adapted to use gaze information (Jacob, 1993). One area in which an average consumer might benefit from eye tracking is in computer games, where gaze direction can add another dimension of input (Smith & Graham,

---

Cite as:

Dorr, M., Pomarjanschi, L., & Barth, E. (2009). Gaze beats mouse: A case study on a gaze-controlled Breakout. <i>PsychNology Journal</i> , 7(2), 197 – 211. Retrieved [month] [day], [year], from <a href="http://www.psychology.org">www.psychology.org</a> .
--

\* Corresponding Author:

Michael Dorr  
Institute for Neuro- and Bioinformatics, University of Lübeck  
Ratzeburger Allee 160, 23538 Lübeck, Germany  
dorr@inb.uni-luebeck.de

2006; Isokoski & Martin, 2006; Dorr, Böhme, Martinetz, & Barth, 2007) Progress in this direction will also be highly relevant to disabled users who lack the dexterity to control the input modalities traditionally used in computer games. Not only could gaming with gaze be enjoyable in itself, but the virtual world of multi-player games might also be one arena where disabled users could meet non-disabled users on an equal footing (Istance, Bates, Hyrskykari, & Vicker, 2008).

However, for a satisfactory gaming experience, it does not suffice to simply replace the mouse with a gaze cursor; usually, changes to the game play will also have to be made. In this paper, we first present an open-source game adapted for control by either a mouse or gaze direction. We then show results from a small tournament that indicate that gaze is an equal if not superior input modality to a mouse for this game. Finally, we will analyze the eye movements made by players and investigate how eye movement strategies differ between novices and experienced players.

## **2. Modification of an Open-source Game**

### **2.1 Breakout**

Breakout was one of the first commercially available video games when it was released in 1976 (Kent, 2001). Its game play was based on Pong, where the player has to move a paddle horizontally to hit a ball that is reflected at the borders of the game area. Breakout now extended this concept by putting bricks in the upper part of the game area which dissolved upon contact with the ball; the goal of the game was no longer to keep the ball in the game as long as possible, but to destroy all bricks (see Figure 1 for a screen shot).

This simple, easy-to-understand game play makes Breakout still appealing today, more than 30 years after it was first sold. Countless clones have been published for various computer platforms (Wikipedia lists 56 “notable” clones alone), with better graphics and game-play-varying extra items, which are released upon the explosion of bricks and either need to be collected with the paddle (“good” extras, e.g. bonus points or an increase in paddle size) or should be avoided (“bad” extras, e.g. freezing the paddle for a short period or a speedup of the ball). The one-dimensional nature of paddle control in Breakout and Pong also makes these games suitable for input modalities that lack the degrees of freedom of keyboard, joystick, or mouse, or use noisier channels, e.g. brain-computer interfaces (Krepki, Blankertz, Curio, & Müller, 2007) or pitch of voice (the Sony SingStar console game). In the following, we will describe our version of Breakout, which was adapted for gaze control.



**Figure 1.** Screen shot of LBreakout2. The paddle at the bottom right can be moved horizontally to prevent the ball from hitting the bottom; the bricks in the upper part of the screen are destroyed on impact of the ball and might release extra items that modify the game.

## 2.2 Implementation

Our gaze-controlled version of Breakout is based on the open-source game LBreakout2 (Speck, n.d.). LBreakout2 is published under the GNU General Public License<sup>1</sup> (GPL), so that the game can be freely modified under the condition that the modifications will only be released under the GPL as well. This open-source approach is especially appropriate for such (currently) small markets as eye movement researchers and games geared towards those with severe motor impairments. Therefore, the source code of our modifications and a binary are available on our web site<sup>2</sup>; we would like to receive feedback and/or incorporate changes made by the community. LBreakout2 is written in C and uses the Simple Media Layer<sup>3</sup> for graphics, sound, and network functionality. We have modified it to work with SensoMotoric Instruments eye trackers, which use an ASCII network protocol sent over a UDP link, requiring no additional libraries. The major change to the source code was to implement a function that waits on a UDP socket for samples from the eye tracker and decodes them; instead of the paddle position being shifted by mouse movements, it is now set in absolute coordinates to the gaze position of the user. Because we do not have access to other eye-tracking equipment, we did not implement an interface to other trackers; however, this modification should be straightforward.

<sup>1</sup> See <http://www.fsf.org/licenses/licenses/gpl.html>.

<sup>2</sup> See <http://www.inb.uni-luebeck.de/tools-demos>.

<sup>3</sup> See <http://www.libsdl.org>.

The part of the program that receives and parses gaze samples consists of a mere 120 lines of code and is quite simple. Implementing the calibration protocol, however, is slightly more complicated because data has to be sent back and forth between the client and the eye tracker. Therefore, a first version of the game used an external tool to calibrate the tracker to the screen before the game was started. Especially for demonstration purposes, where several players take turns, the constant need to shut down and restart the game rendered this external calibration impractical and led us to implement an in-game calibration procedure, initiated by key press from within the game. Furthermore, calibration procedures often use black markers on a light grey background; the background of the game, however, is of a very dark green. This reduction in screen brightness leads to an increase in pupil size compared to pupil size during calibration, which in turn reduces the accuracy of the eye tracking and for some subjects even makes tracking impossible (because their eyelids occlude a part of their pupils). Therefore, we calibrated on a screen of similar brightness as the game's background and increased ambient illumination in the room where experiments took place.

### **2.3 Adaptation of Game Play**

To prevent the ball from going out of play, the paddle needs to be at the same horizontal position as the ball when the ball reaches the lower end of the screen. When the paddle is controlled by gaze, this means that, in principle, the player only needs to look at the position where the ball will meet the paddle. That this is very intuitive might be demonstrated by the following anecdote: During the CeBit trade fair show, we presented our game to a visitor who claimed to have had no experience with computer games at all. After calibration, she started playing and performed very well until, about 2 minutes into the game, she asked when “the whole thing would actually start”. Apparently, she had just constantly looked at the ball (and therefore always hit it with the paddle) without even realizing that the paddle followed her gaze!

Although playing with gaze is very intuitive, players naturally face other challenges in a gaze-controlled setting. A well-known problem for gaze-based user interfaces is how a user should confirm an action (the equivalent of a mouse click). In LBreakout2, a mouse click normally is needed to start the game and release the ball from the paddle. We solved this problem by releasing the ball automatically after 5 seconds when the game is played with gaze.

Another problem is that even the best eye trackers today cannot yet track gaze accurately when the subject makes abrupt head movements or blinks. There are also always calibration errors and, in video-based systems, some camera noise, so that the paddle position might be slightly shifted from the “true” gaze position. This can be highly irritating and must be consciously compensated for by the player, even though there are no fixation targets at the location that needs to be fixated.

Also, by carefully adjusting on which side of the paddle the ball is deflected, the player can control the direction in which the ball is sent off again, which is especially important when only a few isolated bricks remain on the screen and a failure to hit that one remaining target can be very frustrating. Due to tracking noise, this is much harder with gaze than with a mouse, but it seems that gaze players get better at this with some training (also see below).

In the Breakout version we have adapted, bricks that are destroyed sometimes release “extras” that fall towards the bottom of the screen. Once they are collected with the paddle, they alter the game by, for example, increasing the speed of the ball or making the ball explosive (so that several bricks can be destroyed at once). Some of these extras require a reaction by clicking the mouse (e.g. a “gun” that fires brick-destroying rounds), so we removed them from the game using the integrated level editor. Other extras should not be collected by the player because they have a negative impact; to carefully avoid looking at something in a dynamic environment requires a conscious effort and some training on the part of the player – that quick glance to check whether the bad extra has already disappeared could prove disastrous! One extra that is particularly enjoyable in gaze-playing mode, though, is the extra ball. Because of the much higher speed at which the eye can travel compared to the hand, it is possible to keep several balls in play simultaneously. Keeping track of a number of dynamic objects while still maintaining gaze on the ball that is going to reach the bottom of the screen next was a task that our subjects found highly entertaining.

### **3. Pitting Gaze against Mouse**

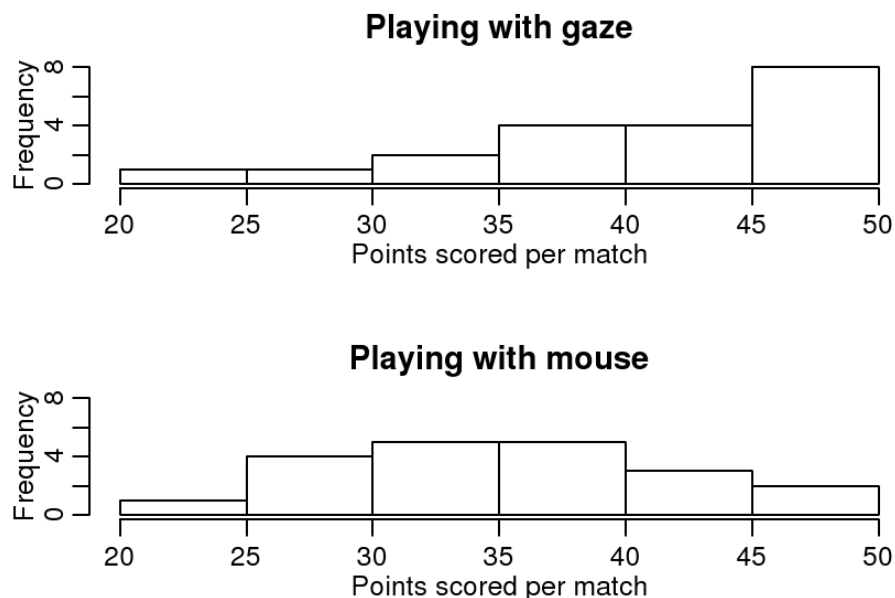
LBreakout2 also offers a multi-player mode with one paddle at the bottom and one at the top of the screen. The goal is to play the balls in such a way that the opponent cannot return them. To make the game livelier, each player can fire up to 3 balls so that up to 6 balls are in the game simultaneously.

To test how well our gaze-based interface fared against the mouse, we set up a little tournament in which pairs of players took turns playing against each other. First, one

player controlled the game with gaze and the other with the mouse, subsequently the roles were reversed.

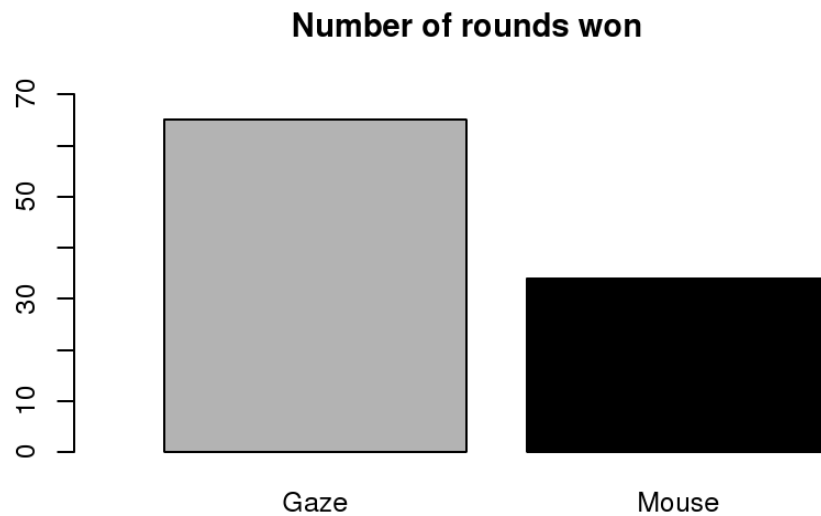
Twenty undergraduate and graduate students from our department volunteered. Four had been previously involved in writing or presenting the game; the other 16 had had little or no eye-tracking experience and had not played the game before. To ensure a fair game, we also matched pairs by their general computer game experience.

Eye movements were recorded with a SensoMotoric Instruments iViewX Hi-Speed tracker running at 240 Hz. After a 5-point calibration, the gaze player was able to try out gaze control for about 15 to 30 seconds (which also served to validate the accuracy of the calibration). Then, the match started. Each match lasted 5 rounds. For every ball that the opponent could not return with their paddle, a player scored 1 point; a round was won by the first player to win 10 points. Each round was set on a different background, i.e. the layout of ball-deflecting bricks in between the players changed. Such bricks close to the player's baseline are a slight disadvantage for the gaze player because it is easier to aim shots exactly with the mouse (see above).



**Figure 2.** Results for the tournament where players using the gaze-controlled game competed against players with the mouse. Gaze outperformed the mouse as an input modality.

The results are shown in Figures 2 and 3. Clearly, playing with gaze yielded a higher score on average (41.95 vs. 36.25). Almost two thirds of all rounds (65 out of 99; one data set had to be discarded because the tracker had lost the pupil temporarily) were won by the gaze player. Gaze control was thus a statistically significant advantage ( $p < 0.0015$ ).



**Figure 3.** Results for the gaze vs. mouse tournament in terms of number of rounds won. 65 out of 99 rounds were won by the player using their gaze; gaze control thus was at a statistically significant advantage ( $p < 0.0015$ ).

## 4. Analysis of Eye Movement Strategies

### 4.1 Methods

Despite the game play's intuitiveness, an obvious question is whether different players might employ different strategies to control the paddle with their eyes and how such strategies might evolve with training. To answer this question, we collected data from 9 subjects; 5 were novices, i.e. had never played a gaze-controlled game before, the other 4 were experts and had at least several hours of gaming experience (two of them were authors, but one of them was naive with regard to the exact analyses to be conducted on the recorded data).

Eye movements were recorded from one eye with a SensoMotoric Instruments iViewX Hi-Speed tracker running at 500 Hz (the tracker is capable of sampling at 1250 Hz, but such high temporal precision was not required for the present study). This tracker requires the head to be fixed using a chinrest. We also have successfully played the game with a 50 Hz SMI RED-X remote tracker, which is obviously better suited for gaming and entertainment because it allows free head movement. However, the accuracy of the remote system is still considerably lower than that of the head-fixed tracker. This was particularly important for the gaze vs. mouse tournament, where no input modality should have an unfair advantage.

Subjects were seated 55 cm away from a screen of 40 cm width and 30 cm height, so that the game screen covered a visual angle of 40 by 30 degrees; at 640 pixels

horizontal resolution, 16 pixels thus corresponded to 1 degree. After a short briefing on the game, subjects were seated at the tracker, the game was started, and an in-game 9-point calibration was performed. The subjects' task was to "Try to collect as many points as possible"; if a subject had lost all their "lives", the game nevertheless continued (but the score was reset to zero). The trial ended when all bricks were destroyed or there were only very few remaining bricks left; because of the difficulty in precisely placing the paddle with gaze, accurately aiming at isolated bricks can be tedious to impossible (note that this premature termination only took place in this experiment where eye movement strategies were investigated; in the gaze vs. mouse tournament, no such unfair advantage was granted to the gaze players). It took subjects between 5 and 7 minutes to finish the level.

The level used for this experiment consisted of 16 rows with 14 bricks each, out of which a randomly drawn 22 (10%) contained extra items. To avoid any bias in extra item selection, the type of extra contained in these bricks was chosen randomly at the beginning of each trial.

It is quite difficult to keep the head still for more than 5 minutes, especially for subjects unfamiliar with eye-tracking experiments; for this reason, our data contained a certain amount of impulse noise (where gaze position briefly changes greatly). We therefore computed the sample-to-sample velocities and discarded the 2% highest-velocity samples, which showed biologically implausible speeds of up to more than 1000 deg/s.

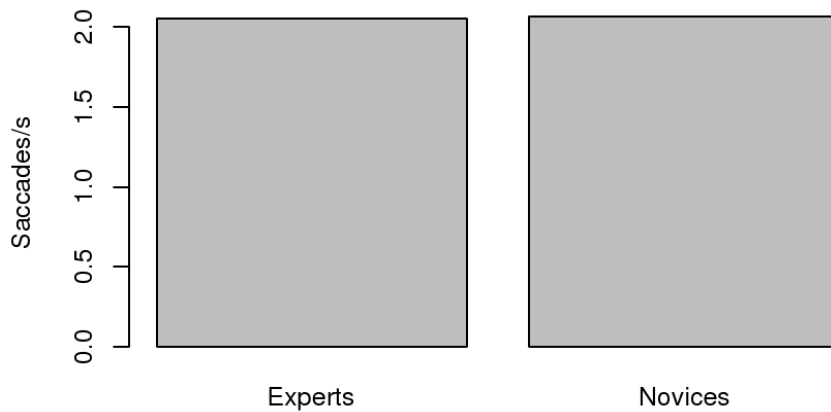
#### **4.2 Number of Saccades**

As can be seen in Figure 4, there is no clear difference in the number of saccadic eye movements made by experts and novices.

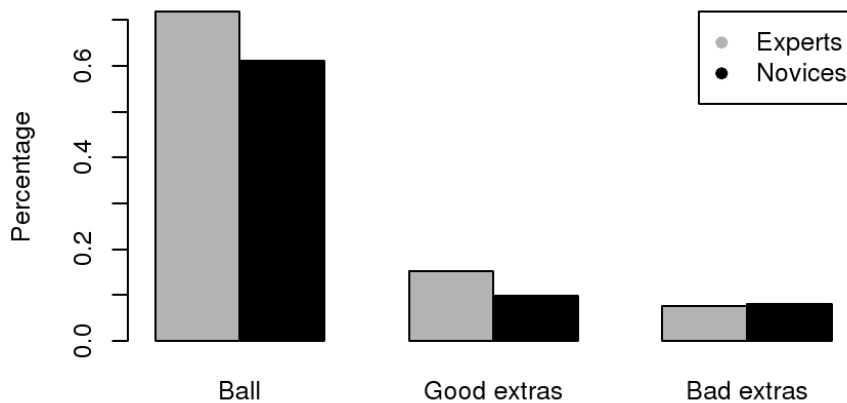
#### **4.3 Focus of Fixations**

We analysed how often experts and novices looked at the ball, which is the most relevant in-game item, and at good and bad extras. For each gaze sample, we determined the in-game item with the smallest Euclidean distance to gaze position on the screen; if the point of regard was more than 80 pixels (5 degrees) away from any ball or item, we classified this sample as "no identifiable focus". The results are shown in Figure 5: clearly, experts spend more time looking at the ball than novices (71.8% vs. 61%); they also spend a lot more time looking at good extras (15.3% vs. 9.8%), but there is no big difference for the bad extras (7.6% vs. 8%). Note, however, that the

latter percentages for the extra items are relative only to the time in which any good (or bad, respectively) items were shown on the screen (18.9% and 15.3% good extras, 10.2% and 9.2% bad extras), so that both experts and novices still spend considerable time “looking” at no item in particular (24.5% for experts, 36.7% for novices). The result for the good extras could be explained by the higher cognitive demands placed on novices compared to the experts; novices might employ a narrower attentional focus and therefore ignore good extras more often. This is at odds, however, with the result for ball-following, which seems to indicate that novices are also more easily distracted from the ball.



**Figure 4.** Saccade rates for experts and novices. There is no significant difference.

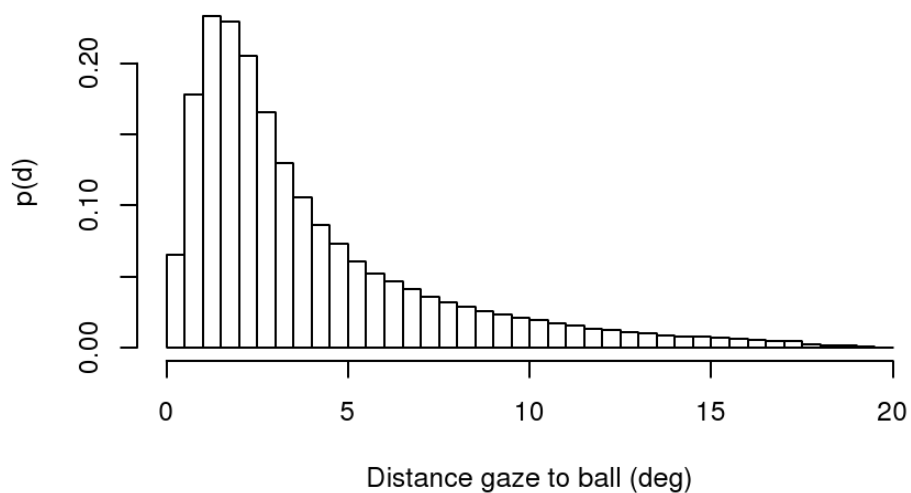


**Figure 5.** Distribution of time spent fixating different in-game items. Experts spend more time looking at the ball and at good extras than novices. Note that percentages for extras are relative to the time where any good (or bad, respectively) items were shown on the screen (between 9 and 19% of overall time). Values missing to 100% correspond to those fixations where no clear fixation target could be identified.

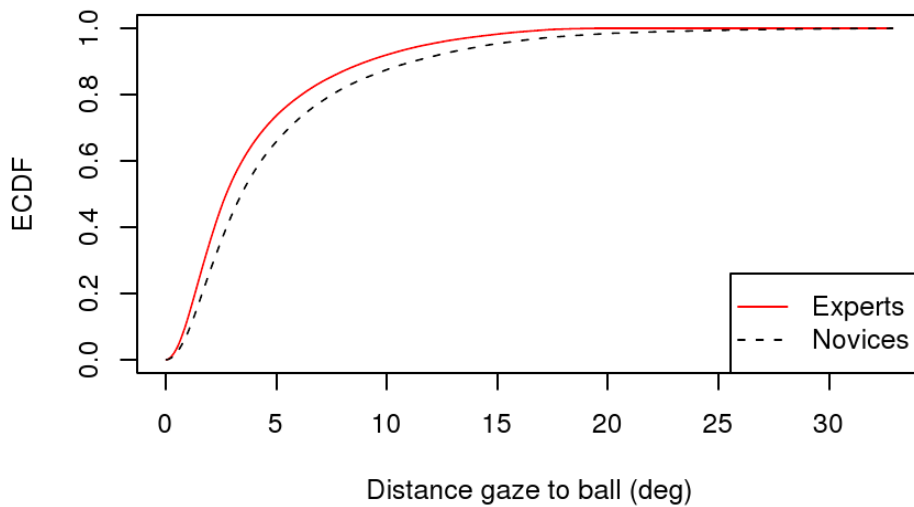
#### 4.4 Distribution of Distance from Gaze to Ball

We now turn towards the question of how close gaze follows the ball. A distribution of distances for the experts can be seen in Figure 6: 50% of all gaze samples are closer to the ball than 2.76 degrees (experts; 3.45 degrees for novices) and 75% of samples are still closer than 5.24 degrees (6.46 degrees for novices). To visualise the difference between experts and novices more clearly, empirical cumulative distribution functions are plotted in Figure 7 and they show a clear difference between the two groups.

This difference is highly significant ( $p < 1e-10$ ) using a corrected Kolmogorov-Smirnov test. The test statistic had to be corrected for the overestimation of sample size introduced by the high sampling rate of the eye tracker: because eye position is sampled at 500 Hz, but significant eye movements occur at a much lower rate, the distance measurements are highly correlated and therefore violate the assumption of independent samples for the statistical test (Weiss, 1989). To correct for this bias, we estimated the amount of statistical dependence by evaluating where the autocorrelation function of the distance measurements had dropped to 0.5; we then subsampled the distance distributions by that factor (averaged over all 9 subjects 118.9 samples, s.d. 28.1).

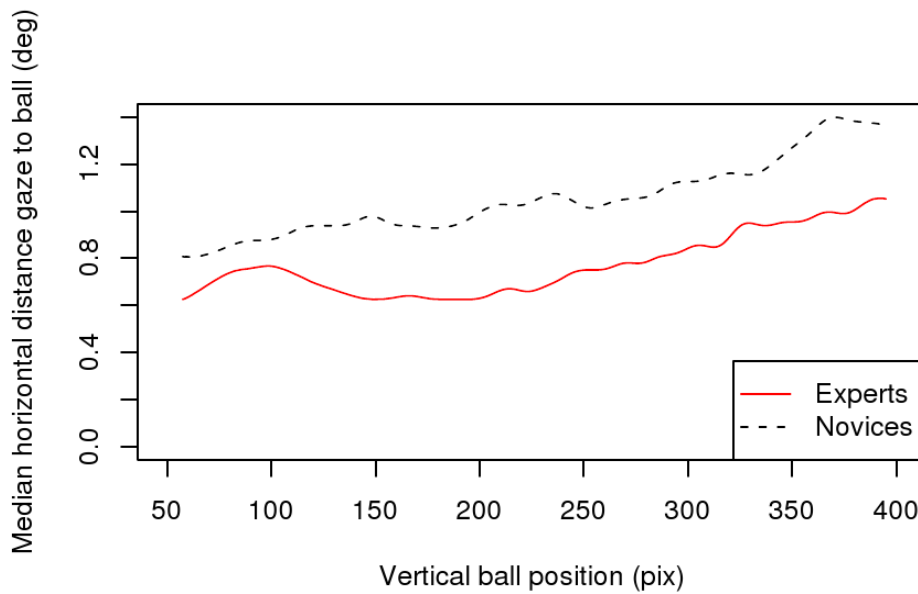


**Figure 6.** Distribution of distance from gaze to ball for the experts. Most of the time, the ball is at least kept in para-foveal vision.



**Figure 7.** Comparison of distribution of distance from gaze to ball for experts and novices. Experts spend more time looking closer to the ball; this difference is highly significant ( $p < 1e-10$ , corrected Kolmogorov-Smirnov test).

A further aspect of novice and expert behaviour we investigated was how ball-following changed across the screen. Strictly speaking, it is only necessary to fixate or closely follow the ball whenever it is close to the bottom, because a failure to do so will result in the loss of a game “life”. When the ball is further up on the screen, the player's gaze is free to look around, for example to collect extra items; nevertheless, the deflection point of the ball at the top of the screen (or the bottom-most row of bricks) might be a worthwhile gaze position because it is most informative about where the ball will come down again (for example, table tennis players use a similar strategy to predict the ball's trajectory, see Land and Furneaux, 1997). We therefore show the median of the distance from gaze to ball as a function of vertical ball position on the screen in Figure 8; note that this plot only shows horizontal distance because it should be independent of vertical ball position, whereas vertical distance might be affected by border effects (e.g. when the ball bounces at the top border of the screen, subjects might fixate slightly below the expected deflection point to maximize the time the ball is close to the centre of fixation). Also note that we plot only those samples where the subject was not obviously fixating another interesting object, i.e. an extra item.



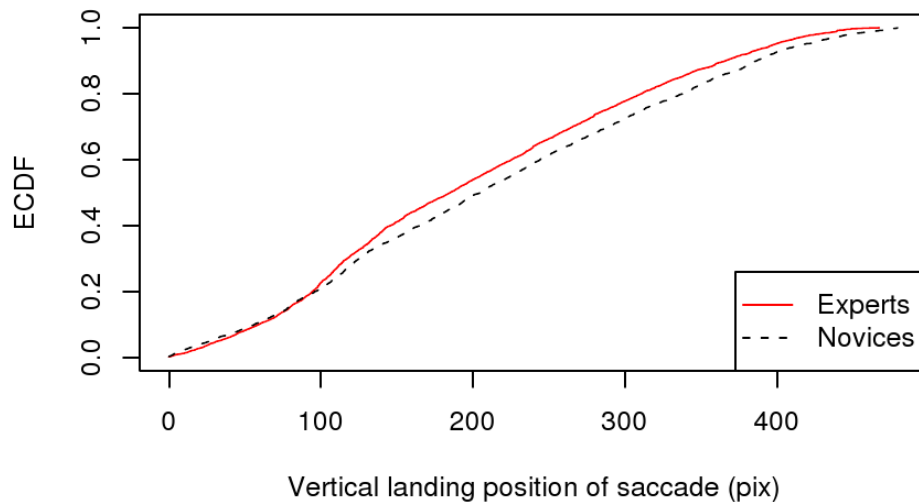
**Figure 8.** Median horizontal distance of gaze to ball as a function of vertical ball position. Ideally, this distance should be minimal at the bottom of the screen; note the local peak in distance for the expert group shortly before the ball needs to be hit by the paddle.

Indeed, novices look closest to the ball only at the bottom of the screen, when they need to hit the ball with the paddle, and horizontal distance increases almost monotonically towards the top of the screen. For experts, however, the picture is not quite as clear: although the distance is at its maximum at the top of the screen, it is relatively small slightly below the middle, and there is even an increase close to the bottom, just before the paddle hits the ball. This local peak close to the bottom was found in all the individual experts' recordings and we can only hypothesize about a possible explanation. Experts might not follow the ball on its way towards the bottom with a pursuit movement, but predict the landing point on the paddle (this sounds very simple; in practice, however, it is quite difficult to maintain fixation in the absence of a fixation target at the bottom of the screen -- and in the presence of a behaviourally relevant, moving stimulus close to the fovea!); this would also help to deliberately hit the ball with one of the sides of the paddle to return the ball in a certain direction.

#### 4.5 Distribution of Saccadic Landing Points

In a final analysis of eye movement strategies, we investigated where experts and novices direct their saccades. Figure 9 shows the empirical cumulative distributions of the vertical components of saccadic landing points: whereas experts and novices seem to make a similar number of saccades towards the bottom of the screen, the gaze of

experts jumps more often slightly above that. This highly significant difference ( $p < 2.6e-5$ , Kolmogorov-Smirnov test) is in line with the difference in horizontal distance to the ball close to the bottom of the screen as depicted in Figure 8: experts seem to employ a different strategy just before the ball hits the paddle.



**Figure 9.** Empirical cumulative distribution function for vertical landing positions of saccades. Experts make fast eye movements towards a region slightly above the bottom line more often ( $p < 3e-9$ , Kolmogorov-Smirnov test).

## 5. Conclusion

We have presented modifications to the open-source game LBreakout2 that allow the game to be controlled with gaze. Even though both the graphics and the game play of LBreakout2 are very simple, our test subjects found “playing with eyes” highly enjoyable.

We also described how playing with gaze affects eye movement statistics and how eye movement strategies change with training. Specifically, we have shown that expert players are fixating a position close to the ball more often, spend more time looking at good extra items, and apparently seem to employ a different strategy to predict where the ball will come down just above the bottom line.

Finally, we have also presented results that show that gaze-based interfaces can be superior to traditional input modalities even for users that have had no previous training with such interfaces. At first, these results seem to be at odds with previous work on gaze-controlled computer games, which found that gaze control does not perform as well as mouse control (Smith & Graham, 2006; Isokoski & Martin, 2006). However,

Breakout has a particularly simple and intuitive game play that makes it ideally suited for gaze playing. Future work will have to address how such natural gaze input can also be successfully used in more complex game scenarios and how games can be designed specifically for gaze input.

## 6. Acknowledgements

Our research has received funding from the European Commission within the project GazeCom (contract no. IST-C-033816) and the Network of Excellence COGAIN (contract no. IST-2003-511598) of the 6th Framework Programme. All views expressed herein are those of the authors alone; the European Community is not liable for any use made of the information.

## 7. References

- Dorr, M., Böhme, M., Martinetz, T., & Barth, E. (2007, September). Gaze beats mouse: a case study. Presented at *the Third Conference on Communication by Gaze Interaction*, Leicester, UK.
- Isokoski, P. & Martin, B. (2006, September). Eye tracker input in first person shooter games. Presented at the *2nd Conference on Communication by Gaze Interaction*, Turin, Italy.
- Istance, H., Bates, R., Hyrskykari, A., & Vickers, S. (2008). Snap clutch, a moded approach to solving the Midas touch problem. In *Proceedings of the 2008 symposium on Eye tracking research & applications, ETRA '08* (pp. 221-228). New York: ACM Press.
- Jacob, R.K. (1993). Eye movement-based Human-Computer Interaction techniques: Toward non-command interfaces. In Hartson, H. R. & Hix, D. (Eds.), *Advances in Human-Computer Interaction* (pp. 151-190). Norwood, NJ: Ablex Publishing.
- Kent, S. (2001). *The ultimate history of video games: From Pong to Pokemon and beyond*. New York: Prima Life.
- Krepki, R., Blankertz, B., Curio, G., & Müller, K.-R. (2007). The Berlin Brain-Computer Interface (BBCI): towards a new communication channel for online control in gaming applications. *Journal of Multimedia Tools and Applications*, 33(1), 73-90.
- Land, M.F. & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 352, 1231-1239.

- Li, D. & Parkhurst, D.J. (2006, September). Open-source software for real-time visible-spectrum eye tracking. Presented at the *2nd Conference on Communication by Gaze Interaction*, Turin, Italy.
- Smith, J.D. & Graham, T.C. (2006). Use of eye movements for video game control. In *Proceedings of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology* (p. 20). New York: ACM Press.
- Speck, M. (n.d.). LBreakout2. <http://lgames.sourceforge.net>
- Weiss, M.S. (1989). Testing EEG data for statistical normality. In *Proceedings of the Annual International Conference of the IEEE Engineering in Volume* (pp. 704-705). Los Alamitos: IEEE Computer Society Press.



# Evaluation of the Potential of Gaze Input for Game Interaction

Javier San Agustin<sup>\*♦</sup>, Julio C. Mateo<sup>\*♦</sup>, John Paulin Hansen<sup>♦</sup> and Arantxa Villanueva<sup>^</sup>

<sup>♦</sup> University of Copenhagen  
(Denmark)

<sup>\*</sup>Wright State University  
(USA)

<sup>^</sup>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](http://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](mailto: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 & Rähä, 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 & Rähkä, 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:

$$IP = \frac{ID}{MT} \quad (1)$$

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:

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

$ID$  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

$$MT = \frac{ID}{IP} \quad (3)$$

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

$$MT = a + b ID \quad (4)$$

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:

$$Throughput = \frac{ID_e}{MT} \quad (5)$$

where  $ID_e$  is determined as follows:

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

$ID_e$  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

$$W_e = 4.133 \times SD \quad (7)$$

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

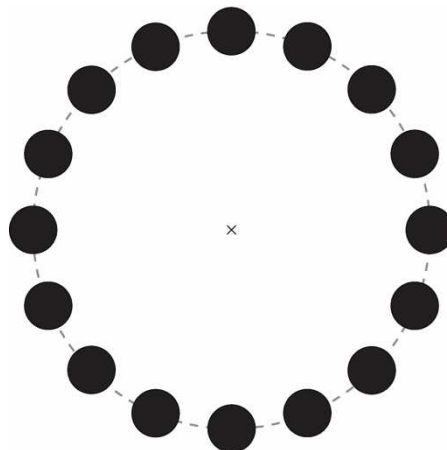
previous experience with joystick devices; 3 had previous experience with eye trackers, and 1 with head trackers.

### *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.

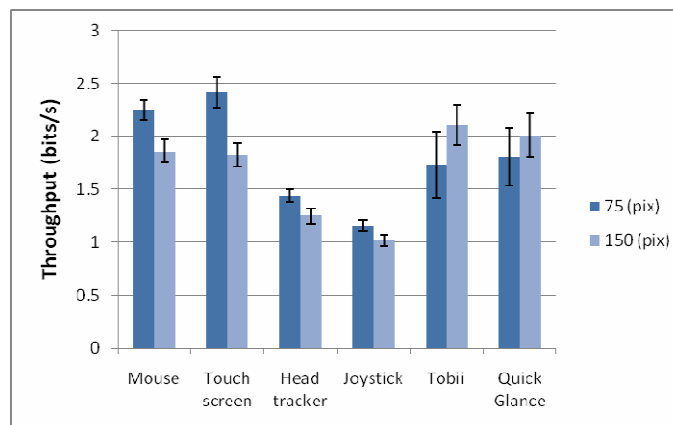
At the beginning of each block, the participant pointed at the X on the center of the screen to indicate he or she was ready to start, triggering the release of the first target. This procedure was repeated at the beginning of each trial to ensure that the starting position of the pointer was at the 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.

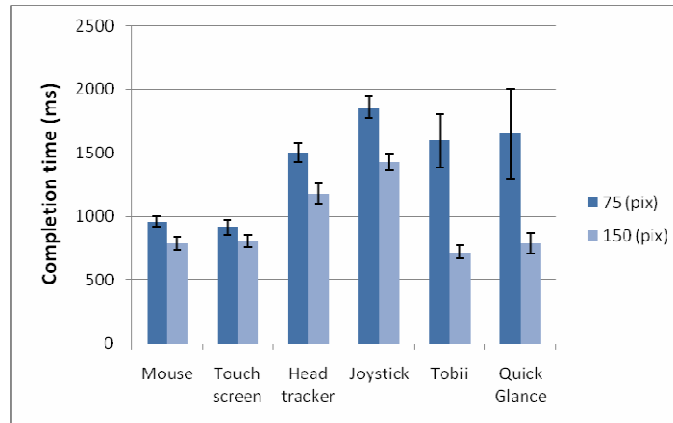


**Figure 2.** Mean throughput of each device for both target sizes. Error bars show standard errors of the mean.

### 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

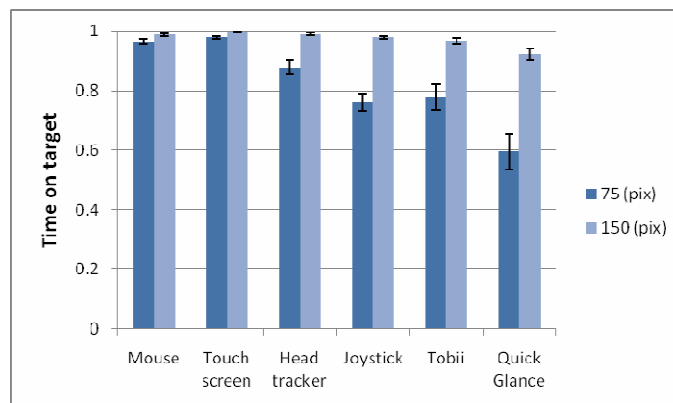
not differ significantly. Size had a significant effect on completion time,  $F(1, 5) = 26.88$ ,  $p < 0.05$ , but type of feedback did not,  $F(1, 5) = 1.41$ ,  $p > 0.05$ . Figure 3 shows the completion time for the different devices and target sizes.



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

#### *Time on Target*

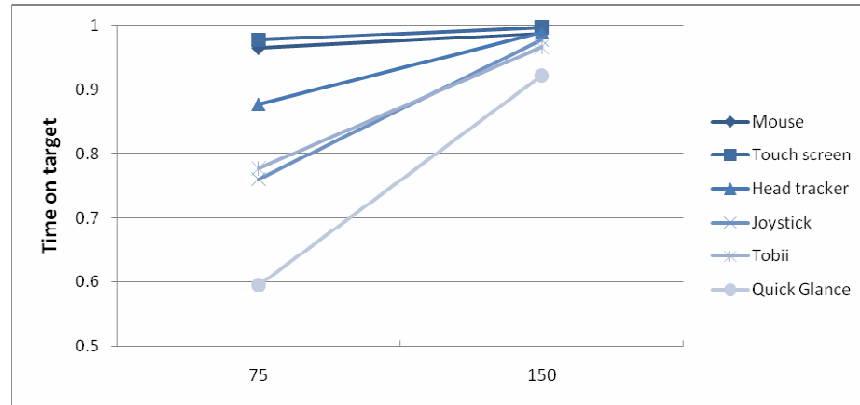
The overall mean time on target (TOT) was 0.90. There was a significant effect of input device on TOT,  $F(5, 25) = 15.06$ ,  $p < 0.05$ . TOT was significantly lower on small targets ( $M = 0.82$ ,  $SD = 0.17$ ) than on big targets ( $M = 0.97$ ,  $SD = 0.04$ ),  $F(1, 5) = 74.77$ ,  $p < 0.05$ . Feedback 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

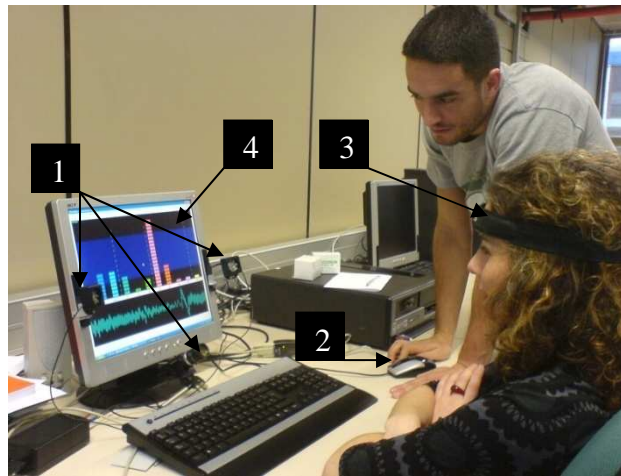
###### *Participants*

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 1024x768 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  $0.5^\circ$  (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.

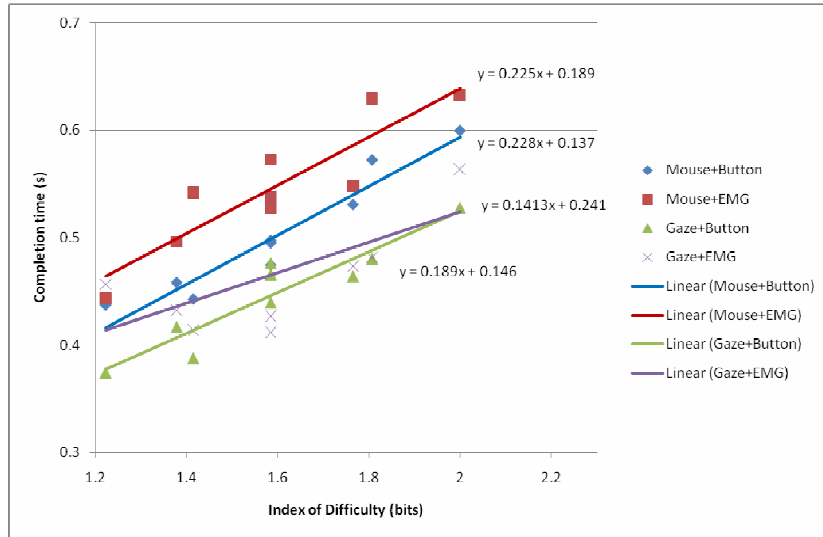
## 4.2 Results

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*

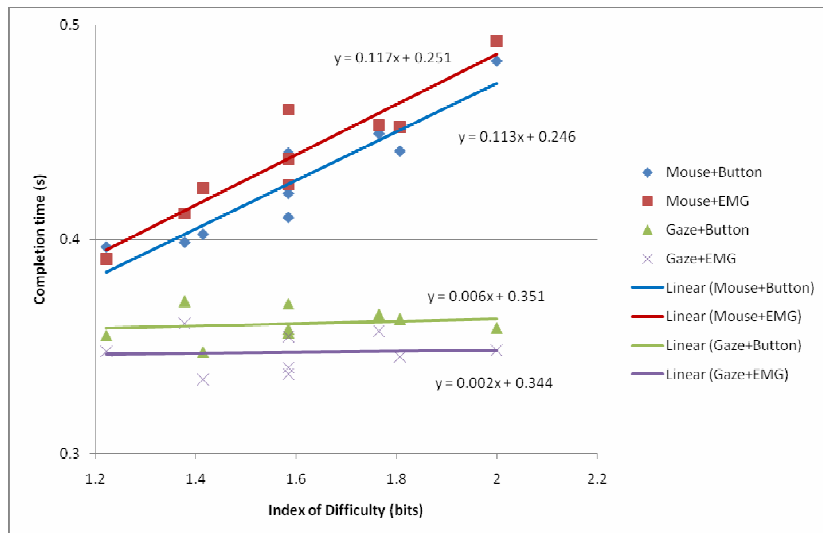
The mean completion times for each combination of size and distance were used to analyze how well the data fitted Fitts' Law. As the index of difficulty (ID) increases, Fitts' Law predicts a linear increase in completion time. Following Equation 4, the regression lines for the four input combinations 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.

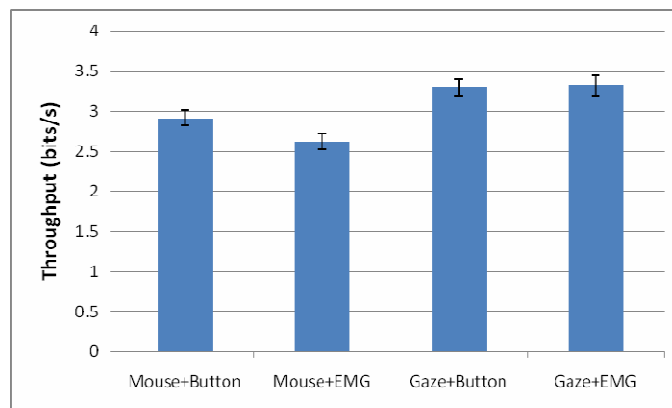
A reanalysis of the data was performed after removing erroneous trials. The regression lines and corresponding equations are shown in Figure 8. When looking at these error-free data, input combinations in which the mouse was used for pointing present positive slopes (slope > 0.11), whereas combinations 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

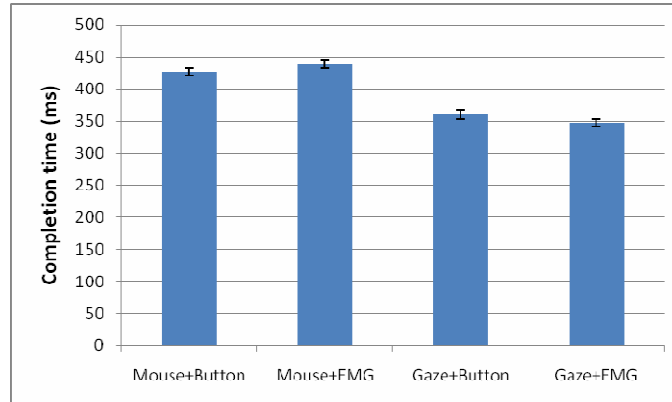
A high error rate was observed in this experiment. Therefore, a correction of the target width was performed by means of the error rate (MacKenzie, 1992). Overall mean throughput was 3.03 bits/s. Mean throughput was higher for gaze pointing ( $M = 3.31$  bits/s,  $SD = 0.78$  bits/s) than for mouse pointing ( $M = 2.76$  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

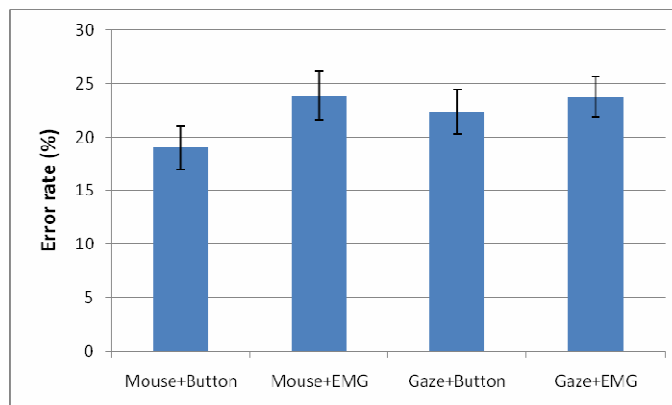
Overall mean completion time was 393 ms. Mean completion time was lower for gaze pointing ( $M = 354$  ms,  $SD = 46$  ms) than for mouse pointing ( $M = 433$  ms,  $SD = 43$  ms),  $F(1, 4) = 29.91$ ,  $p < 0.05$ . The mean completion times for mouse selection ( $M = 394$  ms,  $SD = 57$  ms) and EMG 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 3<sup>rd</sup> *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 *2<sup>nd</sup> 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., & Rähkä, 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.

