SELECTION STRATEGIES

IN GAZE INTERACTION

BY

EMILIE MOLLENBACH

A DOCTORIAL THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF
DOCTOR OF PHILOSOPHY

LOUGHBOROUGH UNIVERSITY

15 JULY 2010

ABSTRACT

This thesis deals with selection strategies in gaze interaction, specifically for a context where gaze is the *sole input modality* for users with severe motor impairments. The goal has been to contribute to the subfield of assistive technology where gaze interaction is necessary for the user to achieve autonomous communication and environmental control.

From a theoretical point of view research has been done on the physiology of the gaze, eye tracking technology, and a taxonomy of existing selection strategies has been developed.

Empirically two overall approaches have been taken. Firstly, end-user research has been conducted through interviews and observation. The capabilities, requirements, and wants of the end-user have been explored. Secondly, several applications have been developed to explore the selection strategy of *single stroke gaze gestures* (SSGG) and aspects of *complex gaze gestures.*

The main finding is that *single stroke gaze gestures* can successfully be used as a selection strategy. Some of the features of SSGG are:

- That *horizontal single stroke gaze gestures* are faster than *vertical single stroke gaze gestures.*
- That there is a significant difference in completion time depending on gesture length*;*
- That *single stroke gaze gestures* can be completed without visual feedback.
- That gaze tracking equipment has a significant effect on the completion times and error rates of *single stroke gaze gestures*
- That there is not a significantly greater chance of making selection errors with *single stroke gaze gestures* compared with *dwell selection*.

The overall conclusion is that the future of gaze interaction should focus on developing *multi-modal interactions* for *mono-modal input*.

ACKNOWLEDGEMENTS

This thesis is based on research carried out at the Applied Vision Research Centre at Loughborough University and in collaboration with the Gaze Group at the IT University of Copenhagen. These groups have extensive knowledge in the areas of both analytical and interactive gaze analysis, as well in the area of eye tracking equipment development.

It has been an interesting and exciting process to conduct and complete this PhD. The challenges could not have been met without the support of many people. First of all I would like to thank Prof. Alastair Gale for giving me the opportunity of doing this work and the trust to complete it on my own terms. I would also very much like to thank John Paulin Hansen who has been my co-supervisor and as offered great support both academically, logistically and by his ability to create an overview of a given situation. Without the support of my partner Martin Lillholm much of this work would never have been completed, and I thank him for bearing with me when times were tough.

I would like to thank all of the people at Loughborough who have supported me. First of all Lindsay Cooper at Loughborough for being a great friend and housing me on several occasions and Sian Phillips and Yan Chen for their friendly presence and making me feel welcome at Loughborough. I would also like to thank Iain Darker for helping to clear my head, especially in the beginning of my PhD. Finally, I would like to thank Sarah Lowe and Ruth Spencer for all of their practical support.

The ITU has been my second home during the making of this PhD. Again I would like to thank John Paulin Hansen for making that possible. A special thanks must go to Henrik Skovsgaard and Javier San Agustin for being fantastic office mates and great friends to have on road trips. I look forward to hopefully working in some capacity with all of the individuals mentioned above.

Finally, I would like to thank my great parents and wonderful family for always believing in me, even when I don't myself.

TABLE OF CONTENTS

PART I

1 INTRODUCTION

1.1 CONTEXT AND MOTIVATION

The general area being addressed in this dissertation is gaze interaction; specifically making selections based on eye movements. Gaze interaction entails using eye tracking technology to place the direction of gaze in relation to a screen or other object of interest, which can then be manipulated by gaze. In the context of this research gaze is viewed mainly as a *sole input modality*. The main challenge of gaze as sole input is to determine when and what the user wants to interact with. In this chapter, a general introduction to this field of research is given.

Initially the basic design constraints of gaze interaction along with the main motivation for this research is introduced; followed by a brief presentation of the theoretical background which leads to the main hypothesis. Finally, there is a description of the methodology and a thesis overview.

1.1.1 MOTIVATION

Gaze interaction has some innate design constraints and affordances (Norman, 1999). The three aspects presented in figure 1 constitute the three different approaches generally taken in the general field of eye tracking research: gaze tracking, gaze patterns and cognition, and end-user directed applications.

Firstly, the development of stable eye trackers is very much of concern to the gaze tracking research community. This type of research entails building hardware and software solutions that extract features from the eyes in order to establish gaze- and eye movement detection. There is a special interest in creating low cost eye trackers which will increase availability of eye tracking to the general public (D. W. Hansen, et al. 2004; San Agustin, 2009; San Agustin et al., 2010).

The second field of research in eye tracking is that of relating gaze patterns to cognitive processes. This is the focus of neuro – and cognitive psychologists and researchers in other areas of both natural and human sciences (Carpenter, 1988; Daniel & Whitteridge, 1961; Land

& B. W Tatler, 2009; Yarbus, 1973). Historically, this has been the most prevalent research field.

Finally, there is the field of applied gaze interaction where the end-user is in focus. Here both the gaze tracking equipment and the cognitive processes which guide search patterns in natural behaviour are perceived as 'black boxes'. The questions asked in this field are not: 'How do we track what we are looking at?' or 'How do we interpret what we look at?' – It is: 'How do we use the movements of our eyes to infer our intention to a system?' As indicated in figure 1 it is the latter of these research areas which constitutes the epistemological standpoint taken in this work.

However, an understanding of the two other research areas is of great importance when designing sustainable gaze interaction:

(1) Eye tracking technology has certain limitations in regard to precision and reliability, which need to be taken into account.

(2) The eyes as input have certain physiological and cognitive constraints which affect how interaction can be implemented.

(3) Depending on the context of the end-user – what and how interaction with gaze can and should be implemented differs greatly.

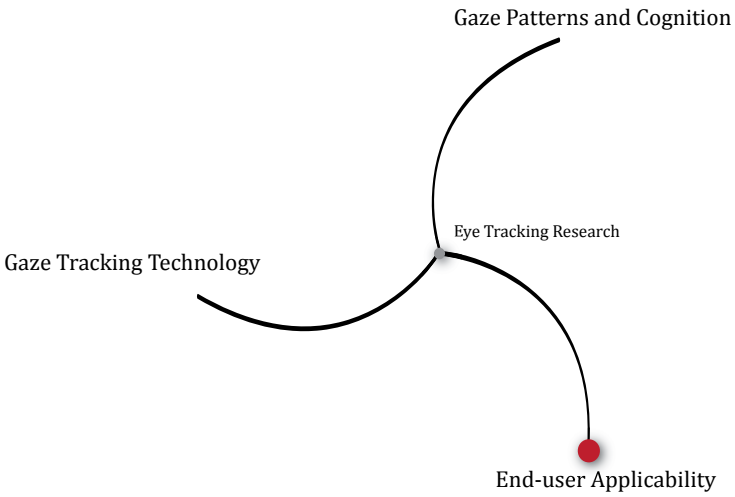


Figure 1: The three main aspects of eye tracking research, the red circle represents the primary placement of this research.

Gaze is used to perceive information from our surroundings. However, in gaze interaction the aim is to be able to convey information from the eyes to a system. The problem with the eyes as input is that they are always 'on'. This problem is sometimes referred to as the *Midas touch problem* (Jacob, 1991a, 1996; Jacob & Karn, 2003). (The mythical figure of King Midas was granted a wish for doing service to a god. He wished that all he touched would turn to gold. When this wish was granted, he quickly found a flaw in his plan, as all he touched, including food and drink, indeed turned to gold and he nearly starved to death). The design challenge in gaze interaction is to determine whether it is the intention of the user to *perceive information* or to *instigate action*. This constitutes one of the main motivations of this research.

Another motivation is to understand the constraints of gaze tracking technology. Inaccuracy is an issue even in the most expensive commercial eye tracking systems. Generally, systems are inaccurate to 1° of visual angle on most commercial systems. Manufacturers claim ½° accuracy but these measurements are often based on ideal circumstances (Porta & Turina, 2008a). In pointing based gaze interaction, where gaze direction is used to substitute a mouse, this inaccuracy affects the size required of on-screen targets; these need to be quite large, if the goal is to reliably affect them by gaze, to compensate for the discrepancy between *intended point of gaze* and *tracking of gaze*.

Finally, motivation also comes from understanding the needs of end-users who can benefit most from eye tracking technology. Gaze interaction has many potential applications; it could be used to supplement existing main stream input modalities enhancing performance in computer-games, 3D modelling programs and in general actions such as scrolling and panning. It could also be employed in any situation where the hands are otherwise occupied, for example a doctor operating, a soldier moving in hostile area, etc. However, the research conducted here has concerned itself with *gaze as sole input* specifically for people with various motor impairments.

For individuals who are paralyzed and with complete or incomplete Locked-in-Syndrome (LIS) (Mollenbach et al., 2008), which renders them unable to interact with a computer in a conventional sense, eye tracking provides the potential for autonomous action. Gaze interaction can mean the difference between users retaining their jobs and maybe more importantly allows them to communicate with friends and family. Gaze control can allow

users to write e-mails, articles and even books, play computer games, and engage in conversations without the need of an assistant carer.

Uncovering flexible and sustainability gaze interaction principles which take into account the nature of eye movements, the limitations of gaze tracking technology and the needs of users with motor- impairments is both the motivation and theoretical foundation for this thesis.

THE END-USER

One end-user who has contributed to this research is the Danish physicist Arne Lykke Larsen, who has had ALS (Amyotrophic Lateral Sclerosis) for almost 10 years. He has contributed through interviews and has also written a Danish book on ALS called 'Rather die of laughter than ALS – 99 truthful stories of what it means to live with Amyotrophic Lateral Sclerosis' (Larsen, 2009). He has advanced ALS, which renders him completely paralyzed except from certain facial expressions and eye movements. However, he is also an Associate Professor in theoretical physics at the University of Southern Denmark. I will let him introduce himself:

'There is not much to say about my childhood. I was naturally a child prodigy with unique mathematical, musical, linguistic, artistic and athletic abilities. But unfortunately no one ever noticed…..

I was diagnosed with ALS at the age of 35 in November 2000. At that time I had already gotten my first wheelchair, but could still walk a few meters with a walker frame.

[...]When I got the diagnosis I went into a state of shock and disbelief -mostly disbelief. The doctors were probably wrong; it had been heard of before. I also clung to the fact that I didn't have all the symptoms of ALS. For instance I still had normal speech. So, it was probably something else which would blow over – just a question of time. Several months passed before I accepted that I had ALS. But I was determined to continue my old life, with work and hobbies, to as great an extent as possible.

Unfortunately, I must admit that the disease has spread so much that I am now more or less paralyzed in the entire body, I can't speak and I have started having problems breathing and am fed through a tube in my stomach. I am getting worse and worse; and each month I can feel myself regressing.

I can still move my head relatively freely, so I can easily control my computer. But besides from that I need help with everything. 'Help with' is incidentally a misleading term in my situation: My own contribution is pretty much non-existent. I live alone in a small 1 bedroom apartment and am now 100% dependent on my personal helpers 24 hours a day.

[...] This has meant that I have been able to continue my work – which is mainly research. At the moment I am also supervising a masters student, with whom I communicate solely by speech synthesis. So I still have my daily 'grind' at the University every afternoon, and also work a lot from home. My spare time is filled with wonderful drives to forests and beaches (though mainly in the summer) or to places of historical or archaeological interest. Visits to friends and family and also frequent trips to the cinema.

Most evenings I use my Bi-PAP (breathing support). Even though I can manage without it for the time being, it provides a nice relief. Besides, it constitutes a super, and indisputable excuse for sitting passively in front of the TV (news, American movies and sport) all night, because there really isn't much else you can do …

I haven't felt any bitterness about my bad luck at any time. For a short while, after I got the diagnosis, I felt sorry for myself. But it didn't help.'

Arne Lykke Larsen

Throughout this thesis excerpts from his book, which I have been allowed to translate, email exchanges and comments from interviews will help to keep the focus of this research on the end-user.

1.1.2 GAZE TRACKING

Most modern eye tracking equipment for interaction is based on one or more high-resolution cameras capturing the eye. Specific software can then extract eye-features that have been enhanced by external light sources. These light sources create a clear reflection on the eye, similar to the 'red eye effect' that can be caused by a photographic flash. Having a bright light directed at the eye is very uncomfortable, so infrared light sources are usually employed. Infrared reflection can be detected by a camera but is not noticed by the human eye (Ebisawa et al., 1989; Morimoto et al., 2000). In spite of the technological leaps that have made gaze

tracking both non-invasive and robust – eye tracking is still cumbersome, inaccurate and expensive. Gaze tracking systems require calibration for every new user and quite often for every new session of use – sometimes even mid-session re-calibrations are necessary. As mentioned, most commercial systems have an inaccuracy of approximately 1° visual angle (Porta & Turina, 2008a) and they also cost more than £6,000-7,000, which makes them inaccessible to many people (San Agustin, 2009).

1.1.3 THE EYE AND EYE MOVEMENTS IN GAZE INTERACTION

The physiology of the eye and the nature of eye movements will be covered in greater detail in Chapter 2. However, a few points are needed to create a context before reading the main hypothesis. As mentioned earlier, one of the major challenges in gaze interaction is the *Midas Touch* problem, that is, the accidental selection of anything that is looked at. To avoid this *dwell time selection* was invented, which is a *fixation based gaze selection strategy* that builds on a familiar button-metaphor. An *on-screen button* can be triggered by having one's gaze *fixated* on it for a prolonged duration of time (a *dwell*). If this extended dwell duration was not implemented clicks would occur every time a user simply glanced at an on-screen button (Midas touch). *Eye fixations* are generally considered to last no less than 100ms and often between 200-400ms (Salvucci & J. H. Goldberg, 2000). On-screen dwell- buttons usually do not activate until after a 400ms *fixation*. Dwell- buttons have been a key building block in the development of gaze interaction (Bates, 2002; Jacob, 1996; Majaranta & K. J Räihä, 2002). However, dwell-time selection can be affected by uncontrollable jitter from small eye movements or from larger involuntary eye movements. The consequence can be that an *intentional fixation* is broken half way through. This paired with the previously mentioned inaccuracy of gaze tracking equipment force *on-screen targets* to be quite large. This greatly limits the amount of information that can be displayed on the screen and impacts both how data can be structured and visualized.

The other major type of eye movement is the *saccade*. *Saccades* are rapid 'jolts' (ballistic movements) that the eye makes between *fixations*. These can move in every direction and recently many theories have emerged on how combinations of eye movements could be used in gaze interaction. These combinations are generally referred to as *gaze gestures* and constitute the main focus of the research conducted in this thesis.

1.2 RESEARCH QUESTIONS

This research investigates *selection strategies* in gaze interaction and different approaches have been investigated and will be discussed. The issue of gaze selection strategy was examined from an end-user point of view, from an interaction design point of view and from a theoretical and empirical point of view. The following research questions will be addressed:

1. Gaze as input has specific interaction design features.
    a. What are these features?
    b. What considerations are important when designing gaze selection strategies?

2. There are many different types of selection strategies.
    a. What are these strategies?
    b. When and how are they best used?
    c. How could they potentially be used in the future?

3. Users with severe motor impairments have special needs.
    a. What are these needs?
    b. How can these be assessed from a design perspective?
    c. How do these affect design of gaze based applications and selection strategy?

4. Gaze gestures constitute a selection strategy.
    a. What is the simplest form of gaze gesture?
    b. What are some of the basic characteristics of simple gaze gestures?
    c. How could these characteristics affect the way gaze gestures are implemented?

1.3 GOALS AND METHODS

The goal of this research is to contribute to the field of gaze interaction; as mentioned specifically the area of gaze interaction that concerns itself with gaze as sole input for users with severe motor impairments. The contribution is two-fold. Firstly, some theoretical perspectives, which could be useful to other researchers, have been formed. Secondly, some

practical guidelines have developed through the empirical investigation conducted in the course of this work – these could potentially be useful for designers of gaze applications

Theoretically, the approach has been to understand existing selection strategies in the field of gaze interaction. By structuring and analyzing the affordances and constraints of these methods an overview of gaze interaction selection strategies is created. From this overview, theories on the current and future state of gaze interaction will be presented and a definition of the basic principles which govern gaze selection strategies proposed.

Empirically two main methods have been used. First an ethnographic design approach was taken in gathering information from the user-group. Interviews were conducted on an individual basis with people with Locked-in-Syndrome, along with several observational sessions where the individuals were engaged in various daily activities such as e-mail writing or communicating with a spouse, relative or carer. Also formal discussions have been undertaken with carers and professionals who work with users with motor impairments.

The second empirical approach was experimental. Several programs have been designed in order to explore the gaze selection strategy of *single stroke gaze gestures* (SSGG): a simple form of gaze gesture which in the case of this research consists of *horizontal* and *vertical* point-to-point eye movements.

1.4 THESIS OVERVIEW

The thesis is divided into four major parts:

Part I: This Introduction

Consisting of Chapter 1; a general overview of the research topic that has been explored in this thesis and function as a reading guide.

Part II: The Background

Consists of two chapters; Chapter 2 is concerned with the physiology of the eye and the technology of eye tracking. Chapter 3 deals with previous work in the field of gaze interaction, specifically creating a taxonomy for different types of gaze selection strategies.

Part III: Empirical Research

Consists of five chapters; Chapter 4 is concerned with the end-user and introduces the design context for the work which will be empirically explored. Chapter 5 consists of two pilot studies that were conducted and introduces many aspects of the experimental methodology that is subsequently used. Chapter 6 is the first of three experiments conducted regarding single stroke gaze gestures, and is mainly concerned with enabling a direct comparison between single stroke gaze gestures and dwell selection. Chapter 7 is the next single stroke gaze gesture experiment, which is concerned with looking at *gesture completion* on different eye trackers. Chapter 8 is the final experiment regarding single stroke gaze gestures and is mainly concerned with exploring gesture completion *with* and *without visual feedback*. Chapter 9 introduces a new iteration of gaze gesture research, which looks at determining a threshold for different levels of gaze gesture complexity.

Part IV: Design, Reflection and Discussion

Consists of the final chapter in which an overall discussion is taken in regard to the experimental design and results presented in the thesis. As well as presenting ideas for future implementations and future research areas.

PART II

2 THE EYE, GAZE AND EYE TRACKING

2.1 THE EYE AND THE VISUAL CORTEX

When conducting eye tracking research, both applied and theoretical, there is a need to understand the basic physiological traits of the eye before it is used for input or analysis. Three main aspects constitute this understanding: (1) the eye as an organ, (2) the connection to the *visual cortex* and (3) the muscles which ensure eye movements. The full complexity of these issues far exceeds the scope of this research and will therefore only briefly be described.

The eye is a light sensitive organ, which absorbs light impulses and directs them to the brain via different pathways. Light is a form of electromagnetic radiation created from the oscillation of electrically charged materials. This is in abundance in our surroundings; which is most likely the reason that humans have developed a sense to register electromagnetic radiation (Sekuler & Blake, 1994).

Visible light only constitutes a very small part of the actual spectrum of electromagnetic energy (other forms being radio waves, infrared and ultraviolet radiation, microwaves etc.). However, visible light's ability to interact with the surface of objects through absorption and reflection allows for a great amount of detail to be conveyed. The structure of the eye (Figure 2) has carefully evolved to map these detailed light impulses to the brain.

Only a few of the components of the eye will be discussed here. The 'white' part of the eyes is the *sclera* and is a protective layer that not only protects the inner workings of the eye, but also defines the curved shape of the eye. The shape of the eye is fundamental both in regard to the movement of the eyes and the way that the *optic nerves* transfer information to the visual cortex.

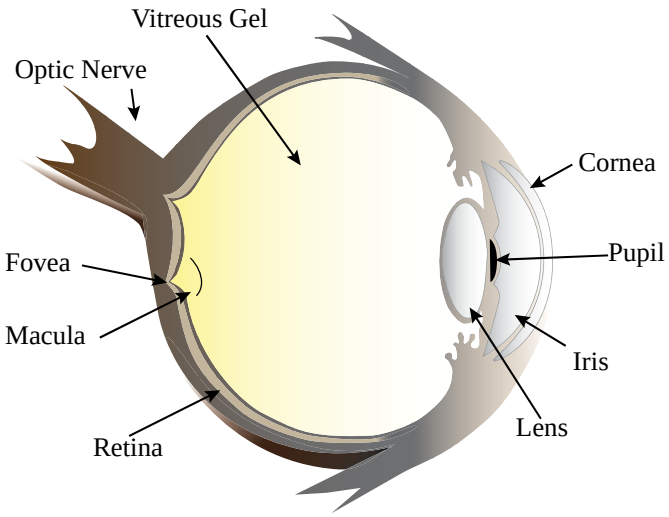


Figure 2: The anatomy of the eye

At the centre front the *sclera* becomes transparent and a small bulge protrudes. This transparent area is called the *cornea*, which constitutes the main light filter. This area is very sensitive and any damage or decrease in transparency to the *cornea* has a detrimental effect on visual perception. The *iris* is multilayered and the pigmentation value of one of its layers defines the colour of the eye. The *iris* surrounds the *pupil*, which is a gap set between two sets of muscles that ensure the expansion and retraction of pupil size depending on light levels. It decreases when the amount of light projected towards the eye increases and vice versa. The size of the pupil varies from person to person and with age it can decrease to less than half of its original size. The *lens* lies behind the *iris* and it (should be) is transparent – the shape determines near or far sightedness and a decrease in transparency (e.g. cataracts) greatly reduces visual perception. The part of the eye with most volume is the *vitreous body*, which consists of a transparent fluid through which light is filtered. This brings us to the inner most part of the eye. Light is funnelled through the *cornea*, *pupil*, *lens* and *vitreous body* to the *retina*. (Sekuler & Blake, 1994)

The process of receiving and transforming light waves into visual information begins in the retina. There are several different ways to sub-divide the retina in order to understand how it is mapped to the visual cortex. The first has been by defining areas of the retina that correspond to the field of view which they relay information from to the visual cortex. This can be done by subdividing the field of view into four areas: *Nasal*, *temporal*, *upper* and *lower* (Figure 3). (Zeki, 1993)

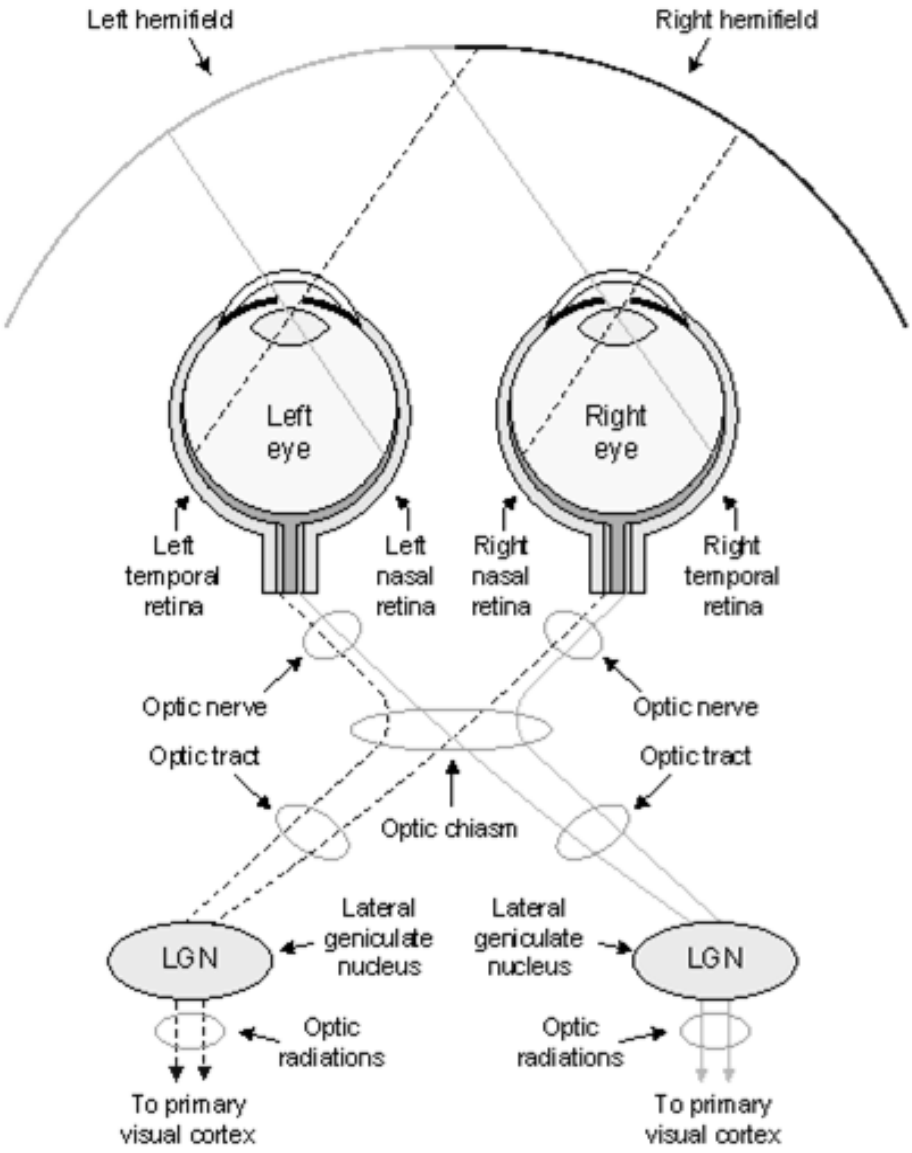


Figure 3: The high level visual pathways of the right and left eye[1].

Because the eye is curved the *nasal area* of the left eye and the *temporal area* of the right eye convey information about the left field of view to the brain; this is called the *left hemifield*. The same principal applies to the right field of view; the *right nasal* and the *left temporal retinal areas* constitute the *right hemifield*. These *hemifields* can then be subdivided again into *upper* and *lower* quadrants depending on whether we are looking up or down. Again because of the curvature of the eye the lower part of the retina transfers information from the upper field of view and vice versa, the upper part looks at the lower field of view (Zeki, 1993).

[1] http://www.diycalculator.com/imgs/cvision-left-to-right.gif

The other way of looking at how information get's transferred through the retina is by looking at a cellular level. There are approximately 130 million light receptive cells in each retina; these can be sub-divided into central and peripheral portions (Hubel, 1963). These millions of cells can be split into two types: *cone* and *rod cells* (Figure 4). *Cone cells* are responsible for our central daylight vision and are placed mainly in the *macula* (c.f. Figure 2), an area that covers a 5° angle of our visual field and which can be referred to as our central vision. However, in the *fovea* (c.f. Figure 2), a small pit in the *macula* area of the retina, cone cells are placed even more densely and are responsible for 1° of visual angle of very high definition viewing. This is also sometimes referred to as central vision, however so as to not confuse things it shall be referred to as *fovic vision* in the context of this research. Rod cells cover the remaining retina and are responsible for our peripheral vision and are active in situations with low illumination, such as night-time.

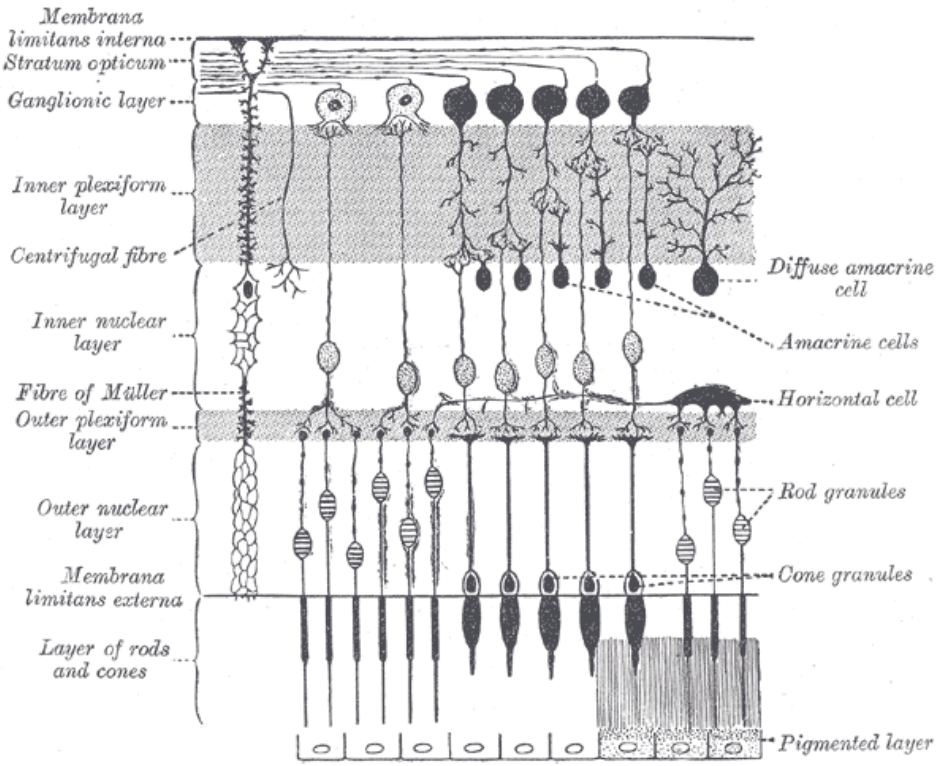


Figure 4: Cross-section of the retinal layers[2].

The journey that light takes after passing through the retina can broadly be described as follows: The *photoreceptor cells* receive light impulses and translate them into electrical impulses. These impulses then move through several layers of neurons until they reach the

[2] http://en.wikipedia.org/wiki/File:Gray882.png

ganglion layer, from here a collective information stream is passed to the *optic nerve*. The impulses are then carried through the *optic nerve*, until they reach the *optic chiasm*. At this point impulses that come from the *nasal area* of the *retina* pass over to the opposite *cerebral hemisphere*. In other words, impulses from the left *nasal retinal area* pass over to the right *cerebral hemisphere* and vice versa. Beyond this point the *optic nerve* becomes known as the *optic tract* and relays information to a subdivision of each *cerebral hemisphere* called the *lateral geniculate nucleus* (LGN) (c.f. Figure 3). From here the journey of light moves to the *visual cortex* with its various processing centres (v1, v2, v3, v4, v5) which are beyond the scope of this research (Zeki & Marg, 1993).

However, some basic observations on the actions of the *visual cortex* are of relevance here. One observation is that by far the largest part of the *visual cortex* is concerned with information received from the *fovea* (i.e., central high definition viewing), which is known as *cortical magnification* (Daniel & Whitteridge, 1961). A consequence of this is that the muscles around our eyes have developed to enable rapid eye movements from one high definition viewing fixation to another, in order to gain detailed information about a larger area. These eye movements are called *saccades* and will subsequently be described in detail.

Another observation about the visual cortex that is of relevance here is the concept of *saccadic suppression* (Matin, 1974), which entails that while our eyes are moving from one *fixation point* to another the visual cortex is not receiving information. A very insightful observation made by Dodge in 1900 describes this effect quite clearly:

`chanced on the observation that when the head was held perfectly still we could never catch our own eye moving in a mirror. One may watch one's eyes as closely as possible, even with the aid of a concave reflector, whether one looks from one eye to the other, or from some more distant object to one's own eyes, the eyes may be seen now in one position and now in another, but never in motion'' (Dodge, 1900, p. 456).

Understanding *saccadic suppression* has key relevance when the use of *saccadic eye movements* for interaction purposes is considered. Without saccadic suppression information could be given throughout an entire viewing session.

Eye movements are rotations of the eye ball. These are supported by six muscles called the *extraocular muscles*, which can be divided into two groups, the *rectus* and the *oblique muscles*.

The four *rectus muscles* account for movement in the horizontal and vertical planes relative to the scene being viewed. The *oblique muscles* rotate the eyeball around the *visual axis* in order to compensate for head movements, figure 5 (Land & B. W Tatler, 2009).

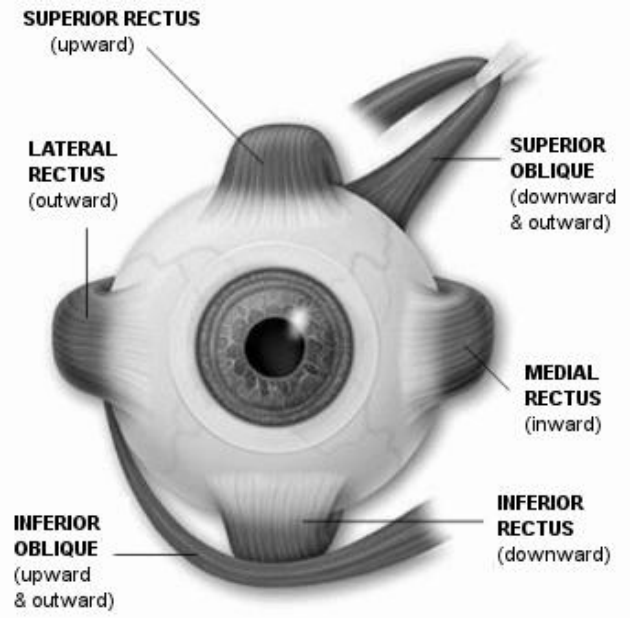


Figure 5: The Muscles of the Eye[3].

Eye movements can be classified as either *conjunctive* or *vergence* eye movements. Here, the term *conjunctive* infers the fact that the eyes move in *conjunction*, which means that the eyes move to the same degree in the same direction, when looking left or right. The consequence of this is that when looking left the *medial rectus muscle* of the right eye contracts along with the *lateral rectus muscle* of the left eye, the opposite is true for the right eyes *lateral rectus muscle* and the left eyes *medial rectus muscles,* which must both relax. The term *vergence* in eye movements covers situations in which the eyes move in opposite directions of each other, for instance inward (Sekuler & Blake, 1994). A good way of illustrating this effect is by looking at a finger held out at arm's length and then slowly moving it towards your nose without taking your eyes off it. By the time it reaches your nose your eyes will have *verged* and have turned inward.

The muscles surrounding the eye ball are incredibly durable and remarkably seem to be greatly spared in degenerative diseases such as Amyotrophic Lateral Sclerosis (ALS) where the individual is otherwise completely paralyzed (Ahmadi et al., 2010). In other instances

[3] http://www.cs.txstate.edu/~uj1001/pictures/eye_muscles.gif

where the state of Locked-in-Syndrome (LIS) is caused by brainstem lesions, the ability to move one's eyes is often also retained to some extent. Furthermore, it also plays an important role in diagnosing the level of cognition and subsequently communication with the individuals (Laureys et al., 2005).

2.2 EYE MOVEMENTS

There are only a limited number of ways the eyes can move and these are fundamentally the same for most human beings (Land & B. W Tatler, 2009). The three basic definitions used to describe eye movements are *saccades*, *smooth pursuits*, and *fixations*. However, other concepts regarding eye movements will also be described here.

2.2.1 FIXATIONS

The fundamental understanding of how our eyes work has been of interest and described in many different ways for centuries. An early understanding of *fixations* was given by Brown in 1895:

`We have before us not a moving panorama, but a series of fixed pictures of the same fixed things, which succeed one another rapidly'' (Brown 1895, page 5)

Today there is a greater understanding, but by no means complete, of the concepts of *fixations*, a term that covers the temporal state in which the eye is relatively still and *'fixed'* on a feature of interest. The use of the phrase 'relatively still' comes from the fact that the eye is constantly moving. The movements that occur during fixations are called *micro-saccades*, *tremors* and *drifts* (Carpenter, 1991) and they scatter around 1˚ of the visual angle (Young & Sheena, 1975). These movements ensure that the eyes are constantly moving and this in turn ensures that new information constantly passes through the retina. Experiments made with stabilized *retinal images* show that objects appear to fade or disappear entirely if the retina is held still:

'For optimal working conditions of the human visual system, some degree of constant (interrupted or uninterrupted) movement of the retinal image is essential. If a test field (of any size, color, and luminance) becomes and remains strictly constant and stationary relative to the retina, it will become and remain an empty field within 1-3 sec. (Yarbus, 1973)

This effect can be experienced by focusing on the central dot in the fuzzy grey field below (Figure 6). If gaze is fixed the boundary of the fuzzy field should disappear.

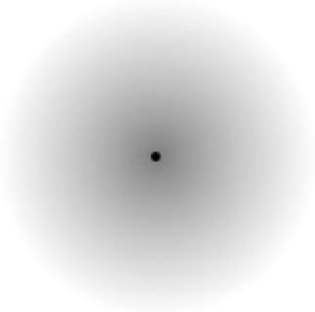


Figure 6: Fixation point, a prolonged fixation will cause outer detail to disappear, illustrating decreased visual input when the retina is held still[4].

During fixations highly detailed information from our central vision is processed in the visual cortex. However, one of the main challenges of eye movement detection and analysis is being able to threshold when fixations *begin* and *end* (Salvucci & Anderson, 2000). Generally, fixations are considered to last no less than 100ms and often between 200-400ms at which time the velocity of the eye is considered to be lower than 20˚/s (Salvucci & J. H. Goldberg, 2000).

Where fixations occur and how long they remain at a specific point is usually an unconscious decision, based on visual stimuli and objects of interest (Land, Mennie, & Rusted, 1999). In other words point of gaze is generally directed by interest and the need for detailed information. However, when designing for gaze interaction a cognitive barrier must be broken, as movement of the eyes needs to be a conscious and deliberate action.

2.2.2 SACCADES

Earlier a fixation was described as a 'relatively still' eye movement separated by *saccades*. Reciprocally *saccades* are the 'very fast' eye movements that are separated by fixations. Saccades are not simply any movements of the eye, for several reasons. First of all, the term

[4] http://www.cns.nyu.edu/~david/courses/perception/lecturenotes/retina/retina.html

eye movements usually cover any motion or lack thereof by the eye. Secondly, as mentioned micro-movements occur during fixations and the eye is therefore never completely still, and finally the eye can also move with a moving object, such as *smooth pursuits,* which will be subsequently described. So saccades are a very specific type of eye movement. A description of saccades before the term was coined can be found in Tscherning's 'Physiologic Optics: Dioptrics of the Eye' from 1900:

``...If, in a railroad train which is going quite fast, we fix a point on the window, the landscape appears confused, the images of its different parts succeeding one another too quickly on the retina to be perceived distinctly. Observing the eyes of any one who is looking at the landscape, we see that they move by jerks. The eyes of the person observed make alternately a rapid movement in the direction of the train to catch the object, and a slower movement in the opposite direction to keep the image of the object on the fovea...'' (B. W Tatler & Wade, 2003; Tscherning, 1900)

Saccades appear to be ballistic in nature. A reason for this is that the motion-path is predetermined and is generally a straight line (although curved saccades are possible) from one point to the next (Zeki & Marg, 1993). As mentioned, *saccadic suppression* occurs during these rapid eye movements and no (or very little) visual information is retrieved. Even though the world is only seen in increments, it is perceived as a smooth visual input. Saccadic suppression is a consequence of the brain not being able to cope with a constant stream of high definition visual input.

Saccades are reactions, forced or unforced, to visual stimuli perceived in areas of lower resolution than that of the fovea; physiologically this is the eyes' way of making up for not being entirely composed of cone cells.

Saccades can last between 30-120ms, and can reach speeds of up to 700˚/s for large amplitudes (Carpenter, 1988). The larger the amplitude the longer but also faster the saccade, one approximation of saccade duration in terms of degrees is 30ms for 5˚ of visual angle and 100ms for 40˚ (D. A. Robinson, 1964). Another approximation is 30ms for a 5˚ visual angle and then an additional 2ms per additional angle (Carpenter, 1988). The main problem with a strict definition of *saccadic speed* and *amplitude* is that it is very dependent on the equipment and protocol used to measure it (Duchowski, 2007). For the purpose of this research the latter

approximation is used. The overall visual angle of saccades is between 5˚ and 40˚ and the head will start rotating for eye movements covering more than a 30˚ angle (Young & Sheena, 1975).

2.2.3 SMOOTH PURSUIT

Where saccades move the eye from fixation point to fixation point – *smooth pursuits* can be described as a fixation-in-motion and occur when the eyes are following a target. The previous example from Tscherning (B. W Tatler & Wade, 2003; Tscherning, 1900)(B. W Tatler & Wade, 2003; Tscherning, 1900), regarding saccades when looking through a train window, can also be said to describe a smooth pursuit motion. The distinction is that a smooth pursuit is a continuous motion, where the eye rotates to follow a target. The reason for this is to avoid blur or saccadic suppression that would otherwise compromise visual perception.

'Under normal conditions, no smooth pursuit movements are possible without the presence of an object moving in the field of vision. Smooth pursuit can begin when objects move at speeds equal to those of the drift of the eye arising during fixation. Satisfactory conditions of perception are possible when the speed of the object does not exceed 100-200 deg/sec.' (Yarbus et al., 1973)

Smooth pursuits require a moving target and following requires a combination of both saccadic and smooth eye movements; the smooth motion stabilizes the object on the retina in order to retrieve stabile information, the saccadic movements perform minor corrections (Rashbass, 1961). For a long time it has been considered that saccades and smooth pursuits had very little to do with one another in the visual cortex. However, recent studies have shown that there are many overlapping features between the two processes (Richard J. Krauzlis, 2004). Usually smooth pursuits are split into two stages. First there is the *open loop pursuit* which is the initial motion to the target, this is saccade-like and usually lasts less than 100ms. After this comes the *closed loop pursuit* where the visual system is constantly trying to adjust the velocity of the retinal movement to that of the moving target (Duchowski, 2007; R. J. Krauzlis & Lisberger, 1994)

Fixations, saccades and smooth pursuits are for the purpose of this research the main distinctions between eye movements. How they relate to each other can be seen in figure 7 where an approximation of the eye movement velocities is shown relative to one another (San Agustin, 2009).

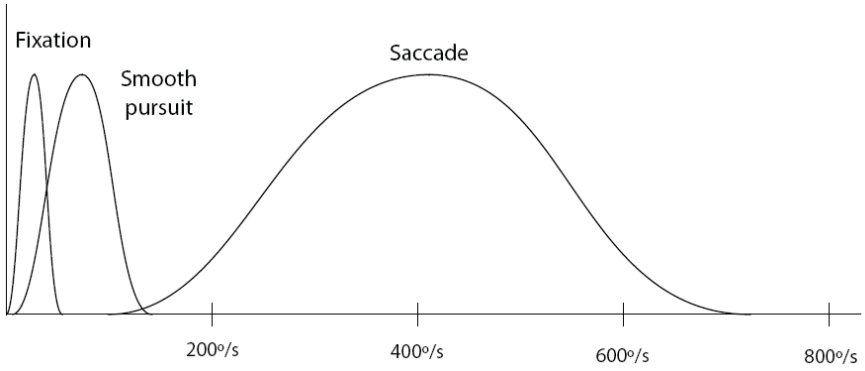


Figure 7: An approximation of the relative speed between fixation, saccades and smooth pursuits (San Agustin, 2009)

2.2.4 OTHER EYE MOVEMENT RELATED ISSUES

Other than these basic temporal states of eye movements, there are several issues regarding eye movements that are relevant when designing for gaze interaction. These are covered in the following section.

THE VESTIBULO-OCULAR REFLEX (VOR)

The *vestibulo-ocular reflex* (VOR) is a mechanism that has the primary function of turning the eye ball in the opposite direction to where the head is moving. The movement of the eye should be made with approximately the same velocity as the head movement; this is done to limit the movement of the world view image by keeping the direction of gaze into space approximately constant. In other words the position of gaze is made independent of head movement (Laurutis & D. A. Robinson, 1986). The importance of this understanding in regard to the research presented here is that the direction of gaze is independent of head movements.

NYSTAGMUS

Nystagmus is an involuntary eye movement where a smooth pursuit or fixation is broken by a involuntary saccadic action in the opposite direction; it is the *vestibulo-ocular reflex* attempting to regain image stabilization. The concept of *nystagmus* was originally defined in the context of involuntary eye movements which can be provoked by body rotation (i.e., spinning on a chair) (B. W Tatler & Wade, 2003).

There are many different types of *nystagmus* that can be a cause of, and therefore used to diagnose, various different types of conditions and diseases. Some of the conditions which can

have pathological *nystagmus* as a symptom are: Head Trauma, Stroke, Multiple Sclerosis, Brain Tumors[5] etc.. Many of the people who could benefit from gaze interaction have *nystagmus* as a symptom and it is therefore important to consider this irregular type of eye movements when designing interactive solutions.

ANTI-SACCADE

The *anti-saccade* is very different in comparison with the other eye movements described here, in that it is not naturally induced either by a healthy or damaged visual system, but requires conscious and deliberate effort. Saccades are, as mentioned, considered ballistic in nature, which means they follow a predetermined path in order to accommodate the brain with specific detailed information – in certain research traditions this type of saccade is also known as a *pro-saccade*. A different approach that has recently been studied in the field of cognitive psychology is the concept of the *anti-saccade* (Huckauf et al., 2005; Kristjansson et al., 2004). The premise is that the eyes can be forced to move in the opposite direction from where a visual stimulus is presented. The *anti-saccade* requires a counter intuitive action to be taken. This particular 'eye-movement' has mainly been explored as an interactive selection method that will be presented again in chapter 3.

2.3 EYE TRACKING

The process of tracking eye movements has been and can still be a very intrusive procedure employing mechanical objects being placed directly on the *cornea* (Figure 8). Quite often a trade-off needs to be made between the cost of the eye tracker, the precision of gaze estimation and the intrusiveness of the equipment.

[5]For a full comprehensive list of diseases http://en.wikipedia.org/wiki/Pathologic_nystagmus

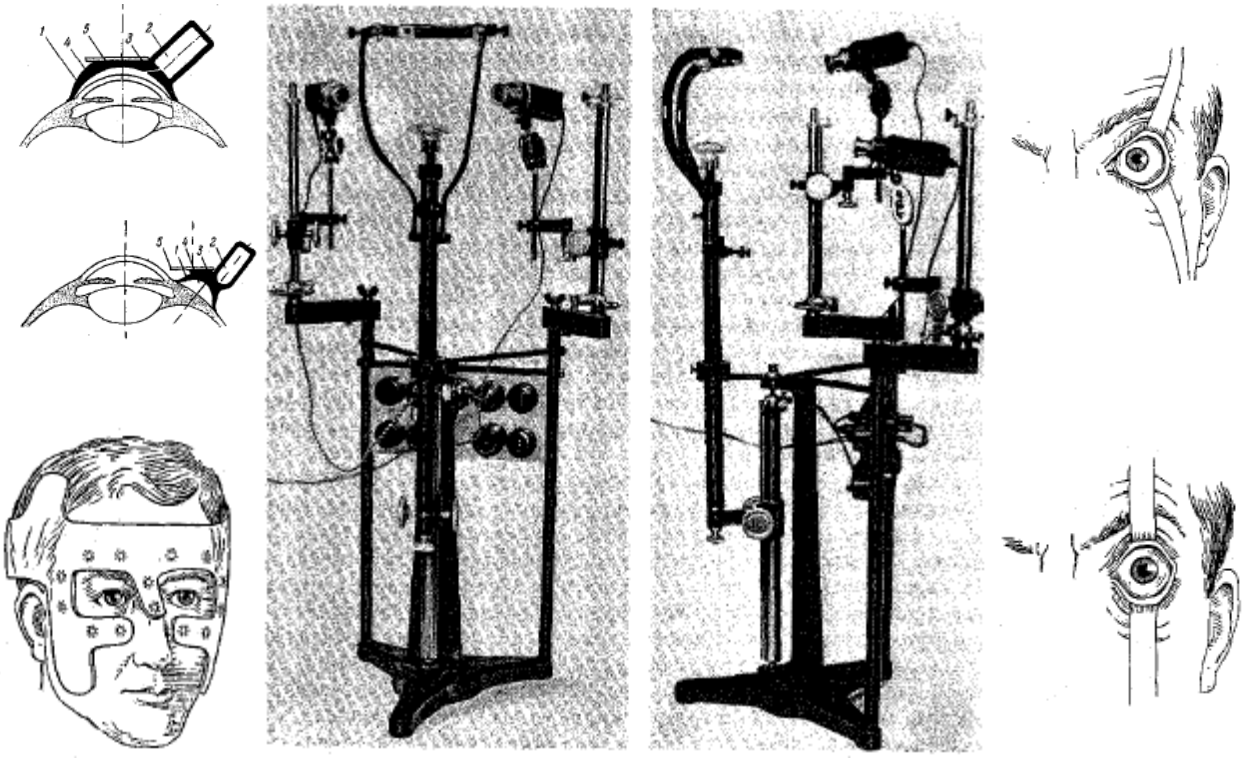


Figure 8. The instruments used in Yarbus's research a. small suction cups placed on the eye

Eye trackers are expensive because they require specialized lenses and specialized software. There is yet to be found a mainstream application for eye tracking which would propel it into the realm of main-stream production costs. As a consequence the price is, as mentioned, at least £6000-7000 and not everyone who needs an eye tracker necessarily has the means available to access one – the tradeoffs are between robustness, precision, cost and availability. A precise, robust, system can be developed at high cost, which is then only available to a few people (San Agustin, 2009). There is a great deal of interest in developing low cost eye trackers with off-the-shelf components (Hansen et al., 2004; San Agustin et al., 2010). This is doable; however, the tradeoffs in this case are between robustness, precision, comfort and price. A system can be developed which is cheap enough for most people to buy, but not precise, robust or comfortable. There are also trade-offs between intrusiveness and precision, by placing different lenses or suction device on the actual cornea – or by placing electrodes around the eyes – very precise reading of eye movements can be made (Young & Sheena, 1975). However, these systems are generally viewed as being too intrusive for everyday use.

Before the invention of eye tracking equipment, one of the earliest methods for examining eye movements was by means of an afterimage (Dodge, 1907; Helmholtz, 1925; B. W Tatler & Wade, 2003). A strong afterimage could for instance be induced by a short bright flash, which would leave a clear mark in a fixed position on the retina that could then be observed by the subject (Yarbus, 1973). Other early studies of eye movements were also done simply by visual observation. A participant would be asked to do a task while placed in front of a mirror. The researcher would then stand behind the subject and observe the movements of the eyes (Newhall, 1928; B. W Tatler & Wade, 2003; Yarbus, 1973).

The modern approach to eye tracking can be split into three: *electro-oculography* (EOG), *contact lenses* and *video-oculography* (VOG). The first, EOG, determines the direction of gaze based on movements of the muscles surrounding the eye. Electrodes are placed around the eye to detect muscle contraction and release. It is a relatively stable system which is invariant to head movements; because the electrodes are fixed to the head and therefore move with it. However, this same fact also means that these systems are viewed as being relatively intrusive (Young & Sheena, 1975).

Special *contact lenses* are still used in eye tracking research, somewhat similar to the Yarbus system. However, the lenses are usually connected to wires, making this approach very uncomfortable and intrusive to use; although newer versions can be wireless. This gaze tracking method is generally confined to laboratory setting (Roberts, Shelhamer, & Wong, 2008).

In *video-oculography* (VOG) a camera is used to record the eye movements of the user. Several different features can then be extracted and used to determine gaze direction. It is the least intrusive of the methods mentioned here. Gaze tracking systems for interactive puposes are generally based on VOG technology. It is this type of eye tracking equipment that has been used in the experimental setup of this research.

There are various different approaches to VOG eye movement detection. One is based on shape recognition. A geometric model of the eye and a similarity measure are used to detect the direction of the eye. The main descriptor features, which the pre-existing models build upon, are the *pupil*, the *iris* or the *corners of the eyes* (Figure 9) (Tian et al., 2000).

Figure 9: Feature point tracking using the iris and shape of eyes (Tian et al., 2000)

Another approach to VOG gaze detection is appearance based. In this technique a model of the eye or regions of the eye are built in real-time. A template of a feature can be created from hundreds of images; the method then seeks to match the template to the feature. There are also hybrid methods, where both shape and appearance features are used in combination in order to create a robust detection. Such combinations could be: Shape and intensity or colours and shape (Hansen & Ji, 2009).

Finally, many eye tracking systems use reflections of *infrared light* (IR) to detect and track eye features and to estimate gaze. IR light with a wavelength of 780-880 nm can be detected by specialized cameras. The reflection differs depending on where *infrared light* sources are placed in regard to the *optical axis* of the camera. The *optical axis* conjoins the different lenses. A light source located close to the *optical axis* of the camera is called *on-axis* light. The captured image shows a bright pupil since most of the light reflects back to the camera. This effect is similar to the previously mentioned red-eye effect that can occur when a flash is used on a camera. Reciprocally, when a light source is located away from the *optical axis* of the camera, which is called *off-axis*, the image retrieved shows a dark pupil (Figure 10). Detecting these features by employing thresholds is a common way of gaze detection (Ebisawa et al., 1989; D. W. Hansen & Ji, 2009; C. H. Morimoto et al., 2000).

.

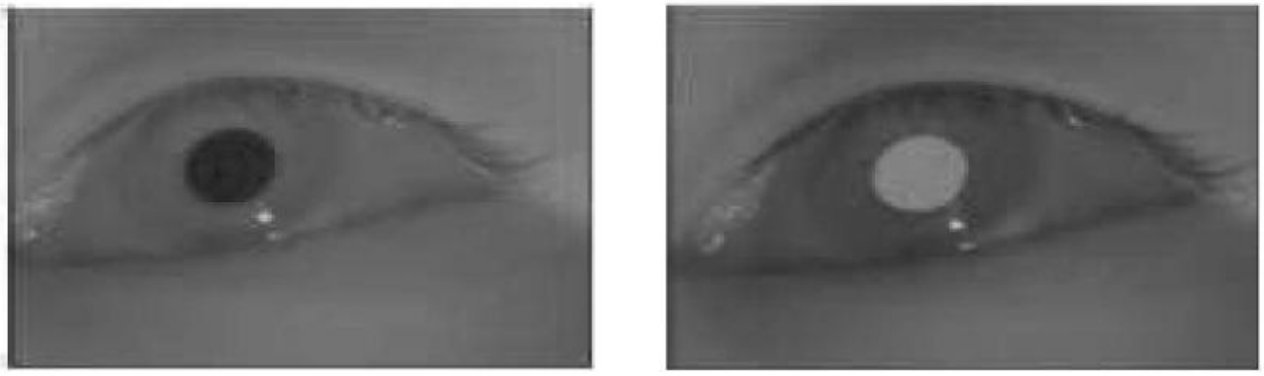


Figure 10: Dark and Bright pupil images (Hansen & Ji, 2009)

VOG is the least invasive method of gaze estimation and is therefore, as mentioned; also the most frequently used in gaze interaction research and commercial systems. Attempts are being made to make eye trackers based on VOG as accessible as possible. Low resolution gaze trackers have been created using standard digital video cameras (Hansen et al., 2004). These systems are not as robust as the commercially produced ones. However, if the accompanying software applications are designed to tolerate noisier input, then the systems work reasonably well (San Agustin, 2009; San Agustin et al., 2010).

The understanding that gaze selection strategies must be able to cope with noisy input, which can come from both the user and the eye tracking equipment, is important if the goal is to make gaze interaction accessible both from a practical and economical point of view.

2.4 EYE MOVEMENTS IN CONTEXT

Most of this chapter has concerned itself with the basic building blocks of the eye, eye movements and eye tracking. The potential combinations of eye movements are more or less endless. However, these have only been explored in a limited number of contexts such as scene viewing, reading and, only relatively recently, in action handling in a natural setting. This has been mainly due to the technological limitations that are only now being overcome. What is known is that eye movement patterns can be very different depending on viewing intent and context. A seminal piece of eye tracking research was done by Yarbus in the 1970's and illustrates this point well. The participants of the study were asked to look at the same picture with different intentions (Figure 11):

1. Free examination of the picture.
2. Estimate the material circumstances of the family in the picture
3. Give the ages of the people;
4. Surmise what the family had been doing before the arrival of the 'unexpected visitor":
5. Remember the clothes worn by the people;
6. Remember the position of the people and objects in the room
7. Estimate how long the "unexpected visitor' had been away from the family.

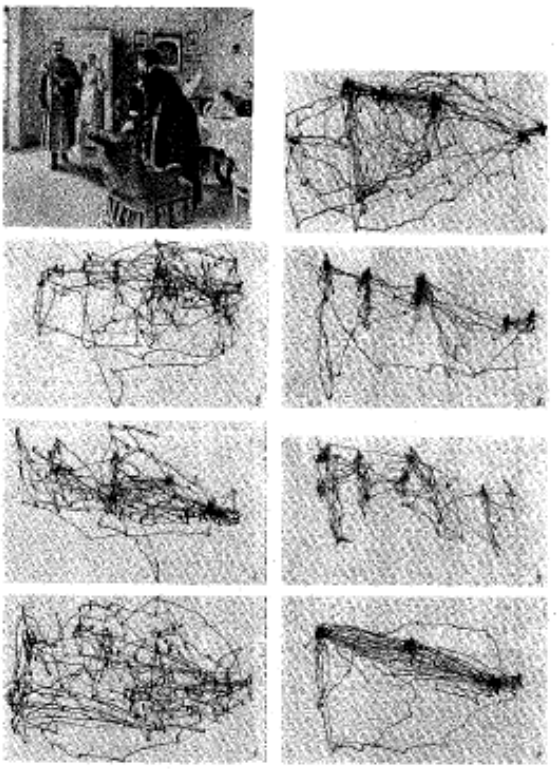


Figure 11: Eye tracking analysis performed on 'Repin's unexpected visitor' (Yarbus, 1973).

Not only are there great differences in the scan paths of searches with different intents, Yarbus went so far as to conclude that people who *think* differently also physically *see* the world differently to a certain degree.

'It may be concluded that individual observers differ in the way they think and, therefore, differ also to some extent in the way they look at things.'(Yarbus, 1973)

A task that has been substantially explored using eye tracking is reading patterns (Reilly & O'Regan, 1998; Starr & Rayner, 2002; Tinker, 1936). Generally fixations occur at every word although small function words are often skipped and saccade size depends on the specific nature task at hand (Kowler & Anton, 1987). In chess, it has been shown that advanced players chunk fixations and saccades together differently from novices, but how the information is being processed is not known (Chase & Simon, 1973). This further substantiates that cognition affects the way scenes are viewed or vice versa. More recently analyses of eye movements during motor-tasks in natural settings, such as driving or playing sports, have been explored (Hayhoe & Ballard, 2005; Land et al., 1999; Land & B. W Tatler, 2009). Each case uncovers an aspect of the intricate role which eye movements have in

cognitive functions, both in guiding *when* and *where* to fixate and in guiding body movements. One example of this is a study on eye movements during the process of making tea by Land et al. (Land et al., 1999) (Figure 12).

Tea making is meant to represent an automated routine activity. The conclusions are that even during habitual actions there is a high level of continuous visual monitoring that the subjects themselves were not aware of. Figure 12 shows the eye movements of three different people in the process of making tea. It shows that clustering of fixations are similar, however the saccade lengths and paths vary.

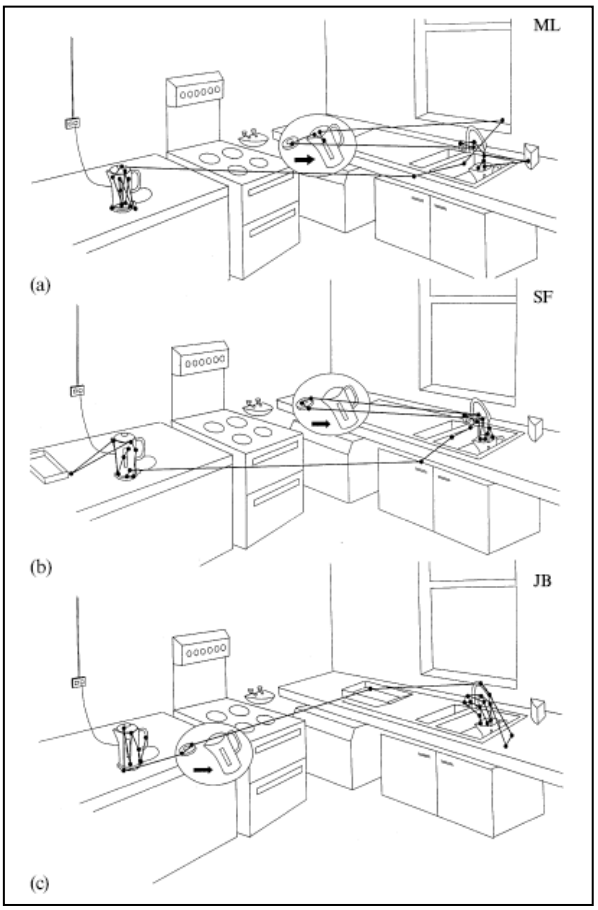


Figure 12: Eye movements during the routine process of making tea from (Land et al., 1999)

The experiment done by Yarbus revealed that saccadic eye movements reflect cognitive processes. *Scan-path* research has shown that not only do gaze patterns reflect decision-making processes – the complexity of the distribution of fixation points depends on the individual intent, context and situation. This is in stark contrast to the relative simplicity of the building blocks that these patterns are made of (i.e., the three temporal states). Generally,

it is the manipulation of the building blocks that constitute the starting point in interactive gaze research. However, the awareness that people see the world differently not just depending on what they are seeing but how they are thinking about what they are seeing is a useful perspective. Interactive gaze control constitutes a fundamental shift in the way the eyes are used and therefore it must also affect the cognitive and physiological processes of *how* and *what* is being seen. It is a challenge when designing gaze control to be able to find solutions to gaze actions that not only are physiologically possible to complete, but that also make cognitive sense.

There are two prevalent approaches in eye tracking research: *diagnostic* and *interactive*. *Diagnostic* is the registration and analysis of eye movements. The *interactive* implementation is a relatively recent addition that has developed as a consequence of the previously mentioned advances in technology.

2.4.1 DIAGNOSTIC GAZE TRACKING RESEARCH

Diagnostic eye tracking has had profound impact on how research has and is being conducted in the fields of cognitive psychology, developmental psychology, experimental psychology, media psychology and industrial design, HCI and usability research, neuropsychology, mental health disorders and ophthalmology. In this research gaze tracking is used as a means to an end, rather than being the end itself.

Yarbus' research showed sequential viewing patterns over specific parts of an image; where the participants focused on different aspects depending on intent. The main challenge in analytical eye tracking is interpreting the recordings of eye movements, because a correct interpretation relies on a *neuro-cognitive* understanding of the visual cortex, which must currently be considered incomplete at best (Zeki & Marg, 1993).

Around the same time as Yarbus presented his theory on contextual viewing, Noton and Stark developed the *scan-path hypothesis*, where they presented similar observations to Yarbus, but with a somewhat different interpretation. They claimed that the sequence of eye movements depended on the *visual pattern* being viewed. Furthermore, they proposed that the *visual memory* for this pattern also depended on the sequence of fixations with which it was viewed, so that the same sequence would be repeated next time the same or similar pattern was

viewed. Especially, this second postulate has been questioned. The fact that fixations tend to fall on or near important details of images is coherent with Yarbus. However, they also showed that even without specific guiding questions, identifiable regions of interest are still fixated upon. More recently the term scan-path has become a frequently used generalized term, without the theoretical notions implied by Noton and Stark (Noton & Stark, 1971a, 1971b). Whether vision is based on gaze directed by *intention*, *pattern recognition* and/or *memory* is beyond the scope of this present research. It speaks to the fundamentals of what, how and where of vision – in this regard Aristotle presented the following postulate:

'assuming, as is natural, that of two movements the stronger always tends to extrude the weaker, is it possible or not that one should be able to perceive two objects simultaneously in the same individual time?' (Hatfield, 1998)

According to Yarbus, and Noton and Stark the answer to this question must be no; it is not possible to perceive two objects simultaneously at the same individual time. However, the speed with which our eyes moves and our brain's interpretation of visual input *create* the impression of viewing multiple objects simultaneously. Even though much insight into the behaviour of gaze has been gained – the correct interpretation is still elusive.

The fundamental understanding of where and what people are looking at taps into thousands of years of philosophical reflection. This curiosity has been instrumental in the development of gaze tracking equipment for diagnostic use.

2.4.2 INTERACTIVE GAZE TRACKING

The technological advance in eye tracking has not only benefitted the field of diagnostic gaze tracking in human sciences. Since the beginning of the 1980's gaze interaction has been a subject for visionaries in human computer interaction; using the technology as a *real time input device*. Bolt (1981, 1982) was one of the first who began to conceptualize how the use of eye-tracking equipment could be implemented in real time (Figure 14). In 1981 he wrote:

'While the current application of eyes at the interface is in a sense a special case, eyes-as-output has a promising future in interactive graphics, especially where the graphics contemplates the possibilities arising out of the orchestrated use of videodiscs.' (Bolt, 1981)

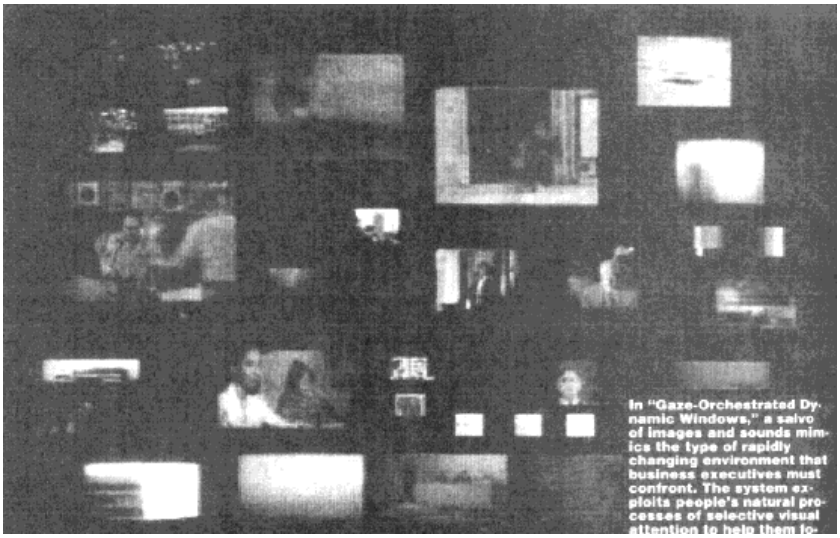


Figure 14: Gaze-Orchestrated Dynamic Windows,(Bolt 1981)

As mentioned previously, gaze as input has the problem of only having one *mode*; the eye tracker recognizes that someone is looking at the screen but not the intention behind the gaze. In other words the eyes are input that is always '*on'* (Midas Touch problem). This creates the risk of conflicting and unintentional selections, because the system cannot differentiate between the intentions of the user:

"Everywhere you look, another command is activated; you cannot look anywhere without issuing a command. The challenge in building a useful eye-tracker interface is to avoid this Midas Touch problem." (Jacob, 1991a)

Another problem in using the eyes as both an input modality and a receiver for feedback from the system is that these constitute two conflicting information streams crossing each other (Bates & Istance, 2002). In this case the potential problem is that it is difficult for the user to differentiate between receiving and giving information to the system. This fact needs to be taken into consideration in the design process.

Finally, two major problems with gaze in navigation and pointing are that users are not used to having objects, on-screen or otherwise, react to sight and that precision control of the eyes, which is required for *gaze selection* and *navigation,* is rarely used in the context of viewing a scene (Jacob, 1991a).

The need for robust, flexible, easily repeatable and sustainable gaze-based target selection remains, as do the difficulties of interactive gaze. The main solution to the issues surrounding eye tracking as input have been *dwell-time selection*; e.g. by dwelling or fixating on a specific object that the system can recognize and that is an intended action by the user(Bates & Istance, 2002). In the following chapter previous work performed in the field of *selection strategies* will be presented in a taxonomy, which has been derived from an analysis of existing methods.

3 GAZE INTERACTION

Gaze interaction deals with using the direction of gaze to navigate and/or make selections in information spaces. In the following chapter, various aspects of gaze interaction are explored; specifically the focus is on how gaze as input can be interpreted. In order to contextualize this some considerations on how input can be interpreted generally in the field of Human Computer Interaction (HCI) will be presented.

Distinctions and definitions of parameters which constitute gaze interaction will be presented throughout this chapter. This is intended to serve as a contribution to a general understanding; as well as to establish why the subsequent experimental work has been done.

One of the main distinctions made in this research is between gaze interaction as an *addition* to existing input modalities and gaze interaction as *sole* input. The reason for this distinction is that the design parameters and applicability of these two input strategies are very different. Gaze as an *addition* mainly deals with the enhancement of navigation processes. Gaze as *sole* input needs to be able to handle navigation and selection processes.

Current gaze selection strategies for sole input can be divided into two main approaches; *dwell-time activation* and *gaze gestures*, both of which seek to circumvent accidental selections. Dwell-time activation consists of prolonged fixations on on-screen targets. Gaze gestures consist of recognisable and repeatable eye movement patterns that can be distinguished from natural search patterns. A taxonomy for dwell and gaze gesture interaction will be proposed in the course of this chapter.

3.1 HCI AND GAZE INTERACTION

HCI has conventionally relied on the motor-skills that our hands and fingers provide. Input devices such as keyboards, the mouse, joysticks, gamepads, touch-screens, etc. have dominated the field (Plaisant et al., 1995). However, as technology, tasks and data structures become more complex and ubiquitous in nature, the need and interest for alternative input increases.

'We can view the basic task of human-computer interaction as moving information between the brain of the user and the computer. Our goal is to increase the useful bandwidth across that interface with faster, more natural, and more convenient communication mechanisms'. (Jacob & Karn 2003)

There are many situations where conventional input devices are inadequate or inappropriate and alternative inputs can be necessary. When researching interaction techniques several parameters affect each other and should be taken into consideration when determining the most appropriate input, selection strategy and visualization. The three basic parameters that can be considered are *task*, *user context,* and *feedback*.

In regard to *tasks* different approaches need be employed when dealing with search-tasks in graphic data representation as compared to searches in purely textual data, or 2D environments versus 3D environments. The *user context* determines not only the necessity for an alternative input device, but also what type is applicable. Physical impairment constitutes a context in which the choice of input device can be a life changing necessity. In other cases, the user might simply be dextrally limited while retaining the need to interact with a system, as mentioned earlier, a doctor performing surgery or a soldier managing weapons. In these cases alternative input can constitute an appropriate choice rather than a necessity. Finally, *system feedback* determines what type of input could be used. There is a vast difference between a user working with the small screen of a mobile device or a large digital whiteboard. Feedback can also be *auditory*, *visual,* or *tactile* which each have a set of affordances and constraints that need to be considered.

There are numerous types of alternative input devices and as many uses. The focus has been on mapping and aiding the natural interaction and communication between human-to-human and from human-to-computer. Several research fields contribute in this regard: ergonomics, HCI, Interaction design and CSCW (Computer supported collaborative work) (Smidth & Bannon, 1992). Overall, the aim has been to broaden the gateway of communication from man to machine so that pervasive and ubiquitous computing can be supported through, e.g., voice recognition, head movement controls, eye-tracking, etc.

As previously mentioned gaze interaction has some inherent issues that entail the need to make a system aware of the intentions behind gaze. In a gaze-controlled interface additional information is needed to inform the system whether our eye movements are an indication of

us wanting to select something or simply observe. The question of selection in gaze-controlled interfaces has, as mentioned, been discussed as far back as in 1981 by Bolt when describing the implementation of the "World of Windows"-display technique:

' There are at least two competing philosophies about what should cause you to be "zoomed-in" upon some window you are looking at: 1) zoom-in automatically, based upon some timing-out of how long you look at ("stare at") some certain window; 2) zoom-in upon the window you are eye-addressing contingent upon a deliberate action via an independent modality (e.g., joystick action; or a word spoken to the system)' (Bolt, 1981).

Recurring themes in the field of ergonomics and HCI are the concepts of intuitive or natural and whether they apply to a given interaction modality or interface solution. Although both terms seem self-explanatory care needs to be taken when using them. In his book, The Humane Interface, Raskin discusses these terms, and proposes using *familiar* or *easily learned* instead:

'Many interface requirements specify that the resulting product be intuitive, or natural. However, there is no human faculty of intuition, as the word is ordinarily meant; that is, knowledge acquired without prior exposure to the concept, without having to go through a learning process, and without having to use rational thought' (Raskin, 2000)

The concern with naturalness of interaction, while interesting and useful as a design parameter, might simply be overrated. Interaction with a mouse is often considered a natural form of interaction.

The assumption is that everyone innately knows how to use the mouse. Raskin tells a story of a Finnish scientist and her first encounter with a mechanical mouse. She begins by turning the object on its head and rolling the ball around, as with a track ball. This has no effect, as the ball is pushing down on all the detectors, thereby cancelling the effect. She discovers this and turns the mechanical mouse the right way up, but instead of placing it on a surface she continues to control the mouse manually, moving the ball with her fingers. When shown how to use the mouse she quickly understood the mapping – easy, but not natural (Raskin, 2000).

In the case of using gaze to gather information the concept of *natural* could very well be applied, but any attempts at using gaze direction for navigation or activation can at best be described as *easily learned* or drawing on a *familiar* analogy. In fact, only between humans

does a ‘natural activation’ occur; as looking at someone is a way of conveying interest and often warrants a response. While it is natural to look at objects of interest before and during interaction, in nature inanimate objects do not respond to gaze.

3.1.1 GAZE AS AN ADDITION TO EXISTING INPUT MODALITIES

An example of exploiting gaze in the context of human-to-human interaction was done by Vertegaal et al. They looked at multiparty mediated communication and collaboration, in other words on-line conference systems. Gaze direction was used as a visual cue to detect who was talking and listening to whom. They concluded that gaze directed cues can ease some of the user ambiguities that exist in these systems (Vertegaal 1999)(Figure 15).

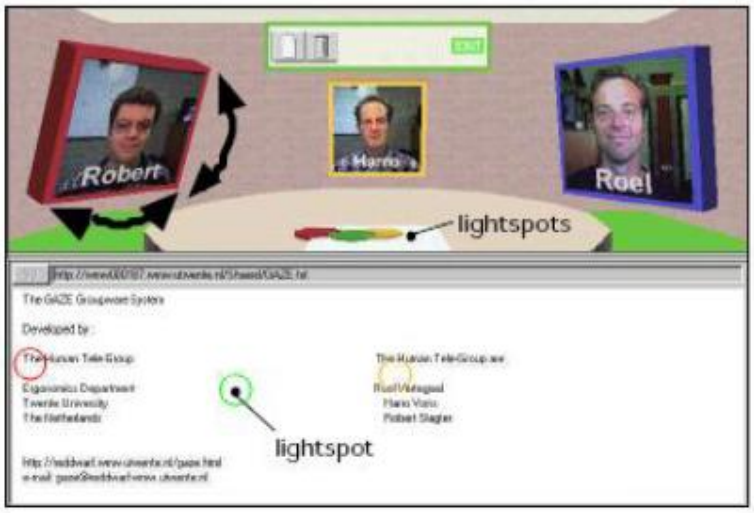


Figure 15: Conference system that rotates depending on gaze direction (Vertegaal 1999).

The limitations of gaze interaction have been the foundation for a sub-field of research within gaze interaction, which regards gaze as an addition to an existing input rather than as a sole input. This was the focus of a study completed by Zhai et al. (1999) in which the possibilities for using the eye tracker as a supplement to the mouse, rather than a substitute, were examined. The concept of MAGIC (Manual and gaze input cascaded) pointing was implemented. The basic idea was to maintain selection and detailed targeting as motor-controlled tasks, only using gaze indirectly by allowing the cursor to follow the point of interest (gaze), so that the cursor would be close to an intended target before the selection process was initiated.

They implemented two different interface and interaction designs, one liberal and one conservative (Figure 16). In tasks, subjects were asked to point and click on targets, which randomly appeared on the screen. The liberal pointer placed the cursor within the vicinity of every object looked at by the user; the user could then decide to select or ignore the target and move on to the next. In contrast, the conservative pointer would not move to the target until the manual input device (a miniature pointing stick as used in laptops) had been set into motion.

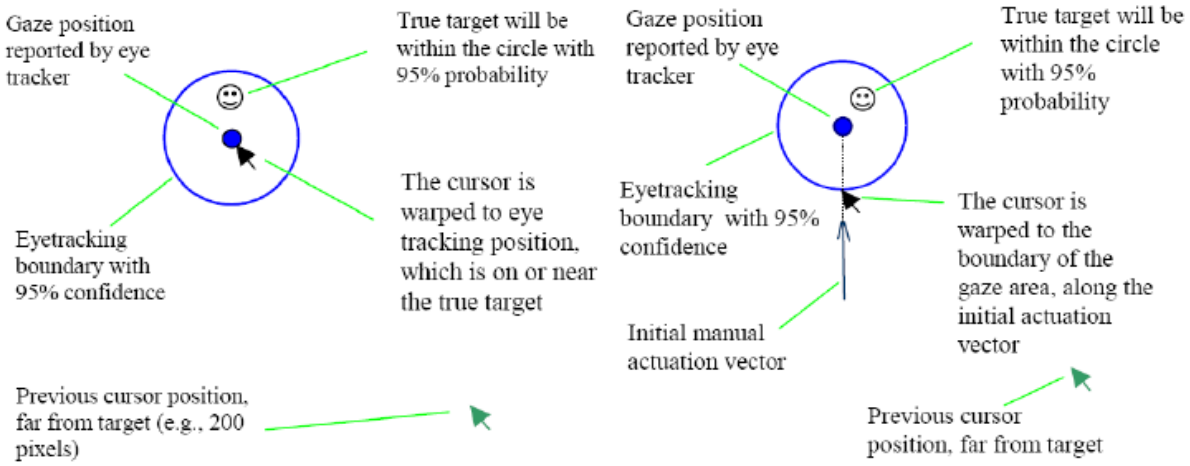


Figure 16: The MAGIC pointer: the liberal MAGIC pointer (left) the conservative MAGIC pointer (right)(Zhai, C. Morimoto, & Ihde, 1999).

The results showed that the liberal pointer was slightly faster than the completely manual input, and the conservative pointer slightly slower than the manual input. Overall, the conclusion was that gaze pointing as a supplement to manual control showed potential.

The advantage of using gaze in navigation is that it extends existing behaviour of orientation and location of *area of interest*. As presented in chapter 2, the muscles that control the eyes offer near fatigue-free pointing (Bates, 2002). Gaze interaction could therefore potentially be of assistance to people who suffer from repetitive strain injuries (Zhai et al., 1999).

Qvarfordt et al. (2005) based their gaze interaction research on observations of eye-gaze patterns in human-to-human dialogue over spatial content. They envisioned that gaze could be an important contributing channel of input in future of multimodal HCI systems. Their iTourist system provided users with information for city trip planning, based solely or primarily on gaze patterns. The system could manage its visual and audio (speech) output and help users to plan an entire trip to a city, including finding hotels and restaurants, only based

on eye movement patterns, most of the time. They reported on erroneous selections and overlaps, but concluded that eye movements combined with speech input could solve these issues (Qvarfordt & Zhai, 2005)(Figure 17).

Figure 17: The Iturist system, the image depicts a user navigating a city map with gaze(Qvarfordt & Zhai, 2005)

Another familiar way of mapping vision to interface control has been by combining gaze with zoom/pan navigation. Bederson et al. describe the motivation behind the zoomable interface of Pad++ as: "tapping into people's natural spatial abilities" (Bederson et al., 2000). The idea of coupling gaze with zooming actions has been explored since eye tracking was first conceived of as a real time input device (Bolt, 1981).

Zoom permits the user to access any level of detail directly from a global view, without the need for entering sub-menus. The geometric integrity of zoom means that proportions within the information space are upheld. The familiarity of visually approaching an object by zooming extends well to gaze interaction. The argument is that the way navigation naturally occurs is by gaining perspective and subsequently focusing in on targets of interest. This therefore makes zooming a familiar mapping of natural navigational.

'In natural scenes, new objects rarely appear abruptly; they are more likely to appear by progressive disocclusion from behind other surfaces (J. J. Gibson et al., 1969).' (Franconeri & Simons, 2003)

When coupled with panning another potential weakness is avoided. Zoom without pan does not enable the user to explore the immediate context of an equally scaled detailed area

without leaving the particular scale level and zooming out and in again. Pan can be defined as planar movements, allowing navigation both horizontally and vertically in the information space. Panning is a well-known form of navigation in information spaces, in regard to its most common implementations (Plaisant et al., 1995)

GaZIR is an example of a gaze interface that allows browsing and searching for images based on implicit feedback. In this system the images that are predicted to be of most relevance are brought out when the user zooms in (Kozma et al., 2009)(Figure 18).

Figure 18: GaZir Image retrival interface, images are presented in consecutive circles (Kozma et al., 2009)

'Implicit feedback should not suffer from the same problem— the intent is not that the user controls the system with eyes, but instead that information is extracted from natural eye movements.' (Kozma et al., 2009)

This issue of directing zoom based on gaze direction, taking advantage of natural search patterns, was also the topic of the author's MSc Thesis. In this work, search tasks were conducted in a multilayered information space using pan and zoom (Mollenbach et al., 2008). It was found that by allowing pan and zoom direction to be controlled by gaze users could successfully navigate through thousands of nodes and find the designated target (Figure 19).

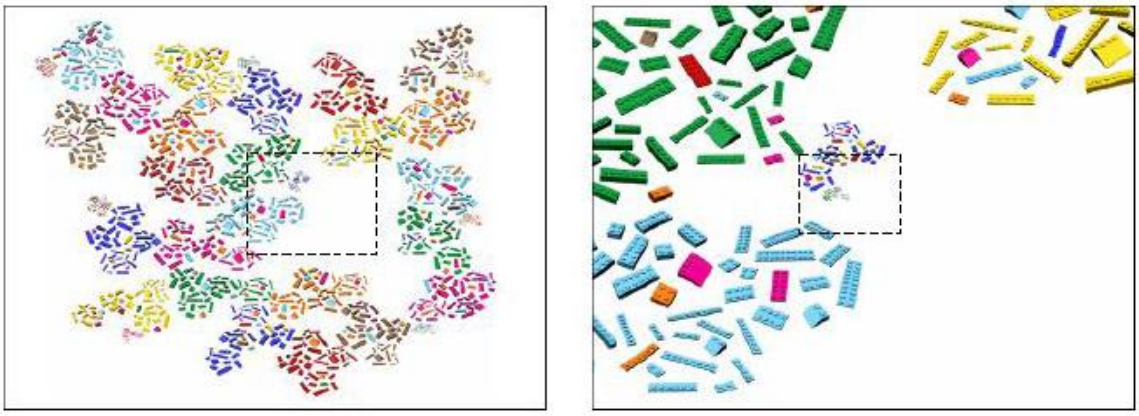


Figure 19: A Zoomable Gaze Interface. The image to the left depicts the overview of nodes. The image to the right shows the process of zooming in on different information layers (Mollenbach et al., 2008)

The research undertaken in the area of gaze as additional input gives insight into both benefits and problems. The main benefit of using gaze in navigational tasks is that it can tap into an existing process. Visual search already occurs as a fundamental part of interaction with visual information spaces. Establishing a visual link between the system and the user therefore broadens the pathway of information flow.

In the examples shown, gaze can be used explicitly to manipulate information spaces. However, research also needs to be done on how gaze direction could be used implicitly to direct information without conscious user involvement, for instance in computer games and web interfaces. Another benefit of gaze as an addition to other input modalities is that it can be implemented as a discrete control. This means that gaze control can be switched on and off depending on context of use and the problems regarding gaze always being 'on' are avoided.

The challenge of gaze as additional input is that, in order to gain the full advantage of employing search patterns, new ways of visualizing and structuring information are necessary. Most existing information structures and layouts are built to accommodate existing input. By changing the construct of input, the relying structures must also change. The incentive to do this will not exist until gaze tracking becomes a mainstream input modality, so currently much of the potential of gaze in conventional HCI is still unexplored.

Gaze input as an addition to existing input devices constitutes a related but ultimately different area of research to the one at the foundation for the empirical experiments that will be presented in subsequent chapters. Gaze as sole input is, as mentioned, a different premise,

from gaze as an addition, and entails the main distinction of especially requiring selections while still allowing navigation and perception to occur through the same channel. However, understanding the parameters of *task*, *context* and *feedback* are the same. Likewise, considering ways of creating gaze selection that build on the mapping of familiar actions also applies.

3.2 GAZE SELECTION STRATEGIES

On a basic level there are three things that need to be taken into consideration when designing for gaze navigation and selection. Most existing implementations cater to one or more of these uses and are: *intention*, *action*, and *perception*.

Intention is, in this case, defined as the perceived interest of the user. When used implicitly it can afford cognitive relief and emergent interaction. In this case it is not necessarily based on directional pointing and also the user feedback comes in the form of subtle output. An example of this could be a webpage that tracks where the user is looking and presents an ad or application data directly within the visual field. When applied explicitly it enables application control, as the *intention* of the user can be translated into *action*.

Action is an explicit use of the intention of the user to map the point of gaze to a direction or selection; for example panning based on gaze direction, dwell-time selection or gaze gesture selection. Implementing these actions should always address the Midas touch problem for dwell- time selection and *accidental gesture completion,* which is the gaze gesture equivalent of the Midas touch problem and will be elaborated upon later.

In this context *perception* speaks to the user's ability to perceive on-screen information. This entails that large quantities of information should be made available if necessary for the task at hand. The user should be able to navigate and perceive the information without any sense of urgency.

One way of accommodating intention, action and perception in gaze-only interfaces is by employing dwell-time selection.

3.3 DWELL-TIME ACTIVATION

The Midas touch problem has traditionally been solved by implementing dwell-buttons on the screen. These function by having the user maintain a fixation in a certain area for a prolonged period of time in order to complete a selection. Fixation activations can be set to last anywhere from 200ms (milliseconds) to 1000ms (Ware & Mikaelian, 1986; Jacob, 1991b; Majaranta et al., 2009) or more, depending on the abilities and preferences of the user. To compensate for the inaccuracy of micro-movements, tremors, potential nystagmus and gaze tracking equipment, these on-screen buttons usually, as mentioned, take up quite a lot of screen space. This is a disadvantage as it limits the amount of information that can be displayed in applications. Also dwell selection sets a time limit on both the user's ability to inspect/navigate an interface and on the speed with which selections can be made.

One of the first evaluations of eye tracking and dwell-time came from Ware in 1986. Three types of on-screen selection techniques were tested. First of all dwell-time selection, where the target was selected by having the user fixate on it: "After experimentation, we adopted a dwell-time of 0.4 seconds.'' (Ware & Mikaelian, 1986). Secondly, the subjects could make selections by looking from an object of interest to an on-screen area that functioned as a button, thereby requiring multiple fixations. Finally, the subjects could complete a selection by fixating on an object while pressing a physical button. Two experiments were conducted; in the first test a structure resembling a menu was employed, while in the second one the subjects were asked to select targets of different sizes. The results of the first experiment showed that dwell-time selection was equal to hardware selection in terms of speed and with lower error rates in the dwell- time condition. The second experiment proved the hardware button to be faster, but dwell-time selection was more consistent and had lower error rates. The on-screen button was by far the most ineffective in both tests. The overall conclusion of this test was that eye tracking showed definite potential.

Jacob presented a study in 1991 where dwell-time was compared with eye pointing and button selection. The test bed was intended to resemble icons on a desktop that the participant then had to select. Mainly findings and observations were presented and a somewhat fast dwell selection time was proposed: '*150– 250 milliseconds dwell time gives excellent results*' (Jacob, 1993)(Figure 20). Statements in this paper name dwell-time as the

'natural' eye movement equivalent of a button-press and has greatly influenced the past two decades of gaze-related interface and interaction design.

The point of view taken in this thesis is that the term 'natural' should not be used in conjunction with dwell-time activation or gaze interaction in general, even though this has been described as one of the main advantages (Bates, 2002). The reason for this has previously been stated and the appropriate terms used in conjunction with gaze interaction should be familiar or easily learnt.

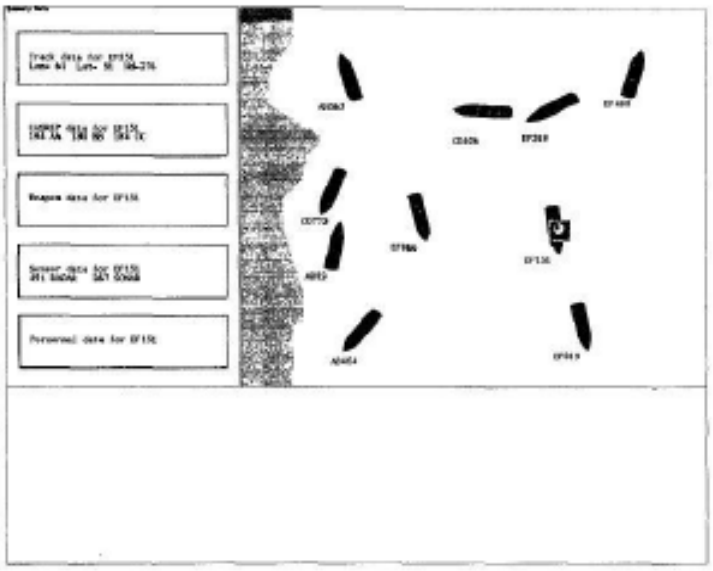


Figure 20: Each boat on the figure represents a target which could be selected by gaze (Jacob,1991)

Regardless of the preliminary findings of Jacob the time range of dwell-time is usually anywhere from 450-1000ms (Qvarfordt & Zhai, 2005), because shorter dwell-times in complex interfaces cause conflict between *intention* and *action* (Midas touch); as a consequence these long fixations can become tiring for users.

The experiments presented above represent some of the fundamental research that constitutes the original methodological approach taken in investigating gaze interaction. The methodology has evolved somewhat from this fundamental approach to being more application driven. Gaze interaction research has in more recent years generally been done in conjunction with solutions to specific tasks, such as type-to-talk applications (Hansen, 2006; Majaranta & K. J Räihä, 2002), environmental control (Shi et al., 2006b, 2006a) and games control (Istance et al. 2008; 2009a; 2010). This has been a consequence of the fact that people

with motor impairments are the main beneficiaries of breakthroughs in this technology, and the intention has been to provide usable software wherever possible. The approach taken in the research presented here is, however, more reminiscent of the fundamental approach. The reason behind this is that as activation strategies evolve there is still a need for a fundamental understanding of these basic building blocks. In the subsequent sections, selection strategies will be presented in conjunction with the applications with which they have been developed.

Figure 21, sums up some of the principles that have been used to implement dwell-time selection. A mixture of the principles is generally used when implementing dwell-time. These main principles are *selection complexity*, *visual feedback* and *activation duration*.

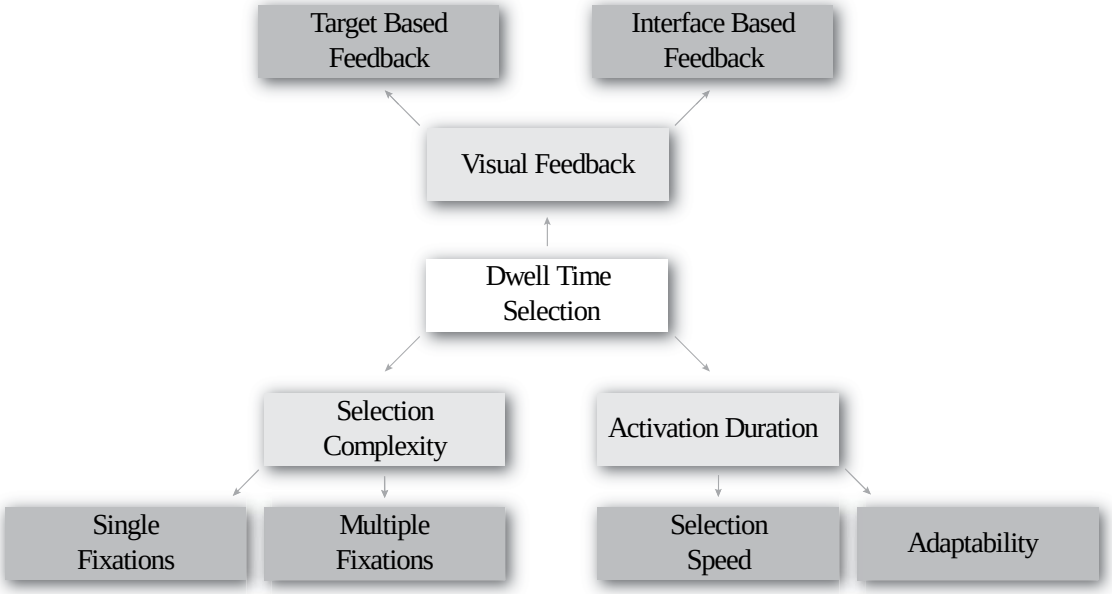


Figure 21: Principles used in the implementation of dwell-time activation

The major advantage of dwell-time buttons is the fact that they rely on the button metaphor that familiarly maps a known form of interaction and visualization. As a consequence recognizable information structures can be created. Dwell-time selection has been used to control applications such as gaze typing systems, computer games, environmental controls, and music programs (Hansen et al., 2006; Shi et al., 2006b, 2006a; Istance et al., 2010; Ward et al., 2000).

Selection complexity speaks to whether a *single fixation* or *multiple fixations* are required to complete a selection. *Visual feedback* deals with how the dwell duration is relayed to the user. This can either be displayed on a dwell-button or the dwell duration can affect the entire

interface, by shifting it towards the target (i.e., zoom). *Activation duration* is a fundamental aspect of dwell-time selection. The speed with which a target can be selected and how this speed can be adjusted are intricate parts of implementations.

As new ways of implementing dwell-time are still being developed, it is intended that this taxonomy could be expanded upon. This taxonomy has been created by analysing existing implementations of dwell-time selection. In the following each of the principles will be further elaborated upon using practical examples.

3.3.1 SELECTION COMPLEXITY

Selection complexity of dwell-time selection speaks to whether activation occurs on the initial fixation or whether multiple fixations must occur.

One type-to-talk system that has been influential and is being used by many end-users is called GazeTalk (Hansen et al., 2006)(Figure 22). This application was specifically developed for people with severe motor impairments and was designed for and with a group of ALS patients in Denmark. The previously mentioned ALS patient Arne Lykke Larsen, who is still working after more than 10 years with ALS, uses this system with a speech synthesis to supervise students, give talks and generally conduct conversations with whomever he wishes.

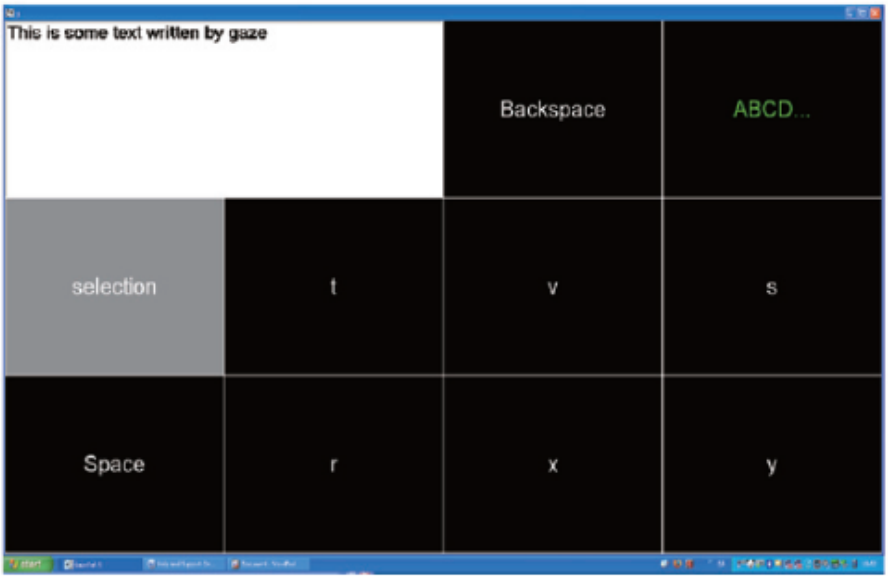


Figure 22: The GazeTalk interface, the image shows the user choosing the predicted word 'selection' in the mid-left side of the screen to complete the sentence that is displayed in the top-left field (Hansen et al., 2006)

GazeTalk is an example of a system built on *single fixation* selections. The system allows users to communicate by spelling out each word. Letters are initially selected in stages. Firstly, groups of letters are displayed in a grid of dwell-buttons. When one of these buttons is selected the individual letters of the particular group are displayed in a grid of dwell-buttons. The intended letter can then be selected. After this the user can select another letter from the same group or return to the main screen and select a new group of letters. Word prediction, which is displayed on a dwell-button of its own, allows for speedy completion of words. Dwell-time can be adjusted, however not in real time. It was found that:

'Most users start with a dwell time setting at more than one second, but some experienced users can control dwell time settings that are less than half a second. Most users like to have longer dwell times when they are tired in order to avoid unintended selections.'(Hansen et al., 2006)

An example of a *multiple fixation* activation process was given in NeoVisus (Tall, 2008) (Figure 23). At the initial fixation an expansion of the target would reveal a second target, which had to be subsequently fixated upon in order to complete a selection. The intention of this was to limit potential erroneous activations by not allowing any selection to occur on initial fixation. If the user looks away from the dwell-button in all other directions than that of the 'rolled-out' second dwell- button, the system was reset and the second dwell point is retracted.

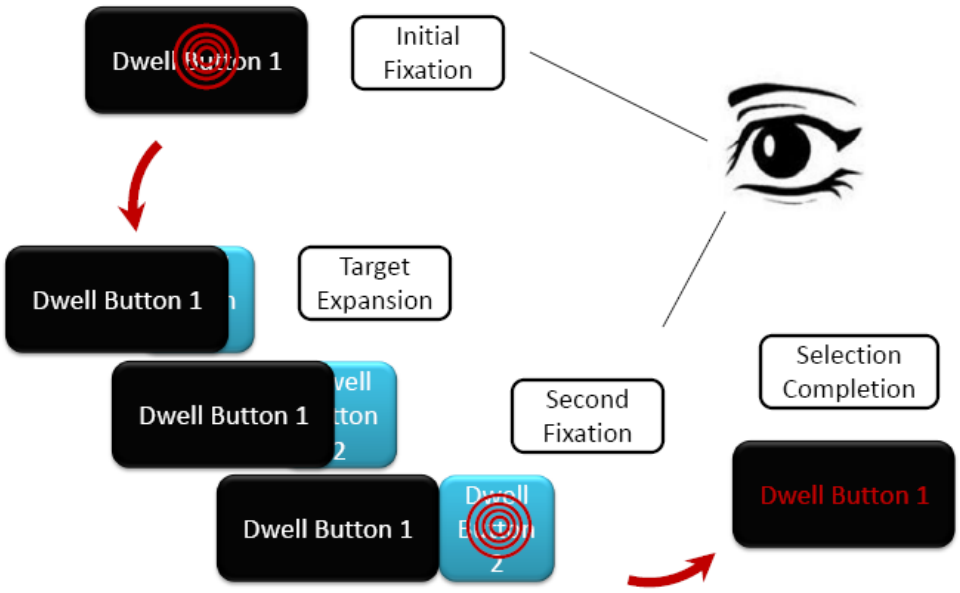


Figure 23: Adapted from NeoVisus, gaze selection components. A short dwell duration leads to the expansion of a second target. By dwelling on the second target the selection can be completed (Tall, 2008)

Whether single or multiple fixations are appropriate depends on what *task* is being performed, what *context* the user is in, and the *feedback* opportunities that the system can provide. The duration of single or multiple fixations end up as being approximately the same; shorter durations can be applied when using multiple fixations, but then more of them have to be made. If the buttons are of a significant size, single fixations require less precision from the system and possibly they also represent a lower cognitive load for the user. However, the single fixation activation does not solve the Midas touch problem, it simply delays it. The multiple fixation activation circumvents this issue, in that a combination of fixations is required, rendering the initial fixation of no consequence.

Selection complexity can be surmised as the difference between *one* action having *one* consequence and *several* actions having *one* consequence. As mentioned the empirical work done in the context of this thesis concerns itself with *single stroke gaze gesture* (SSGG) which can be compared to the single dwell duration, as they both represent *one* action with *one* consequence. The discussion on selection complexity is also very much present in the area of gaze gesture interaction and will be explored shortly.

3.3.2 VISUAL FEEDBACK

Visual feedback is concerned with how the activation process is presented to the user. For instance, sometimes the visualization is fixed to the target, other times the entire interface moves towards the target by zoom, which can in turn be discrete or continuous.

In the GazeTalk system (Hansen et al., 2006), which was introduced earlier, the visualization of dwell occurs on the target itself: a light shaded border appears around the target that is being viewed and begins to contract from the border towards the centre of the target. A selection is completed when the lighter shade fills out the entire target[6]. If at any time the subject looks away from the target, the selection process is discontinued and either the same or a new target can be selected. This feedback allows the user to be aware of what he or she is potentially selecting. Variations on this type of target-based feedback are used in many

[6] http://www.youtube.com/watch?v=uu5r_HVJT_w

applications and systems. For instance most of the applications for the Tobii Communicator system, which is a commercial eye tracking system, use this type of feedback (figure 24)[7].

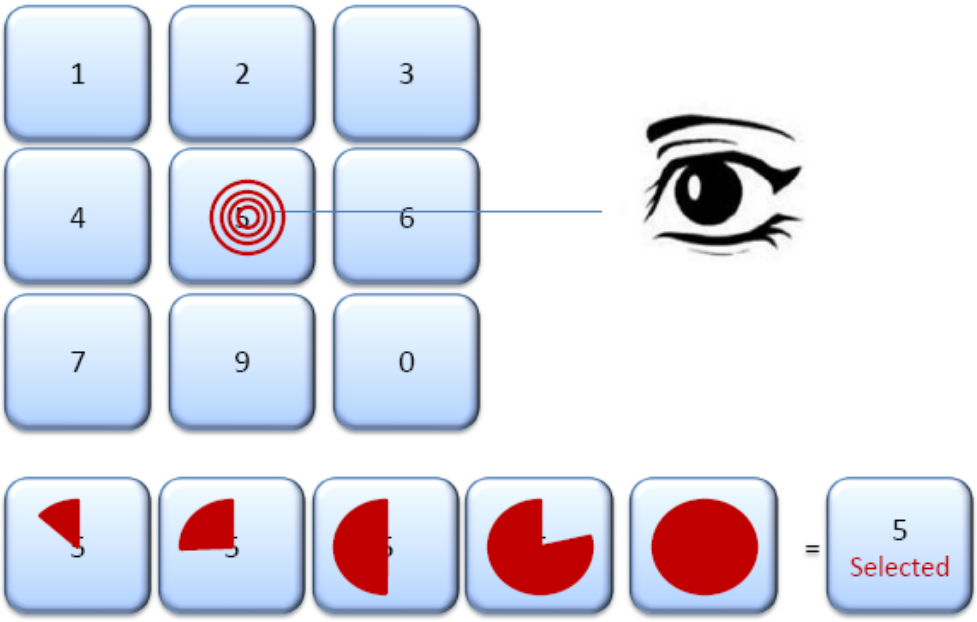


Figure 24: Windows Adapted from the Tobii communicator. The target '5' is dwelled upon and a red circle begins to expand as shown in the bottom sequence. When the circle is full the selection process is completed.

As mentioned earlier, directing zoom by gaze has been explored as an approach that combines natural search patterns with other inputs from the mouse or keyboard. However, this principle has also been explored in gaze selection. Here the dwell visualization does not occur on the target, but affects the entire interface by shifting it towards the target or by magnifying the area surrounding the target.

Both continuous and discreet approaches to zoom have been explored. An example of a continuous zoom selection interface is StarGazer (Hansen et al., 2008) (Figure 25). Pan and zoom were used to navigate the interface. The idea was to allow the user to navigate through graph structured data; in this case a gaze typing system was implemented. The inaccuracy of both eye movements and eye tracking systems were compensated by having target size increase to a level where there was no doubt that a selection by the user was intended. This approach was especially beneficial when employed with low-resolution eye trackers or on small-size displays, since it allows *more* selectable objects to be displayed on the screen than the accuracy of the gaze trackers would otherwise permit.

[7] http://www.tobii.com/assistive_technology/products/vs_communicatior.aspx

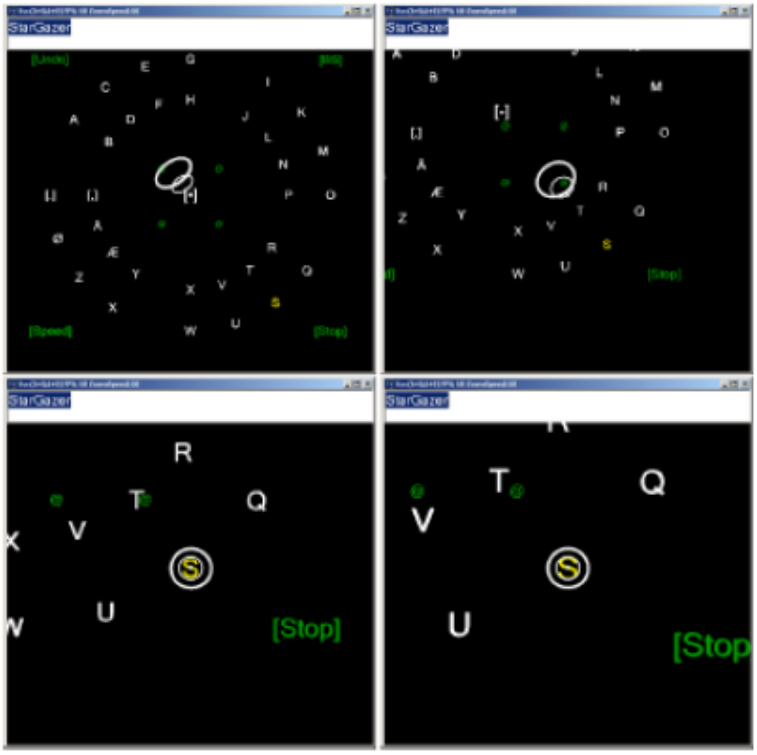


Figure 25: The Stargazer Interface, the cursor indicates the direction of gaze. The interface continuously zooms towards the user in the direction of gaze. In the images shown above the process of selecting the letter 'S' is depicted. (Hansen et al., 2008)

The disadvantage of a continuous zoom interface is that the constant movement of the interface in its entirety makes it difficult for the user to orientate where to go next, as any perceptive action causes the interface to start shifting.

Discreet target-based magnification where only a part of the interface is affected has also been explored. The advantage of this selection type compared with continuous zoom is that it allows the user to magnify only the surrounding area of the target (Skovsgaard et al., 2008)(Figure 26).

In this work, it was found that the two-step magnification had a much higher hit rate with small targets than conventional dwell-time selection. This argues for the technique's ability to handle larger amount of data being displayed on the screen, thus avoiding the traditional issue of dwell-time activation where the size of the targets limits the amount of on-screen information available. Visual feedback not only affects how the user can interact and perceive information on the screen, but also how the information is structured. Two-step magnification can be useful if the intention is to deal with the existing windows, icon and menu structure

known from current layouts of desktop environments and text editing. However, gaze specific applications that incorporate a static or dynamic visualizations are often developed to take full advantage of gaze, by not only providing feedback to the user in a gaze appropriate way but also by designing the underlying information structure in a way that makes it readily accessible to gaze input.

Figure 26: Two-step magnification technique. The square represents the area which will be magnified by using dwell. (Skovsgaard et al., 2008).

The examples above illustrate how the selection process of dwell can be visualized. This constitutes one of the main challenges in most gaze selection strategies. Because not only does the feedback to the user entail content, but the process of selecting that content also must be visualized. One of the experimental parameters that has been experimentally explored and will be presented in Chapter 8 is completing the selection process without visual feedback to the user.

3.3.3 ACTIVATION DURATION

In most of the applications that have been presented so far a dwell-time of approximately 400-500ms has been employed as a standard, but quite often it needs to be much higher for people with nystagmus, large jitter and/or tremors. In the settings of most gaze applications the dwell-time duration can be adjusted. This is a relatively cumbersome process that is usually handled by the carers in the case of an impaired individual.

However, many users who are dependent on gaze tracking become expert users and could most probably reduce the activation completion time required to perform a selection as they become acquainted with an application. Also there is a difference in the abilities of users with complex needs from day to day, sometimes even hour to hour. Therefore, there is a need to implement a dwell activation time that can be adjusted by the user during their use of the application.

In a recent study conducted by Majaranta et al. (2009) participants used a system that allowed them to dynamically change the activation time while typing on an on-screen keyboard. This was done repeatedly over 10 sessions. Activation times went from an initial average of 876ms to an average of 287ms (Majaranta et al., 2009).

However, there might also be situations where tasks, context and feedback require a fixed dwell-time. Such a task situation could be a gaze based computer game in which the objective was to fixate for longer durations of time. A user with cognitive impairments could also accidentally readjust the dwell-time to a point where it inhibited his or her use of the system. If the feedback was from a multiuser interface, having standard settings that apply to all users is useful.

Activation duration is of importance in dwell interfaces because it greatly affects the user experience. One aspect of gaze gesture interaction is that these actions can be completed without fixed completion times.

As mentioned, the previously presented taxonomy of dwell-time interaction was derived from analysing existing selection methods. Creating new implementations of dwell could happen by adjusting and/or altering any of the three principles: selection complexity, visual feedback and activation duration.

Whatever principles are used to implement dwell activation, be it single or multiple fixations, static or dynamic visualizations, with fixed or adjustable activation duration speeds, dwell selection has and will be a fundamental part of gaze interaction. However, dwell activation has a set of constraints which make it an insufficient tool to solve all of the complex tasks that could be of interest for users to be able to complete. These constraints have to do with the basic principles in the presented taxonomy. Single and complex dwell selections have each

their constraints. Single dwell requires a relatively long dwell duration and only limited number of targets can be visualized on the screen. Complex dwell also has the constraint of being limited by screen space, and requires a high level of precision pointing by the user. The visual feedback of the selection process constrains the amount and type of information which can be displayed on the screen, whether it is target or interface based. And activation duration constrains the time with which a task can be completed. Because of these constraints the field of gaze interaction has sought to find alternatives to dwell time selection. The other gaze interaction principle which has been explored is gaze gestures.

3.4 GAZE GESTURES

Gaze gestures are a more recent addition to the *interaction vocabulary* of gaze and many possible implementations of gaze gestures have been and need to be explored. *Strokes* are the foundations of gaze gestures. A stroke is the motion between two *intended fixations*; it is not necessarily the same as a saccade, which is the eye movement between *any* two fixations. A *stroke* can be completed even if jitter causes small saccades to take place between the two intended fixation points. Potentially, gaze gestures hold many advantages as a selection method:

1. Speed: gaze gestures can potentially be very fast. As presented in chapter 2 saccades can cover a 1° and 40° visual angle and last between 30-120ms (Duchowski, 2007). This is substantially faster than a standard dwell-time duration of approximately 300-500ms (Majaranta et al., 2009).

2. Screen real estate*:* gaze gestures need not take up much screen space. The initiation and completion fields of the gaze gestures could be transparent, allowing for more on-screen information that is unaffected by gaze. However, as will be shown in the subsequent sections, gaze gestures have mainly been implemented with semi-transparent or opaque *fix-points*, which still infringe on the issue of screen real estate. A fix-point is a visual point on the screen, which functions as a navigational target. either opaque or semi-transparent (Istance et al., 2009; Istance et al., 1996; Vickers et al. 2008; Istance et al., 2010; Istance et al. 2009b; Wobbrock et al., 2008).

3. No Midas touch problem: if gaze gestures are implemented in a way that they are easily distinguishable from navigational eye movements, the initial point of gaze is of no consequence, which means the Midas touch problem is avoided.

However, gaze gestures have their own set of innate problems. The main issue being *accidental gesture completion*: a potential overlap between natural search patterns and the eye movement patterns needed to complete a given gaze gesture. Accidental gesture completion is the equivalent of the Midas touch problem for dwell activation. Another issue with gaze gestures is the cognitive load that they might induce. Dwell activation is generally a direct response to information presented on the screen, whereas gaze gestures often rely on the user to be able to remember specific eye movement patterns and their consequence with no immediate feedback. It will be shown that purely visualization-based gestures are also possible. However, these can only be used in conjunction with specialized software. The diversity of gaze gesture applicability is only just being uncovered and in the following section considerations regarding existing implementations will be presented.

3.4.1 THE GENERAL CONCEPT OF GESTURES IN HCI

A useful definition of gestures in general comes from Raskin's 'The Humane Interface':

'a gesture is a sequence of human actions completed automatically once set in motion' (Raskin, 2000, p.37)

As a consequence of this definition a gesture can consist of any repeatable and recognizable bodily motion that can be robustly recognized as separate from normal physical action.

Stylus based text entry is an area where gestures have been greatly explored. The main similarity between gaze and stylus based interaction is that they both consist of one pointer connecting with the interface. Unistrokes (Goldberg & Richardson, 1993) was a *gesture* alphabet designed for the stylus. It was based on (Figure 27)(a.) five basic shapes that could be (b.) rotated in four directions and (c.) completed from different directions. Other examples of this type of stylus alphabets are Cirrin (Mankoff & Abowd, 1998) and Grafitti (Blinkenstorfer, 1995).

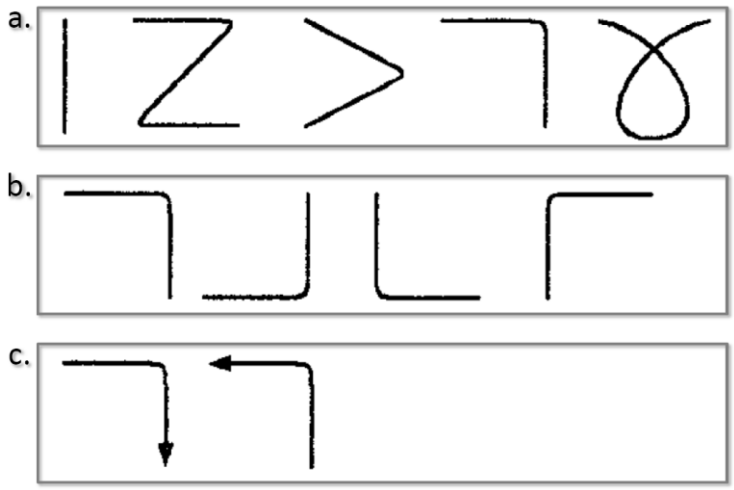


Figure 27: (a.) The basic Unistroke shapes, (b.) one shape rotated, (c.) one shape completed in two different directions (Goldberg & Richardson, 1993).

Unsurprisingly, stylus based text entry has served as an inspiration for gaze gesture based text entry. Perlin's QuickWrtiting (Figure 28) has done this to a great extent and there are two reasons for this. In this implementation the stylus does not need to be lifted from the screen and the motions used are continuous. These two characteristics of the system mirror the requirements for a gaze-only input (Perlin, 1998).

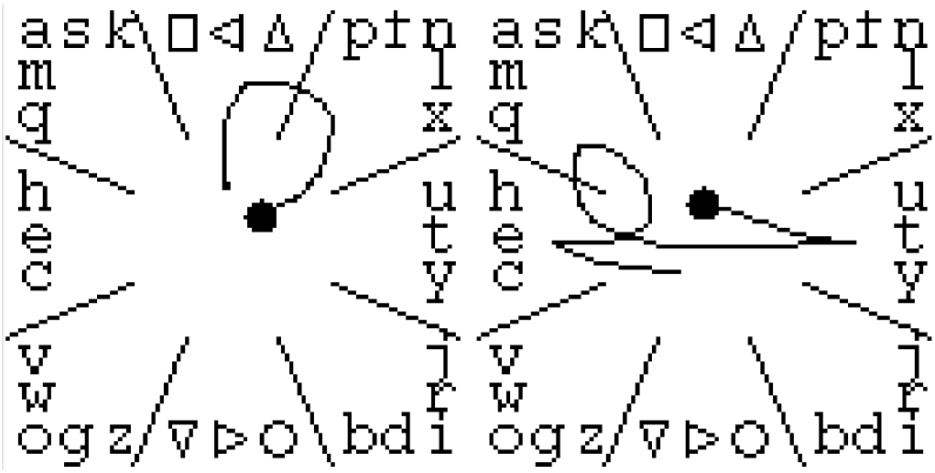


Figure 28: Stylus gestures in Quickwriting (Perlin, 1998)

The system was designed so that characters could be selected by moving from the centre of the layout to different zones around the edge. The simplest gesture consisted of moving from the resting zone in the middle of the layout directly to an outer zone and then back again. In this way the middle letter of a group could be selected, in the presented layout (figure 29) 'a', 'n', 't', 'i', 'o', 'e' can be selected by simple centre-to-zone gestures. If a 'q' were to be selected

(Figure 29), a gesture would be formed by moving from (a) the centre resting area to (b) the zone twice removed to the left of the group with the 'q' in it, then from here to (c) the zone which actually contains the letter 'q' and then (d) back to the centre again.

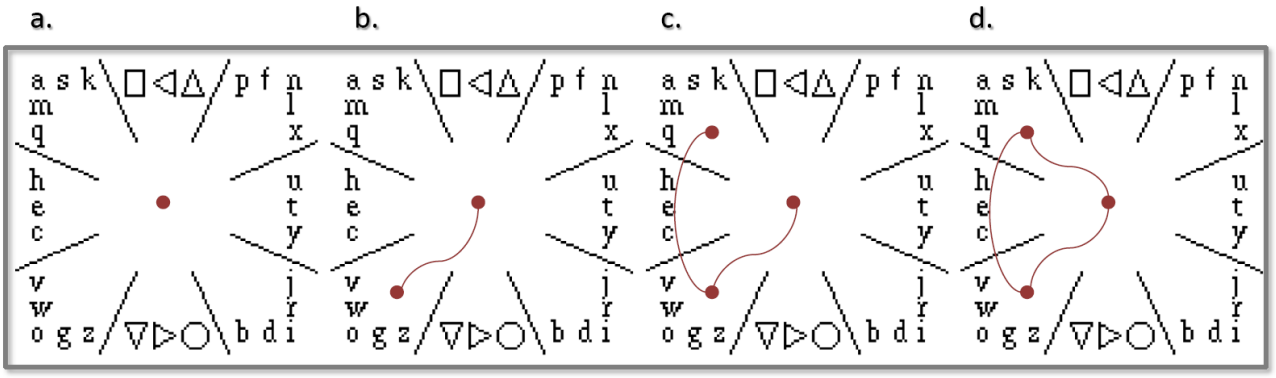


Figure 29: The Sequence of selecting a Quickwriting letter, in this case the letter 'q'. Adapted from Perlin (1998)

Gestures have recently become a mainstream interaction principle with the introduction of multi touch finger gestures for mobile devices, laptops and pads. Some of the most common finger gestures are (a) *tapping*, (b) *swiping* and (c) *pinching* (Figure 30)[8]. Tapping resembles a button press and as with mouse interaction both single and double taps mean different things. Swiping is often used to instigate a scrolling action and converse/inverse pinching allows the user to zoom in and out of the interface.

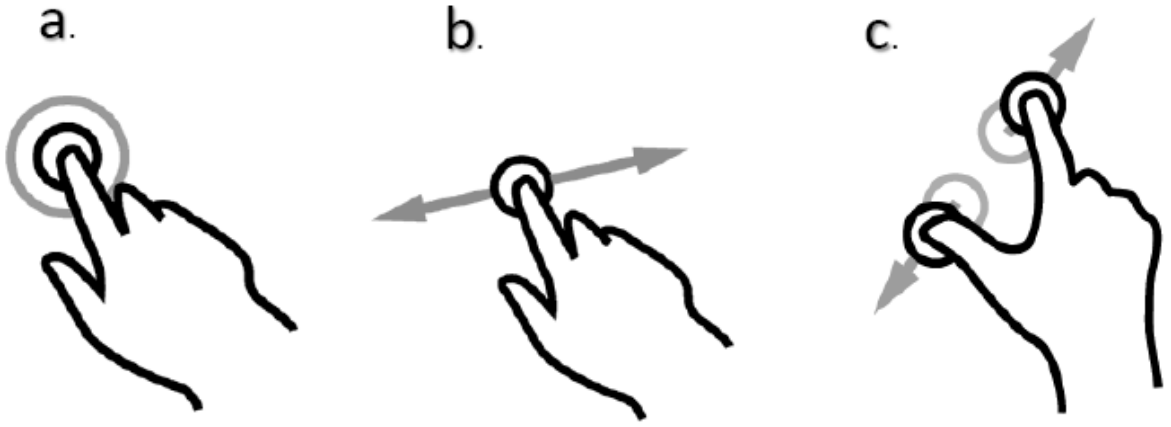


Figure 30: Tapping, swiping and pinching gestures on a touch sensitive surface

Various types of physical gestures have also been explored when considering users with motor-skill impairments because they allow for individually adapted interaction. For example,

[8] http://www.danrodney.com/mac/multitouch.html

head gestures consisting of simple horizontal and vertical motions have been used as switch control gestures (Keates & P. Robinson, 1998).

3.4.2 GAZE GESTURE INTERACTION

Gaze gestures have a shorter history than dwell-time but a similar range of principles have already been applied in various applications. Istance et al. (2010) have proposed a definition of gaze gestures that also applies in this research and does not conflict with the general definition of gestures presented by Raskin:

A definable pattern of eye movements performed within a limited time period, which may or may not be constrained to a particular range or area, which can be identified in real time, and used to signify a particular command or intent. (Istance et al., 2010)

A similar taxonomy to dwell-time activation can be applied to gaze gesture selection (Figure 31). The main difference is mainly fixation duration, as gesture selection implementations only use very 'short' dwell-time duration, if any. The basic principles used to define gaze gestures in the context of this work are *finite gaze gestures*, *continuous gaze gestures* and *complexity of strokes*.

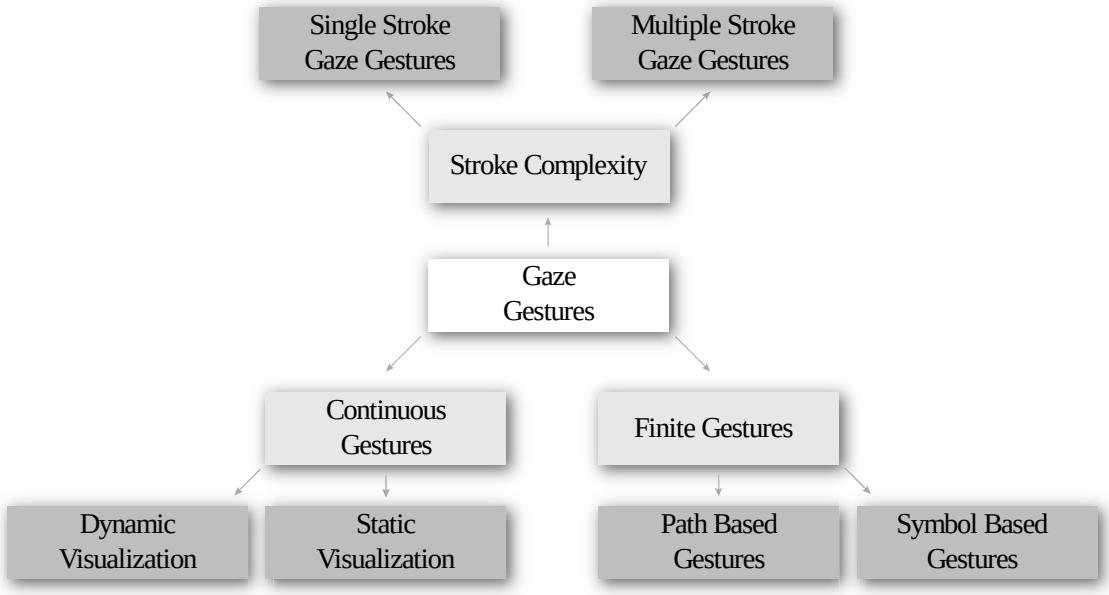


Figure 31: The principles used in implementing gaze gestures

Stroke complexity is mainly a concern for *finite gestures* where multiple or single strokes are used to initiate activation. However, specific strokes can also be incorporated into continuous

gestures where these signify explicit actions. *Finite gestures* represent eye movement patterns in which a defined shape is completed. These can be based on abstract shapes in which the eye movements follow a discernable shape, but one with no innate meaning, these are defined as *path based gestures*. *Symbol based gestures* also follow a specific shape; however these either have an intrinsic meaning or representation. *Continuous gestures* are not based on specific shape, but are derived through a response to either a dynamic or static visualization.

This taxonomy has been derived by examining existing implementations of gaze gestures. In the following section each of the principles described in Figure 31, will be explained by exploring practical examples.

3.4.3 STROKE COMPLEXITY

Strokes are the foundations of many types of gaze gestures. Earlier a stroke was defined as the motion between two intended fixations. The reason why intended fixation points are interesting for gaze gestures, despite fixations not necessarily being a direct function in gaze gestures, is that they ensure a distinction between inspection eye movements and selection eye movements. In other words, even a very short fixation on a specific initiation area can be an important way of avoiding accidental gesture completion.

A *single stroke gaze gesture* (SSGG) is the simplest form of finite gesture and is defined as the motion between two intended fixation points to complete a selection. A *complex gaze gesture* is the motion between three or more intended fixation points. SSGG are the main focus of the experimental research conducted and subsequently presented in this thesis. The main assumption of this type of gesture is that there will be a high likelihood of accidental gesture completion.

Complex gaze gestures have the advantage of increasing the interaction 'vocabulary' of gaze greatly. However, this brings with it both cognitive and physiological difficulties. Cognitively it is difficult to remember a large number of gestures and physiologically it is difficult to create and complete them. There is a need for determining a *complexity threshold* for the number of finite shapes and the number of strokes they can contain; that can be both remembered and repeated in a sustainable way.

3.4.4 FINITE GAZE GESTURES

The shape of gestures is mainly relevant when discussing complex gestures. There are many different principles under which gaze gesture patterns can be created. The two which have been explored so far are symbolic and path based patterns. Naturally, symbolic gestures also require the user to follow a path. However, the distinction lies in the definition of a symbol:

'[Symbol] – something that stands for or suggests something else by reason of relationship, association, convention, or accidental resemblance' (www.merriam-webster.com)'

A path based gesture is simply a form which can be completed by gaze – abstract or otherwise; it could be the eyes covering the pattern of a line, a triangle, a square, circle, etc. These are simply paths that have no innate meaning inferred onto their completion. A symbol is a representation of a meaning. In the area of gaze gestures the most obvious example of this is that of gesture patterns resembling a character in an alphabet. Much of the innovation in gaze interaction has happened in the field of text input and this has also already been the case in the research done into gaze gesture selection.

In 2000, Isokoski et al. presented the MDITIM text input system, which was based on gestures being completed between large off-screen targets, Figure 32.

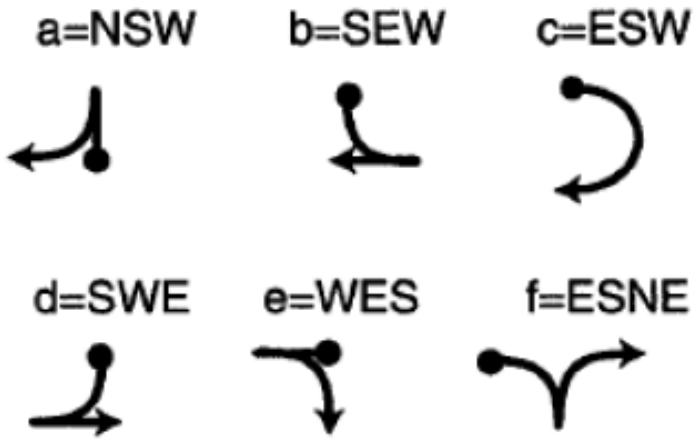


Figure 32: Example of MDITIM gaze gestures, characters were completed by completing the patters shown above between off-screen targets (Isokoski, 2000)

This was compared to Morse code, a gaze controlled version of QuickWriting and a Cirrin circular keyboard layout (Mankoff & Abowd, 1998). One of the main observations was the distinction between having *few targets* that require *many activations* per character, and having *many targets* that require *fewer activations* per character (Isokoski, 2000). This exemplifies one of the issues concerning implementation of gaze gestures. Simple gestures are

easy to repeat, but there are only a limited number (similar to few targets). Complex gestures (similar to many targets) are plentiful and can describe an entire alphabet. However, they require more precision when being completed and increase the cognitive load. Furthermore, at present they have only been implemented with on-screen visualization that is not conducive if the goal is to minimize the on-screen real estate of selection strategies and potentially incorporate other selection capabilities.

Woobrock et al. (2008) presented EyeWrite, a gaze controlled text-input system, based on a text input system for PDAs called EdgeWrite, Figure 33. The user could map out letters by combining the four corners of a square in various ways. Their results showed that the complexity of the alphabet proved very difficult. In their study they compared EyeWrite to an on-screen keyboard. The average input speed for the on-screen keyboard was approximately 6 wpm (words per minute) and for EyeWrite it was only around 2 wpm. EyeWrite was also more error prone, which most likely has to do with the complex nature and high number of gestures (Wobbrock et al., 2008).

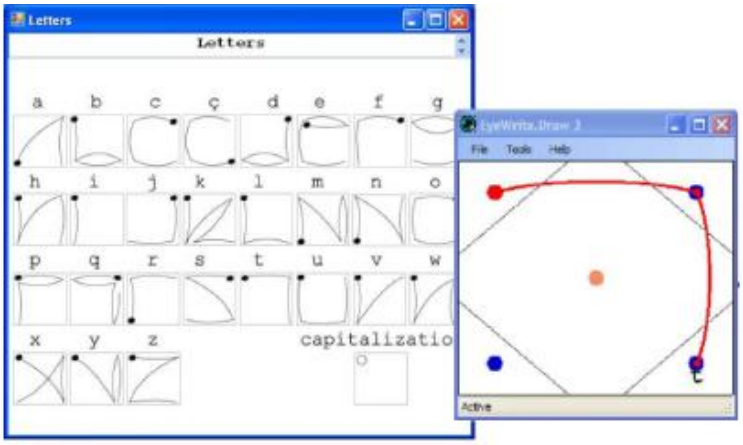


Figure 33: Gaze gesture in the EyeWrite System, characters were completed by glancing at fix-points in different sequences (Wobbrock et al. 2008)

Porta and Turina. (2008) developed the EyeS system, which had users enter letters by fixating on fix-points in various sequences. Here the gaze patterns were designed to be symbolic by having them resemble the shape of the character that was being completed, Figure 34. They reported a 6.8 wpm average for experienced users and show that there was a significant difference between novice and experienced users. Of relevance to the research presented here

they also made suggestions of using these types of gaze gestures for shortcuts in a windows interface (Porta & Turina, 2008a).

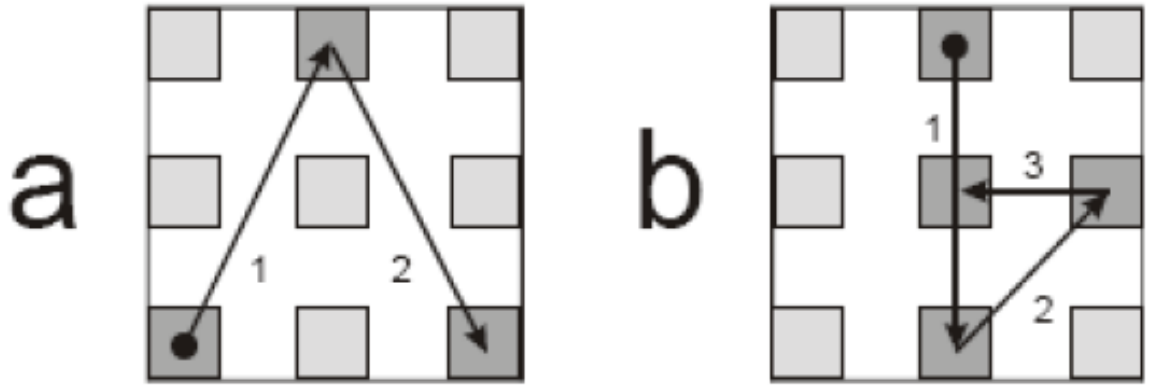


Figure 34: Example of gestures in the EyeS system, again selections were completed by gesture completion based on fix-points (Porta & Turina, 2008a)

These all are examples of symbol based gaze gesture solutions. The main advantage of this principle is that a larger number of gaze gestures could become easier to remember because they have become contextualized and build on familiar representations (the alphabet). The disadvantage is the higher cognitive load, which could overwhelm some users.

Path based gestures have also been explored on several occasions. Heikkilä and Räihä (2009) presented research regarding how gaze gestures could potentially be used in a drawing programme. This represents one of the first examinations of how gaze behaves in situations where the user follows an intended shape path with the eyes. The overall conclusion of the research was that there was not a significant difference between the time it took to complete the path of large shapes compared to the time it took to complete the path of smaller shapes. This was found to be because the eye movements of participants would reach higher speeds when completing the contours of the larger shapes. Overall they found that it was very difficult for users to follow the paths properly, especially curved ones, Figure 35 (Heikkilä & Kari-Jouko Räihä, 2009).

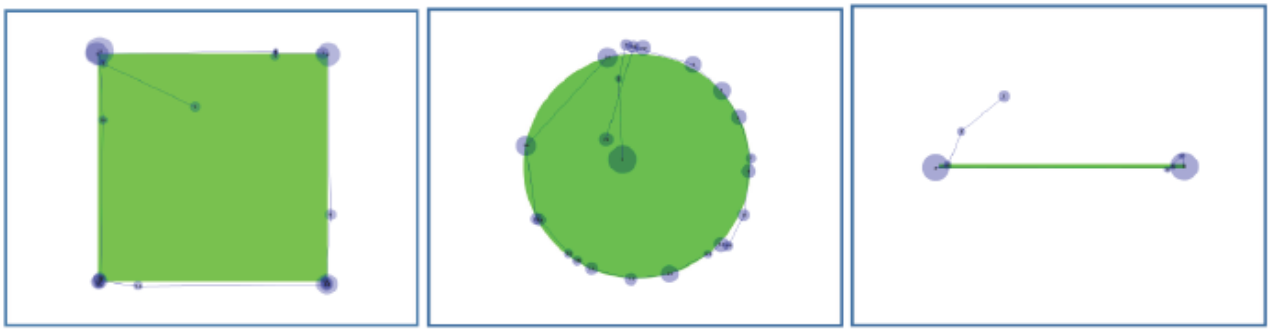


Figure 35: The gaze paths of a slow but accurate users following the contour of a shape (Heikkilä & Kari-Jouko Räihä, 2009)

Another example of abstract path based gestures is the research done by Drewes and Schmidt (Drewes & Schmidt, 2007)(Figure 36). One of their main motivations for researching gaze gestures was the deduction that accuracy becomes less of an issue in this type of selection strategy, because gaze is not used for pointing and the completion of a gesture is, in this case, relative to the eye movements rather than on-screen fix-points.

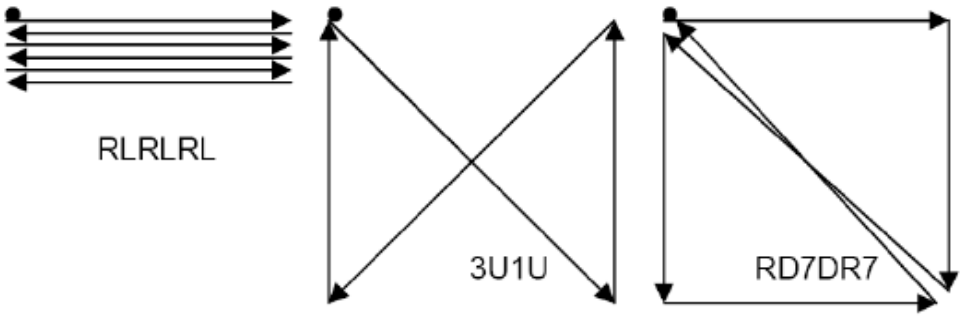


Figure 36: The three gestures used in the experiment (Drewes & Schmidt, 2007)

A set of gaze gestures was developed based on a combination of existing set of mouse gestures available for the Firefox browser[9] and the gestures in EdgeWrite (Wobbrock et al., 2003), which was presented earlier in this Section. Figure 36 shows the three gestures that were used in the experiment. A short dwell-time was implemented to help differentiate between eye movements.

They asked participants to complete the gestures on different backgrounds and found that on backgrounds with text, tables or web pages, even complex gaze gestures could be completed reliably. They also found that neither background nor the complexity of the gesture had a

[9] http://optimoz.mozdev.org/gestures/

significant impact on the completion time. Gesture completion time only depended on the number of segments, in other words the number of strokes that needed to be completed. They recorded an average stroke completion time of between 550-600ms and also found that only the right-left-right-left-right-left (RLRLRL) gestures had a risk of causing accidental gesture completion during the natural search patterns of web browsing.

In 2008, Istance et al. proposed a novel approach to gaze selection, called *snap clutch*. This method was implemented as a modal interface that allowed the user to control a character in the computer game 'Second Life'. The initial trials compared mouse control, dwell and the snap clutch principle. Snap clutch was implemented so that the user could move between four different modes by looking off- screen in four directions. The results showed that selection times almost reached that of the mouse and had fewer errors compared to dwell- time (Istance et al., 2008; Vickers et al., 2008). An important element of the work undertaken by this group is the research done on task-analysis, the concept of modes and mode changes in gaze-only interfaces (Istance et al., 2008), which contributes to the overall understanding of when and how different gaze selection strategies should be implemented.

In another project regarding the computer game 'World of Warcraft', Istance et al. (2009) compared a series of tasks performed with mouse/keyboard and gaze respectively. The preliminary results show surprisingly small differences between the two input modalities (Istance et al., 2009b).

Based on this research and inspired by Perlin's Quickwriting setup, Istance et al. (Istance et al., 2010) have set about using abstract path based gaze gestures to enable game control in the setting of the online massive multiplayer game 'World of Warcraft'. Their main motivation is to allow users with severe motor impairments to be able to participate in online communities on equal footing with other players. In the most recent implementation of gaze gestures for game control they researched the innate qualities of their design both separately and in the game of World of Warcraft.

Initially they examined 12 individually recognizable gaze gestures, which were to be completed by looking from the centre of the screen to the different fields in various 2- and 3-stroked combinations, Figure 37.

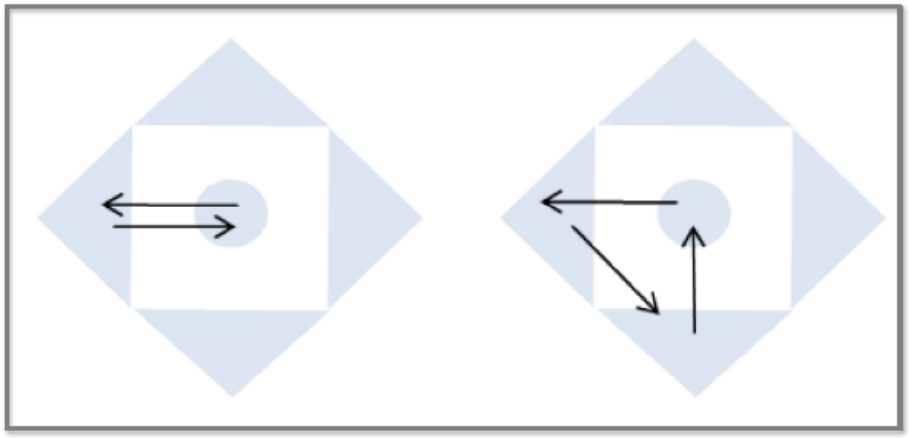


Figure 37: 2 and 3 stroked gestures for game control in WOW (Istance et al., 2010)

The centre of the screen was chosen as a starting point, because the player's avatar in World of Warcraft is always placed there. This meant that the player could continue to monitor their avatar in the game, while making selections. This has in turn been one of the main arguments for using gaze gestures in this type of gaming situation and not dwell-time activation, e.g. dwell-buttons would force the user to look away from their avatar for a prolonged period of time, which is not ideal in a real time multi-player game, where monitoring the avatar in a, for instance, a multiplayer combat situation is required. All gestures had to begin and be completed at the centre of the screen within a 2 second time period in order to be deemed valid. The experiment looked at 2 and 3 stroked gestures, horizontal/vertical versus oblique selection patters and initiation stroke (i.e., whether a left or right eye movement started the gesture.) (Figure 38).

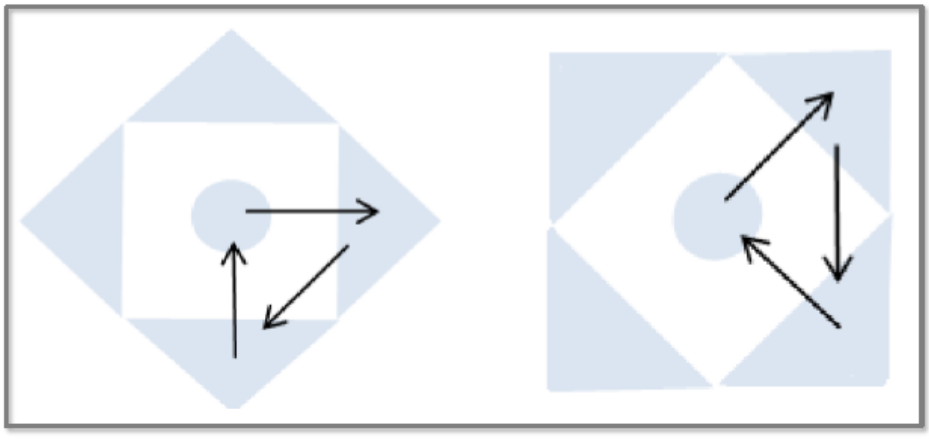


Figure 38: Horizontal/Vertical gestures and Oblique gestures (Istance et al., 2010)

They found that there was a significant difference between the completion time of 2 and 3 stroked gestures. Two-stroke gestures on average took 490ms, which is comparable to the usual dwell activation time. Three-stroke gestures took much longer, 880ms. Whether the initial stroke was horizontal/vertical or oblique only had a small effect on 2 stroke gestures and no effect on 3 stroked gestures. For 2 and 3 stroke gestures they found that there was a general effect of gaze gestures starting with a leftward move being faster than those starting with a movement to the right. However, this effect was mainly derived from 3 stroke gestures, as there was virtually no effect from the 2 stroke gestures.

In a subsequent experiment these gaze gesture types were implemented in World of Warcraft. Four 2 stroked gestures were chosen to represent the 'w','a','s', and 'd' keys often responsible for forward, left, backward and right movement in 3D environments. Three-stroke gestures were implemented as controls for various attack- and spell commands (Figure 39). They found that all 12 participants, who participated in the study, could use gestures to navigate and perform actions in the game. However, locomotion proved quite difficult during attacks, but easier during long movement. Gaze gestures also functioned well for discreet spell casting and only 2 stroke gestures were completed accidentally.

Overall they reported a stroke time of 247ms/stroke for 2 stroke gestures and 293ms/stroke for the 3 stroke gestures. One of their main conclusions (which is completely coherent with the overall conclusion of the research conducted here) was that gaze gestures are a valid gaze selection method, but that its true potential lies in combining it with other gaze interaction techniques.

Figure 39: The diamond shaped gaze gesture layout implemented into World of Warcraft.

The *single stroke gaze gesture* (SSGG) that will be explored experimentally in chapter 5-9 is in the context of this research defined as a finite gesture. However, these simple gaze gestures can also be used as part of continuous gesture patterns. This principle will be subsequently presented.

3.4.5 CONTINUOUS GAZE GESTURES

Much of the research on gaze gestures deals with strokes being tied to some kind of visualization on the screen. In the EyeS system (Porta & Turina, 2008b) and in the World of Warcraft implementation (Istance et al., 2010) just described, gestures are completed by focusing on various combinations of semi-transparent fix-points on the screen.

Another type of gesture is not based on predefined finite shapes. This gaze gesture is defined and shaped by the visual layout of information, dynamic or otherwise, on the screen. This will be referred to as *continuous gaze gestures*. Path-based gestures are often finite patterns, whereas continuous gestures rely on continuous eye movements.

The most classic example of this is Dasher (Ward et al., 2000)(Figure 40), which is also one of the first gaze gesture interfaces. As with other gaze contingent interfaces, Dasher uses magnification as feedback to indicate which letter is being selected. Initially 27 characters are placed in a column to the right of the screen. To select the letter 't' the user looks at the letter

‘t’ in the right column. The size of the letter begins to increase and move towards the left. Once the letter crosses the line dividing the screen, it is ‘selected’. The user can reverse and ‘deselect’ a chosen letter at any time by looking at the left side of the centre line. Other than being a continuous magnification based gaze contingent typing application, Dasher is known for having a very well integrated letter probability prediction. This probability is visualized by increasing the size of the most probable subsequent letters. If ‘t’ is the first selected letter, then the letter ‘h’ is presented as the closest and one of the largest subsequent letters. After that ‘e’ becomes the most prominent letter and so forth[10].

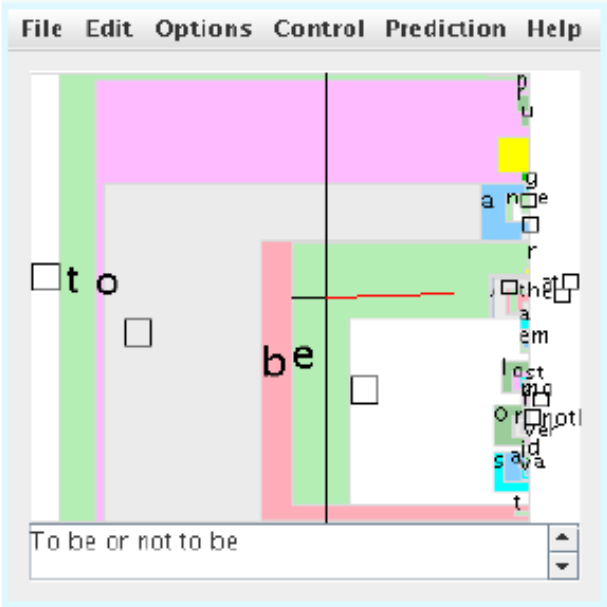


Figure 40: The Dasher Interface (Ward et al., 2000)

Dasher is not based on fixation as with other zoom interfaces such as StarGazer (Hansen et al., 2008). StarGazer requires the user to focus on one target, until this target encompasses most of the screen. In Dasher, the targets are always moving. In this type of interface stroke speed is not interesting because the continuous motion means that there is a constant mix of fixations, saccades and smooth pursuits. In their research Ward et al. (2000) show that despite an initial steep learning curve, Dasher has the potential of being an extremely fast text input method. They refer to one subject who used Dasher for a prolonged period of time and reached text entry speeds of 34 wpm (Ward et al., 2000). As a comparison, research of typing speeds on an on-screen QWERTY keyboard show average typing speeds of approximately 22 wpm

[10] http://www.inference.phy.cam.ac.uk/dasher/TryJavaDasherNow.html

(Mackenzie et al., 1994). However, typing speeds on a regular keyboard can for professional typists reach as much as 60-70 wpm, when copying text (Brown 1999).

Another example of visualization based gestures is presented by Bee et al. (2008). Again, inspired by Perlin's QuickWriting, they explore a form of continuous writing and argue that this is the best suited text entry method for gaze. The user looks at the centre of the visualization, where there are several groups of letters and characters. By looking at one of these, the letters and characters contained in that group become highlighted and an individual character can be selected. They make the distinction between *typing*, *gesturing* and *continuous writing* (Bee & André, 2008) (Figure 41).

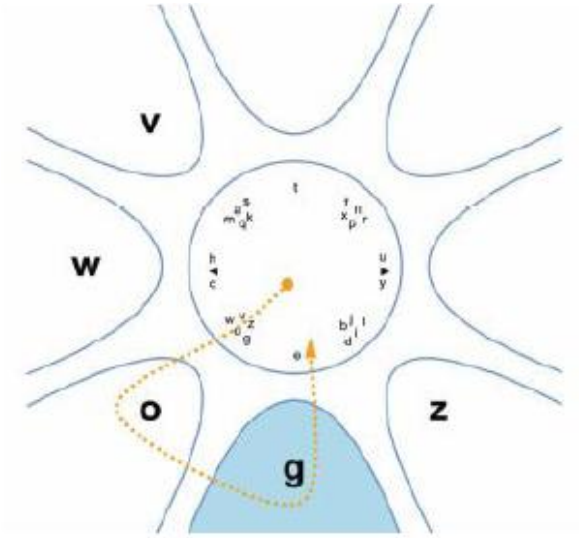


Figure 41: Example of a gesture in the gaze adapted QuickWriting, (Bee & André, 2008)

As mentioned earlier, in the current research, the distinction is made between *finite gestures*, *continuous gaze gestures* and *complexity of strokes*. These researchers compared their gaze adapted version of QuickWriting (which achieved 5 wpm) with an on-screen keyboard (which achieved 7.8 wpm). With practice users could achieve up to 9.5 wpm (Bee & André, 2008).

Gaze gestures have been a way of creating gaze selection that is not necessarily dependent on dwell-time activation. This idea of dwell-free interfaces was explored in three different applications by Urbina & Huckauf (2007). The gestures used in their work are all continuous and tied to the changing visualizations on the screen and not individual character shapes or sequences explicitly. In the pEYEdit interface the principle of expanding menus was explored. Each slice contained a group of letters – when selected, by crossing the outer border of the

slice, the group would expand into a new pie where each slice had only one letter, which could then be selected based on the same selection principle (Figure 42a).

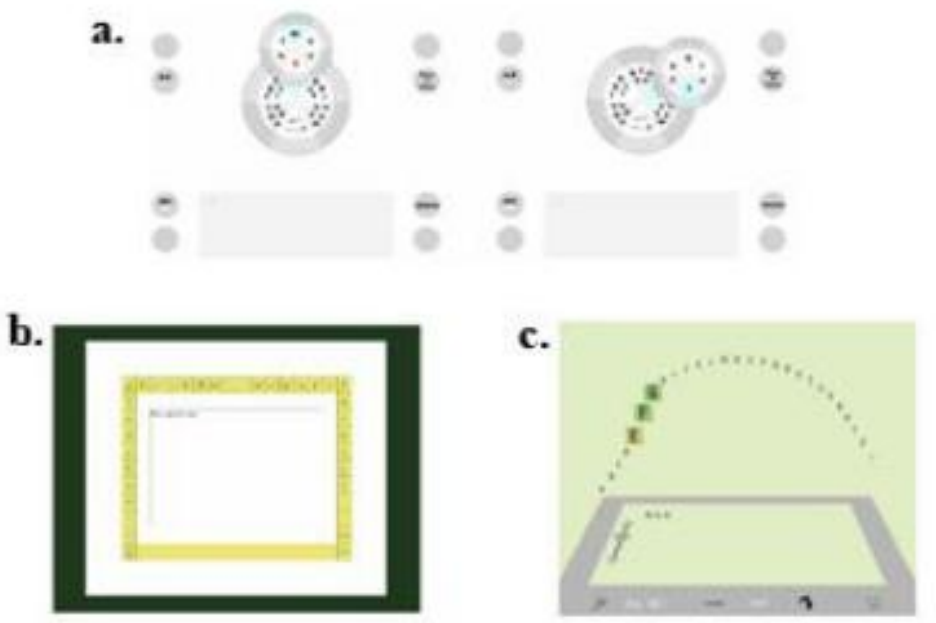


Figure 42: a. The pEYEdit interface; b. The IWrite Interface; c. The StarWrite Interface (Urbina & Huckauf, 2007)

In IWrite the characters were placed in a frame on the screen and selection was facilitated by short strokes from the intended character to the outer frame, which functioned as an on-screen button but without dwell (Figure 42b).

And finally, StarWrite allowed the user to drag letters from a half-circle onto a text-field (Figure 42c). These three selection methods were compared to a QWERTY on-screen keyboard and Dasher without character prediction (Ward et al., 2000). They had both novice and advanced users test all the systems. The advanced users had the following results: QWERTY 15,8 wpm; Dasher (without prediction) approx. 6,5 wpm; pEYEdit 10,9 wpm; IWrite 11.4 wpm and with StarWrite 8,4 wpm (Urbina & Huckauf, 2007).

In a more recent study, Urbina et al. (2010) presented a modified version of the pEYE system. A dwell-controlled version of the interface was compared to the border crossing gesture based original. And also a third pie expansion was added, which provided users with the most probable next letters (Figure 43). They found that participants preferred to use the gesture based selection method, as this was faster than the dwell selection method. 8.06 wpm was achieved with gesture selection borders and 4.71 wpm with dwell-time. Also, the letter

predictive method when adding the third pie allowed an increase in typing speed, with an average of 13 wpm and the fastest speed towards 17.2 wpm.

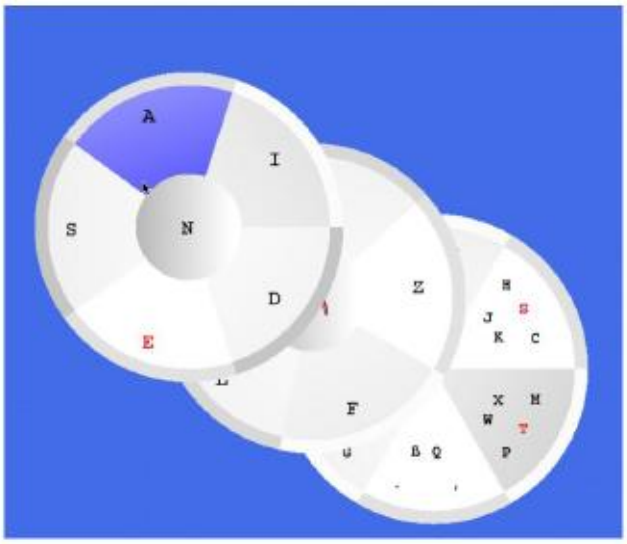


Figure 43: The modified pEYE interface with third predictive expansion. (Urbina & Huckauf, 2010)

A final example of this type of continuous gestures was the typing interface developed by Morimoto and Amir. (2010) (Figure 44). Two complete QWERTY keyboards where presented on the screen. Two different eye movements where used to focus and select a target. A short dwell duration (150ms) informed the system of which target was in focus and a saccade across the opposite keyboard completed the selection. They called this selection strategy *context switching*. Their results show varying text entry speeds and error rates. However, their fastest subject reached an average of 20 wpm with an error rate of less than 2% on the last of eight sessions.

Figure 44: Context switching keyboard, selections are made by switching between the two keyboards(Morimoto & Amir, 2010)

Continuous gaze gestures use eye movements to guide the user through the interface by giving visual feedback that corresponds to the consequence of a given action. The disadvantage of these techniques is that they are very task and application specific. In other words there is no way to extract any general approach from these techniques because they cannot be applied to existing software such as games, text entry programs or environmental control system, developed by a third party. This does not mean that it is the opinion of the author that these methods are not useful. On the contrary, the research done into this field is important to understand how gaze-only applications can and should be designed.

As with the taxonomy of dwell selection, each of the principles in this gaze gesture taxonomy has its own set of constraints. Again, single and complex gaze gestures each have their own set of constraints. Single gaze gestures have the constraint of being limited in the number of possible completion-directions; also it is constrained by having a possible high number of accidental gesture completions, as large saccades could potentially be mistaken for intended gestures. Complex gestures have a cognitive and physiological constraint; a threshold for this needs to be set. Finite gaze gestures that are symbol based have the constraint of requiring a high level of cognition from users, as well as, precision pointing. Pat based finite gestures have the constraint of not having a familiar mapping between action and consequence. Finally, all continuous gestures have the constraint of requiring specialized software. However, gaze gestures afford a new dimension in gaze interaction, the affects of which are still to be seen.

3.4.6 CONCLUSION

In this chapter a taxonomy regarding the two main gaze selection strategies (dwell and gaze gestures) has been proposed. The foundation for this taxonomy has been previous work done in the field of gaze interaction. The concentric circles used in figures 21 and 31 were intended to indicate that each of the elements in the subcategories affect each other. It is to be considered a dynamic taxonomy, which will hopefully expand as more gaze selection strategies come to light. However, the goal was to surmise the existing fundamentals of each selection strategy. For dwell-time implementations these basic principles were deemed to be: *selection complexity*, *fixations duration* and *visual feedback*. For gaze gestures they were considered: *stroke complexity*, *continuous gestures* and *finite gestures*.

Both dwell-time activation and gaze gestures have innate affordances and constraints. Future research could concern itself with the implementation of both explicit dwell activation and gaze gestures. The combination of the two could create multi-modal interactions for mono-modal input; a concept which will be revisited again in the concluding chapter.

There are three factors which have not been uncovered in the previous work regarding gaze interaction and it is the three factors which form the foundation for the experimental research which has been conducted.

The first is the concept of the *single stroke gaze gesture* (SSGG). This has not previously been explored despite it being a fundamental building block of both finite and continuous gaze gestures. The second is gaze selection without visual selection feedback. As mentioned, selection process feedback is a constraint of gaze interaction. An experiment was design to discover whether or not a gaze selection principle could be completed without feedback. Finally, there is a need to determine what the upper and lower boundaries of selection complexity are in regard to gaze gestures. These are the three overall issues which have been explored experimentally and will be presented in Chapter 5 through 9.

However, before experimental empirical evidence is presented the end-user will be introduced along with the presentation of some 'no-tech'/lo-tech approaches to gaze communication.

PART III

4 THE END-USER

Assistive Technologies are becoming increasingly commonplace, serving to enhance the quality of life for users by providing a wide range of environmental controls. However, smart homes, which are truly ubiquitous, require complex systems to be managed by the user. This in turn creates the challenge of providing adequate and appropriate user-centric interfaces; which is particularly important when the user is either elderly or suffers from physical and/or cognitive impairments which restrict their limb movements and/or ability to interact and control applications.

Understanding the *needs* and *capabilities* of the end-user is important when designing for gaze-only input; as those who can benefit the most from such strategies have very diverse needs and capabilities. A basic grasp of the design principles which have been applied to other input devices and communication strategies is useful. Low- or no-tech solutions have been, and are being, used to a great extent by individuals with motor impairments. In regard to the research presented in this thesis, acquiring an understanding of these approaches was intended to ensure the design of sustainable gaze selection strategies.

In a report developed for the COGAIN (Communication by Gaze Interaction) network of excellence an estimated 571,250 people in Europe could benefit from gaze interaction (Level, 2004). Not all of these potential users would choose eye tracking as there are many different types of assistive technologies, but according to the market report:

'A market share on 5 – 10 percent might be realistic, meaning the market size in EU should be about 28,000 – 57,000 persons.' (Level, 2004)

Until now the term 'users with motor impairments' has been used without further definition. It covers a range of people who all have in common, that eye movements are their only or strongest means of interaction. The input capabilities of the end-user and how they relate to gaze as input can be viewed as a spectrum. At one end of the spectrum there are people suffering from complete Locked-In-Syndrome (LIS); in this context eye movements are often their only controllable form of interaction. At the other end are people who could benefit from gaze interaction, but for whom it is not an absolute necessity.

The spectrum mentioned above forms the foundation for a useful classification of conditions and input modalities. The requirements regarding gaze interaction can be categorized depending on the end-users' ability to provide sole or multiple inputs to the system. As a design classification, gaze input can be grouped into three:

First of all, gaze interaction can be implemented based on a *mono-modal input* principle, as *sole* input. Users who require this design consideration can generally solely convey reliable input to a system by the use of eye movements.

Secondly, gaze can be implemented as a *bi-modal input*, where gaze interaction can be done in conjunction with another input such as a switch control of some kind. Users who require these types of designs might have a controllable finger twitch or have some reliable head movements.

And thirdly, gaze can be viewed in the context of several inputs, where the design premise becomes *multi-modal*. In this context users might have complete or mildly restricted physiological input control. This user group will only use gaze if it provides a substantial benefit compared to other inputs for certain tasks, as it is not strictly necessary. Figure 45 relates all of these conditions to the three part classification above.

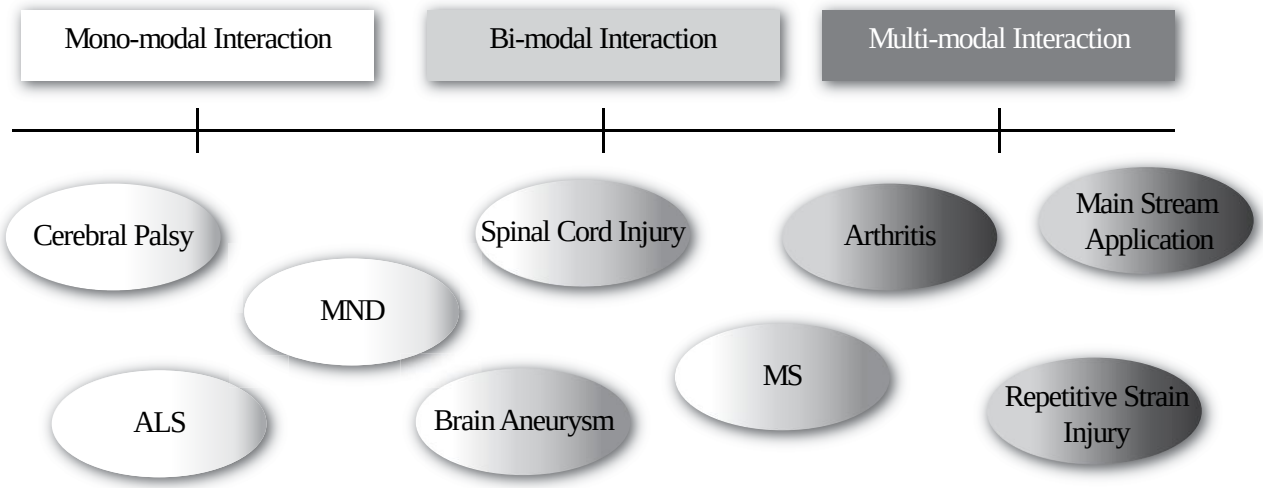


Figure 45: Modality classification of gaze for various user groups. Physiological conditions have been ordered in the end-users ability to interact with gaze as mono-, bi- and multi-modal input.

4.1.1 GAZE AS SOLE INPUT

Motor Neuron Disease (MND) and Amyotrophic Lateral Sclerosis (ALS) are two types of diseases which cause almost complete paralyses and cerebral palsy often causes spastic paralysis, where gaze interaction can represent the only motor control which does not elevate spasms (Buchholz & Holmqvist, 2009).

Motor Neuron Disease is an overall term for a group of diseases which selectively affect the motor neurons in the body. The motor neurons are a type of cell which conveys information regarding voluntary muscle activity throughout the body. A disease in these cells can cause varying levels of paralyses, affecting the ability to speak, walk, breathe, swallow and limb control. Some of the symptoms are: *spasticity*, *muscle atrophy* and *potential cognitive changes*. The disease often fluctuates between periods of rapid degeneration and plateaus of relatively stable symptoms. 50 % of patients die within 14 months of diagnosis, after which the survival rate becomes significantly skewed and people can live for decades with the diseases. Most famous is probably Stephen Hawking who has lived with MND for 40 years[11].

ALS is a form of MND. There is slight confusion even within medical circles as to when to use MND or ALS, as these are sometimes used synonymously. However, there are many types of MND of which ALS is one. People with ALS are often spared cognitive changes although these can rarely occur. Overall, muscle atrophy gradually affects the motor system and eventually the individual will need both breathing and feeding aids. Of relevance to the research conducted here is that the *ocular motor system* is generally spared in much of the disease progression. However, long time sufferers also experience reduced ocular function (Distal et al., 2008).

The term Cerebral Palsy covers different non-progressive motor impairment disorders. The disorder is a permanent state which affects the development of movement. There are several classifications of the disorder. *Spastic cerebral palsy* is the main characterization of this is lack of control of bodily movements. A less prevalent type is *ataxic cerebral palsy* in which the main symptoms are tremors and reduced muscle strength. *Athetoid cerebral palsy* affects the muscle tone – controlling movement requires a lot of focus and concentration and involuntary movements also occur. *Hypotonic cerebral palsy* causes the individual to appear limp and

[11]http://www.hawking.org.uk/index.php?option=com_content&view=article&id=51&Itemid=55

capable of little if any movement. Varying levels of cognition can accompany the different types (Krigger 2006).

For people who suffer from MND, ALS and cerebral palsy, eye movements will most likely at some point become their sole source of physiological output; either to another human being or to a digital system which can interpret these eye movements. For this reason they have been placed on the far left of the spectrum in figure 45.

The natures of these diseases are progressive and/or episodic. This means that the requirements of the end-users are not constant. It would be of benefit to the design of gaze attentive systems that those who implement them began considering ways of creating gaze systems which can adapt to *multi-, bi- and mono-modal input*. In this way the end-user can become familiar with the system, before requiring gaze as *mono-modal input*.

4.1.2 GAZE IN COMBINATION WITH OTHER INPUT MODALITIES

The diseases and conditions which affect motor skills are many and can present themselves in a multitude of ways. The implications of the conditions, in regard to human computer interaction, are difficult to classify because each individual end-user differs in progression of disease and presentation of symptoms. In the following sections several conditions will be described and most of them fall into two and sometimes all three of the modality groups depending on progression and the severity, or lack thereof, of the condition.

A *brain aneurysm* is caused by the wall of an *artery* being weak, in which case a dilation of blood can cause a ballooning effect from the artery, it is not until it potentially ruptures that it usually has any effect. A small unchanging aneurysm has very few if any symptoms. However, if a large aneurysm ruptures it can cause death or permanent disability. Physiologically varying degrees of paralysis can occur, as well as cognitive impairments (Asari & Ohmoto, 1993).

Spinal cord injury is usually a consequence of *accidents*, *tumours* or *various diseases*. It affects the brain's ability to transmit signals through the *spinal cord*. The consequences range from complete motor impairment to relative control of motor function depending on the severity of the injury. Many people can recover to varying degrees (Bracken et al., 1997). This is another situation where gaze attentive systems should be able to move from one to multiple

modalities. That way as the recovery process occurs other inputs could be added to the same system.

Multiple sclerosis is a disease which affects the brain's ability to send signals through the spinal cord. Some of the symptoms for MS are: muscle weakness and spasms, problems with coordination or balance, problems with speech and/or swallowing, visual problems including potential *nystagmus*. While there is no cure for MS, there are treatments which slow its progression. In the initial phases of the disease the symptoms occur in episodes, these episodes increase in length and strength as the disease progresses (Chandler et al., 1997; Sadovnick & Ebers, 1993).

Depending on the severity and nature of the conditions above, end-users can require any of the modality types presented in figure 45.

At the far right of the spectrum in figure 45 are the conditions which could potentially benefit from gaze input. Included in this group is a category called *main stream*. Main stream gaze interaction is, as mentioned in chapter 3, a design field unto itself, where the main focus is to use eye tracking to enhance the existing input stream. However, there are also contexts in which gaze interaction should be applied, not because it is the only form of input, but because it is the best input to alleviate strain. *Arthritis* and *repetitive strain injury* are two of such conditions.

There are many different types of arthritis which affect the joints of the body, mainly as a consequence of age. One of the symptoms may be loss of *dextral function* and pain (Verstappen et al., 2004). *Repetitive strain injury* is a condition which affects both muscular ability and the nervous system and is often caused by the continued repetition of a task. Symptoms include non-particular pain in hand, arm, shoulder and back. These conditions could potentially be aided by the introduction of gaze interaction in mainstream computing. However, this is beyond the scope of the research which will be presented here.

4.2 END-USER RESEARCH

The focus of this research has been on gaze as mono-modal input, mainly because the need for innovation in this area would not just be beneficial, it is necessary. The overall aim has been to develop designs which not only support the end-user directly by allowing a higher degree of autonomous control; but also to enables individuals in the immediate environment.

Not only has the research criterion been based on mono-modal input, it dealt specifically with individuals who are cognitively acute. The design approach has been to attempt to understand the entire context of the end-user's environment and subsequently take this context into account in the design process.

To determine what types of gaze interaction*s*, systems, and applications are important and possible for end-users to engage with, several end-users investigations were undertaken. Other than observing and engaging with users at conferences, three distinct end-user sessions have influenced the course of this research and are described in the following section.

4.2.1 1ST USER

The first and most influential interview and observation session was done in February 2007 with an individual who will be referred to as MC. At the COGAIN conference in 2006 MC gave a presentation using a set of pre-typed phrases which he could initiate by eye control and were conveyed through speech synthesis to the conference audience.

He had suffered a brain aneurysm, which had caused almost total paralysis. This made it necessary for him to have care around the clock. In the previously presented classification he was able to use bi-modal input. He still had the ability to move one hand slightly and thereby control a switch. He also had slight head movement control and facial expressiveness, although these were not used as input because they were not reliable. Additionally, the aneurysm had caused him to have some uncontrollable head movements, severe jitter and nystagmus, which all made it very difficult for him to focus on on-screen targets.

He used a myTobii eye tracker to cover part of his communication, particularly writing e-mails and so was familiar with eye tracking and the need to first be calibrated for a particular eye tracking system. During the visit several attempts were made to complete the calibration

process with him, with only relative success, as although it was possible with considerable effort to complete a calibration; soon after he would move and the calibration would be lost necessitating re-calibrating. A calibration positions the cursor on the screen. This is done in regard to the placement of the user and eye tracker.

He had a myTobii 1750 available for use, which had been bought with charitable funds raised by friends who had skydived on his behalf. However, even this top of the line system had issues with head-pose invariance (i.e., being able to track the movements of the eye independently from head movements and positioning).

As a consequence of there being problems with the calibration, on the day in question, a low tech communication aid was used which allowed him to communicate. Discussions and conversation where conducted using an E-tran communication system (Figure 46).

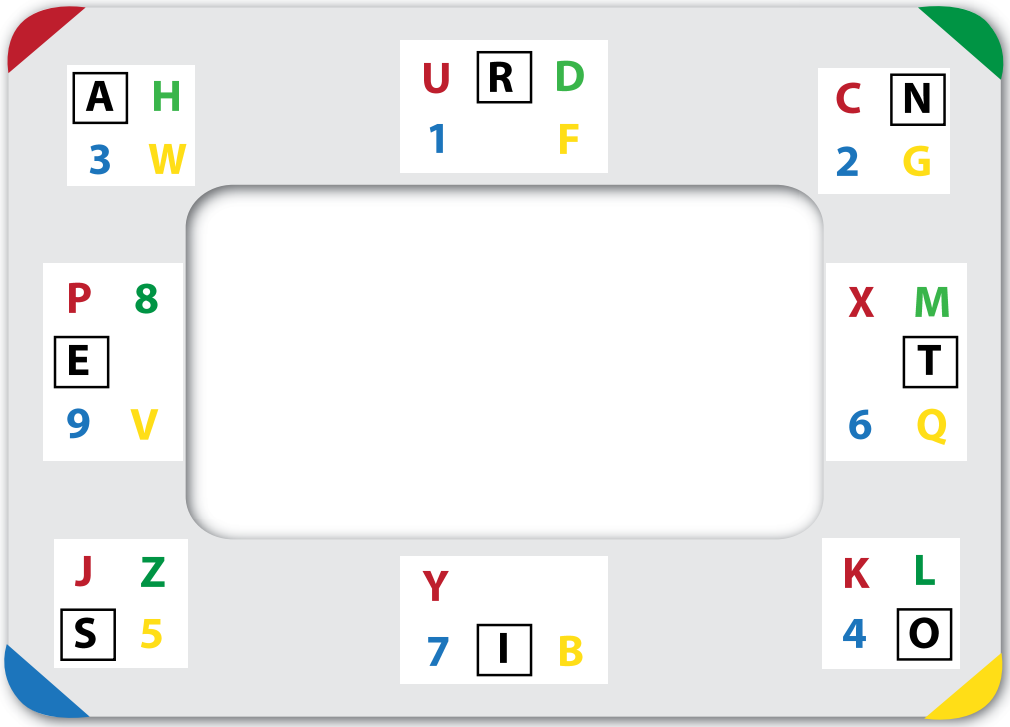


Figure 46: Adapted version of The E-tran communication system

There are many different versions of this type of communication device. Generally it is a board made from Plexiglas or cardboard, and allows for communication to occur between two people: the person holding the board and the person communicating. This is done via eye movements. A cut-out in the centre allows the 'holder' to look through the board from the

back and see the eye movement direction of the 'communicator'. Depending on the cognitive abilities of the person using the system and the general context, any manner of items can be placed on the board, in various ranges of complexity. In its simplest form 'yes' or 'no', or icons of happy or sad smiley faces can be used. In this case letters and numbers were placed in groups around the board (an adapted version of this is shown in the figure 46). Words were spelt out by selecting individual letters. These were selected by looking at a group of letters. In each group there was one letter which could be selected at first glance. In figure 46 these letters are portrayed with a black box around them, e.g. the letter 'A'. This representation was applied to the most frequently used letters. Other letters were selected by looking at a group of letters and then subsequently looking to the corner which matched the colour of the letter.

An example of the character selection process is shown in figure 47 where the sequence which is needed to spell the word 'dog' is shown. To select the letter 'D' the user looks at the top middle group and then to the green corner. To select the 'O' the user looks to the bottom right corner and dwells there. The letter 'G' is selected by looking at the group in the top right and then to the yellow corner.

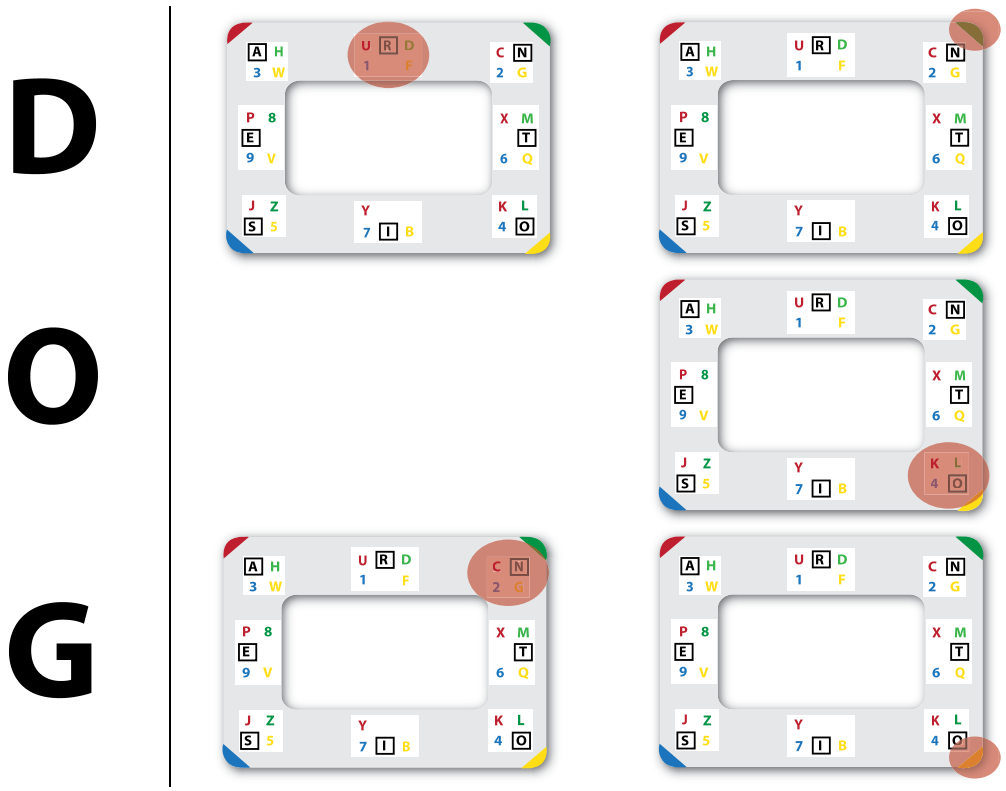


Figure 47: Spelling the word dog on an E-tran system.

The potential ambiguity which could occur if this were a digital system is cancelled by the fact that the interpreter of information is another human being. Issues which are of great concern in gaze communication system development, such as word and sentence prediction are completely overruled by the human-to-human mutual ability to infer meaning from very little information.

During the conversation with MC the carer, who was familiar with him, could understand and predict context and meaning before he completed any sentence. As a consequence, this was a very efficient way of conducting gaze based communication between a user and someone who was familiar with the user's needs and likely conversations.

However, as MC mentioned, he still appreciated the autonomy and control which the myTobii system afforded him. During one of the conversations he coined the concepts of *local* and *global communication*. He stated that the needs are different for *local communication* where the user would be talking to carers and family members compared to *global communication* where the user would be communicating long distance through email or chat to others including strangers unfamiliar with his communication difficulties

The most important observation was that of his eye movement patterns in general. One of the main problems which MC had was with fixation; as mentioned he had occasional nystagmus and large eye jitter. This greatly inspired the idea of using simple large eye movements, which would not be affected by jitter, for gaze interaction. In other words *single stroke gaze gestures* which could be completed in spite of jitter and nystagmus.

4.2.2 2ND USER

The second visit was with an individual who will be referred to as BBJ. He has had ALS for 12 years and was in the later stages of the disease. This meant that his eye movements had become more restricted. In the previously proposed classification he could only use *mono-modal* gaze input. He has written a book, called: Is it better in heaven? (Jeppesen, 1999) as well as several articles using gaze interaction. He was on a ventilator, was fed through a tube and needed 24 hour care provided by several carers. He used a QuickGlance eye tracking system primarily for written communication, which he tired from quickly. Just as MC, he used a human-to-human gaze based communication system.

This system, however, purely relied on memory. It required that both he and the person he was communicating with could memorize a grid of letters. This spelling grid consisted of grouped letters in rows (Figure 48).

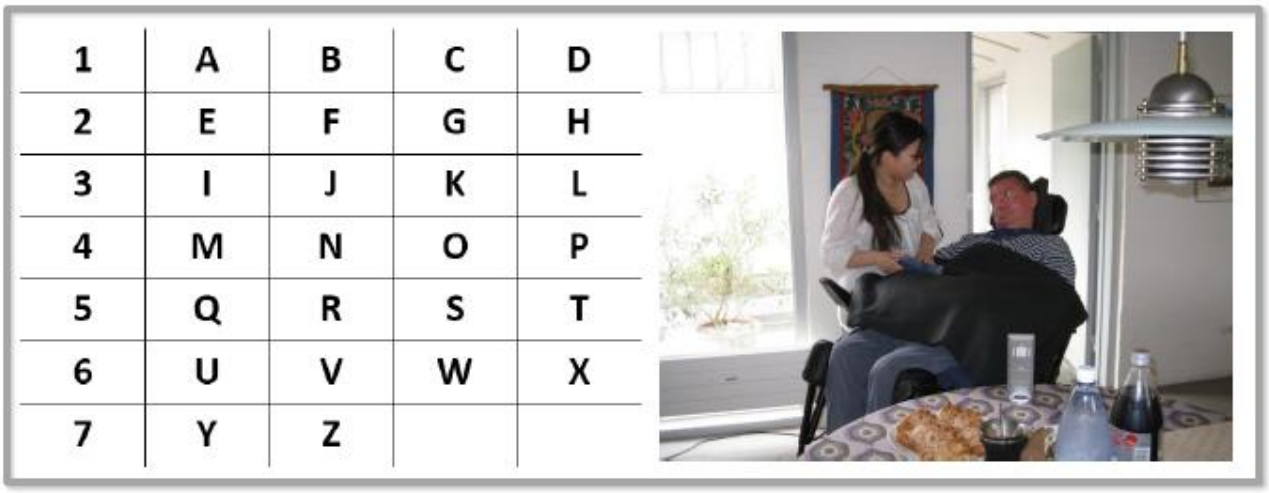


Figure 48: A possible visualization of a spelling grid and BBJ using the spelling grid with one of his carers

The person he wanted to communicate with identified the intended character by first listing the numbers of the groups allowing him to select a row (i.e., 1,2,3). The selection was made by completing a clear and distinct upward eye movement. Then the carer proceeded to go through the individual characters of that group until the preferred one was found – after which the process was repeated in order to select the next character, and so on.

This way the 'interpreter' and the 'communicator' spelled out each word. Again a large part of the communication was based on the interpreter's ability to predict meaning and finish sentences. This approach was very reminiscent of the scanning techniques used in some switch controlled interfaces (Shein et al., 1991; Istance et al., 1996).

The reason for the visit to BBJ was part of another project in which a low cost eye tracker was to be tested (San Agustin et al., 2010). The low cost eye tracker was set up in the participant's living room. The screen image was projected onto a smooth wall and the camera was placed in close proximity to the participant's eye and held in place on a fixed adjustable arm. The main focus of the trial was the calibration process and the participant's ability to interact with a gaze controlled application (in this case GazeTalk) (Figure 49).

Figure 49: BBJ using the low cost ITU gaze tracker.

Several calibration tests were conducted, which revealed useful information regarding the ITU gaze tracker and about the user's ability to interact with the system. The trial showed that completing a calibration was consistently achievable and subsequently it was shown that, even with a poor calibration, the participant was able to interact with the software application. However, a few attempts at using the application were needed before the right settings were found. The participant spent 20 minutes using the system adjusting, writing and correcting missed selections. After this he was too tired to continue.

During the trial he also kept communication going through the helper which allowed him to comment while he was using the system. His main comments regarding the system were that it would be an excellent solution for people with limited funds.

Two main observations were relevant in regard to the present work. At one point he commented that he believed that the difficulties he had calibrating could have been due to the fact that his eyes moved slower and lacked the precision they once had. This was an effect of late stage ALS, and emphasized the fact that systems and selection strategies created for people with motor impairments, especially those with degenerative conditions, need to be able to adapt to the changing capabilities of the end-user. A direct quote from BBJ as interpreted by his carer was: 'Det kunne bruges den gang mine øjne var gode' – ' It could have been used when my eyes were good'

The other observation of interest was given by his wife and was in regard to the set-up of the system rather than the system itself. The wall projected image made it possible for everyone in the room to follow what BBJ was saying, without the need for having it go through speech synthesis. She believed that this setup could be useful in social circumstances, such as Christmas or Birthday celebrations. It was quite clear that for interpersonal conversations the memorised grid alphabet was still preferred.

Just as MC had made a distinction between *global* and *local communication*; the distinction which was made here was between *social* and *interpersonal communication*.

4.2.3 3RD USER

Arne Lykke Larsen, who presented himself in the introduction, has had ALS for approximately 10 years. In the classification used here he is mainly a user of *mono-modal input*. He could also be a *bi-modal* user as he also to a certain extent had control over various facial expressions and slight head movements, as of last time we spoke. However, these potential physiological inputs were not used.

Arne had a breathing apparatus and was fed through a tube in his stomach. When asked if the introduction of his book could be used in the beginning of this thesis he wrote back and said: 'but that was written so many years ago; probably in 2004. Subsequently I have got much worse than described there, and also much better!' He was asked what eye tracking meant to him in his everyday life, this is his reply written in GazeTalk (Hansen et al., 2006):

'Eye tracking and gaze interaction means everything for me in my daily life. I am completely paralyzed by ALS/MND and cannot speak, so I use gaze interaction, in average, 10 hours a day. Gaze interaction and speech synthesis means that I can go to work every day, as associate professor in theoretical physics at the University of Southern Denmark, I can give talks at the university and at international conferences, I can speak with my secretary, with colleagues and students, I can discuss with other physicists, I can do computations and write scientific papers, I can write emails to my foreign collaborators and I can have an almost normal working life.

At home, it means that I can speak with my personal assistants, I can speak with family and friends, I can speak with the therapists that come to my home and I can live an almost normal

life. Last year, I had a book published about my life with ALS/MND; most of the book was written using gaze interaction. So gaze interaction means everything for me!'

An interview was conducted with Arne in October 2009. In preparation for this many wrong assumptions about how to conduct such an interview were made. Some large boards were created, one with 48 words which were believed to be relevant to the topic which was to be discussed, namely present and future implementation of gaze interaction technology. The other board had 28 images depicting different functions which were thought to potentially relate to future applications of gaze interaction (Figure 50). These sets of information were created with the intention of fulfilling two functions.

Primarily, the function was that of *mind maps* (T. Buzan & B. Buzan, 2006) which were intended to guide and inspire the conversation. The reasoning was to refer to numbers on one of the mind maps and then use the numbers system to group and prioritize the different elements. On the day of the interview each item was detachable from a board and could therefore be physically grouped.

The second function was to ease the burden of conversation. This was inspired by the work done by Buchholz et. al. (Buchholz & Holmqvist, 2009) in which they have developed some tools for assessing and communicating with people with complex needs.

Regardless of the intention, this turned out to be the completely wrong way of approaching the interview situation with this particular individual. A large part of the autonomy which Arne still has is based on his ability to maintain a conversation, in which he can freely respond and initiate topics.

Figure 50: The two mind maps intended to be used in the interview session with Arne Lykke Larsen.

It became clear that Arne had no interest in playing the mind map 'game' which was presented and with very good reason. First of all the work which inspired this approach (Buchholz & Holmqvist, 2009) dealt with people who suffered from both physiological and cognitive difficulties. Arne has all of his cognitive capabilities intact.

The second reason was that when he uses the eye tracker and GazeTalk with speech synthesis his language flows quite easily and it was in his interest to keep the conversation flowing without it being tied to the framework of mind maps. The most important lesson in regard to interview technique, which was made during this visit, was how not to be tied down by assumptions. The assumption made here, was that an interview was to be conducted with an individual with complex needs and the preparation was done accordingly.

In actual fact the interview was with a highly intelligent theoretical physicist, who did not see the point in tedious mind maps. However, a four hour conversation regarding the future of eye tracking was conducted and it was clear from this conversation that a few distinct wishes were important in regard to future implementations. The two main wishes from an ALS patient using an eye tracker daily were:

Removal of the screen: Even though he had the ability to carry on a conversation, the screen posed a divide between him and the person with which he was speaking. This was something

which was very inconvenient especially in social circumstances with more than one person, where the direction of gaze is usually used to signify attention and interest to a specific individual.

Mobility: Being able to control the wheelchair using gaze in a fluent and safe way. Arne had switches at the back of his head with which he could control the wheelchair, as he still had slight movement of his head. However, the strain of moving his head was much greater than that of moving his eyes, so gaze controlled mobility was of interest. The issue of removing the screen still applied, for instance – he did not see any point in being able to control a wheelchair when going through a park, if he was not able to actually look at the surrounding environment at the same time.

4.2.4 PRACTITIONERS

Other people who have greatly influenced the direction taken in this work have been the experts who work with the end-users on a daily basis, assessing, developing and evaluating systems as the technology evolves. Two institutions have in particular played a role. Primarily the ACE Centre in Oxford which is dedicated to dealing with individuals with complex needs, and has a great deal of experience in the field. Several discussion sessions were held with Mick Donegan at the ACE Centre among other places, over the past few years. At the time he was affiliated with the ACE centre and was one of the leading international practitioners in eye tracking and disabilities. These conversations dealt with four main topics: *diversity of needs*, *feedback*, *sustainability,* and *mobility*.

Diversity of needs had to do with the many different types of end-users which he had worked with over the years who could benefit from eye tracking. The acknowledgement that an individual with cerebral palsy, which would often be accompanied by varying degrees of spasticity and cognitive ability, has different capabilities and needs to those of an individual who has suffered a high spinal cord injury, which often constitutes a prolonged, but stable condition and which rarely affects cognitive ability. In turn the needs of individuals with spinal cord injuries are different from those of people with degenerative diseases such as ALS, where their capabilities change at varying points during the progression of the disease. When gaze controlled selection strategies and applications are developed they should ideally be able to cater to all of these situations.

The discussions on *feedback* dealt with supporting general navigation and conveying to the user where he/she was at any given point in the application. It is also important to convey what the consequences of a given activation could be and how to reverse that activation. If gaze was the primary source of input for a user it should be the main focus of any design to minimize ambiguity, to ensure smooth navigation and system response.

Sustainability was discussed as an issue of vital importance. Much of the research done in the area of gaze interaction focuses on selection speeds and optimizing processes (including the research which will subsequently be presented). This has been done mainly because matrices such as *selection speed*, *task completion time*, *words per minute* and *error rate* are quantifiable and measurable, which make them easy to analyze and draw conclusions from. A concept such as sustainability is more diffuse and requires a different set of matrices which can be subjective and must be observed over a period of continued use. *Fatigue*, *perceived ambiguity* of actions and *consistent repeatability* could be some of the future measurements of gaze interaction.

Finally, *mobility* was discussed on various occasions. For chronic wheelchair users there is a great need to be able to control movement. With users for whom gaze is the primary input, this is however a difficult task, due to the previously mentioned inaccuracy of gaze trackers and the nature of eye movements, which can be erratic. Any form of ambiguous instruction given to the system, can in the case of mobility have dire consequences.

The other institution which was visited was the West Midlands Rehabilitation Centre; where a meeting was held with Dr. Clive Thursfield who is a Consultant Clinical Scientist and an expert in electronic assistive technology. This discussion dealt with a range of existing technologies and possibilities for future implementation. The main points which were derived from this conversation was again the existence of diversity and the need for input devices to be able to be combined in order to create flexible solutions.

A discussion regarding the many possibilities for alternative inputs was held. Many of the existing technologies such as switch control, gaze control, blink control, speech recognition, EEG and EMG input, facial recognition, breath input etc. constitute separate research fields – with good reason, as the technology in many cases is still being developed. But from a practitioner perspective the real need was to be able to combine these different input

modalities to fit the capabilities of the individual. In other words, if a user has eye movement control as well as facial expressiveness, a combination of *gaze control*, *facial recognition* and *EMG switches* could provide the user with the kind of flexible input which could allow for control of complex systems. In regard to the previously stated taxonomy, the focus should be more on *bi-modal* and *multi-modal* gaze interaction.

4.2.5 GENERAL REFLECTIONS ON END USER NEEDS

These visits with end-users and experts, along with participation in conferences which have dealt with various topics regarding assistive technology, has given a perspective on not only what considerations must be taken into account when designing for users with special needs, but also what their goals and wishes are which allows the work which has been conducted here to be put into a larger context.

The information acquired during the visits and conversations with experts have resulted in an overall understanding of what the needs of end-users with Locked-in-Syndrome can be. Overall six key areas need addressing when developing for assistive technology and gaze interaction for people in a locked in state:

1. The ability to point at objects of interest in the user's environment.
2. The ability to interact with objects in the environment.
3. The desire to control a multitude of applications.
4. Taking special considerations when design peer-to-peer communication systems
5. Facilitating easy digital communication
6. Creating gaze controlled wheelchair control (Figure 51).

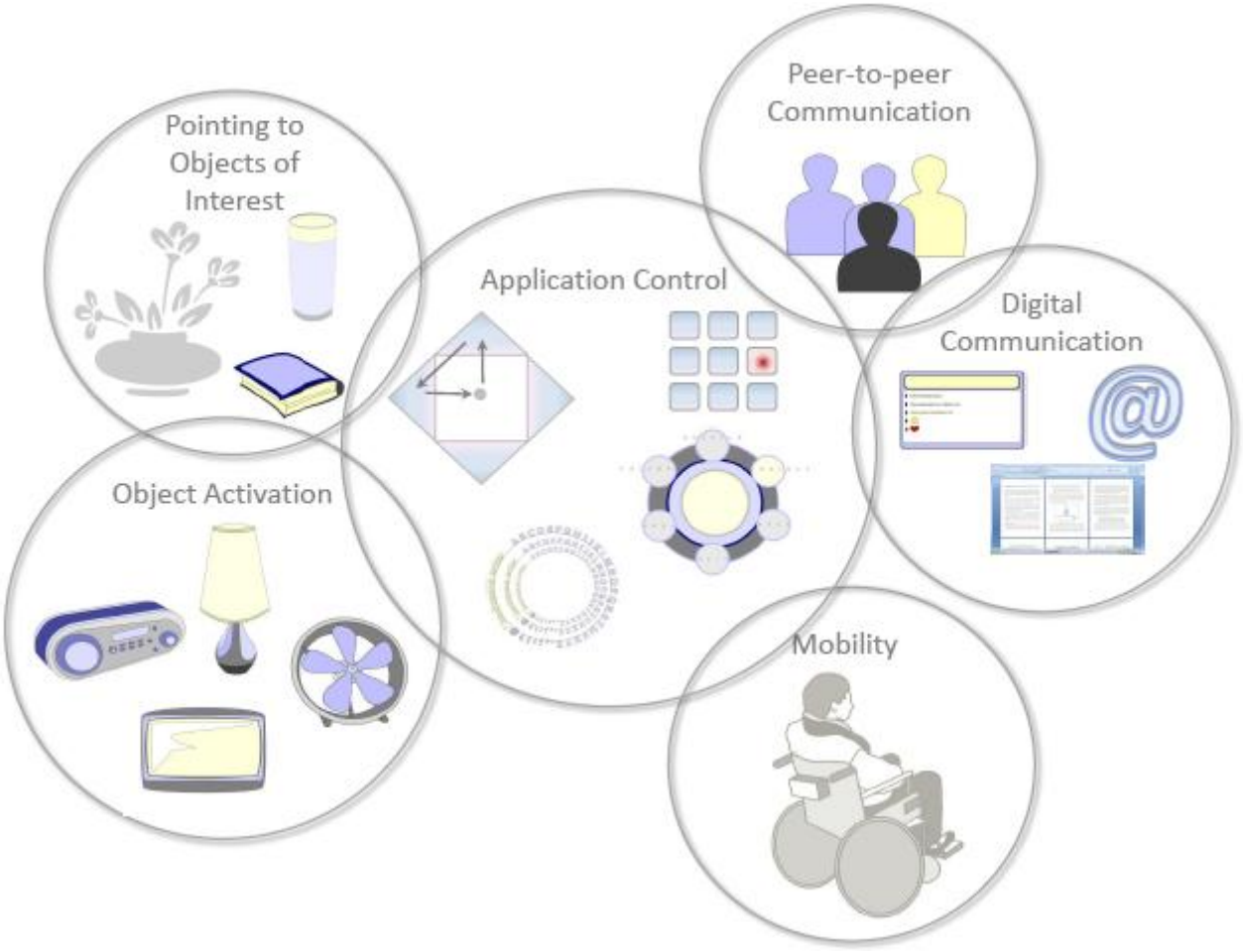


Figure 51: The 6 key areas which must be considered when developing gaze interaction.

Pointing at inanimate objects in the environment is a useful and simple way of inferring meaning on to inanimate objects or to guide the hand and sight of someone else. Pointing to a glass of water can infer the meaning of being thirsty. Pointing to a shelf in a bookcase can guide the hand of someone to pick up a specific book or CD. Eye tracking systems which place the gaze direction of the user in their direct environment, such as the ART system, could facilitate direct object eye pointing (Shi et al., 2006a). If every item in the user's environment could be recognised by image recognition the user could simply eye point (look) at an item of interest and the system could convey this interest to the user's surroundings. This form of interaction could also translate into environmental control tasks, such as turning on the TV or a lamp simply by looking at the object.

Communication facilitated with gaze, such as the earlier mentioned 'Type-to-talk' – systems, allow users to communicate solely by gaze. These systems all interpret the intention of the

user and some use speech synthesis to convey this intention to the user's surroundings. However, a need was expressed to make peer-to-peer interaction not as dependent on a screen and also the low-tech version of gaze communication had some definite advantages, such as the human prediction abilities, which have yet not been matched digitally. The users also expressed an interest in being able to use a multitude of applications, mainly work-related applications which would allow them to continue their jobs; as well as the wish for entertainment in the form of computer games and web-browsing. Finally, there was an interest in being able to control a wheelchair, but particularly without it infringing on the user's view of the environment in front of, and around, them.

In order to accommodate all of these wishes, gaze interaction must become more flexible, sustainable and reliable. The reliability is down to the technological advances in the field of developing eye trackers and beyond the scope of this work.

However, increasing the vocabulary of gaze selection strategies can allow a higher level of complexity in applications and provide more flexible interaction and, therefore, a wider audience assisting with the cost factors mentioned earlier.

The intention has been to develop selection strategies which could be integrated with dwell time to enable sustainable gaze interaction solutions.

The research which will subsequently be presented is in regard to the concept of *single stroke gaze gestures* (SSGG) for activation control. They could be used for various global commands, e.g. controlling zoom interfaces, flicking through websites etc. If gesture based interaction is to be implemented successfully, the way information is structured and visualized needs to be rethought.

Research presented in this chapter has been published:

Mollenbach, E., A. G Gale, and J. P Hansen. "The Assistive Eye: An inclusive design approach." *Contemporary Ergonomics* (2008).

Mollenbach, E., J. P Hansen, and A. G Gale. "From On-screen Navigation to Real World Interaction," The Scandinavian Workshop on Applied Eye-Tracking**,** 2008.

San Agustin, J., H. Skovsgaard, E. Mollenbach, M. Barret, M. Tall, D. W Hansen, and J. P Hansen. “Evaluation of a low-cost open-source gaze tracker.” In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, 77–80, 2010.

5 PILOT STUDIES ON SINGLE GAZE GESTURES

In Chapter 3 *strokes* were established as the basic building block of gaze gestures. A further distinction that needs to be made is between: *single stroke gaze gestures* (SSGG) and *complex gaze gesture* (Figure 52). A SSGG is defined here as the motion between two intended fixations to complete activation. The complex gaze gesture is in this context defined as the motion between three or more intended fixation points. Earlier the three main principles of gaze gestures were defined as being stroke complexity, finite, and continuous gaze gestures. Figure 53 shows examples of finite SSGG and finite complex gaze gestures.

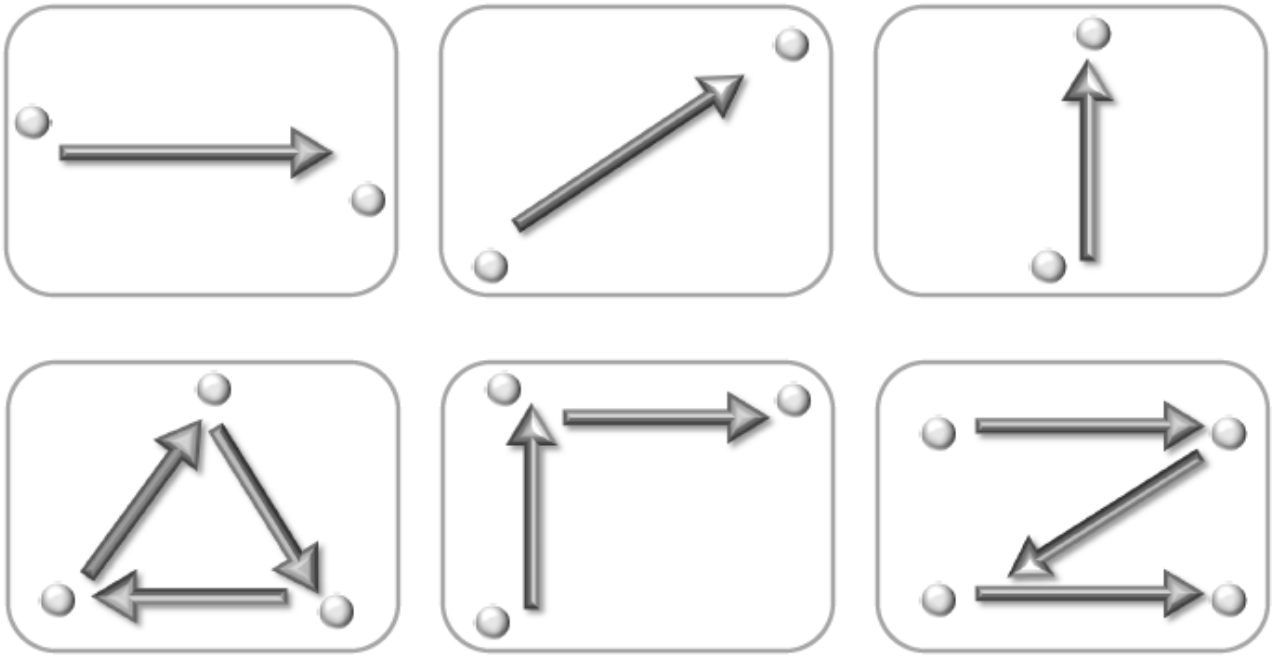


Figure 52: Illustration of single stroke gaze gestures and finite complex gestures.

Complex gaze gestures have the advantage of greatly increasing the interaction vocabulary of gaze. However, this brings with it both cognitive and physiological difficulties. Cognitively it is difficult to remember a large number of gestures and physiologically it is difficult to create and complete them (Porta & Turina, 2008a). SSGG have the disadvantage of being 'single'; like dwell selection, they represent one action. The great advantage of the SSGG is that it is easily learnt and sustained; both cognitively and physiologically it requires less of its user than both dwell selection and complex gaze gestures.

Continuous visualization-based gaze gestures are often a combination between fixations, saccades and smooth pursuits, and could therefore be construed as complex in nature. However, SSGG can be an intricate part of the selection process used to signify specific actions; by having the SSGG occur at specific places in the layout. The pEYE's system is the clearest example of this, where a SSGG across the border of a pie instigates an expansion of the system. The same activation is used in the IWrite interface, where a SSGG across the outer rim confirms the selection (Urbina & Huckauf, 2007) and this is also the case for the contextual switching presented by Morimoto and Amir (Morimoto & Amir, 2010).

Stroke gestures on mobile devices, such as finger-strokes in different directions, have become extremely popular because they are easily achieved and sustained even in the relatively noisy context of mobile use. Translating the principle of stroke gestures into a gaze selection method has been the focus of this research.

Single stroke gaze gestures could be used for overall navigation (i.e., switching from a communication application to an environmental control application). This would be equivalent to toggling in Windows that is performed by the 'Ctrl+tab' function or the 'Spaces' functionality on the Mac that allows users to switch between various desktop environments. SSGG could also be implemented as various forms of modal commands; such as simulating the Enter-key or spacebar, whose designated action depends on which mode the system is in (Raskin, 2000). An example of a mode dependant command is the spacebar causing either a gap between words in text editing applications or a 'jump' action in some avatar based computer games (e.g. World of Warcraft).

In the field of finite gaze gesture research no one has looked at SSGG as a selection method, even though single stroke gaze gestures are the basic components of all finite gaze gestures. The main reason for the lack of research is, most likely, the assumption that they are too error-prone. There could be a high correlation between natural search patterns and the single point-to-point gaze gesture. This correlation between natural visual search patterns and gaze gesture activation is the equivalent of the Midas touch issue for dwell selection, as explained in chapter 2 and 3. This potential overlap is referred to as accidental gesture completion. Accidental gesture completion is defined as the potential correlation between perception and selection eye movements, which can result in faulty selections. For gaze interaction purposes

it is desirable to have a low level of accidental gesture completion thereby minimizing the number of erroneous selections.

In subsequent sections research on the SSGG will be presented. This will contribute to the understanding of whether or not SSGG can be employed as a selection method and what the innate properties of this interaction element are. In Chapter 3 the potential advantages of gaze gestures were presented as being: *Speed*, *freeing up on-screen real-estate*, and *avoiding the Midas touch problem*.

5.1 INITIAL PILOT STUDY

5.1.1 DESIGN CONSIDERATIONS

The goal of this pilot study was to investigate whether SSGG represented a feasible selection strategy, what kind of feedback the user might need, and to examine whether selection times could be comparable to standard dwell selection completion times (i.e., 450-1000ms). Two pilot studies were conducted in this regard; the first is presented in this section.

Two main design decisions regarding the layout of the interface were made. First of all the centre of the screen was unaffected by gaze. Secondly, the initial point of gaze in itself should have no effect on the system. The solution became to have only a rim around the edge of the screen that was affected by gaze. This idea was further substantiated in research done by Tatler who found a bias tendency to fixate at the centre of the screen, regardless of saliency patterns (Tatler, 2007). In other words, when viewing images on a screen the periphery is rarely inspected, even with high saliency objects placed there. As a consequence, using the periphery of the screen, particularly in a sequence of multiple fixations, can potentially decrease the amount of accidental selections, because fixations rarely occur there. Furthermore, search patterns rarely initiate from there, thereby decreasing the risk of accidental gesture completion.

Most of the initial design considerations in the pilot studies were in regard to feedback; how to provide feedback to the user as to what was being selected. Initially, different rim designs were considered, and in particular rim complexity was of interest. The idea was that several

fields could exist on each side of the frame and that combining these fields in various ways could result in different activations (Figure 53)

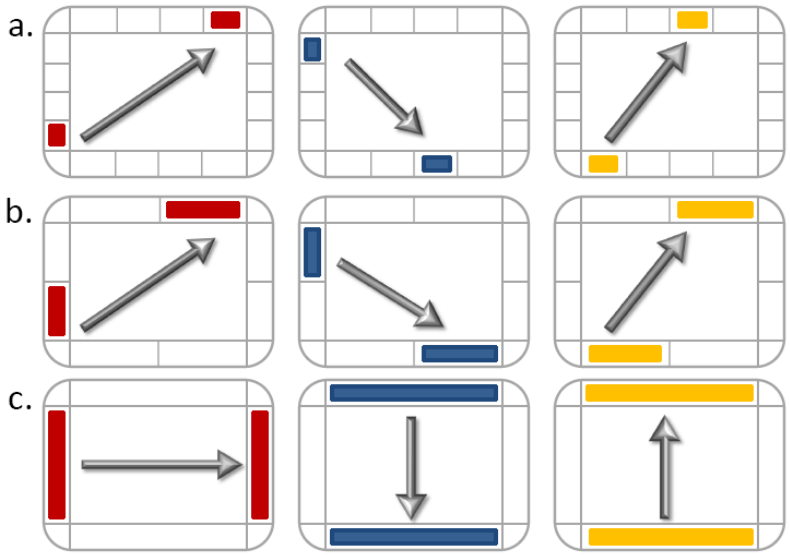


Figure 53: (a.) shows an example of a rim design that had four fields on each side. (b.) an example of a two field rim design. (c.) the single rim design

The basic concept was that a selection could be initiated in any field. Depending on the subsequent completion field the consequences of the SSGG would differ.

One of the main challenges, in regard to feedback, was to provide navigational information to the user both in both the central and peripheral vision, so that the user would be clear on the consequences of a given action. In other words, because the selection was to be completed by looking from one edge of the screen to another, it was important for the user to be aware at the initial point of gaze what the consequences of a given gesture would be. In order to facilitate this, a local indicator that was intended to afford a small local overview of possible selections was designed.

Figure 54 shows the local indicator design visualized in the initial layout. (a.) The full interface is shown with the red circle indicating where the user is looking. The arrows show that from this initial selection field three potential activations are possible: yellow, red and blue. (b.) Shows the local indicator, which skates along the border of whatever field was being viewed. This is intended to grant a central vision overview. (c.) A stationary indicator was also present at the potential completion field, to provide the user with a peripheral mark to follow.

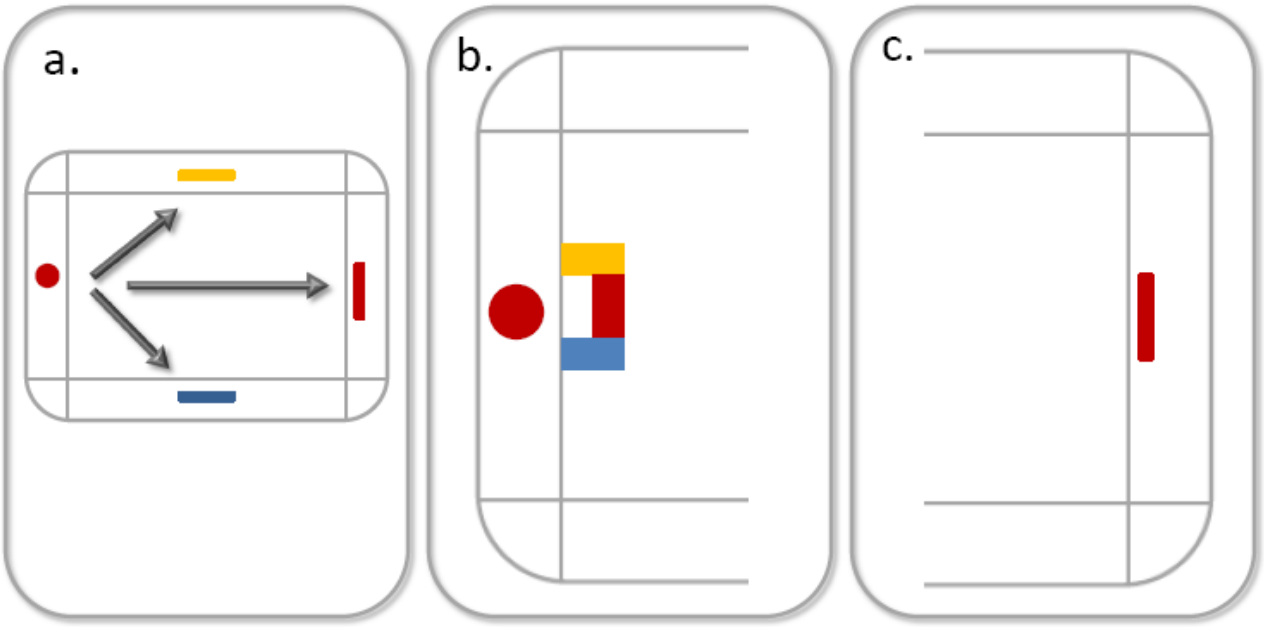


Figure 54: Visualization of initial design considerations: a. is the complete interface; b. is the local indicator; c. is the peripheral indicator.

These were some of the initial ideas; however, not all of them made their way into the experimental implementation. The design of *multiple initiation fields*, with *multiple completion fields* was abandoned; as it would move the focus of the investigation from exploration of an interaction principle to the design of a particular layout. It was therefore decided to keep the design as simple as possible. Only *one* initiation field per side and only *one* completion field per initiation field (directly opposite) was used.

5.1.2 DESIGN IMPLEMENTATION

The actual test environment consisted of the unaffected centre of the screen in which tasks were to be displayed, and four single *initiation* and *completion* fields at the edge of the screen. Three feedback principles were implemented: the *permanent initiation indicator*, the *central vision indicator* and the *peripheral vision indicator* (Figure 55).

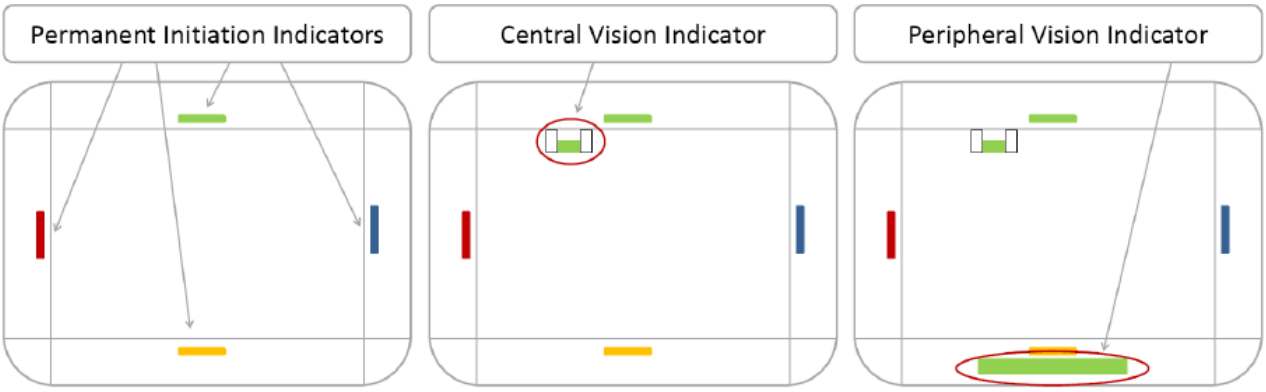


Figure 55: Three feedback indicators used: Permanent initiation feedback, central vision feedback, peripheral vision feedback.

The *permanent initiation indicator* showed which colour could be selected from that initiation point. In other words, the first step to select the colour red was to look into the left field, to select green an initial glance in the top field was required, blue was initiated to the right and yellow in the bottom field.

When an *initiation field* was glanced at two other feedback processes would occur. First, a *central vision indicator* would appear and skate along the border of the selection field. Even though the idea of multiple completion points had been abandoned, the design of the local indicator was the same. It indicated what side of the screen should be looked at and what selection the particular SSGG would complete. The second feedback was the *peripheral vision indicator*. When an initiation field was glanced upon a matching coloured square would appear at the opposite side of the screen, and would remain there until the selection was either completed or cancelled. This was intended to provide a peripheral directional guide.

There are two reasons why colour selection was at the foundation of these tasks. Colours are abstract and carry very little intrinsic symbolism. The distinction between the early abstract research done (Ware & Mikaelian, 1986; Jacob, 1991a) and the more contextual, application driven approaches seen in recent research (Istance et al., 2010; Hansen et al., 2006; Majaranta & K. J Räihä, 2002) was introduced in Chapter 3. There are advantages to both approaches; viewing a selection strategy derived of any known context minimizes the amount of confounding variables. However, by not viewing the selection strategy in a well-known context it becomes very difficult to make comparisons with existing methods. Because the selection strategy was new the approach taken was more fundamental. Therefore an abstract

exploratory design approach was taken in this research, where the focus remained primarily on the selection strategy itself rather than on finding task-specific solutions.

The nature of the task in this experiment was, as mentioned, abstract. Four different colours could be selected by the four different SSGG. An eye movement from the left to right side of the screen caused the colour red to be selected. Moving the eyes from top to bottom resulted in a green selection; from right to left the colour blue would be selected and from bottom to top selected the colour yellow. Figure 56 shows the initial task screen, where 8 different coloured boxes were presented to the participant. Selections were made in the sequence in which the coloured boxes were presented.

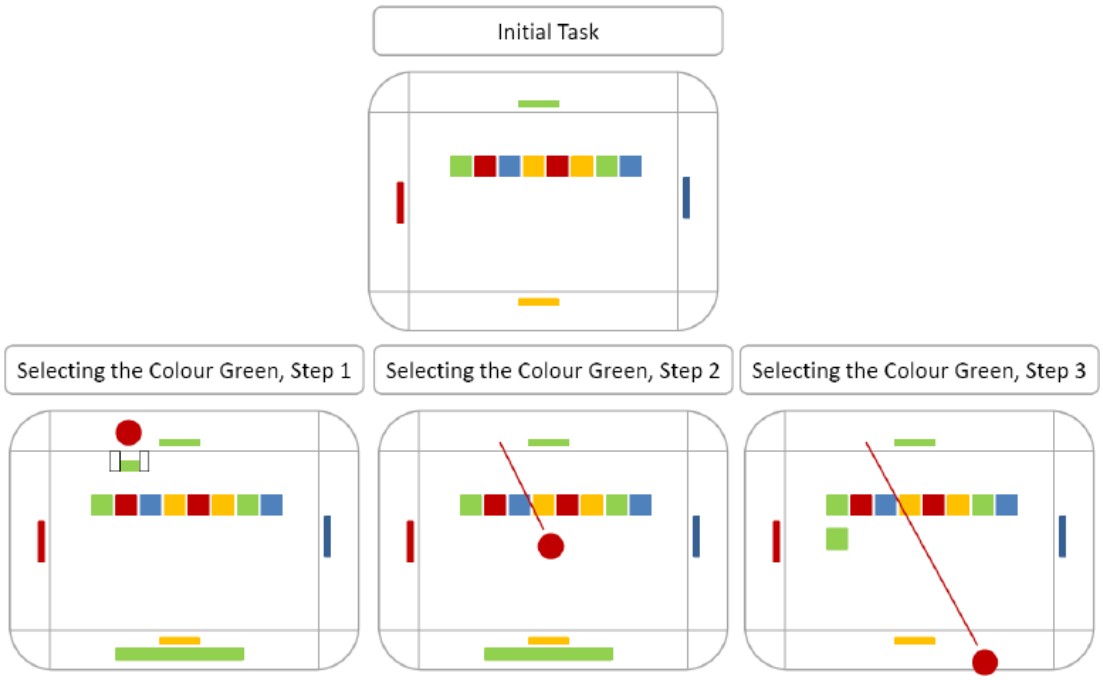


Figure 56: The initial *task screen* and a visualization of the selection process of the colour green

A visualization of the selection process of the colour green is also shown:

- Step 1: the user looks at the green initiation field, the central vision indicator and the peripheral indicator were both on.
- Step 2: gaze would move out of the initiation field and the central vision indicator would be turned off. An initiated action was set to be completed within 1000ms. At this point 3 possible actions could be taken:

- The first was to simply remain in the centre of the screen until the 1000ms timer ran out. The peripheral indicator was visible while the timer was running, so the participant could know when the timer had run out and the selection was cancelled.
- The second course of action was to enter one of the adjacent fields or re-enter the green field, in which case the timer would be reset, the peripheral indicator turned off and the current activation would be cancelled.
- The third and final option was to complete the selection within the 1000ms by looking to the opposite selection field.

- Step 3, if the selection was completed and it corresponded to the colour in sequence a small box appeared under the just selected colour.

Another visualization of this is shown in figure 57, where the colour red is selected as the 5th selection.

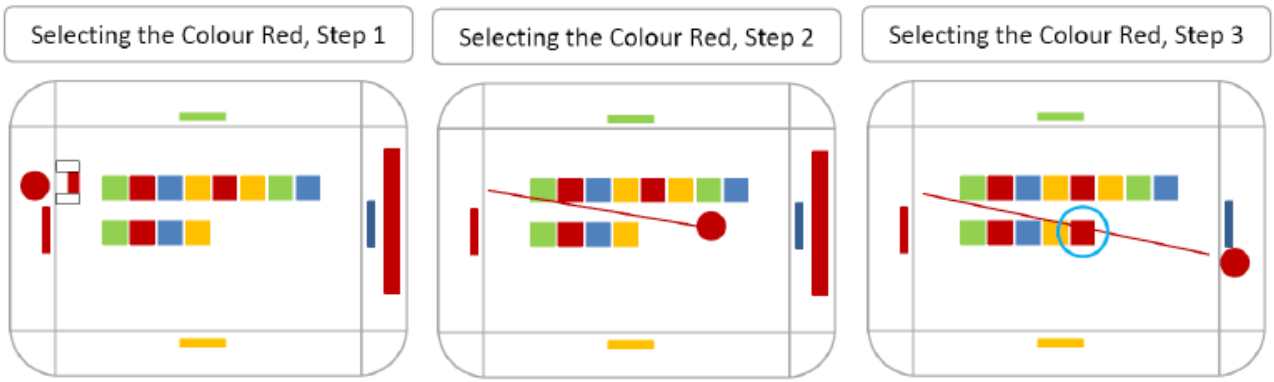


Figure 57: Selecting the colour red as the fifth selection.

5.1.3 EXPERIMENTAL DESIGN AND RESULTS

The task was to select eight colours in the same sequence. The application was written in Java using the Eclipse editor. The experiment was conducted using a QuickGlance 3 eye tracker with a 13 inch monitor and a screen-resolution of 1024x768 running at 20 frames/sec. Two participants completed the initial pilot study in which they selected the sequence of 8 colours 10 times. Each direction was completed twice for every eight selections. Making a grand total of n=160 observations and n=40 per selection direction. *Selection completion times* were registered as the time which elapsed between exiting a selection field and entering a completion field. The boundaries of the selection fields were defined by coordinates. So in

other words *selection completion times* could be described as lasting from the last gaze coordinates of an initiation field to the registration of the first gaze coordinates in the completion field. A timer event was fired when the boundary of an initiation field was crossed and ended when gaze moved across the boundary of the completion field, the elapsed time was registered (Figure 58).

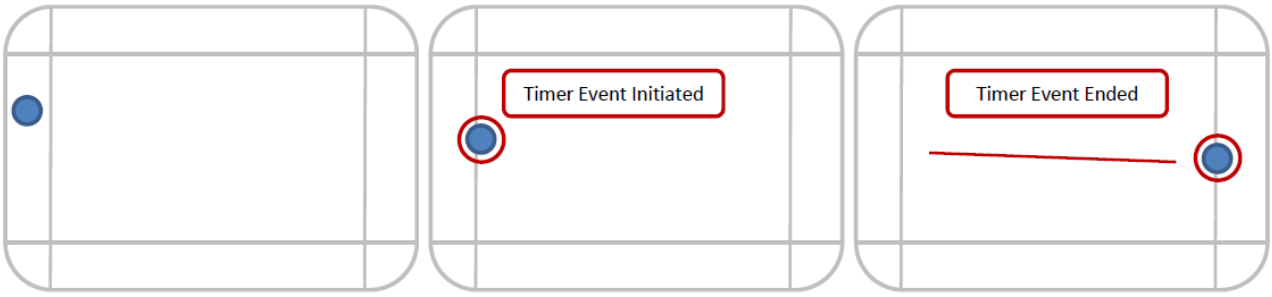


Figure 58: The timer event for *selection completion time*.

If a gaze gesture was initiated and the subject subsequently looked at an adjacent field, the selection process would be cancelled and the system reset. If after initiation the subject simply looked at the centre of the screen for more than 1000ms the selection was cancelled and the system reset (Figure 59).

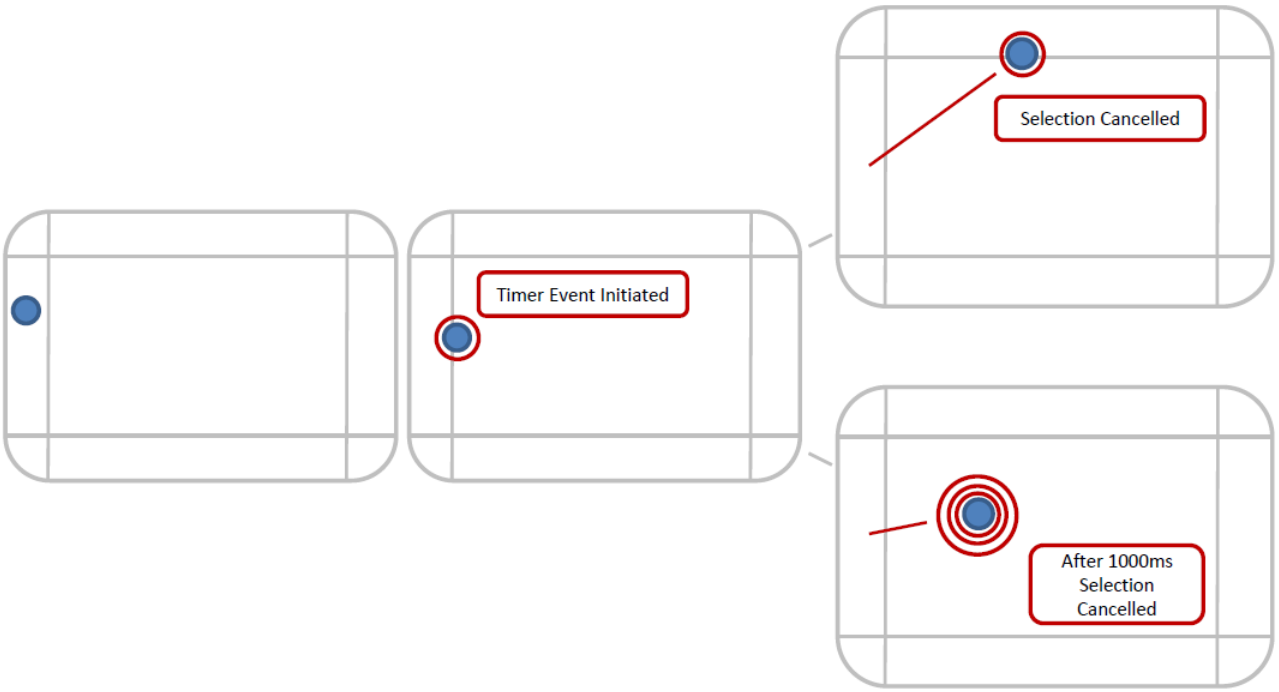


Figure 59: The process of selection cancelation.

The two null hypotheses were:

1. There was no difference between *selection completion times* for SSGG in four different directions (Right/Left; Left/Right; Top/Bottom; Bottom/Top).
2. There was no difference between overall *horizontal* and *vertical* SSGG. It was also of interest to see how the grand mean of this type of selection would compare to an assumed dwell selection time of 400ms.

A factorial analysis was conducted. The independent variable was *selection direction* with four levels (Left/Right; Right/Left, Top/Bottom, Bottom/Top). *Selection completion time* was the dependent variable measured in milliseconds (ms).

The data were then examined to see whether or not the distribution was normal. A Kolmogorov-Smirnov test showed that the data was negatively skewed and significantly deviated from a normal distribution. Logarithmic-, squared- and reciprocal transformations were attempted but the assumption for normal distribution which is (among others) required when using parametric statistics was violated. Also Maucley's test showed that sphericity could not be assumed and this was further substantiated by the Greenhouse-Geisser test where $\varepsilon = 0.543$ (this is a measure between 0 and 1, the further from 1 it is the more likely that sphericity cannot be assumed, even with a corrected *F*-ratio). For these reasons a non-parametric analysis was chosen instead[12].

OVERALL DIRECTIONAL EFFECT

After ranking the original scores, Friedman's non-parametric ANOVA for repeated measures was used to evaluate the differences in *selection completion time* depending on *selection direction* (Figure 60). The four mean values based on n=40 observations were (Table 1):

[12] This is the statistical methodological approach which will be used to analyse all data in this thesis: The data will be explored and the appropriate statistical methods applied. However, the description of the process will not always be as elaborate as it will be for these initial studies. Also all plots will be based on mean values, with error bars representing standard error of mean.

Right / Left	Left /Right	Top /Bottom	Bottom/Top
333ms	349ms	224ms	330ms

Table 1: Overall directional mean values of *selection completion times*.

There was an overall significant effect between *selection completion times* in the four different directions of left/right; top/bottom; right/left and bottom/top. $X^2_r = 60{,}791$ (3, n=40), $p < 0.01$. A Bonferonni adjusted post hoc analysis revealed that only the top/down motion was significantly faster than the other directions.

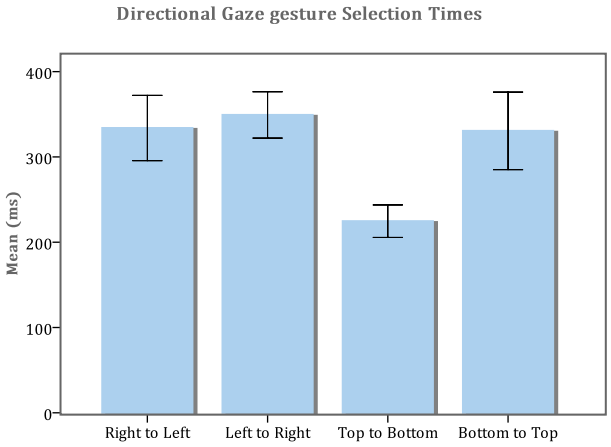


Figure 60: Analyses of selection completion time based on selection direction. Error bars represent standard error of mean

HORIZONTAL AND VERTICAL EFFECT

After ranking, the Wilcoxon signed-ranks test was used. Overall horizontal and vertical *selection direction* observations were compared based on *selection completion time* measured in ms (Figure 61).

The two mean observations were based on n = 80 observations per result were: Horizontal = 341ms and Vertical = 277ms.

There was a significant difference between the overall horizontal and vertical conditions $Z = -5{,}049$; $p < .001$. The *horizontal* SSGG were significantly slower than the *vertical* SSGG.

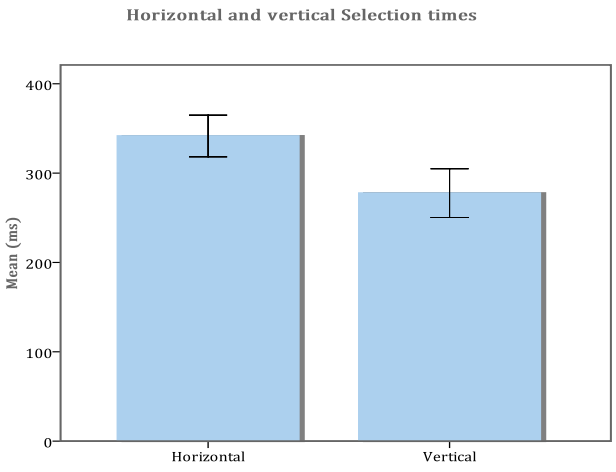


Figure 61: Analysis of overall Horizontal and Vertical gaze gestures based on Selection completion time. Error bars represent standard error of mean

The *selection times* were considered in regard to an estimated 450ms dwell selection threshold. The grand mean was 309ms with a median value of 266ms. These both fall below the 450ms threshold. The 25th percentile was 204ms and the 75th percentile was 344ms.

DISCUSSION

Both null hypotheses were rejected, as there was a significant difference in directional *selection completion times*, both within the four different directions and within horizontal and vertical selections. The first observation was that completing a gesture from the top to the bottom of the screen was significantly faster than any other SSGG. Whether this trend would be systemically significant would require a larger scaled study. Determining which directions of gestures are most effective has potential implications on how gaze gestures and their related interfaces should be implemented.

The second observation made was in regard to general direction. These initial results showed a significant difference between vertical and horizontal gestures, in favour of the vertical eye movements. This was a carryover effect from the top/down gaze gesture being so fast. However, it could also have been the result of the difference in distance needing to be covered: the screen-resolution was 1024x768, which meant that horizontal movements were longer than vertical ones.

Finally, the general result of 310ms was interesting for two reasons. First of all it was relatively fast compared to an existing dwell-time selection standard of approximately 400ms-1000ms, and there were several indications that the selection time could improve further. The fastest selection was 125ms and the slowest was 734ms. Because of these outliers it was relevant to also look at the median value instead, which was 266ms. Also, in the context of saccadic eye movements, using approximately 300ms to move from one side of the screen to the other, within 10˚ of visual angle, was a very long time. There could be several reasons for this; the relatively low frame-rate of the QuickGlance system (20 frames/sec) could be a factor. The program was using a Windows timer to register timer events and this could be causing a delay in the registration of entry and exit points. Also participants could be making multiple saccades during the course of completing activation.

Even though it was only a small, preliminary study, the conclusion was that there were enough issues of interest to warrant further investigation.

The next step was to conduct a larger scaled experiment in the same test environment, in order to investigate whether or not the initial trends found in the pilot could be substantiated. Also a larger scaled experiment could give insight into whether or not this experimental design was the most appropriate way of assessing SSGG.

5.2.1 EXPERIMENTAL DESIGN AND RESULTS

12 participants (5 female) participated in the study. The task was to complete eight SSGG selections in a specific order repeatedly; the order of the sequence was the same for each repetition. Initially, the task was completed 5 times for practice after which it was completed 10 times. This resulted in n=80 observations per participant and N = 960 for the total dataset. The experiment was conducted using the same QuickGlance 3 eye tracker with 13 inch monitor and a screen-resolution of 1024x768, running at 20 frames/sec. After the experiment the users were asked which activation direction they preferred.

The two null hypotheses were:

1. There was no difference in *selection completion time* for the four directions of SSGG.
2. There was no overall difference between *horizontal* and *vertical* SSGG. Again it was of interest to see if SSGG would be faster than an assumed dwell time selection of 450ms.

A factorial analysis similar to the previous experiment was carried out.

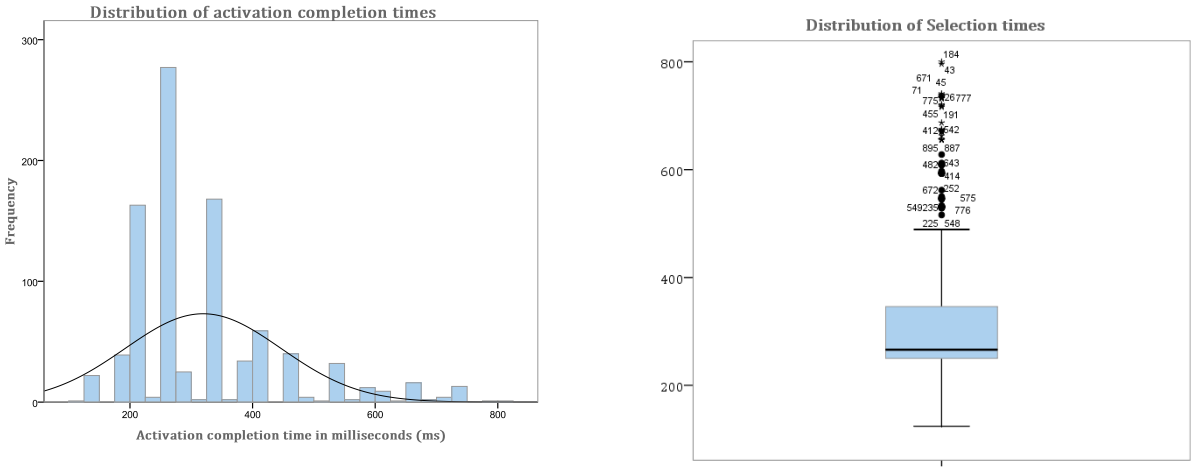


Figure 62: Distributions of scores of selection completion times. To the left, a histogram of scores, with a distribution curve. To the right, a box plot with the median displayed as the black line and outliers at the top.

A Kolmogorov-Smirnov test confirmed what was already apparent from the histogram and box plot above (Figure 62); which was that the data were negatively skewed and significantly deviated from a normal distribution. Logarithmic-, squared- and reciprocal transformations were attempted but the assumption for normal distribution was violated. Sphericity could also not be assumed as Mauchley's test was p <.05. However, the Greenhouse-Geisser correction was $\varepsilon = 0.955$ which means that a corrected F-ratio could have been used, if the data had been normally distributed. Due to these circumstances non-parametric tests were also used to analyse these data.

OVERALL DIRECTIONAL COMPARISONS

The data were analyzed in regard to *selection directions* after ranking, Friedman's non-parametric repeated measured ANOVA with a Bonforonni adjusted post hoc comparison was used; *selection direction* was the main factor and *selection completion time* the dependent variable measured in milliseconds (Figure 63).

The four mean values based on n=240 observations were (Table 2):

Right / Left:	Left /Right	Top /Bottom	Bottom/Top
337ms	330ms	355ms	312ms

Table 2: Mean *selection completion time* base on *selection direction*.

There was an overall significant effect between *selection completion times* in the four different *selection directions*. $X^2_r = 31{,}455$ (3, n=240), $p < 0.01$. A Bonferonni adjusted post hoc analysis revealed that the top/down motion was significantly faster than the bottom/top activation. There was no significant difference between any of the other *selection directions*.

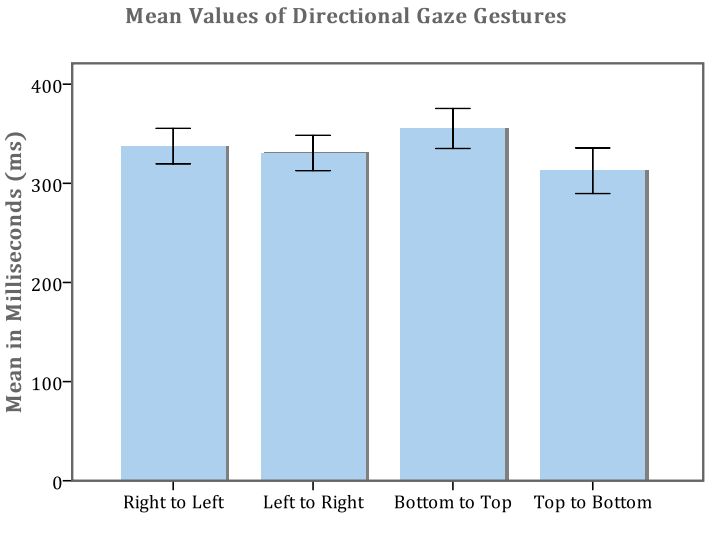


Figure 63: Analyses of *selection completion time* based on *selection direction.* Error bars represent standard error of mean.

HORIZONTAL AND VERTICAL COMPARISON

The data were again ranked and analysed using the Wilcoxon signed-ranks test to compare the overall *selection completion times* of *horizontal* and *vertical* SSGG (Figure 64).

The two observations over mean values based on n = 480 observations per were: *Horizontal* = 334, 07ms and *vertical* = 334, 02ms.

There was no significant difference between the overall horizontal and vertical conditions $Z = -0{,}778$; $p = 0.437$.

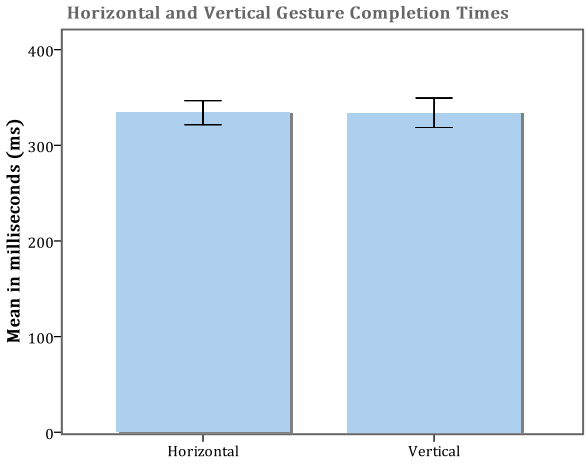
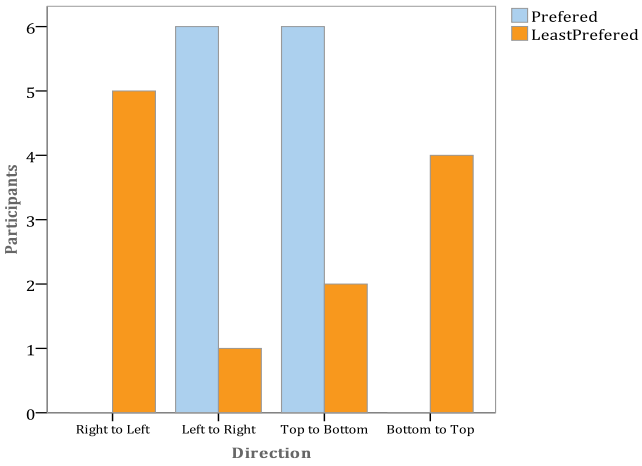


Figure 64: Analysis of overall Horizontal and Vertical gaze gestures based on Selection completion time. Error bars represent standard error of mean

PREFERENCES

The participants were presented with questionnaires and asked about their preference of direction. They were asked which selection direction was their least and most preferred (Figure 65).



Of the 12 participants who took part in this experiment, none preferred the Bottom/Top motion or the Right/Left motion either. They were split down the middle with regard to whether they preferred the Top/Bottom or Left/Right gestures.

Figure 65: User preferences based on *selection direction.*

S*election completion times* were evaluated in regard to an estimated 450ms dwell time threshold. The mean (μ) value was 334ms with a median value of 267ms which both fell below the 450ms threshold. The 25th percentile was 250ms and the 75th percentile was 391ms. Overall, 83% of all selection times fell below the chosen 450ms threshold.

5.3 DISCUSSION

The first null hypothesis was rejected as there was a significant difference in directional activation completion times within the four different directions, with the difference occurring mainly in regard to top/bottom gesture.

The second null hypothesis is retained as there was no significant difference between overall horizontal and vertical selections.

The first observation was that the top to bottom SSGG was significantly faster than the other directions. This mirrored the results found in the initial pilot study and could indicate that this activation could be assigned to actions which require rapid repetition. Half of the participants also identified this as their preferred gaze gesture direction; which was indicative of this potentially being an effective gesture direction.

The second observation was that there was no overall significant difference between horizontal and vertical SSGG. This indicates that there might not be any real benefits in implementing particular actions to different directions depending on frequency of use.

The preferences found here reflect the reading directions in western culture. It could be interesting to examine the preferences of Chinese or Arabic users as these could be different. At least there is some indication that there might be an attention bias based on lateralized brain functions. In other words certain gaze patterns could be preferred based on conditional responses that are either culturally or physiologically induced.

The grand mean for all SSGG was 334ms and the median was 267ms. both these results indicate that as a gaze selection strategy then SSGG can match dwell selection speeds of 450ms. However, the question still remains from a physiological point of view: why were they not faster?

There are some issues with the experimental design that need to be taken into account when understanding the results above and also in understanding the design of the subsequent experiments. What was learnt during the two pilot studies was first and foremost an understanding of how to design, conduct and analyse this type of scientific experiment. Because of the learning curve involved, the validity of the actual results should be questioned, as some of the choices regarding especially the design of the experimental test-bed were erroneous to begin with.

5.3.1 PROBLEMS CONCERNING THE EXPERIMENTAL DESIGN

There were several design issues that will subsequently be described and that formed the foundation of the next iteration of experiments.

A main concern in regard to the experimental design was that the gaze gesture selection had not been compared directly with dwell-time selection within the same task framework. In other words the results of this experiment were compared to an estimated dwell-time average of 450ms, which was derived from literary references and practitioners' advice. However, this could be viewed as an unfair comparison as a faster dwell-time might have been sufficient in this task. How to compare dwell-time and gaze gesture selection in a meaningful way is a general question throughout this research.

Another experimental design issue was that the order of selections was the same each time the user repeated the task, this could have caused a bias towards certain selection patterns, as these could have been based on memory rather than discovery and some selection patterns could have been easier to remember and complete than others. It was impossible to conclude whether the initial significant effect found in regard to the Top/Bottom gestures was real or whether it was due to the placement of the target in the sequence. Also, any abstract comparison to real world tasks is difficult, as most applications require more navigation and discovery, for instance looking at documents, websites, using a word processing application, game etc.

The decision to leave out dynamics in the task, in other words nothing moved or required the user to explore the entire screen, meant that the SSGG activations were done with virtually no interference from natural exploratory eye movements. Again, most real world tasks have

elements of orientation and navigation unless the user is working in a very well known interface.

Some design choices turned out to be unnecessary and created more confusion than clarity. The local indicator that was implemented to give people an indication of potential moves within their central field of vision was unnecessary as the actions are simple. Also the fact that the target zone lit up to indicate when and where selections were possible (peripheral indicator) seem needless again because of the simplicity of the actions.

The major advantage of SSGG was their simplicity; creating an overly complicated interface defeats the purpose. Also, a critical interface mistake was made in allowing the cursor to be visible. This would occasionally cause the user to follow the cursor rather than focus on the intended fixation point. Furthermore, the initiation and completion fields were of equal thickness but as mentioned earlier the resolution was 1024x768 requiring longer strokes for the horizontal selections as the distance covered was longer.

There was no error measure in this initial design. This made it impossible to examine how often faulty selections occurred due to accidental gesture completion. As this is the equivalent of the Midas touch problem for gaze gestures, this must be included in any thorough examination of gaze gestures.

There were also technical difficulties. The eye tracker used had a frame-rate of approximately 20 frames per sec, which was relatively low by modern standards. Also, there was quite a lot of difficulty in calibrating people, calibration was attempted with 20 people, but only 12 actually managed to complete the experiment.

Other experiments will be presented which have sought to address the issues presented in this last section. Even though these reflections indicate that the results presented in this chapter are not sufficiently supported by the experimental design; the experience of conducting the experiment, yielded a great deal of information both qualitatively and quantitatively, which have benefitted future designs which are described in subsequent chapters.

Elements of the research presented in this chapter were published in:

Mollenbach, E., J. P Hansen, M. Lillholm, and A. G Gale. “Single stroke gaze gestures.” In *Proceedings of the 27th international CHI conference on Human factors in computing systems*, 4555–4560, 2009.

6 SINGLE STROKE GESTURES AND DWELL TIME SELECTION

This experiment was designed to facilitate a direct comparison between two types of *single stroke gaze gesture* (short and long), and *dwell selection*.

6.1 EXPERIMENTAL DESIGN CHANGES

The main intention was to develop an experimental test environment which took into consideration the previously mentioned shortcomings of the initial design. The elements which needed correcting were:

- Simplifying the interface and removing the cursor
- Introducing dynamics in the selection process to force more elaborate visual search patterns
- Presenting targets in a random order
- Creating a direct comparison to dwell time selection
- Making sure the distance covered was equal for both horizontal and vertical actions
- Introducing longer and shorter SSGG
- Introducing multiple measures

The abstract nature of the task was retained. This was done because the intention was to be able to investigate the fundamentals of SSGG.

6.1.1 SIMPLIFYING THE INTERFACE

As mentioned the focus of the design of the pilot experimental test environment was feedback and ensuring that the interface supported navigation. However, while conducting the experiment it quickly became clear that all of the extra navigational tools were unnecessary, because of the simplicity of the actions. So both the *central* and *peripheral indicators* were removed from the interface and the *permanent indicators* were exchanged with outlines of the initiation areas. In other words the *initiation field* for the colour red, which could be selected by looking from left to right, had a red outline and likewise for the other initiation colours (Figure 66).

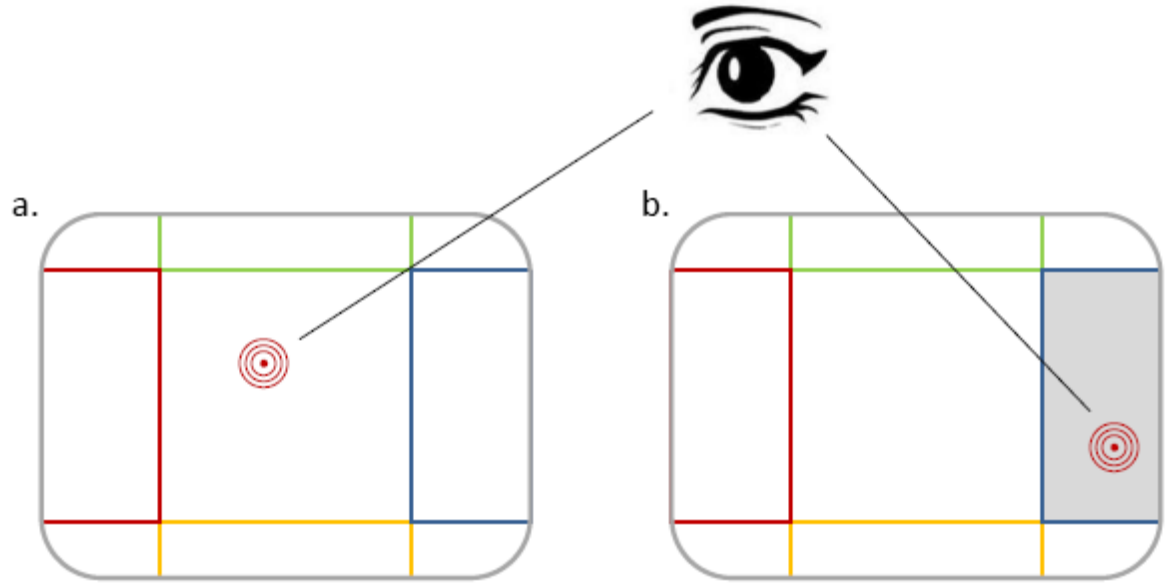


Figure 66: Simplified single gesture interface. To the left the user is looking at the centre of the screen and no feedback is given. To the right the user is looking at an initiation field, which switches to light grey as feedback.

In the pilot design the cursor had been visible in order to provide feedback. This also turned out to be more confusing than beneficial as participants occasionally would start following the cursor rather than look at their next target. When looking at the centre of the screen (a) no feedback would be given and the only dynamic feedback given to the user was that (b) a field would shift to a light grey both when it was used as an initiation point and as a completion point.

6.1.2 INTRODUCING RANDOMNESS AND DYNAMICS

The targets in the pilot design were static in respect to how they were situated on the screen and the sequence in which they had to be selected. In this iteration of the experimental design the targets were introduced randomly at the top of the screen after which they proceeded to descend down the screen at a constant speed. The object of the task was to select the target which had fallen the furthest before it disappeared at the bottom of the screen (Figure 67). There were two main reasons for implementing this type of task. First of all it introduced a time pressure incentive, so that participants would complete the selections as quickly as possible. Secondly, the intention was to explore whether this selection strategy would work in an environment with dynamic content.

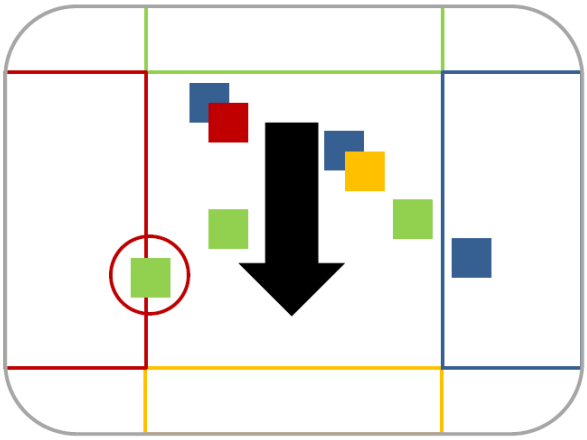


Figure 67: Dynamic and random presentation of the targets in the new tasks. The arrow represents the direction in which the targets are falling. The circle around the green target indicates that this is the target up for selection.

6.1.3 DESIGNING A COMPARISON TO DWELL

In the pilot studies the grand means were simply considered in regard to a standard dwell selection time of 450ms. However, recent research has shown that if users are allowed to dynamically adjust their dwell selection time they could go from an average of 876ms to 282ms per activation in a typing application (Majaranta et al., 2009). It was therefore imperative to try and create a test environment in which SSGG and dwell selection of varying *selection completion times* could be compared. Four dwell-buttons were placed at the bottom of the screen on the dwell interface. The task was the same for both selection methods (Figure 68).

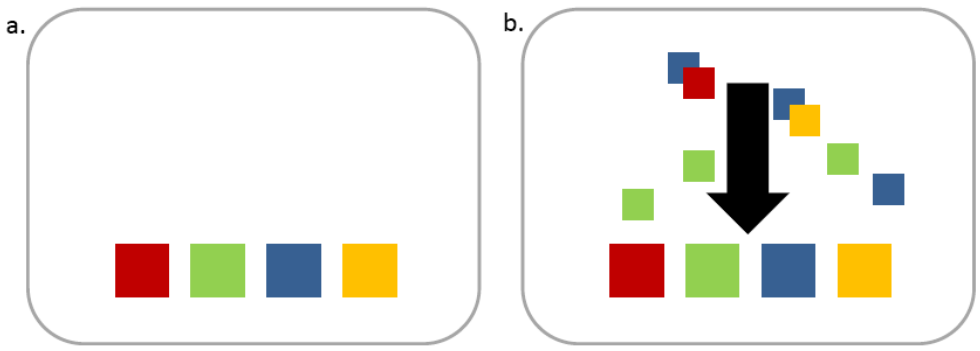


Figure 68: The basic layout of the dwell selection interface and the task

Targets were selected by fixating on the dwell-button which corresponded in colour to the current target. Feedback regarding the selection process was provided on the dwell-button. When a dwell-button was fixated upon then a small white circle would appear at the centre and proceed to expand for the duration of the dwell period; when the circle 'filled out' the

dwell-button the selection was complete (Figure 69). Looking away from the target at any point would break the activation process, which then needed to be started again, if the selection was to be completed. Dwelling longer on a target than the dwell period would initiate a new selection, if maintained then a new selection would be completed.

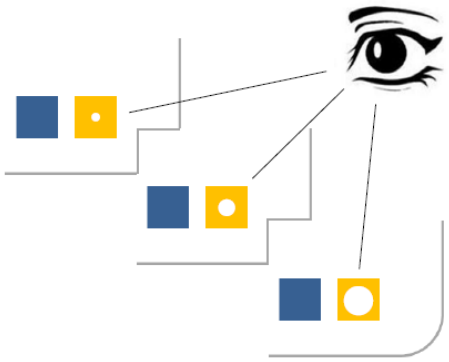


Figure 69: The dwell time activation feedback. By maintaining a fixation on the target the white feedback circle would increase in size until it filled out the entire dwell button.

Several different *dwell selection times* were implemented in order to determine what the ideal threshold for dwell selection would be in this particular interface.

6.1.4 EQUAL LENGTH OF GESTURES AND INTRODUCING LONG AND SHORT GESTURES

The horizontal selection fields were made larger to ensure that there was an equal distance to cover when completing *horizontal* and *vertical* SSGG.

To establish whether it had actually made a difference in the first place, gesture selection fields of different sizes were implemented. *Long* SSGG required the user to cover 70% of the screen and *short* SSGG only required movements to cover 40%. This resulted in another experimental factor which was the potential difference between *long* (Figure 70a) and *short* SSGG (Figure 70b).

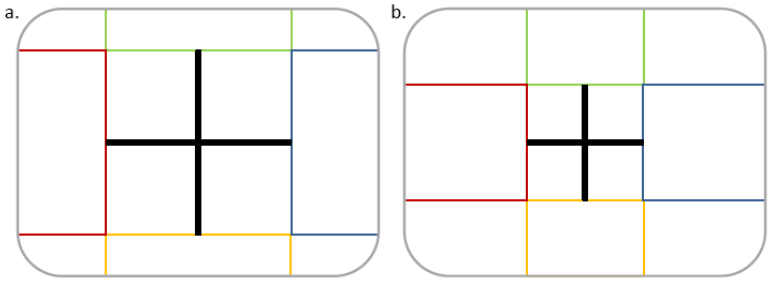


Figure 70: Illustration of long (to the left) and short (to the right) gaze gestures.

6.1.5 INTRODUCING TEST MULTIPLE VARIABLES:

Finally, a major change in the design was the introduction of several more dependent variables.

Selection completion time was still measured, for the gaze gesture conditions, by recording the time of exit from an initiation field to the point of entry in a completion field. Also this action still had to be completed within a 1000ms timeframe or the activation would be cancelled. Dwell-time selections were fixed at 100ms, 300ms, 400ms, 500ms, 600ms, and 1000ms.

Task completion time was a measure of the elapsed time between successful activations, which could include erroneous selections and missed targets.

Two types of error measurements were introduced and applied to both gaze gesture and dwell selection. The first error type was wrong selections, referred to in future as *selection error*. This measured how often the user completed a selection which did not match that of the current selection task; in other words if the task was to complete a green selection, then any other selection except a green selection was considered an error. *Selection error* was intended to be an indication of the level of accidental gesture completion.

The second error type was the number of targets which were missed, referred to in future as *target error*. As mentioned the object of the task was to select a target before it disappeared at the bottom of the screen. If, however, the target was not selected then it was considered a *target error*. This was intended to be an indication of how well the selection method worked in regard to the inherent time pressure in the task as well as being indicative of ease of use.

6.2 COMPARISON OF DWELL, LONG AND SHORT GAZE GESTURES

The purpose of this experiment was to compare dwell selection and SSGG of different lengths to each other.

The participants were introduced to the test environment and task framework before beginning the experiment proper. Twelve participants took part (6 female). The experiment was conducted on a Quick Glance 3 eye- tracker with a sampling of 20 frames/sec. The application was written in Java and developed on the open source Eclipse editor. Each participant was calibrated successfully and none of them were colour-blind. Before each trial

the participants were allowed to practice with the interaction methods for approximately 10-15 minutes.

The experiment was balanced based on two overall parameters: *selection methods* (*Dwell selection [D], long SSGG [LG] and short SSGG [SG]*) and whether or not dwell selection started at 100ms and then proceeded to 1000ms or vice versa. The order of the various tests was conducted as shown in table 3. This was done in order to examine whether there was a potential learning effect.

Participant	1. Activation Strategy	2. Activation Strategy	3. Activation Strategy	Dwell Time Starting point
1	LG	D	SG	100ms
2	SG	LG	D	1000ms
3	D	SG	LG	100ms
4	LG	D	SG	1000ms
5	SG	LG	D	100ms
6	D	SG	LG	1000ms
7	LG	D	SG	100ms
8	SG	LG	D	1000ms
9	D	SG	LG	100ms
10	LG	D	SG	1000ms
11	SG	LG	D	100ms
12	D	SG	LG	1000ms

Table 3: The balanced design of the experiment

Each participant completed 20 selections for *long* SSGG, *short* SSGG and for each of the fixed *dwell selection completion times* of which there were 6; this meant that each participant completed 160 selections. Therefore overall there were n=1600 observations.

The independent variable in this experiment was *selection method*, which had eight levels: *Long* SSGG, *short* SSGG and six dwell-time activation levels (100ms, 300ms, 400ms, 500ms, 600ms, 1000ms).

The dependent variables were *selection completion time*, *task completion time*, *selection error* and *target error*.

The null hypotheses which this experiment sought to reject or retain were:

- Null 1: There was no difference between the activation completion times of *long* SSGG and *short* SSGG.
- Null 2: There is no learning effect, in terms of activation completion, associated with beginning the trial with *long* SSGG or *short* SSGG.
- Null 3: There was no difference between the *task completion times* of any of the eight conditions.
- Null 4: There was no difference in the amount of *selection errors* which have occurred during any of the conditions.
- Null 5: There was no learning effect, in regard to *selection errors*, associated with beginning the trial with *long* SSGG or *short* SSGG.
- Null 6: There was no learning effect, in regard to *selection errors*, associated with beginning the experiment with 100ms and then progressing up to 1000ms or starting with 1000ms and progressing down to 100ms
- Null 7: There was no difference in the amounts of *target errors* which occurred in the various conditions.
- Null 8: Combining *selection completion time* and *selection error* into a new measure would show that there was no difference in the *augmented selection times* between the various conditions.

6.2.1 SELECTION COMPLETION TIME EFFECTS

The direct comparison of *selection completion times* was only relevant when looking at the SSGG conditions, as the *selection completion time* was fixed in the dwell selection conditions.

Before statistical methods were chosen the data were explored to determine the appropriate course of action. First of all the data were examined to determine the distribution.

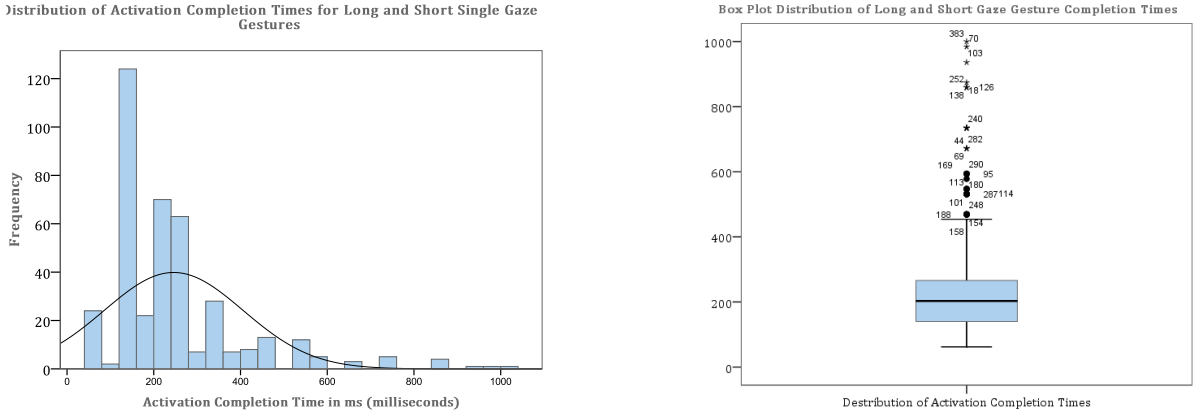


Figure 71: Normal distribution and box plot of activation completion time for long and short gestures

A Kolmogorov-Smirnov test confirmed what was already apparent from the histogram and box plot above (Figure 71); which was that the data were negatively skewed and significantly deviated from a normal distribution. Transformations of the data were done, but with no significant results. Maucley's test was also significant which meant that sphericity could not be assumed. Because these assumptions were not met non-parametric analyses were chosen instead.

After ranking the scores, the Wilcoxon signed-ranks test was used to compare the means (Figure 72). *Selection method* was the independent variable with two selected levels (*short* SSGG and *long* SSGG) and *selection completion time* was the dependent variable measured in ms (milliseconds). The mean values of the two conditions were based on n = 200 observations per result. The grand mean for both *selection methods* = 245ms; for *Long* SSGG = 301ms and *Short* SSGG = 190ms.

There was a significant difference between *long* and *short* SSGG Z = -8, 020; p <.001. *Short* SSGG were significantly faster than the *Long* SSGG.

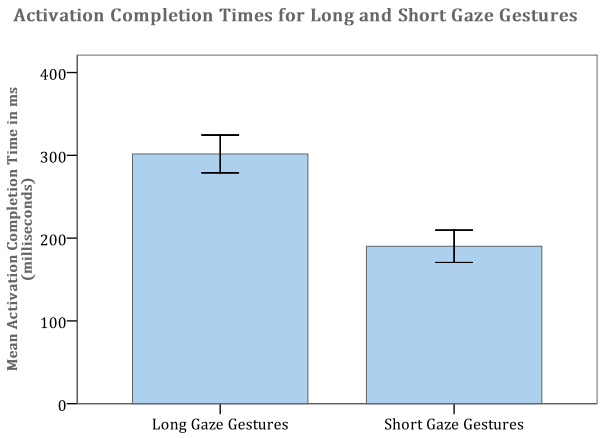


Figure 72: Mean of activation completion times for long and short gaze gestures. Error bars represent standard error of mean

6.2.2 LEARNING EFFECT IN REGARD TO SELECTION COMPLETION TIMES

Some of the participants started the experiment with *short* SSGG and some started with *long* SSGG. The scores were grouped depending on first and last *selection method* (Figure 73). The data were the same as for the analyses above and did not comply with the requirements for applying parametric statistics. Two Wilcoxon signed ranks tests were applied based on overall n = 100 observations per result.

There was no significant effect between *activation completion times* for participants who started with the *long* SSGG, compared to those who had *long* SSGG as their second *selection method*. Z = -0,362; p = 0,717. There was a significant difference between those who had *short* SSGG as their first and second *selection method*. Z= -2,884, p= 0.004. Those who had *short* as their second SSGG condition had significantly faster *selection completion times.*

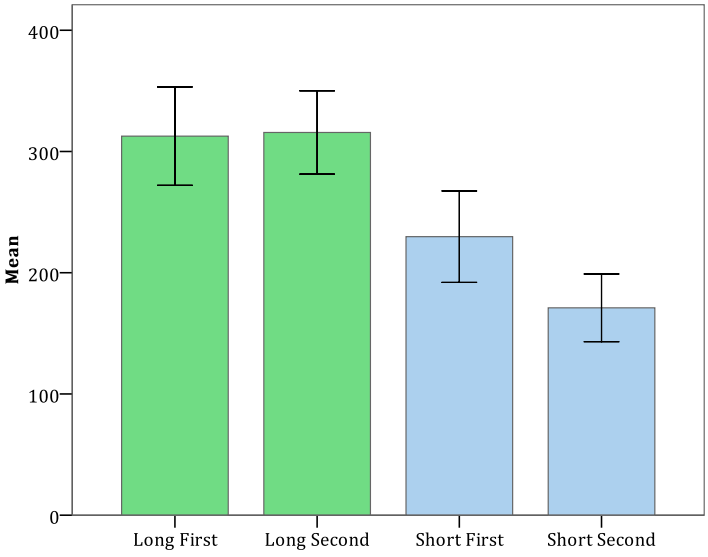


Figure 73: Learning effects based on activation completion times for long and short gaze gestures

6.2.3 TASK COMPLETION TIMES

Task completion time was the elapsed time between two successfully completed selections. These were accumulated for each *selection method*. The analysis was therefore based on one observation per selection strategy per person, in other words n=10 observations per result (Figure 74).

The assumptions required to conduct a parametric statistical analysis were not met, so non-parametric methods were used. Friedman's non-parametric ANOVA was used for the overall analysis and multiple paired Wilcoxon signed rank test for post hoc analysis.

There was an overall significant effect between the task times $X^2_r = 48{,}683$ (7, n=10), $p < 0.01$. Overall *long* SSGG took significantly longer than all other conditions, except *short* SSGG. Also *short* SSGG took significantly longer than most of the *dwell selection* conditions except 600ms and 1000ms. There was no significant difference in *task completion time* between 100-500ms or between 500-600ms *dwell*. 100-400ms was significantly faster than 600-1000ms and 500ms was significantly faster than 1000ms *dwell activation*.

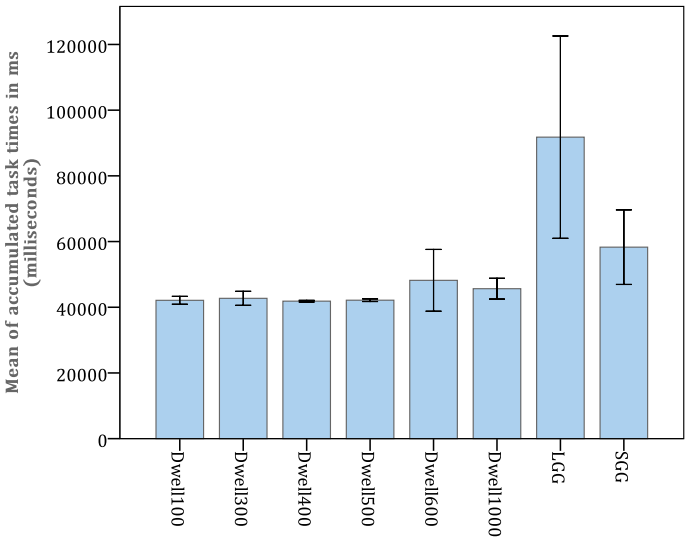


Figure 74: Task completion times for all conditions. Error bars represent standard error of mean

6.2.4 SELECTION ERRORS

The *selection error* was a completed selection which did not correspond to the action required to select the target. These were accumulated in each selection condition and the analysis conducted was based on these accumulated scores which consisted overall of n=10 observations per result (Figure 75).

The skewed distribution of the data, along with not complying with the assumption of sphericity meant that non-parametric methods of analysis had to be employed. Friedman's non-parametric ANOVA was used to determine the overall effect. Multiple Wilcoxon signed rank tests were used to establish where the significance was. The average selection errors for all of the conditions were (Table 4):

Dwell 100	Dwell 300	Dwell 400	Dwell 500	Dwell 600	Dwell 1000	Long SSGG	Short SSGG
49.7	19.7	13.3	12.6	6.9	0.7	10.4	4.8

Table 4: Average number of selection errors for all conditions

There was an overall significant effect between the number of *selection errors* $X^2_r = 41{,}370$ (7, n=10), $p < 0.01$. 100ms had significantly more selection errors compared with all other conditions. 1000ms has significantly fewer *selection errors* compared to all other conditions. The rest of the differences were non-significant.

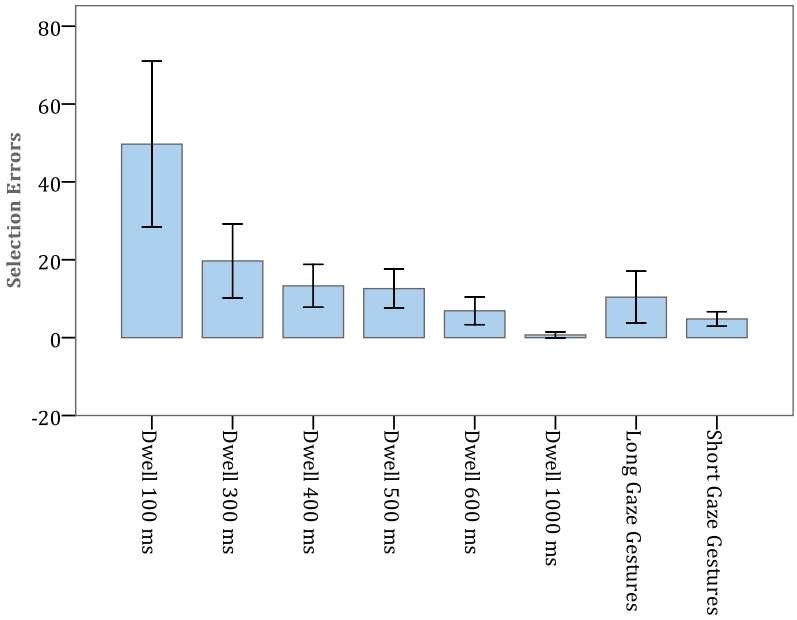


Figure 75: Mean selection errors for all conditions. Error bars represent standard error of mean

6.2.5 LEARNING EFFECT BASED ON SELECTION ERRORS FOR GAZE GESTURES

The data were divided to see whether or not it had made a difference in the number of *selection errors* whether the participant started with *long* SSGG or *short* SSGG. Because the number of errors was accumulated the analyses below was based on n=5 observations per result (Figure 76). The data did not comply with the assumptions required for the use of parametric statistics, so Friedman's non-parametric ANOVA was used.

There was no significant overall effect between the *selection error rates* of the various conditions. $X^2_r = 1{,}00$ (3, n=5), $p = 0{,}807$.

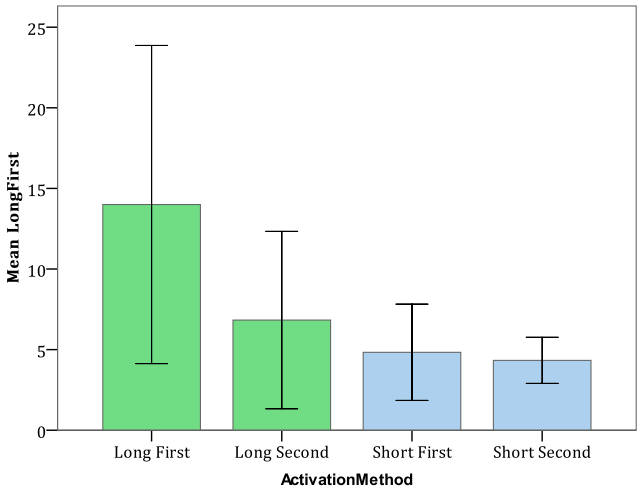


Figure 76: Error based learning effect for long and short gestures.

6.2.6 LEARNING EFFECT BASED ON SELECTION ERRORS FOR DWELL TIME ACTIVATION

This test was done to indicate whether there was a potential increase or decrease in the number of *selection errors* depending on which *dwells selection condition* began the trial.

The data were divided based on which participants started with 100ms and those which started with 1000ms. Because the measure was accumulated the analyses below were based on overall n = 5 observations per result (Figure 77). The data did not comply with the assumptions required for parametric statistics so non-parametric methods were used. Multiple Wilcoxon signed rank tests were used to determine whether there was a significant difference between pairs of *selection errors*. In other words 100ms, was compared between those who started with 100ms and those started with 1000ms.

None of the Wilcoxon signed rank showed any significant difference in learning effect based on the number of *selection errors* between the conditions which started with 100ms and those which started with 1000ms. *p> 0.05* for all cases.

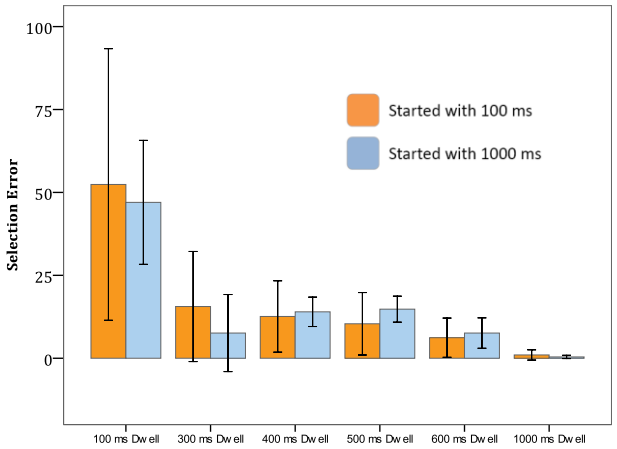


Figure 77: Error based learning effect for dwell time activation. Error bars represent standard error of mean

6.2.7 MISSED TARGET ERROR

The *target error* was a measure of how many targets crossed the screen without being selected. This number was accumulated for each selection condition and the analysis was based on n= 10 observations per result (Figure 78). The distribution of data was skewed and sphericity could not be assumed, therefore non-parametric methods of analysis were used. Friedman's non-parametric ANOVA was used and multiple Wilcoxons signed rank tests were performed.

There was an overall significant effect between the number of *target errors* $X\ ^{2}_{r}=$ 37,216 (7, n=9), p < 0.01. There were a significantly higher number of *target errors* in the *long* SSGG condition, compared to all other conditions except the *short* SSGG. There was no significant correlation between any other conditions.

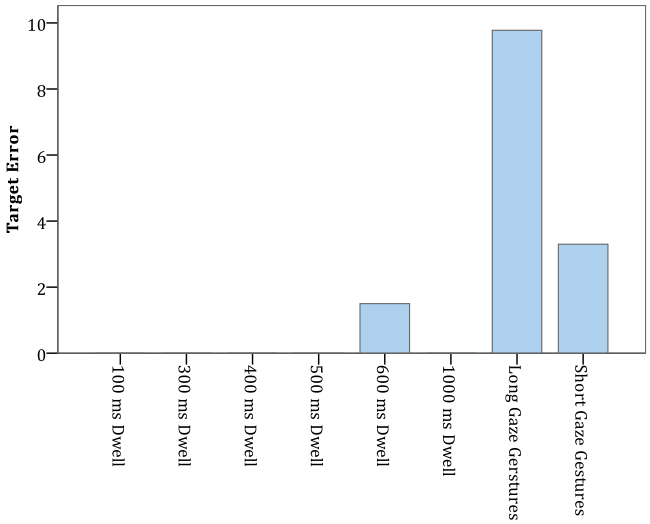


Figure 78: Missed target error for all conditions.

6.2.8 COMBINED ACTIVATION COMPLETION TIME AND SELECTION ERROR

Both *selection completion time* and *selection error* were insufficient in establishing a comparative foundation between *dwell selection* and SSGG. Therefore, a combination measure of the two was created to express *selection errors* and *selection completion times* in a single measure. The premise for this single measure was to convert *selection errors* into milliseconds (ms) which could then be added to the average *selection completion times* of the various selection strategies. An *augmented selection time* consisting of combined *selection error* and *selection completion times* would thereby be created.

The amount of time (ms) which a *selection error* would add was dependent on *selection method*. Each *selection error* would add the equivalent of an *average selection completion time*. In other words, when using the *dwell selection* condition of 100ms each average *selection error* would add 100ms. In the 600ms condition 600ms per average *selection error* would be added and in the *long* SSGG condition the average *selection completion time* would be added for each average *selection error* etc. The penalty represented the average time it would have taken to complete the erroneous activation.

Combining *selection error* and *selection completion time* was intended to reveal how much time the participant at a minimum had to spend when completing both successful and unsuccessful selections. This *augmented selection completion time* makes SSGG and *dwell selection* comparable. Equation 1 is a generic expression of this.

$$\hat{\mu} = \mu(\varepsilon + 1)$$

Equation 1: A generic expression of the combined activation completion time and selection error measure

The result of this equation was an *augmented selection completion time* which is represented by '$\hat{\mu}$'. The original average, which was fixed for *dwell-time selection* and estimated for *gaze gestures*, is represented by 'μ'. This is then multiplied by the average number of errors 'ε'. The '+1' represents the successful selection required.

The *augmented selection completion time* was calculated for all instances and the analysis was therefore based on n = 200 observations per result (Figure 79). The distribution of data was skewed and sphericity could not be assumed, Therefore non-parametric methods of analysis were used. Friedman's non-parametric ANOVA was used and multiple Wilcoxon signed rank tests were performed.

There was an overall significant effect between the augmented averages $X^2_r = 1289{,}698$ (7, n=200), $p < 0.01$. All of the conditions were significantly different from one another. The *augmented selection completion time* for *short* SSGG was the fastest condition, followed by 1000ms dwell time condition and then *long* SSGG.

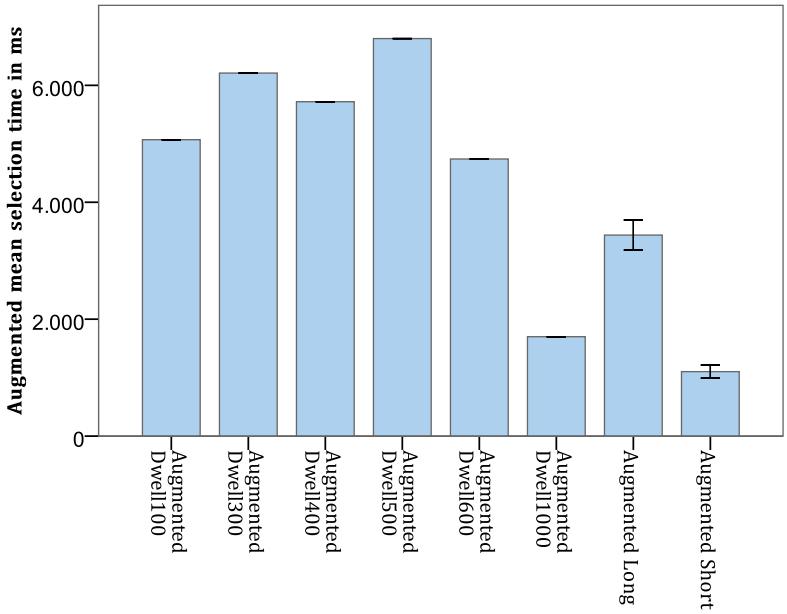


Figure 79: Augmented mean values for all conditions, error bars represent standard deviation. The reason that there are no error bars in the dwell conditions is because the calculations where based on a fixed dwell time and a fixed average error rate, which meant that there was no deviation.

It was found that this measure was only relevant for selection strategies which had a number of *selection errors* which was less than, the number of selections which needed to be completed. A generalized expression of this is (Equation 2):

$$\alpha = \frac{\varepsilon}{S} < 1$$

Equation 2: A generic expression of the threshold for a usable *selection method*

Where 'ε' is again the average number of errors. '*S*' is the total number of selections and 'α' is threshold for whether or not an *augmented completion time* should be calculated. In other words, if 'α' is bigger than 1, it means that for every successful selection made, the likelihood of making an error is more than one. This basically means that the selection method should be discarded for practical puposes.

The calculations of α-levels for the eight conditions are as follows (Table 5):

Dwell 100	Dwell 300	Dwell 400	Dwell 500	Dwell 600	Dwell 1000	Long SSGG	Short SSGG
2,535	1,035	0,715	0,68	0,395	0,085	0,57	0,29

Table 5: α levels for each of the conditions.

6.3 DISCUSSION OF THE DWELL, LONG AND SHORT GESTURE EXPERIMENT

The results which have just been presented answer the null hypotheses which were presented earlier in the chapter. In the following each of the results will be discussed in accordance with its original null hypothesis.

6.3.1 NULL 1: THERE WAS NO DIFFERENCE BETWEEN THE *SELECTION COMPLETION TIMES* OF *LONG* SSGG AND *SHORT* SSGG.

This null hypothesis was rejected as there was a significant difference in the activation completion time between *long* SSGG and *short* SSGG.

This was not expected for two reasons. First of all there is no immediate physiological reason for this finding – on the contrary longer saccades reach higher velocities than shorter ones, see Chapter 2. Secondly, a screen at normal viewing distance only represents 10° of visual angle, which means that the actual difference in length was very small. A general observation was that there seems to be a difference between naturally occurring eye movements and intentional eye movements used for activation.

During the experiment it was observed that when participants found gaze gesture interaction difficult, they became tense and tried to 'force' the selections by starring hard at the screen. This was the case for both *long* SSGG and *short* SSGG, however for the latter the distance being forced was shorter. It could potentially be interesting to see if an experienced user's ability to relax would decrease *selection completion time* and *selection error* rate.

The difference in *selection completion times* could be a consequence of the interface design. Even though the distance between horizontal and vertical activations had been adjusted for, the adjustment had caused the horizontal fields to be larger than the vertical fields. This could have an effect as the targets were easier to hit.

The grand mean for all gaze gestures was 245ms. Theoretically, these *selection completion times* could have been much faster. As mentioned earlier they can cover between a 1° and 40° visual angle and last between 30-120ms (Duchowski, 2007). One of the reasons for this could be found in the eye tracking equipment. The experiment was conducted on an eye tracker which had a frame rate of 20 frames/sec. The main consequence of the low frame rate was that it caused a delay in the registration of the gaze point. This affected both the user experience and the data collection.

The user's experience was affected by experiencing a delay in the system's response time. In other words the user would look at something but the system did not register it. This could have caused frustration because the user had to repeat an already completed selection.

The system delay was not so much a problem for the dwell-time interface because the length of fixation covered several frame grabs. The observation was that for registering fast saccadic motion low frame rate was potentially a huge problem.

Another technical issue with the QuickGlance eye tracker was the *smoothing factor*. A *smoothing factor* means that the gaze tracking software takes 'x' number of frames and estimates the gaze position as an average of the estimated gaze positions in the frames. Smoothing was applied to the system by default. This was done in order to avoid the system registering every jitter and renders a smooth output to the user, again very conducive to dwell selection interaction. As a consequence the system had difficulty in registering fast

movements across the screen, because the cursor placement could potentially end up somewhere in the centre.

As mentioned, some participants found that they tired quickly because they felt that they had to 'force' the selections, this could of course also have been an effect of the before mentioned hardware issues, but the unfamiliarity of the interaction method could also have played a part. Dwell-time selection has the advantage of mapping the existing metaphor of on-screen buttons.

6.3.2 NULL 2: THERE IS NO LEARNING EFFECT, IN TERMS OF *SELECTION COMPLETION TIMES*, ASSOCIATED WITH BEGINNING THE TRIAL WITH *LONG* SSGG OR *SHORT* SSGG.

This null hypothesis was both accepted and rejected in a sense. There was no learning effect in the *long* SSGG condition between those who started and finished with *long* SSGG. However, there was a difference between those who started and ended with *short* SSGG as *selection completion times* were significantly faster for those who had *short* SSGG as their second *gesture strategy*.

Essentially this meant that those who went from *long* SSGG to *short* SSGG had faster selection times. This makes sense as it generally seemed to be easier to complete the *short* SSGG. Perhaps having become used to completing *long* SSGG first, it was easier to adjust 'down' to *short* SSGG.

The possibility that SSGG becomes faster as an individual becomes more relaxed and trained in the *selection method*, is very likely. However, to fundamentally establish those types of potential benefits a longitudal study should be conducted.

6.3.3 NULL 3: THERE WAS NO DIFFERENCE BETWEEN THE *TASK COMPLETION TIMES* OF ANY OF THE EIGHT CONDITIONS.

This null hypothesis was rejected as there was an overall significant difference in task times. *Long* SSGG had the longest *task completion times* after which came *short* SSGG and 1000ms.

These *task completion times* were indicative and a consequence of several different things. First of all the long *task completion time* of *long* SSGG makes sense if *target error* was also taken into consideration. More targets were missed in this condition than in any other. The main reason for this was the previously described system delay, which hugely affected the

rate of successful completions. Participants were very often forced to repeat intended selections multiple times in order to complete a successful activation. The targets continuously moved down the screen at a predetermined speed. Therefore, targets would often be missed not because the action was not being completed correctly, but because the system was not registering it. Once a target had been lost it was very difficult to 'get back in the game' as the subsequent target was right after and could easily be lost. This created a stressful experience for the participant, which only made it even more difficult to complete selections. This was also the case only to a lesser degree for *short* SSGG. The length of *selection completion time* in the 1000ms *dwell-time condition* was the main reason for the prolonged *task time* as very few errors of any kind were made during this type of activation.

The combination of a higher number of *target errors* and prolonged *task completion times* could be indicative of SSGG being less effective than *dwell selection*. However, it could also simply mean that SSGG are not appropriate for time pressured tasks. Or that there might be a substantial advantage to be found in eye tracking equipment which has been designed for saccade based interaction, rather than fixation based interaction.

6.3.4 NULL 4: THERE WAS NO DIFFERENCE IN THE AMOUNT OF *SELECTION ERRORS* WHICH HAVE OCCURRED DURING ANY OF THE CONDITIONS.

This null hypothesis was rejected as there was an overall significant difference in the amount of *selection errors* which occurred in the different conditions. 100ms had significantly more *selection errors* compared with all other conditions. 1000ms had significantly fewer *selection errors* compared to all other conditions.

Even though the 1000ms dwell condition had an average error rate of 0.7 it was an impractical *selection method* because it was strenuous to sustain lengthy fixations. The long *selection completion time* was also reflected in the before mentioned *task completion times*.

In terms of *selection completion time* the 100ms *dwell condition* was innately the fastest of the *selection methods*. However, it would be useless in most applications due to the high number of *selection errors*. The average *selection error* for the 20 selections was 49.7. This meant that for every successful selection – 2.5 errors are made. If this was hypothetically transferred to the selection of characters in a typing application nothing would ever be written, because for

every successful character entry 2.5 corrections would need to be made and for every correction needed to be made 2.5 corrections would need to be made and so on.

A general observation which was made, but which was difficult to support statistically, other than the fact that it was based on statistical insignificance, was that there seemed to be a plateau of *selection errors* between the 400ms-500ms *dwell-time condition*. This was interesting in regard to the use of a standard *dwell selection time* of 450ms that was often employed (Qvarfordt & Zhai, 2005). For practical purposes there seems to be an assessment that states that this was the range where the efficiency of *selection completion times* and the disadvantage of *selection error* rates balance each other out.

6.3.5 NULL 5: THERE WAS NO LEARNING EFFECT, IN REGARD TO *SELECTION ERRORS*, ASSOCIATED WITH BEGINNING THE TRIAL WITH *LONG* OR *SHORT GAZE GESTURES*.

This null hypothesis was retained as there was no *selection error* based learning effect associated with initiating or ending the trial with *long* SSGG or *short* SSGG.

6.3.6 NULL 6: THERE WAS NO LEARNING EFFECT, IN REGARD TO *SELECTION ERRORS*, ASSOCIATED WITH BEGINNING THE EXPERIMENT WITH 100MS AND THEN PROGRESSING UP TO 1000MS OR STARTING WITH 1000MS AND PROGRESSING DOWN TO 100MS

This null hypothesis was retained as there was no *selection error* based learning effect within the conditions, regardless of whether the trial had been initiated with 100ms or 1000ms.

6.3.7 NULL 7: THERE WAS NO DIFFERENCE IN THE AMOUNTS OF *TARGET ERRORS* WHICH OCCURRED IN THE VARIOUS CONDITIONS.

This null hypothesis was rejected as there was a significant difference in the number of targets lost depending on condition. There were a significantly higher number of *target errors* in the *long* SSGG condition, compared to all other conditions except the *short* SSGG. There was no significant correlation between any other conditions.

The only condition which really caused users to miss targets was *long* SSGG. As explained earlier there was a time pressure element to the task caused by the dynamics of the falling squares. A consequence of this was that once participants had missed a target they had very little time to select the next target.

Closer examination of the data revealed that there were no problems during the first five selections and showed that once the participants had 'lost' a target, they would subsequently 'lose' multiple targets while trying to regain control.

As previously explained there are various issues regarding *long* SSGG; among others the system set-up.

6.3.8 NULL 8: COMBINING *SELECTION COMPLETION TIME* AND *SELECTION ERROR* INTO A NEW MEASURE WOULD SHOW THAT THERE WAS NO DIFFERENCE IN THE *AUGMENTED SELECTION TIMES* BETWEEN THE VARIOUS CONDITIONS.

This null hypothesis was rejected as there was an overall significant difference between the *augmented selection completion time* scores. All of the conditions were significantly different from one another. Considering both *selection completion speed* and *selection errors – short* SSGG was the fastest condition, followed by 1000ms *dwell-time condition* and then *long* SSGG.

There are a few things which can be observed in regard to these results. Within the dwell conditions there was an *augmented selection time* plateau around 500ms-600ms, this was a shift from the *selection error* distribution where the plateau was around 400ms-500ms. This could be viewed as further support of the practically applied 450ms as a *dwell-time standard*, as it falls below the point where the cost of an error becomes too high in regard to the number of errors which were completed. The 1000ms condition was, as expected, proven to be the most lengthy *selection method*; even though this condition had the lowest *selection error* rate by far. Put in other terms, the cost of making an error in the 1000ms *dwell activation condition* was higher than the benefit of rarely making one.

The α-levels of this experiment show, that 100ms and 300ms dwell should be discarded as practical selection methods in the context of this experiment, because the number of errors made per successful selection was more than one.

The best overall selection type was the *short* SSGG, which meant that the combined cost of making successful and erroneous selections was the lowest of all the conditions.

Surprisingly, there was a significant difference between *long* SSGG and *short* SSGG which was not expected. The expectation was also that dwell selection would have far fewer *selection errors* than SSGG, because it built on familiar mapping, which was not the case.

The main error of interest was the *selection error*, because this error dealt with the two main design issues regarding both dwell and gaze gesture selection. *Selection errors* in dwell-time selection speak to how well the implementation deals with the Midas touch problem. In other words a high number of *selection error,* meant a high level of unintentional selections (i.e., a high degree of Midas touch). The same principle applied in the *gaze gesture selection methods* where *selection errors* reflect how much accidental gesture completion there was between search patterns and gesture patterns. It was expected that the number of accidental gesture completions would be very high for SSGG, as their simplicity might easily correlate with navigational search patterns. However, there was no significant difference in number of *selection errors* between the SSGG and the *dwell selection conditions* between 300-600ms. As this is the range in which dwell selection is generally implemented this is also the range which was the most relevant to compare with. In this experiment there was not a significant difference between the level of Midas touch and accidental gesture completion.

In order to determine whether or not the frame rate and *smoothing factor* had an effect on *selection completion times* the next experiment compares different eye tracking systems; more specifically eye tracking systems with higher sampling frequencies. This was done by conducting the gaze gesture part of the previous experiment on the QuickGlance 3 eye tracker (20 frames/sec) and on a Tobii T50 (50 frames/sec). The *smoothing factor* on the Tobii systems was implemented differently. They use a fixation detector to predict when to apply smoothing, so smoothing should only occur while the user was fixating. The following experiment will examine these points. The experimental design is essentially the same as the experiment just presented, with only a few minor modifications.

7 SINGLE STROKE GAZE GESTURES ON DIFFERENT EYE TRACKERS

This was the second SSGG experiment which was inspired by the pilot studies and the *dwell*, *long* SSGG and *short* SSGG experiment. The focus was to compare SSGG on different eye trackers in order to determine what kind of affect this might have on *selection completion times* and overall error rates.

7.1 EXPERIMENTAL DESIGN

In the previous experiment the focus was on comparing SSGG with dwell selection. *Selection direction* was not one of the independent variables, which was examined. However, the pilot studies showed that there could potentially be a difference in *selection completion time* or *selection error* depending on direction. Whether or not this was the case it could influence the way gaze gestures (*single*, *complex* and *continuous*) should be implemented in future designs. The experimental setup was the same as previously described.

The participants were introduced to the test environment and task framework before beginning the experiment. Nine participants took part in the study (four female) all of which had normal or corrected to normal vision. None of them were colour-blind. Five had previous experience with eye tracking. The application was written in Java and testing was completed on the same QuickGlance 3 as previous, using a 13 inch monitor (20 frames/sec) system, and a Tobii 1750 using a 17 inch monitor (50 frames/sec).

The independent variables were:

Input device with two levels: QuickGlance3 and Tobii 1750

Selection method with two levels: *long* SSGG and *short* SSGG.

Selection Direction with four levels: Left to Right, Right to Left, Top to Bottom and Bottom to Top.

Each participant had to complete 20 successful selections 3 times in each condition: QuickGlance long SSGG, Quickglance short SSGG, Tobii long SSGG and Tobii short SSGG.

The dependent variables were:

Selection Time: the time from when the user exits the initiation field and enters the opposite field. These results should not be viewed as saccade times, but as selection times.

Selection Error: A full completed selection which does not respond to the current target.

Target Error: Targets which descent the screen and pass through the other side without being selected.

The experiment was designed to explore the following null-hypotheses:

7.1.1 FACTOR 1, INPUT DEVICE:

- Null 1: There was no overall difference in *selection completion* time on either *input device.*
- Null 2: There was no significant difference in *selection completion times* for *selection direction* on either *input device*.
- Null 3: There was no difference in *selection completion time* for *selection method* on either *input device*.
- Null 4: The different capabilities of various *input device* had no effect on the number of *selection errors*.
- Null 5: There was no difference in the number of *target errors* on either *input device*.

7.1.2 FACTOR 2, SELECTION METHOD

- Null 6: There was no significant difference in *selection completion times* of *selection methods* based on results from both *input devices*.
- Null 7: There was no difference in *selection errors* for *selection method* based on results from both *input devices*.
- Null 8: There was no difference in *target error* depending on *selection method*.

7.1.3 FACTOR 3, SELECTION DIRECTION

- Null 9: There was no overall difference in *selection completion time* for *selection direction* including results from both *input devices*.

- Null 10: There was no difference in *selection completion time* for overall horizontal and vertical selections based on results from both *input devices*.
- Null 11: There was no difference in number of *selection errors* based on *selection direction* with results from both *input devices*.

There are more combinations of independent and dependent variables which could possibly form the foundation for other null hypotheses. However, the above mentioned were considered the most interesting.

7.1.4 FACTOR 1: INPUT DEVICE

OVERALL SELECTION COMPLETION TIME BASED ON INPUT DEVICE

Overall it was of interest to see if there were significant differences in *selection completion time* between the two *input devices.*

The data were grouped overall depending on *input device*. The results below were based on n = 1080 observations per result.

Maucley's test of sphericity was insignificant which meant that sphericity could be assumed. By logarithmically transforming the data, a normal distribution was achieved based on the Tukey method; so parametric statistics could be used for this analysis (Figure 80).

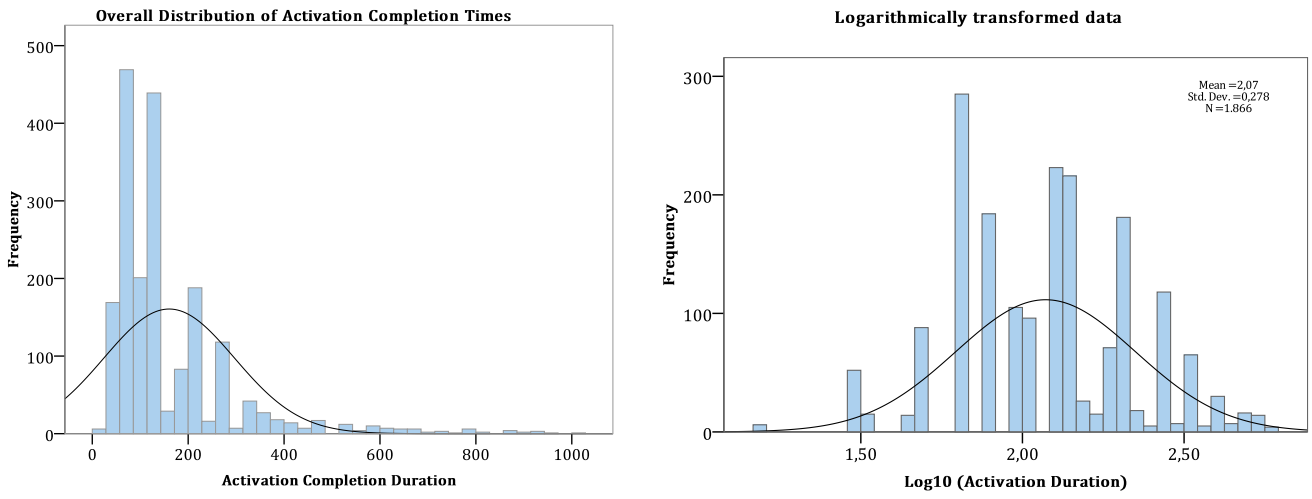


Figure 80: Distribution of activation completion times on both eye trackers. To the left, the original scores. To the right, the logarithmically transformed scores.

A t-test was used to compare *selection completion times* on both *input devices* (Figure 81). The grand mean for all *gesture selection completion times* was 160ms.

The mean *gesture selection completion time* on the QG was 216ms and on the Tobii it was 105ms.

There was a significant difference in *selection completion time* between the QuickGlance and the Tobii, $t(1079) = 23{,}457$; $p<0.01$. *Selection completion times* were significantly faster on the Tobii compared with the Quickglance.

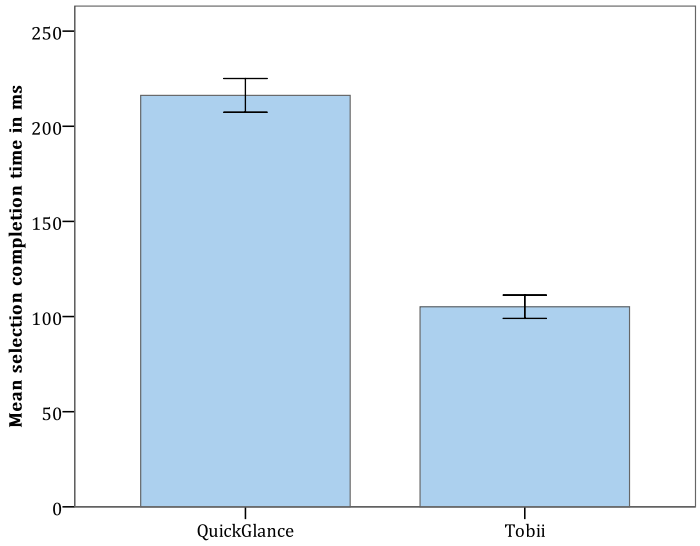


Figure 81: Overall *selection completion times* on the two input devices. Error bars represent standard error of mean

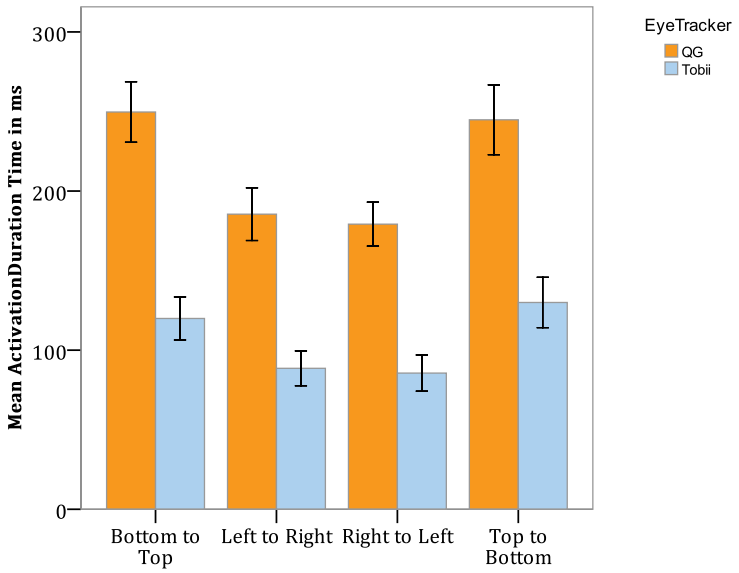
SELECTION COMPLETION TIME OF SELECTION DIRECTION BASED ON INPUT

Selection direction effects on either eye tracker were analysed in two ways. First, each of the four directions was compared to its counterpart on either eye tracker, to establish if one particular direction was faster on one system or another. Secondly, the four directions were compared within the same eye tracking system, to establish whether there was a similar pattern of *selection direction* completion times on both machines.

The data were grouped depending on *selection direction* and *input device*. The random entry of targets in the task meant that there was a slight difference in the number of observations; therefore some observations were left out. Overall, the analyses below were based on n=240 observations per result. The assumptions required to conduct parametric statistics were met, as the sphericity measure was insignificant and normal distribution could be achieved with a logarithmic transformation. To analyze the *selection completion times* for the four different *selection directions* across the two *input devices*, four repeated measures t-tests were conducted (Figure 82). The grand mean for all *gesture selection completion times* was 160ms. For the individual conditions their means were (Table 6):

Input	Bottom to top:	Left to right:	Right to left:	Top to bottom:
QG	249ms	185ms	179ms	244ms
Tobii	119ms	88ms	85ms	129ms

Table 6: Mean directional completion time on both eye trackers



There was a significant difference between all of the *selection completion times* in each condition on both machines. Bottom to top: t (239) = 12,071; p<0.01. Left to right: t (239) = 10,589; p<0.01. Right to left: t (239) = 11,325; p<0.01. Top to bottom: t (239) = 9,090; p<0.01. All of the *selection completion times* on the Tobii eye tracker were significantly faster than on the QuickGlance 3 eye tracker.

Figure 82: Differences in activation completion time depending on selection direction across the two eye tracking systems, error bars represent standard error of mean

To analyze the difference in *selection completion time* based on *selection direction* within each of the *input devices* two repeated measures ANOVAs were conducted (Figure 83). A Bonforonni adjusted pair-wise analysis was used to determine where any potential differences were.

The grand mean for SSGG on the QuickGlance was 214, 781ms, the individual means were (Table 7):

Bottom/top	Left/right	Right/left	Top/bottom
249ms	185ms	179ms	244ms

Table 7: Mean directional *selection completion times* for the Quickglance

The grand mean for the SSGG on the Tobii 1750 was 105ms, the individual means were (Table 8):

Bottom/top	Left/right	Right/left	Top/bottom
119ms	88ms	85ms	129ms

Table 8: Mean directional *selection completion times* for the Quickglance

There was a significant difference in *selection completion time* based on *selection direction* within the QuickGlance eye tracker condition F (1, 3) = 22,017; $p < 0.01$. Left/right and right/left were significantly faster than top/bottom and bottom/top SSGG. There was a significant difference in *selection completion time* based on *selection direction* within the Tobii 1750 eye tracker condition (1, 3) = 12,736; $p < 0.01$. Again left/right and right/left were significantly faster than top/bottom and bottom/top.

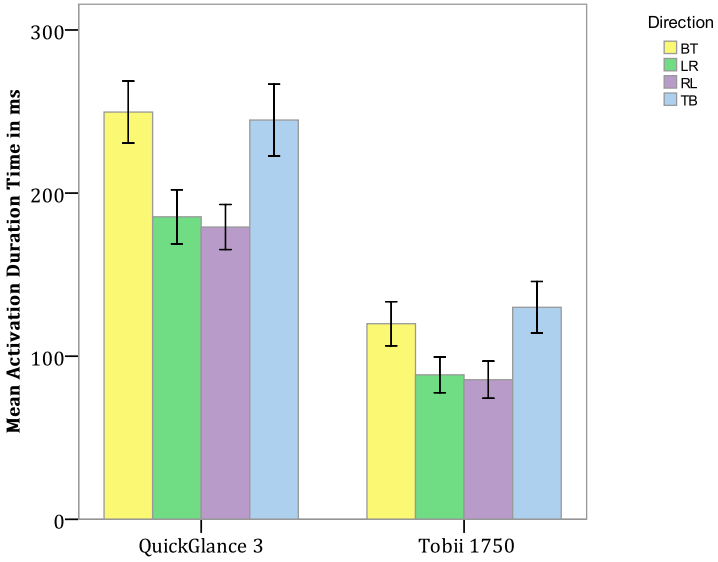


Figure 83: Differences in activation completion time depending on selection direction within the two eye tracking systems, error bars represent standard error of mean

SELECTION COMPLETION TIME FOR LGG AND SGG BASED ON INPUT DEVICE

As with the analysis of *selection direction*, the potential differences in the *selection completion time* of *long* and *short* SSGG (*input method*) depending on *input device* was analyzed in two ways.

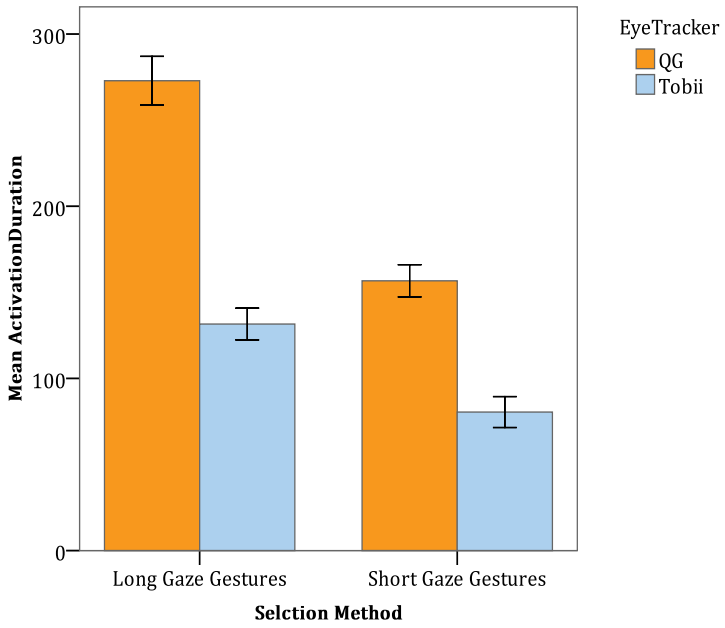
First the difference in *selection activation time* across the two *input devices* was analysed and subsequently the difference in *selection activation time* was analysed within the two *input devices*. The data were grouped depending on *selection method* and *input device*. Some of the observations were left out, overall the analyses below was based on n = 480 observations per result.

Maucley's test of sphericity was again insignificant which meant that sphericity could be assumed. By logarithmically transforming the data a normal distribution was achieved; so parametric statistics could be used for this analysis.

To analyze the difference in *selection completion time* depending on *selection method* across the two *input devices* two repeated measures t-tests were done (Figure 84).

The means for *long* SSGG were: QuickGlance 272ms and Tobii 131ms.

The means for *short* SSGG were QuickGlance 156ms and Tobii 80ms.



There was a significant difference in *long* SSGG on the QuickGlance and Tobii respectively, t (479) = 17,875; p<0.01. *Long* SSGG on the Tobii were significantly faster than on the QuickGlance system. There was also a significant difference between *short* SSGG on the two eye tracking systems respectively, t (479) = 11,974; p<0.01. *Short* SSGG were significantly faster on the Tobii gaze tracker than on the QuickGlance.

Figure 84: Differences in activation completion time depending on selection method across the two eye tracking systems, error bars represent standard error on mean

In order to analyze the difference in *selection completion times* for *selection method* within each of the *input devices*; two repeated measures t – tests were conducted (Figure 85).

The means for the QuickGlance eye tracker were: *Long* SSGG 272ms and *short* SSGG 156ms

The means for the Tobii 1750 eye tracker were: *Long* SSGG 131ms. and *short* SSGG Tobii 80ms.

There was a significant difference between *long* SSGG and *short* SSGG on the QuickGlance t (479) = -14,872; p<0.01. *Short* SSGG were significantly faster compared to *long* SSGG. There was also a significant difference between *long* SSGG and *short* SSGG on the Tobii 1750, t (479) = -8,080; p<0.01. *Short* SSGG were significantly faster than *long* SSGG.

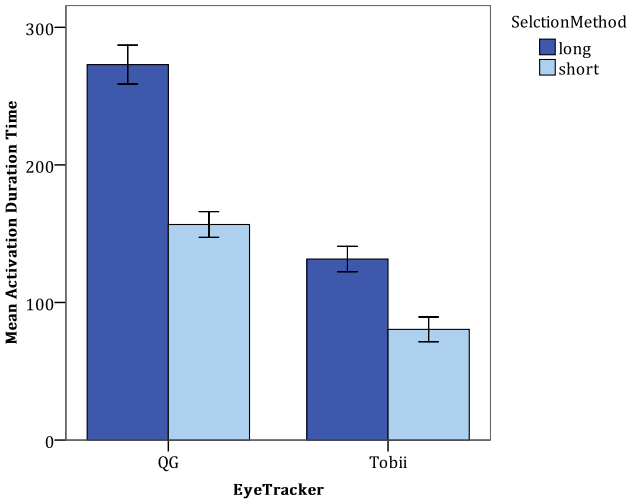


Figure 85: Differences in activation completion time depending on selection method within the two eye tracking systems, error bars represent standard error on mean

OVERALL SELECTION ERRORS BASED ON INPUT DEVICE

The number of *selection errors* was the number of times a participant completed a selection which did not match the target.

Maucley's test of sphericity was insignificant which meant that sphericity could be assumed. By logarithmically transforming the data a normal distribution was achieved; so parametric statistics could be used for this analysis (Figure 86).

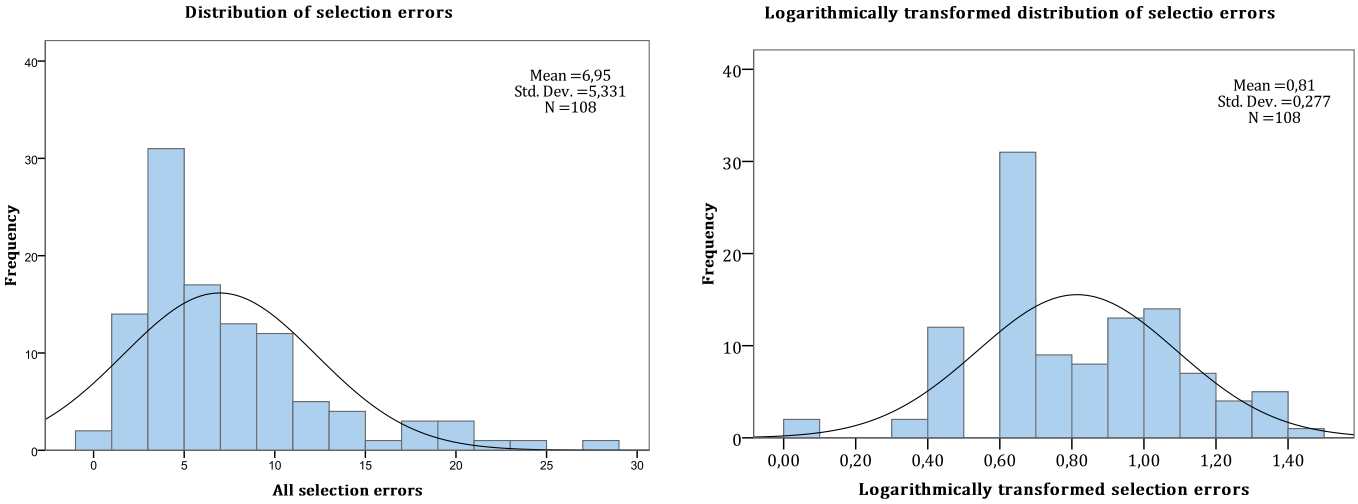


Figure 86: Original and logarithmically transformed distributions of selection errors on the two eye trackers

The results were analyzed using a repeated measures 2-tailed t-test with *input device* as the independent variable and *selection error* as the dependent variable measured in accumulated observations (Figure 87).

The mean values for the two conditions were based on n = 54 accumulated observations per result. The overall mean value for *selection errors* on the QuickGlance was 8,61 and the overall mean for the Tobii was 5.31.

There was a significant difference between the numbers of *selection errors* on the QuickGlance compared with the Tobii. t (53) = 3,271; p<0.01. There were significantly fewer *selection errors* on the Tobii.

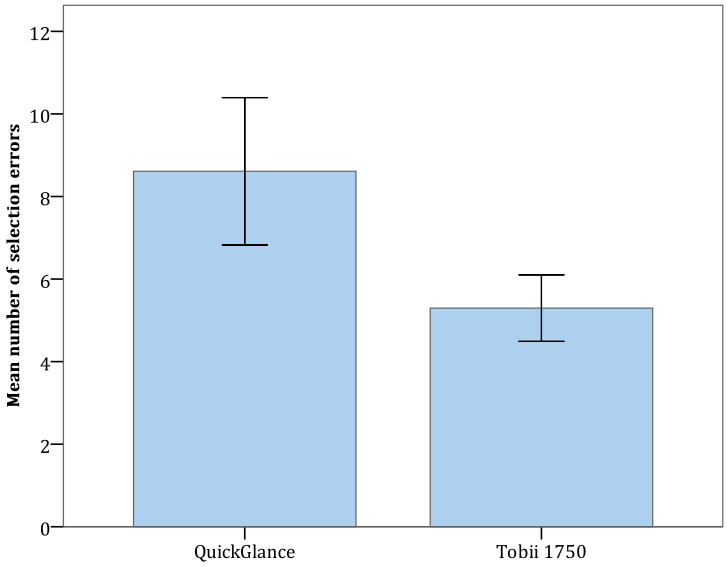


Figure 87: Overall selection errors on the two eye tracking systems, error bars represent standard deviation.

OVERALL TARGET ERRORS BASED ON INPUT DEVICE

T*arget error* was the number of targets which crossed the screen without being selected.

The data were explored and did not fit the assumptions required to complete parametric analyses. After ranking the scores a Wilcoxon signed-ranks tests was used where *input device* was the independent variable with two levels (QuickGlance and Tobii 1750) and *target error* was the dependent variable (Figure 88).

The mean values of the two conditions were based on n = 54 observations per result and were: Mean *target error* on the QuickGlance = 4,44 and on the Tobii = 0,11.

There was a significant difference between the numbers of *target errors* on the QuickGlance compared with the Tobii. $Z = -3{,}104$, $p<0.01$. There were significantly fewer *target errors* on the Tobii.

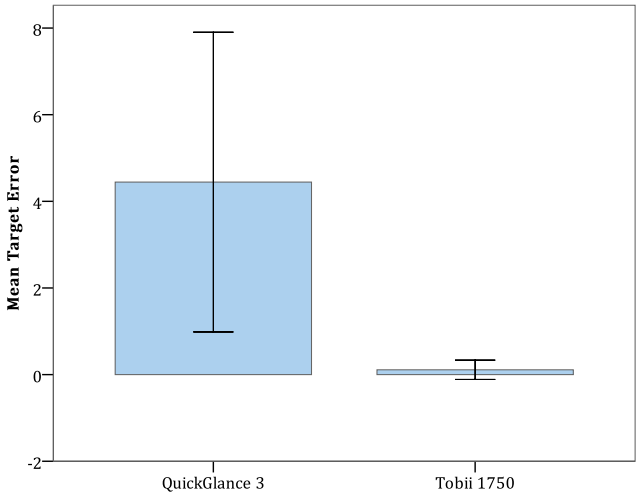


Figure 88: Overall missed target errors on the two eye tracking systems, error bars represent standard deviation.

7.1.5 FACTOR 2: SELECTION METHOD

OVERALL SELECTION COMPLETION TIME FOR LGG AND SGG

This analysis was done to substantiate the significant difference in *selection completion time* which was found in the *dwell*, *long* SSGG and *short* SSGG experiment.

Maucley's test of sphericity was insignificant which meant that sphericity could be assumed. By logarithmically transforming the data a normal distribution was achieved; so parametric statistics could be used for this analysis (Figure 89).

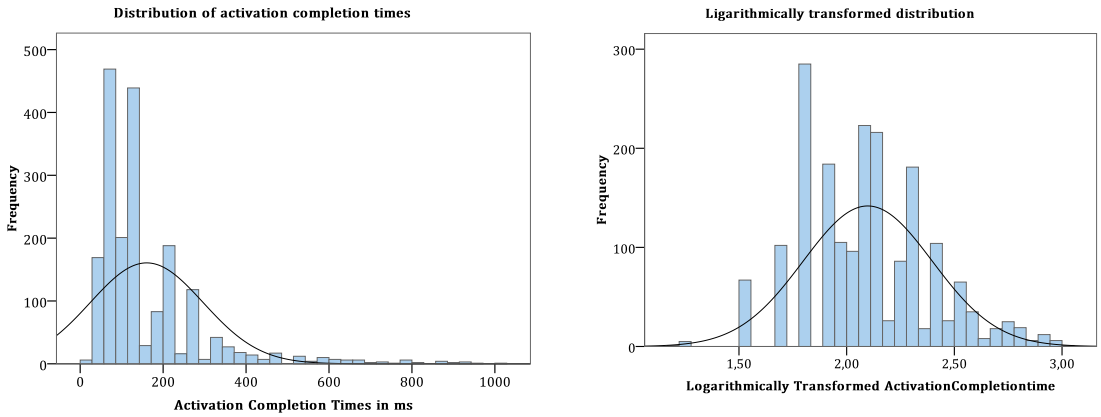


Figure 89: Original and logarithmically transformed distributions of selection completion times

The results were analyzed using a repeated measures 2-tailed t-test with *selection method* as the independent variable and *selection completion time* as the dependent variable measured in ms (Figure 90).

The mean values for the two conditions were based on n = 960 observations per result. The overall mean value for *short* SSGG was 118ms and the overall mean for *long* SSGG was 202ms.

There was an overall significant difference between *short* SSGG and *long* SSGG in all four directions on both machines t (959) = -16,299; p<0.01. *Short* SSGG were significantly faster than *long* SSGG.

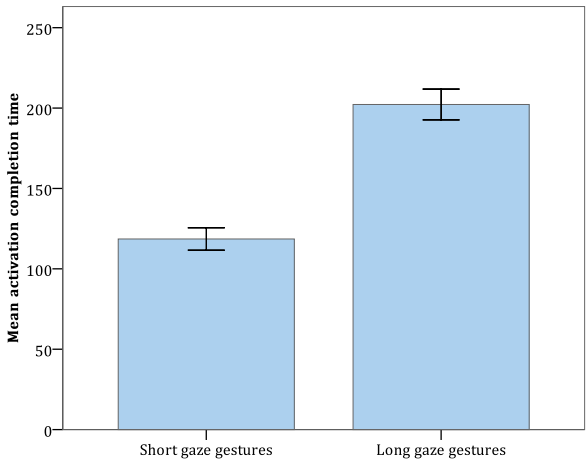


Figure 90: Overall difference in activation completion times based on selection method. Error bars represent standard error of mean.

OVERALL SELECTION ERROR BASED ON LGG AND SGG

In the *dwell*, *long* SSGG and *short* SSGG experiment there was no significant difference in the number of *selection errors* based on *long* SSGG and *short* SSGG. It was of interest to see whether this was also the case here.

Maucley's test of sphericity was insignificant which meant that sphericity could be assumed. By logarithmically transforming the data a normal distribution was achieved; so parametric statistics could be used for this analysis (Figure 91).

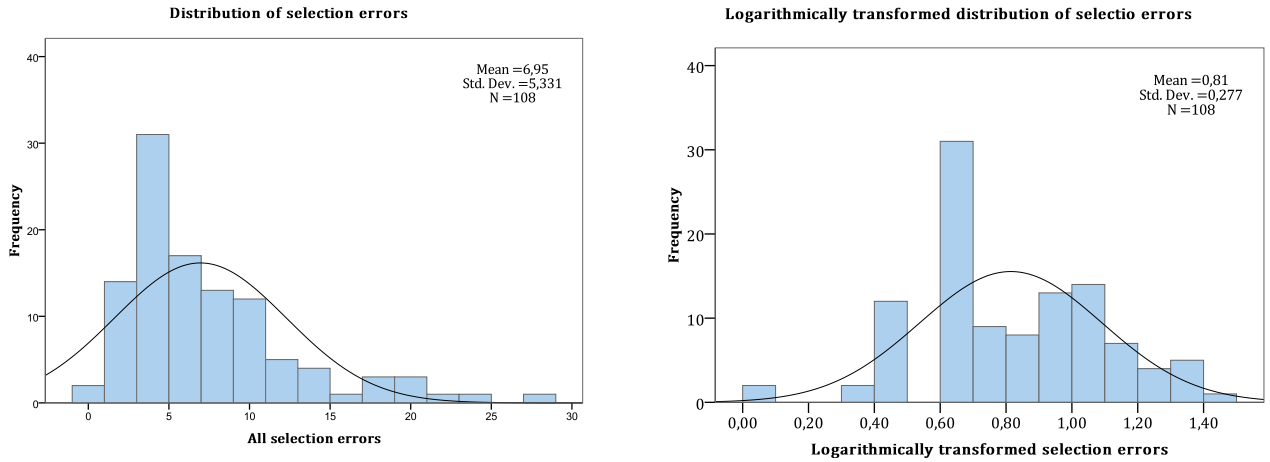


Figure 91: Original and logarithmically transformed distributions of selection errors

The results were analyzed using a repeated measures 2-tailed t-test with *selection method* as the independent variable and *selection error* as the dependent variable measured in accumulated observations (Figure 92).

The mean values for the two conditions were based on n = 54 accumulated observations per result. The overall mean value for *selection errors* in *short* SSGG was 7,43 and the overall mean for *long* SSGG was 6,48.

There was no overall significant difference in the number of *selection errors* produced using *long* SSGG and *short* SSGG in all four direction on both machines t (53) = 0,941; p = 0,351.

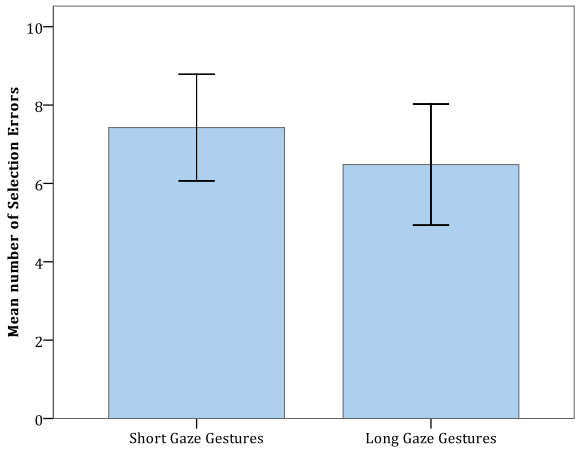


Figure 92: Overall differences in number of selection errors based on selection method, error bars represent standard error of mean.

MISSED TARGET ERROR BASED ON LGG AND SGG

The number of *target errors* was the number of targets which crossed the screen without being selected.

The data were explored and did not fit the assumptions required to complete parametric analyses. After ranking the scores a Wilcoxon signed-ranks tests was used where *selection method* was the independent variable with two levels (*long* SSGG and *short* SSGG) and *target error* was the dependent variable (Figure 93).

The mean values for the two conditions were based on n = 54 accumulated observations per result. The overall mean number of *target errors* for *short* SSGG was 1,17 and the overall mean for *long* SSGG was 3,39.

There was no overall significant difference in the number of *target errors* based on *selection method*. Z = -1,084, p = 0,279.

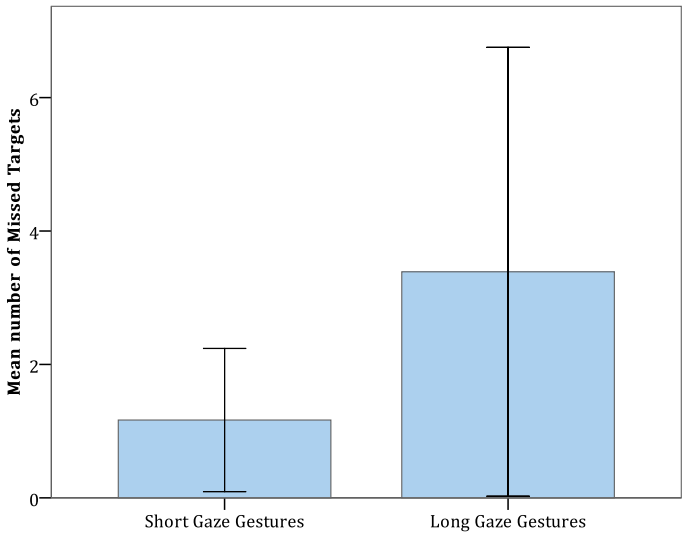


Figure 93: Overall differences in missed target errors based on selection method, error bars represent standard error of mean.

7.1.6 FACTOR 3: SELECTION DIRECTION

OVERALL SELECTION COMPLETION TIME BASED ON SELECTION DIRECTION

It was of interest to see whether there was an overall difference in *selection completion time* based on the four directions across both *input device* and *selection method*.

The data set was explored and Maucley's test of sphericity was insignificant which meant that sphericity could be assumed. By logarithmically transforming the data a normal distribution was achieved; so parametric statistics could be used for this analysis.

A one way ANOVA was conducted with *selection direction* as the independent variable with four levels (Left/Right; Right/Left; Top/Bottom; Bottom/Top) and *selection completion time* as the dependent variable measured in ms (Figure 94).

The mean values were based on n = 480 observations per result. The overall mean values were (Table 9):

Left /right	Right /left	Top /bottom	Bottom/top
137ms	132ms	187ms	184ms

Table 9: Mean directional *selection completion times*

There was an overall significant difference in *selection completion time* depending on the four directions. F (1, 3) = 34m163; $p < 0.01$. Right/left and left/ right were significantly faster than top/bottom and bottom/ top.

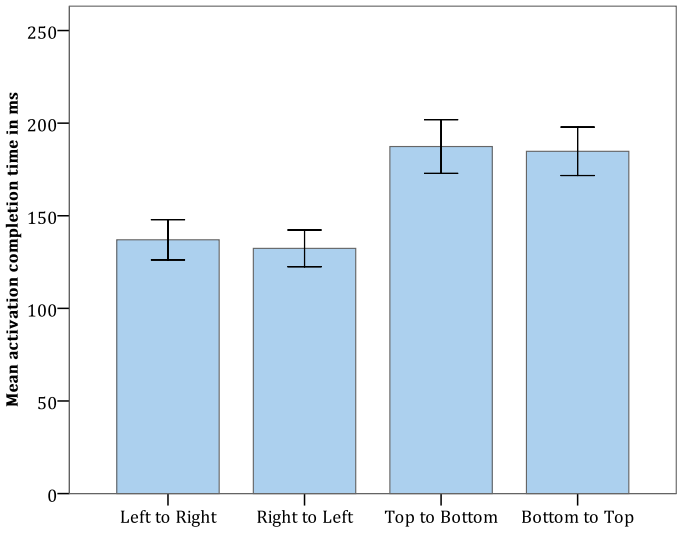


Figure 94: overall selection time based on selection direction. Error bars represent standard error of mean.

OVERALL SELECTION COMPLETION TIME FOR HORIZONTAL AND VERTICAL

As a consequence of the results presented above there seemed to be an interesting observation to be made in examining the overall horizontal and vertical *selection completion times*.

As it was the same dataset as the one used above, it was known that the assumptions for parametric statistics were upheld.

A *t*-test was used to examine the results of the overall horizontal and vertical *selection completion times*. *Selection direction* was the independent variable (horizontal and vertical) and *selection completion time* was the dependent variable measured in ms (Figure 95).

The mean values were based on n = 1024 observations per result. The overall mean horizontal activation completion time = 137ms and the overall mean vertical activation completion time was 184ms.

There was an overall significant difference in *selection completion time* depending on *horizontal* or *vertical gestures* t (1023) = -9,390; p < 0.01. *Horizontal* SSGG were significantly faster than *vertical* SSGG.

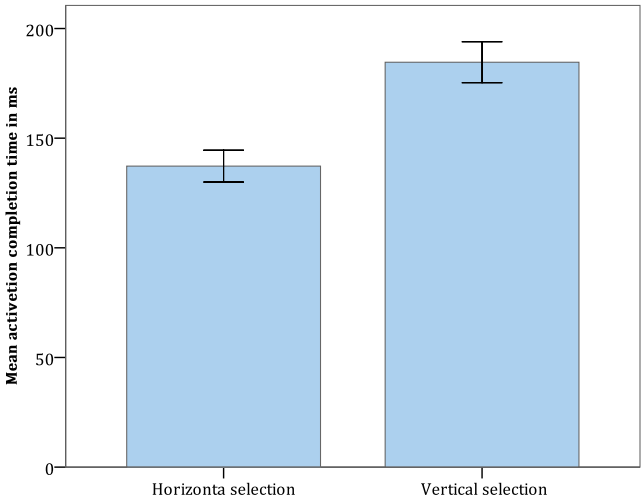


Figure 95: Overall selection times based on horizontal and vertical selections. Error bars represent standard error of mean.

SELECTION ERROR BASED ON FOUR DIRECTIONS

The *selection error* has until now been an accumulated measure for either *input device* or *selection method*. However, when analysing directional *selection error* the accumulated values are split up according to direction. Therefore the number of *selection errors* will at first glance appear smaller, than those previously presented.

The data were explored and did not fit the assumptions required to complete parametric analyses. After ranking the scores Friedman's non-parametric ANOVA was used where *selection direction* was the independent variable with four levels (Left/right; Top/Bottom; Right/ Left; Bottom/ Top) and *selection* error was the dependent variable (Figure 96).

The mean values were based on n = 108 observations per result. The four mean values for selection errors were: Left to right = 1,5; Top to Bottom = 1,55; Right to Left =1,78; Bottom to Top = 2,10.

There was no overall difference in *selection error* depending on the four different directions of selection $X^2_r = 7$, 605 (3, n=108), p = 0.055.

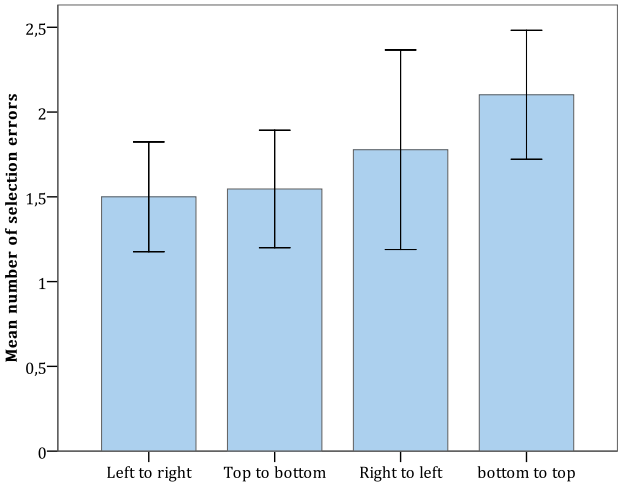


Figure 96: Selection error based on selection direction. Error bars represent standard error of mean.

7.2 DISCUSSION OF THE SINGLE GAZE GESTURES AND DIFFERENT INPUT

There were many other combinations of data which could have been examined. However, the above presented results represent those correlations which are considered of most relevance to gaze interaction design. As with the discussion section in the *dwell*, *long* SSGG and *short* SSGG experiment, each of the null hypotheses will be discussed.

7.2.1 FACTOR 1, INPUT DEVICE:

NULL 1: THERE WAS NO OVERALL DIFFERENCE IN *SELECTION COMPLETION* TIME ON EITHER *INPUT DEVICE*.

This null hypothesis was rejected, as there was a significant difference in *selection completion times* based on the two *input devices*. SSGG on the Tobii tracker were significantly faster than on the Quickglance.

This was both expected and unexpected. It was expected due to the higher frame rate and more appropriate smoothing algorithm of the Tobii. A consequence of this finding is the acknowledgement that saccade based gaze interaction might require different technological considerations than fixation based interaction. However, it was also slightly unexpected because the Tobii had a 17 inch monitor and the QuickGlance was only using a 13 inch monitor. This difference in size was thought to compensate slightly for the higher frame rate of the Tobii, however, this was not the case.

The mean SSGG *selection completion time* on the Tobii, which was 105ms, reflects more clearly the high speed potential of this type of *selection strategy*.

NULL 2: THERE WAS NO SIGNIFICANT DIFFERENCE IN *SELECTION COMPLETION TIMES* FOR *SELECTION DIRECTION* ON EITHER *INPUT DEVICE*.

This null hypothesis was rejected in two ways. Each of the four *selection directions* were compared to their counterparts on either machine. In this case the results showed that all of the *selection completion times* for each *selection direction* individually were faster on the Tobii than on the QuickGlance. This further substantiated that there was an overall significant difference between *selection completion times* on the Tobii and QuickGlance.

The other way in which this null hypothesis was rejected was by looking at the *selection completion times* within each *input device* to see whether there were any mutual differences on in *selection direction*.

Here it was established, which will also be substantiated later, that left/right and right/left movements were faster than top/bottom and bottom/top movements. This was an unexpected result.

In the pilot study it was found that top/bottom *selection completion times* were significantly faster. The exact opposite was shown in this experiment, where there seems to be a trend towards *horizontal* SSGG being faster than *vertical* ones. One of the reasons for this could be the difference in static and dynamic targets as well as the random target introduction in this second iteration of the experimental design. The static targets in the pilot study could have presented the top/down action with an unfair advantage due to placement in the selection sequence. In other words it might have been easier and faster to complete that specific *gesture* based on the *preceding gesture*. Another reason could have been the difference in length between horizontal and vertical gestures in the pilot study, as results indicate that gesture length makes a significant difference in selection completion time.

This observation will be further substantiated and elaborated upon later in this section.

NULL 3: THERE WAS NO DIFFERENCE IN *SELECTION COMPLETION TIME* FOR *SELECTION METHOD* ON EITHER *INPUT DEVICE*.

This null hypothesis was rejected in two ways. First there was a significant difference in *selection completion times* based on *selection method*. When comparing *short* SSGG on the Tobii and QuickGlance, they were significantly faster on the Tobii and the same was true for *long* SSGG. This was expected and further substantiates the initial observation that SSGG were faster of the Tobii than on the QuickGlance.

Secondly, there was a significant difference between *long* and *short* SSGG on each *input device.* On both systems *short* SSGG were significantly faster than *long* SSGG. This was not expected, but it substantiated the observations made in the *dwell*, *long* SSGG and *short* SSGG experiment. However, the assumption was that the significant difference in *long* and *short* SSGG on the QuickGlance, were mainly due to the technological restrictions of the system, such as the low frame rate and the automatic smoothing. It had therefore been expected that when the same experiment was conducted on a Tobii tracker, this pattern would not be present. However, it was – this was an indication that there might be a physiological reason which has to do with altered eye movement behaviour when dealing with interactive content, compared with eye movement patterns in perception.

NULL 4: THE DIFFERENT CAPABILITIES OF VARIOUS *INPUT DEVICE* HAD NO EFFECT ON *SELECTION ERRORS*.

This null hypothesis was rejected, as there was a significant difference in the number of *selection errors* on either *input device*. There were far fewer *selection errors* on the Tobii. There are at least two possible reasons for this.

First of all the higher frame rate and different smoothing algorithm on the Tobii made selections easier, so it was not so necessary for the participants to repeat SSGG in order to repeat selections. This implies that some of the *selection errors* on the QuickGlance were caused in the process of repeating.

The second reason could be due to the difference in monitor size. As mentioned the Tobii used a 17 inch monitor, compared to the 13 inch monitor of the QuickGlance. The longer distance on the Tobii could mean that accidental selections were less likely to occur.

NULL 5: THERE WAS NO DIFFERENCE IN *TARGET ERRORS* ON EITHER *INPUT DEVICE*.

This null hypothesis was also rejected, as there was a significant difference between the number of *target errors* on the two *input devices*. There were far fewer *missed targets* on the Tobii than on the QuickGlance.

The main reasons for this have already inadvertently been expressed. The faster response time of the Tobii made gaze gestures much easier to complete. In the *dwell*, *long* and *short* SSGG experiment, participant would often get slightly lost in the selection process, in the sense that when they started losing targets and they found it very difficult to 'catch up' again. This also occurred a few time in this experiment on the QuickGlance, however, no one had that experience with the Tobii tracker. The few incidents where targets were lost were mainly due to the participant having looked away, and they found it relatively easy to 'catch up' with the targets again.

7.2.2 FACTOR 2, SELECTION METHOD:

NULL 6: THERE WAS NO SIGNIFICANT DIFFERENCE IN *SELECTION COMPLETION TIMES* OF *SELECTION METHODS* BASED ON RESULTS FROM BOTH *INPUT DEVICES*.

This null hypothesis was rejected. As there was a significant difference between *long* and *short* SSGG based on all the results for all *selection directions* on both *input devices*.

This was of course expected, as the *selection methods* had already been examined based on *input device*. However, it was of interest to see how great the difference was in the overall means. For *short* SSGG it was 118ms and the overall mean for *long* SSGG was 202ms. This is a substantial difference and further emphasizes that there is the potential to use *long* and *short* SSGG to alternate between tasks in specially designed interfaces.

NULL 7: THERE WAS NO DIFFERENCE IN *SELECTION ERRORS* FOR *SELECTION METHOD* BASED ON RESULTS FROM BOTH *INPUT DEVICES*.

This null hypothesis was retained as there was no significant difference between the amounts of *selection errors* for *long* or *short* SSGG respectively. This corresponded with the results found in the *dwell*, *long* SSGG and *short* SSGG experiment.

However, this result was not expected. It was expected that there would be a higher number of *selection errors* associated with completing *shorter* SSGG. The reason for this was the general assumption that the smaller, simpler and more centralized a gesture was on the

screen, the more likely it was to cause accidental gesture completion and therefore a higher number of *selection errors.*

The number in this case was only slightly higher for *short* SSGG and not enough to be significant.

NULL 8: THERE WAS NO DIFFERENCE IN *TARGET ERROR* DEPENDING ON *SELECTION METHOD*.

This null hypothesis was retained as there was no significant difference between the number of *target errors* between the *long* and *short* SSGG conditions. This was also mirrored from the *dwell*, *long* SSGG and *short* SSGG experiment and was therefore expected.

7.2.3 FACTOR 3, SELECTION DIRECTION

NULL 9: THERE WAS NO OVERALL DIFFERENCE IN *SELECTION COMPLETION TIME* FOR *SELECTION DIRECTION* INCLUDING RESULTS FROM BOTH *INPUT DEVICES*.

This null hypothesis was rejected as there was a significant difference between *selection completion times* based on *selection direction*. There was a significant difference between the two horizontal conditions left/right and right/left and the two vertical conditions top/bottom and bottom/top. There was no significant difference within those two groups.

As mentioned earlier this was an unexpected result. In the pilot study it was speculated that one of the reasons that the top/bottom condition was significantly faster than the horizontal conditions, was that the distance between the vertical fields were closer to each other, compared with the horizontal ones. This speculation was most probably correct as a trend has been revealed that there was a significant difference in *selection completion time* depending on lengths of gesture. Therefore, having the horizontal selection fields further apart than the vertical ones would have made a difference in the initial pilot study. Another consideration, in regard to the top/down significance in the pilot study, which caused a change in the experimental design was the static targets and fixed nature of the pilot study. This could, as previously mentioned, very well have unfairly benefitted the top/bottom SSGG.

These arguments support the validity of this current experimental design. The horizontal trend could therefore indicate systemic eye movement behaviour. Whether this difference has

to do with cultural reading patterns or whether horizontal eye movements are fundamentally faster requires further investigation.

NULL 10: THERE WAS NO DIFFERENCE IN *SELECTION COMPLETION TIME* FOR OVERALL HORIZONTAL AND VERTICAL SELECTIONS BASED ON RESULTS FROM BOTH *INPUT DEVICES*.

This null hypothesis was rejected, as mentioned above. As horizontal SSGG were significantly faster than vertical ones. The mean *selection completion time* was for *horizontal gestures* = 137ms and for *vertical gestures it* was 184ms. It could be argued that even though this is a significant difference in a statistical sense, it might have smaller implications for practical application.

However, if *horizontal gestures* are systemically faster than *vertical gestures* this principle could influence the way interfaces for *single*, *complex* and *continuous gestures* should be designed.

NULL 11: THERE WAS NO DIFFERENCE IN NUMBER OF *SELECTION ERRORS* BASED ON *SELECTION DIRECTION* WITH RESULTS FROM BOTH *INPUT DEVICES*.

This null hypothesis was retained, as there was no significant difference between the numbers of *selection errors* depending on *selection direction*. Overall, the lack of significant difference in *selection error* was mainly an indication that the experiment and the selection patterns in the experiment were well balanced.

There was an expectation of a bias towards *vertical gaze gestures* being more error prone. A reason for this assumption was that the targets were descending from top to bottom. This could potentially have forced a higher number of vertical searches; which could, in turn, have caused *accidental gesture completion* and therefore a higher number of *selection errors*. However, this was not the case.

7.3 GENERAL DISCUSSION OF SINGLE GAZE GESTURES AND DIFFERENT INPUT

There are three main observations which can be derived from this experiment. First of all there was a systematic overall difference between the two input devices. Secondly, the results regarding *long* SSGG and *short* SSGG substantiated the results found in the *dwell*, *long* SSGG and *short* SSGG experiment. And finally, the most surprising, and therefore most interesting, results were the seemingly systematic difference between horizontal and vertical SSGG.

The overall significant difference between the systems (QuickGlance, Tobii) led to the observation that system set-up does have an effect on the *selection completion speed* and *selection error* rates of SSGG selections. This is a natural consequence of gaze gestures being based on saccades. Eye trackers, particularly those designed for gaze interaction*s*, have as mentioned been focussed on reliable solutions for fixation based interaction. However, as a consequence of these findings it was clear that consideration should be taken from a technical perspective to support saccadic interaction to a greater degree.

The findings regarding *long* and *short* SSGG support the findings in the *dwell*, *long* SSGG and *short* SSGG experiment. This was not surprising, as both experiments were conducted in the same test environment. However, if this principle of interaction were to be further substantiated and explored in other contexts, it could make a useful addition to the gaze interaction design toolbox. As mentioned in the discussion of the previous experiment various zones of interaction depending on gesture length, could be implemented.

The findings regarding the *selection completion times* based on direction were the most interesting. In the pilot study top/bottom SSGG had been significantly faster. The exact opposite was shown in the experiment with the two eye trackers; *horizontal gaze gestures* were faster than *vertical* ones. There can be several reasons for this. First of all, as the sequence of targets was repeated in the same order; the top/down action might have had an unfair advantage. It might have been easier and faster to complete the top/down gesture based on the preceding action in that order. As a consequence of this, a second observation can be presented. This was that eye movements may behave differently depending on static or dynamic content being visualized on screen. If Yarbus (Yarbus et al., 1973) stated that people's eye movements behaved differently dependent on intent, the same could be applied

in reverse, namely that eye movements behave differently depending on the dynamics of the content being presented.

Finally, this leads to a general consideration which applies to all of the observations made in the course of these experiments – very little is known about the way eye movements behave during interactive tasks. Eye movement research has focussed on dealing with perception eye movements on images, in actions, while reading etc.. If the vocabulary of gaze is to be increased with sustainable actions, it could be useful to know how gaze behaves during these actions.

It is important to point out that the *selection completion time* measurement in these experiments should not be viewed as an attempt to model saccadic eye movements. Neither of the eye tracking systems used here are precise enough. The results represent *patterns* of *selection completion times*, *task times*, *selection errors* and *target errors*. And in understanding these patterns there are several issues which require addressing, some of them regarding the experimental design. The simplification of the interface compared with the pilot study was a definite improvement in the sense that it was more easily explained and understood by the participants. However, the choice to show the initiation colours as the contours of the selection fields was in hindsight probably not the best representation. The contour represents the edges of the selection fields this meant that quite often the participant would be looking at the edge of the selection field; in order to navigate. This in turn implies that natural jitter and any form of skewed calibration could have effected whether or not the selection field would be activated.

The most interesting result, which concerned direction of gestures, might also be the most uncertain. The results indicate that direction does affect eye movements. However, this could simply be a consequence of the experimental design having slightly larger horizontal selection fields. The horizontal fields were increased so that the distance between vertical and horizontal selection fields was the same as a consequence the horizontal selection fields were slightly bigger. In the next experiment this was remedied.

Research in this chapter was published in: Mollenbach, E., M. Lillholm, A. Gale, and J. P Hansen. "Single gaze gestures." In ETRA *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, 177–180, 2010.

8 SINGLE STROKE GESTURES WITH AND WITHOUT VISUALIZATION

This experiment was designed to explore aspects of single stroke gaze gestures (SSGG) which relate to potential future implementation of this gaze selection strategy.

One of the ideas for future implementations of SSGG was to be able to complete them *without* any visual feedback, thereby freeing up screen real-estate to be used for something else. Potentially, this could also be a layout principle which allowed dwell and gaze gestures to be implemented simultaneously in the same interface; having dwell-buttons in the centre of the screen and transparent gesture selection fields around the rim. As a consequence, the main focus of this experiment was to investigate transparent initiation and completion fields. However, this was also a repetition of the *dwell, long and short* SSGG experiment, with some adjustments. These adjustments were intended to highlight some of the issues which have already been explored, as well as creating an opening for new aspects of potential gaze gesture implementations.

8.1 DESIGN CONSIDERATIONS

The first design feature which was reconsidered was the size of initiation fields. The fact that the initiation and completion fields for the *horizontal* SSGG were larger than those in the *vertical* direction could have biased the results towards faster *horizontal selection completion times*. As a consequence the fields for both *long* and *short* SSGG were implemented with equal size; as well as maintaining the principle of equal distance between *horizontal* and *vertical selections* (Figure 97).

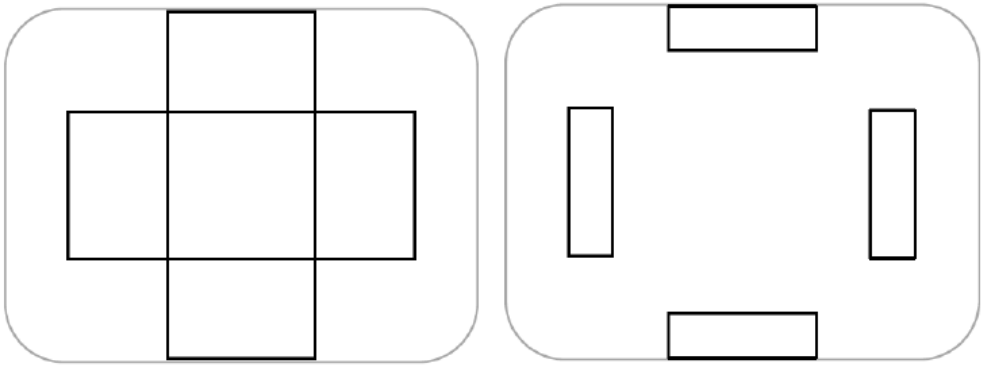


Figure 97: The new implementation of short (left) and long (right) SSGG with equal size initiation and completion fields.

Secondly, in the previous experimental design the targets in the task appeared all along the horizontal axis at the top of the screen (Figure 98). This was initially done in order to force perception, inspection and navigational eye movements to occur any place on the screen. However, this type of task interface layout might not be the best way of visualizing information for this type of selection strategy. It could potentially be more beneficial to present information away from the activation fields, in other words only in the centre of the screen.

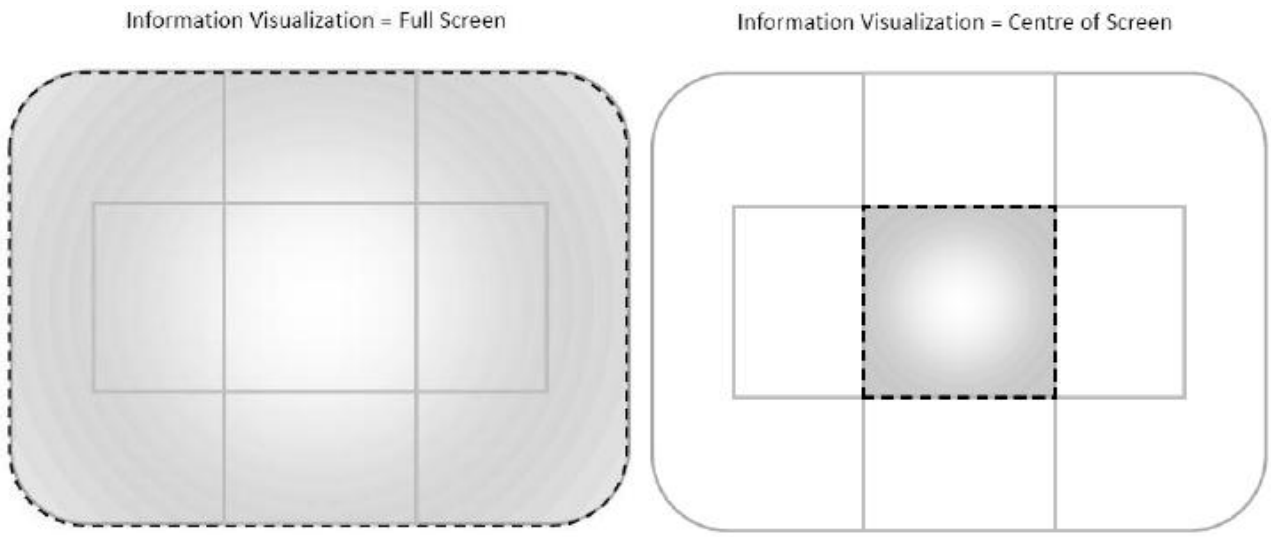


Figure 98: Full screen information visualization vs. Centre of screen information visualization.

There were two reasons for this. First of all Tatler (Tatler, 2007) expressed that there is a bias towards looking at the centre of the screen, regardless of saliency. It therefore makes sense to display information visualization in the centre.

Secondly, presenting targets all over the screen might cause unnecessary accidental gesture completion, by having targets in selection fields, and therefore an overinflated number of *selection errors*. Because of this the task area was limited to the centre part of the screen. Targets were presented at the top of the centre field and disappeared at the bottom (Figure 99). This infringes on the concept of freeing up screen real-estate. However, the previous experiments have already shown that SSGG can be completed while utilizing the entire screen. This experiment sought to establish a potential ideal condition, especially for gestures without visual feedback.

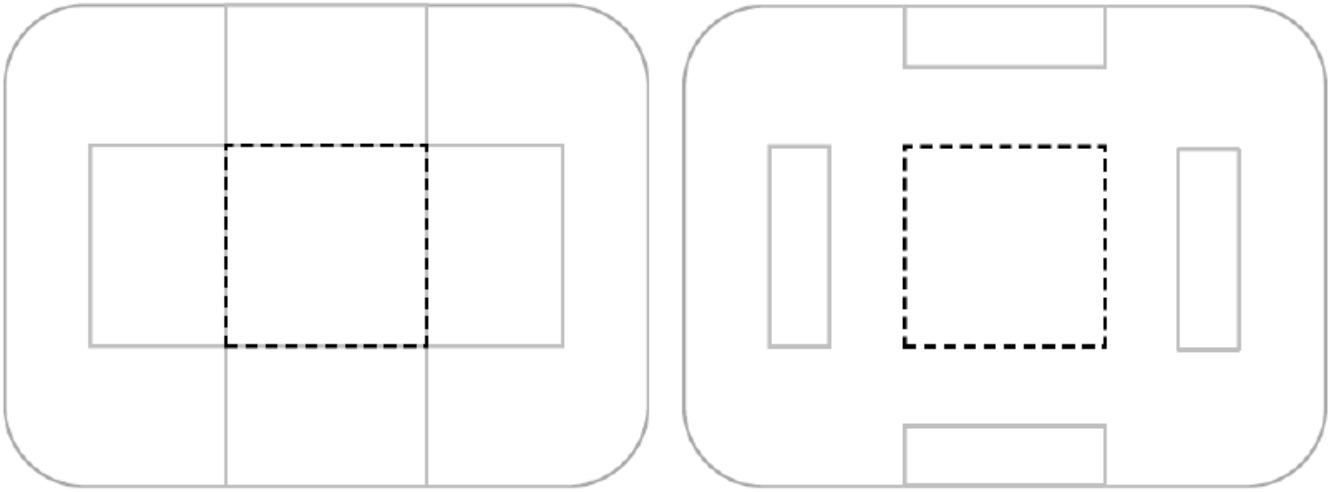


Figure 99: Centre visualization field were targets were presented at the top and disappeared at the bottom.

The targets in the previous tasks had coloured squares which descended the screen. These were abstract and there was no innate mapping between the gesture action and the target. In other words there was no connection between the colour yellow and a top/bottom SSGG. In order to create this connection between target and action the coloured squares were exchanged with arrows pointing in one of the four SSGG directions. An arrow in the gesture interfaces represented a 'full' gaze gesture; the SSGG should start at the beginning of the arrow and end at the arrow point. In other words an arrow pointing to the left meant a SSGG action from right/left (Figure 100). The task was again to select the target which was furthest in its decent down the screen.

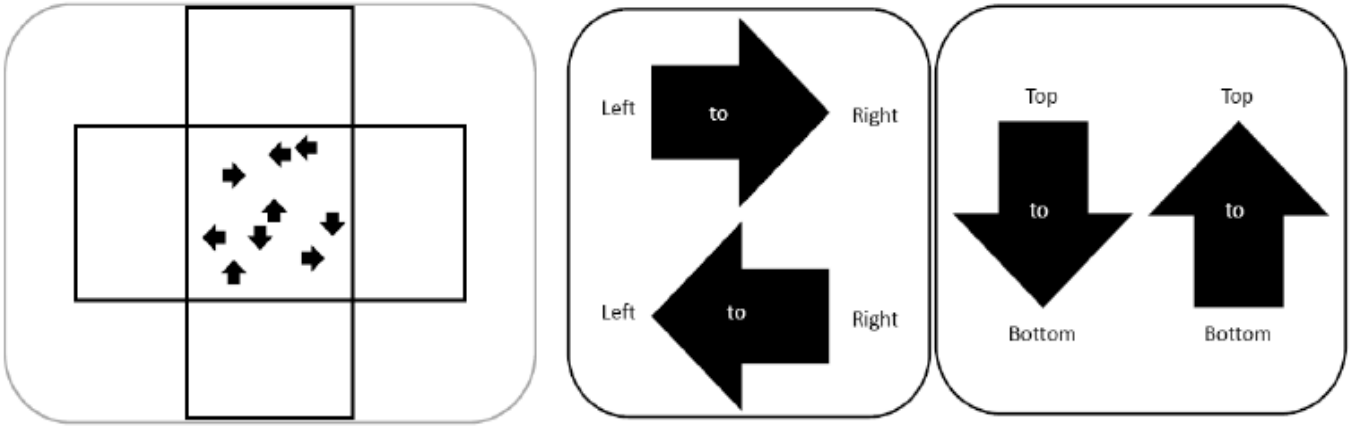


Figure 100: New target design. Each target symbolising a complete gesture action

The experiment was designed to explore two overall selection strategies again, SSGG and *dwell selection.*

The SSGG part of the experiment was split into four selection strategy tasks: *long* SSGG *with visualization*, *long* SSGG *without visualization*, *short* SSGG *with visualization* and *short* SSGG *without visualization* (Figure 101). The arrow shaped targets afforded the transparent initiation and completion fields as all of the required task information was implied in the target shape.

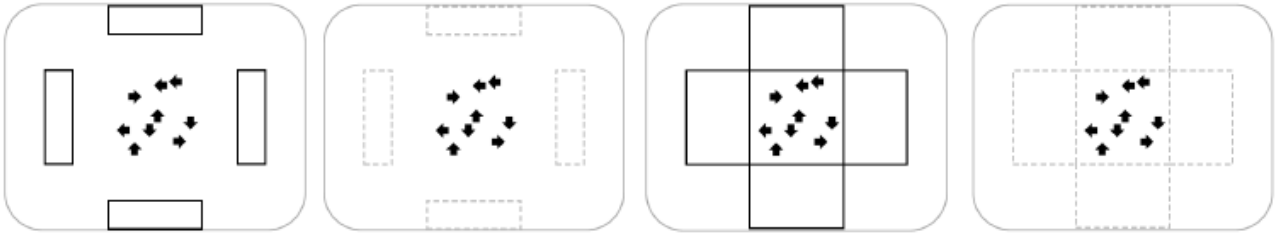


Figure 101: The SSGG tasks with and without visualization.

The *dwell selection* part of the experiment was also implemented slightly differently compared to the previous experiments. The change of target visualization from colours to arrows meant that the original layout of the dwell buttons at the bottom of the screen was inappropriate, because of the implied directional information of the arrow targets. In other words the cognitive load of finding the right arrow button without any visual mapping would be too great compared to the mapping which would occur in the SSGG conditions.

Because of this the dwell-buttons were placed in a circle around the centre of the screen, where they matched the direction in which their target arrow was pointing. In other words an arrow pointing to the left meant selecting the dwell-button to the left. The selection feedback to the user was again given on the dwell-button (Figure 102). Dwell selection was implemented with five selection completion times: 25ms, 50ms, 100ms, 200ms, 400ms; these fixation times were chosen to explore a range of dwell *selection completion times* which were more comparable with SSGG.

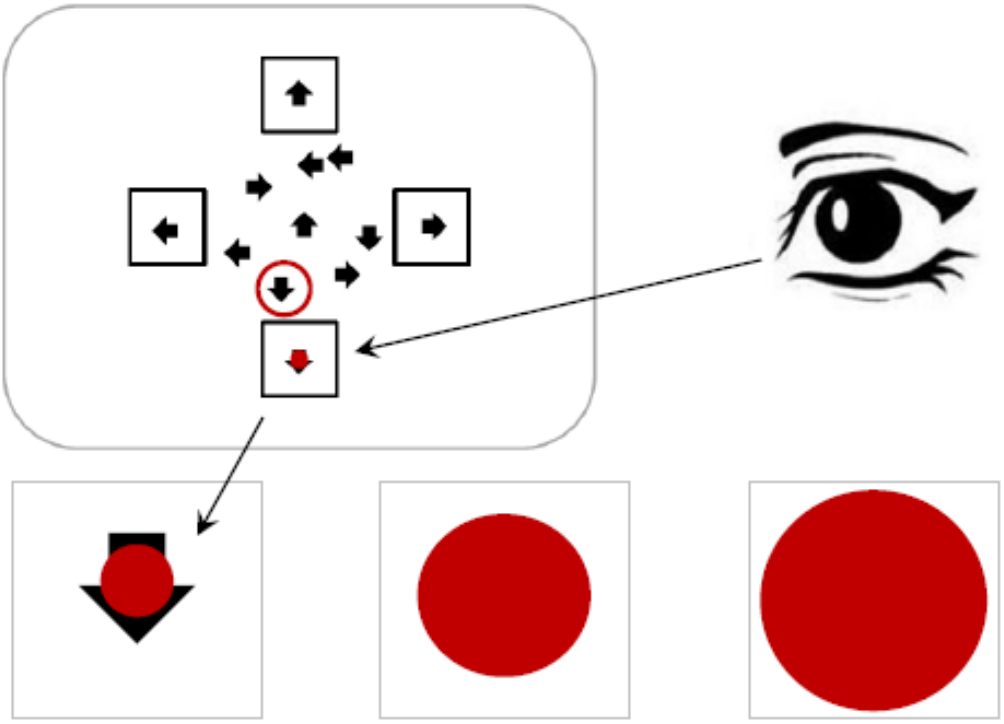


Figure 102: The dwell selection implementation. Targets were pointing to their matching buttons and feedback was given on the buttons. The red circle represent the feedback which was given to the user. It increased in size as the user was fixating on the target.

8.2 EXPERIMENTAL DESIGN AND RESULTS

This experiment sought to examine the difference between SSGG *with* and *without visualizations*; as well as further elaborating on the comparison between SSGG and *dwell selection*.

The task was to select targets as they descended the screen. The targets were arrows which indicated which gesture should be completed in the SSGG conditions and which dwell button should be fixated upon in the dwell conditions.

SSGG were completed by looking from an initiation field on one side of the screen to the completion field on the opposite side of the screen, this was to be done within a 1000ms timeframe. If a selection was not completed within that timeframe the system would reset. The feedback was given to the user in three ways depending on three overall conditions. Firstly, selection field feedback was given in the SSGG conditions where visualization was on. The selection fields were visible and shifted to a light grey colour when looked upon. Secondly, in the *no- visualization* SSGG conditions, the only feedback given to the user was that

the target would disappear when it had been selected. Finally, in the *dwell selection* condition the feedback was given on the dwell button as shown in figure 102.

Short SSGG required the user to cover 30% of the screen and *long* SSGG required the user to cover 60% of the screen. These were shorter than the distances in the previous experiment. The reason for this change was twofold. First it was of interest to see if the difference in completion times of gesture lengths was still significant with relatively shorter distances. The second reason was that these distances suited the target area which was implemented.

Participants were introduced to the test environment and had the opportunity to try SSGG and *dwell selection* before the actual experiment was conducted. Eighteen participants (9 female) completed the experiment. Seven had used an eye tracker before. The application was written in Java and the experiment was completed on a LC technologies eye follower system running at 120 frames per second. Once again the experiment was completed on yet a different eye tracker, with an even higher frame-rate. The question is whether there is an upper boundary where a higher frame-rate no longer effects the *selection completion time*.

The experiment was balanced by having participants complete the three overall conditions (*long* SSGG, *short* SSGG and *dwell selection*) in different order. Furthermore the *dwell selection* condition altered between beginning with 25ms and beginning with 400ms. All of the SSGG conditions started by having visualization on and subsequently off. This was based on the notion that gaze gestures without feedback need to be learnt.

The independent variables were:

Selection method: Overall there were nine different *selection methods,* based on four separate parameters; first, the overall *selection methods* of SSGG and *dwell selection;* secondly, for SSGG, *selection length* (*long* and *short* SSGG); thirdly, for SSGG, *with* and *without visualization* and finally, for *dwell selection*, five increments of *fixation duration time*.

The nine levels of *selection method* were therefore: *Long* SSGG *with visualization*, *long* SSGG *without visualization*, *short* SSGG *with visualization*, *short* SSGG *without visualization* and *dwell selection:* 25ms, 50ms, 100ms, 200ms and 400ms.

Each participant had to complete 20 successful selections per level of *selection method*.

Selection direction was the other independent variable with four levels: left/right; right/left; top/bottom; bottom/top.

The dependent variables were:

Selection Time: the time from when the user exits the initiation field and enters the opposite field measured in milliseconds (ms)

Task completion Time: the time elapsed between each successful selection. Both accumulated and individual task times were recorded and measure in milliseconds (ms).

Selection Error: A full completed selection which does not respond to the current target.

Target Error: Targets which descent the screen and pass through the other side without being selected.

The null-hypotheses which this experiment seeks to reject or retain are as follows.

Factor 1, Long and Short SSGG:

> Null 1: There was no difference between the *selection completion time* for *long* and *short* SSGG.
> Null 2: There was no significant difference between *task completion times* of *long* and *short* SSGG.
> Null 3: There was no difference in the number of *selection errors* between *long* and *short* SSGG.
> Null 4: There was no difference in the number of *target errors* between *long* and *short* SSGG.

Factor 2, Visualization:

> Null 5: There was no overall significant difference between *selection completion times* of SSGG *with* or *with visual feedback*.
> Null 6: There was no overall significant difference between *task completion times* of SSGG *with* or *without visual feedback*.
> Null 7: There was no overall significant difference between number of *selection errors* for SSGG *with* or *without visual feedback*.

Null 8: There was no overall significant difference in the number of *target errors* for SSGG *with* or *without visual feedback*.

Factor 3, Dwell Selection:

Null 9: There was no overall significant difference between the *task completion times* between the different *dwell conditions*.

Null 10: There was no overall difference in number of *selection errors* for the different *dwell conditions*.

Null 11: There was no overall difference in the number of *target errors* for the different *dwell conditions*.

Factor 4, Overall and Combined measure:

Null 12: There was no overall difference in *task completion time* between any of the conditions.

Null 13: When using the *combined measure* on all of the nine *selection method* conditions there will be no significant difference between calculations.

8.2.1 LONG AND SHORT SSGG

SELECTION COMPLETION TIME FOR LONG AND SHORT SSGG

As with the previous experiments it was only relevant to compare *selection completion times* for the SSGG conditions. This first analysis was based on *long* and *short* SSGG both *with* and *without visualization*.

The data were explored and Levene's test for homogeneity of variance was found to be significant, so the variance between the *long* and *short* SSGG datasets respectively was too great to employ parametric statistics. Also the data were not normally distributed.

After ranking the scores a Wilcoxon signed ranks test was used to compare the means. *Selection length* was the independent variable with two levels *long* and *short* SSGG and *selection completion time* was the dependent variable measured in ms (Figure 103). The mean values were based on n=720 observations per result. The grand mean for the two *selection methods* was 152ms. For *long* SSGG = 186ms and *short* SSGG = 118ms.

There was a significant difference between *long* and *short* SSGG $Z = -12{,}915$; $p < .001$. The selection completion time for *Short* SSGG was significantly faster than for *long* SSGG.

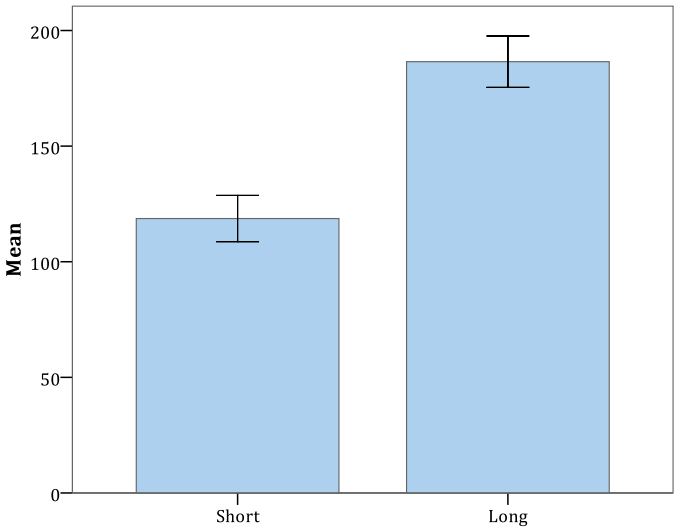


Figure 103: Mean of selection completion times for long and short gaze gestures. Error bars represent standard error of mean

TASK COMPLETION TIME FOR LONG AND SHORT SSGG

Task completion time was the elapsed time between two successful selections. In the previous experiments this had been a cumulative measure. However, in this experimental design *task completion times* were recorded for each selection separately.

The data were explored and neither the assumptions of homogeneity of variance or normal distribution were met, so a non-parametric analysis was conducted.

After ranking a Wilcoxon signed ranks test was used to compare the means. *Selection length* was the independent variable and *task completion time* was the dependent variable measured in ms (Figure 104). The mean values were based on n=720 observations per result. The grand mean for both conditions was 1949ms. For *long* SSGG the mean was 2151ms and for *short* SSGG the mean was 1747ms.

There was a significant difference in *task completion time* between *long* and *short* SSGG $Z = -3{,}671$; $p < .001$. *Task completion times* were significantly faster for *short* SSGG than for *long* SSGG.

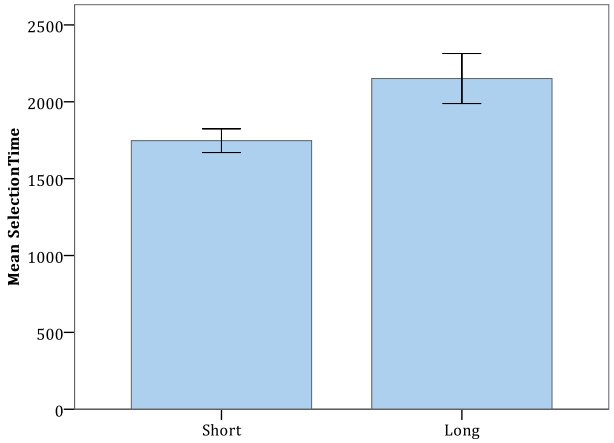


Figure 104: Mean of task completion times for long and short gaze gestures. Error bars represent standard error of mean

SELECTION ERROR FOR LONG AND SHORT SSGG

Selection error was the number of times a selection was completed which did not correspond to the current target. This was seen as an indication of accidental gesture completion.

The data were explored and neither the assumptions of homogeneity of variance or normal distribution were met, so a non parametric analysis was conducted.

A Wilcoxon signed rank test was completed to compare the *selection error* means of the two conditions. *Selection error* was measured as both an accumulated number and for selection independently. In this case the means were based on the accumulated value (Figure 105). The number of observations is therefore n=36 per result. The grand mean was 4,31 *selection errors*. For *long* SSGG the mean was 4,72 and for *short* SSGG it was 3,89.

There was a no significant difference between the number of *selection errors* for *long* and *short* SSGG $Z = -0{,}306$; $p = 0{,}759$.

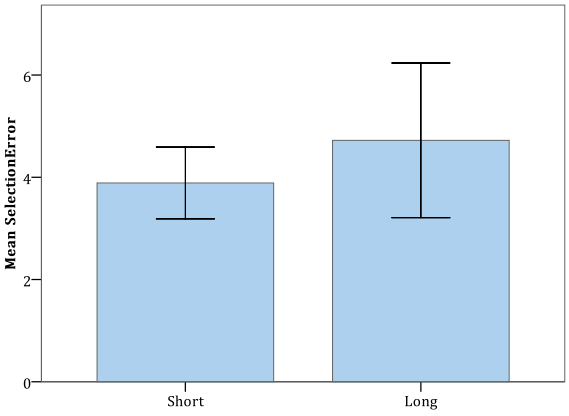


Figure 105: Mean of selection errors for long and short gaze gestures. Error bars represent standard error of mean

TARGET ERROR FOR LONG AND SHORT SSGG

Target error was the number of targets which crossed the target field without being selected.

The data were explore and outliers were removed (± 3 standard deviation) based on a linear regression. The reason for this was that a qualitative inspection of the data indicated that there was one outlier, which deviated greatly from the rest of the data. Levene's test of homogeneity of variance was significant and the data was not normally distributed, so non-parametric methods had to be applied.

A Wilcoxon signed ranks test was used to compare the two means, which were based on n=34 observations per result (Figure 106). The grand mean for both *long* and *short* SSGG was 0.08 *target errors*. The mean for *long* SSGG was 0,19 and for *short* SSGG it was 0,0.

There was a no significant difference between the number of *target errors* for *long* and *short* SSGG $Z = -1{,}633$; $p = 0{,}102$.

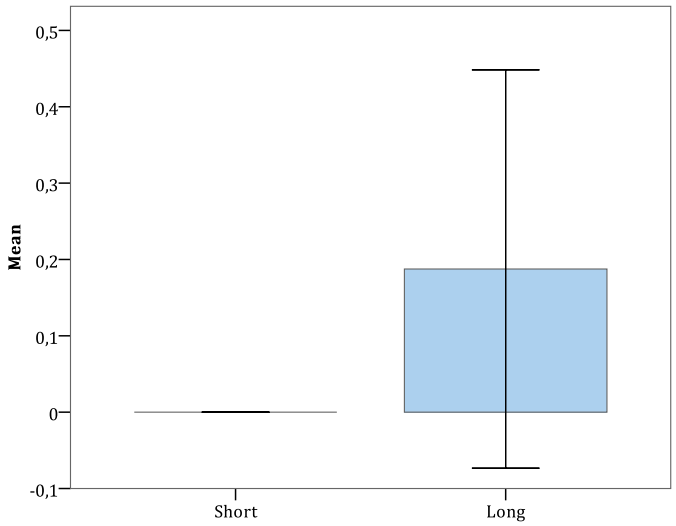


Figure 106: Mean of target errors for long and short gaze gestures. Error bars represent standard error of mean

8.2.2 VISUALIZATION

SELECTION COMPLETION TIMES WITH AND WITHOUT VISUALIZATION

As previously explained it was the main focus of this experiment to explore what effect removing the visual feedback for the SSGG would have.

The data were explored and the assumptions for the use of parametric statistics were not met, so a Wilcoxon signed ranks test was used to compare the means (Figure 107). The grand mean for both types of visual feedback was the same as the grand mean for *short* and *long* SSGG 152ms. *With visualization* the mean was 190ms and *without visualization* 115ms.

There was a significant difference between the *selection completion times* for SSGG with and without visualization $Z = -11{,}401$; $p < 0{,}001$. SSGG *without visualization* were significantly faster than SSGG *with visual feedback*.

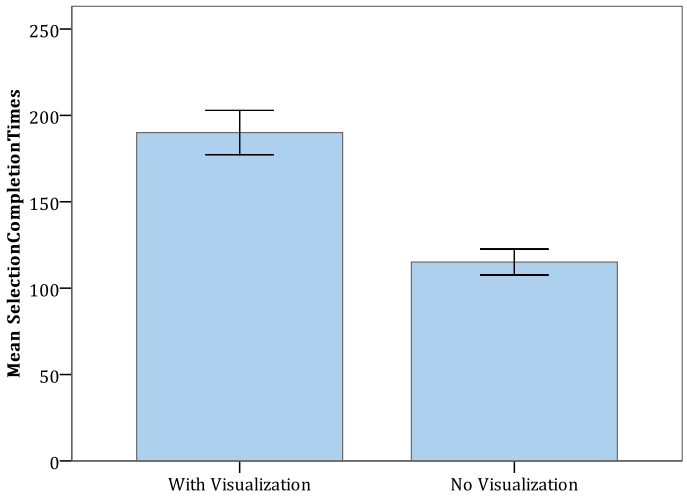


Figure 107: Mean of selection completion times for gestures with and without visual feedback. Error bars represent standard error of mean

TASK COMPLETION TIMES WITH AND WITHOUT VISUALIZATION

Task completion time was a combined measure of the selection action itself; as well as any faulty selections and the time spent on navigational search

The data were explored. Levene's test for homogeneity of variance was found to be insignificant p = 0,145. By squaring the data a normal distribution was achieved (Figure 108). Therefore parametric statistics could be used to analyze the data.

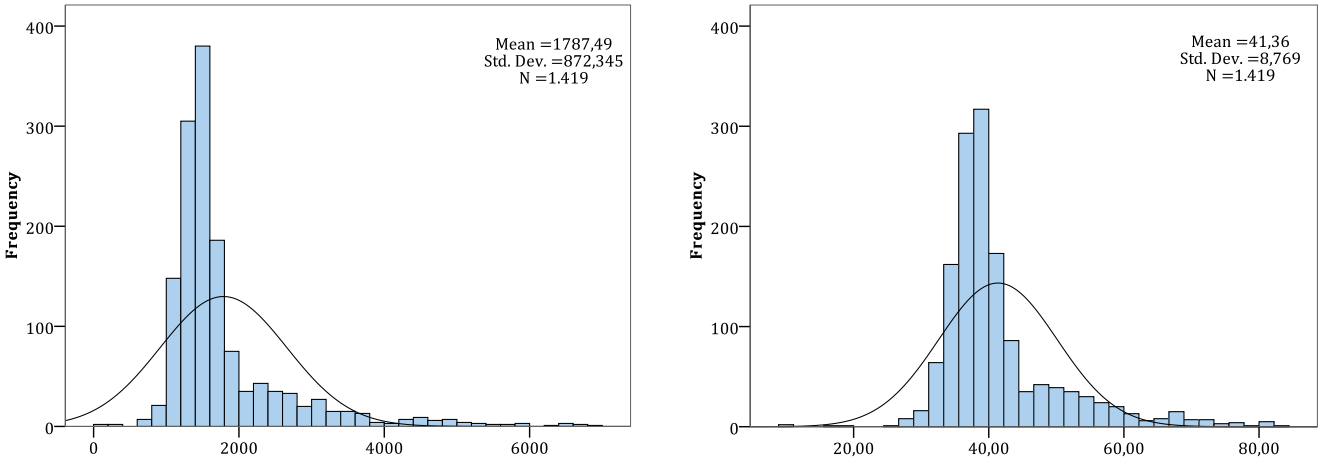


Figure 108: Original and square rooted distributions of task completion times.

A two-tailed repeated measures t-test was used to analyse the data (Figure 109). Each of the mean results was based on n=720 observations. The grand mean for *task completion time with*

and *without visualization* was 1949ms. For SSGG *with visualization* the mean was 1940ms and for SSGG *without visualization* the mean was 1958ms.

There was no significant difference in *task completion time* for SSGG *with* or *without visual feedback*. t (719) = -0,199; p = 0,843.

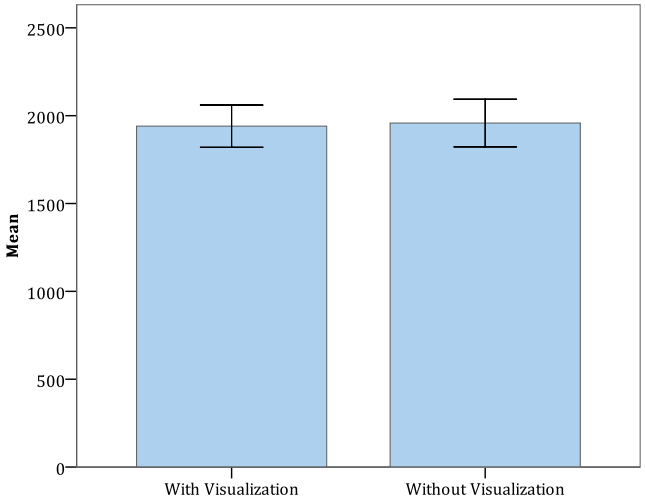


Figure 109: Mean of task completion times for gestures with and without visual feedback. Error bars represent standard error of mean

SELECTION ERROR WITH OR WITHOUT VISUALIZATION

It was of interest to see whether the removal of the visual feedback would have an effect on the number of *selection errors* and thereby the amount of accidental gesture completion which would occur.

The data were explored and Levene's test for homogeneity of variance was found to be insignificant p = 0,602. After logarithmically transforming the data a normal distribution was achieved and parametric statistics could therefore be preformed (Figure 110).

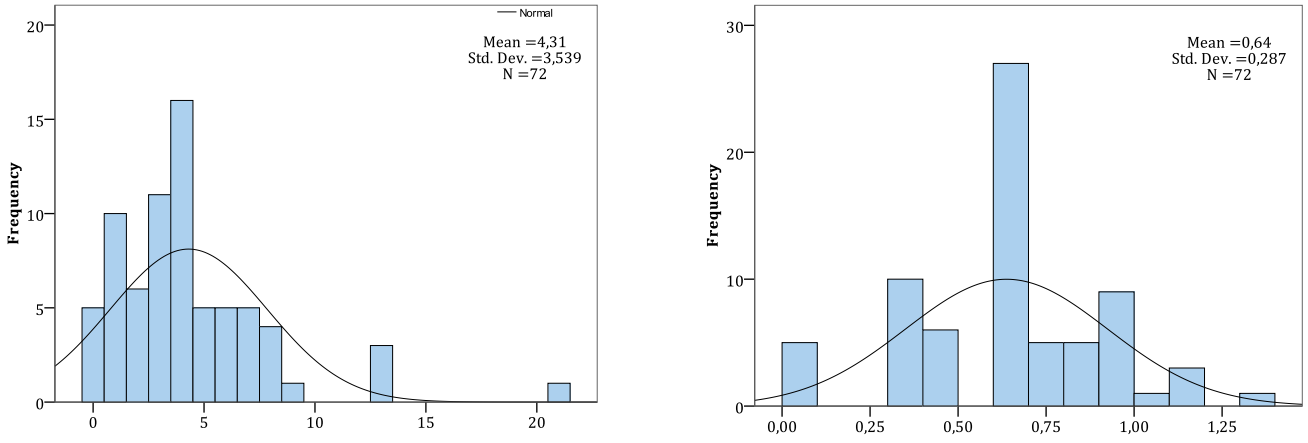


Figure 110: Original and logarithmically transformed distributions of selection errors.

A two tailed repeated measures t-test was used to compare the means (Figure 111). The mean values were based on n = 36 observations per result. The grand mean for *selection errors* of both types of visual feedback was 4, 31. For SSGG *with visualization* it was 4,03 and for SSGG *without visualization* it was 4,58.

There was no significant difference in the number of *selection errors* for SSGG *with or without visual feedback*. t (35) = -0,626; p = 0,535.

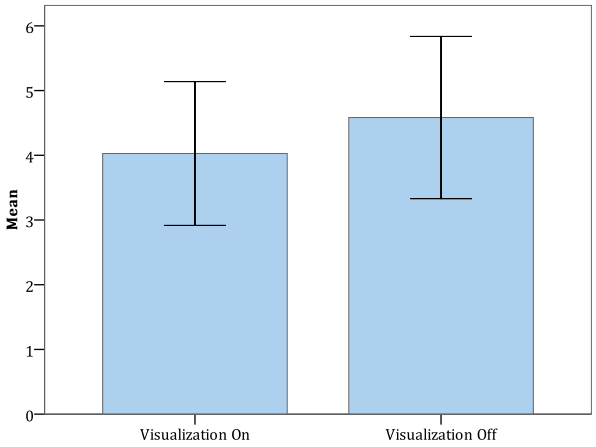


Figure 111: Mean of selection errors for gestures with and without visual feedback. Error bars represent standard error of mean

TARGET ERROR WITH OR WITHOUT VISUALIZATION

This analysis was based on the number of targets which were not selected.

The data were explored and some outliers were removed (± 3 standard deviation) based on linear regression. Even though Levene's test for homogeneity of variance was insignificant p = 0,063, a normal distribution could not be achieved so non-parametric statistics were used.

A Wilcoxon signed ranks test was used to compare the means. Each of the mean results was based on n=35 observations (Figure 112). The grand mean for number of *target errors* for both types of *visualizations* was 0,08. The mean for SSGG *with visualization* was 0,63 and for SSGG *without visualization* it was 0,97.

There was no significant difference between the number of *target errors* for SSGG *with and without visualization* $Z = -0{,}542$; $p = 0{,}588$.

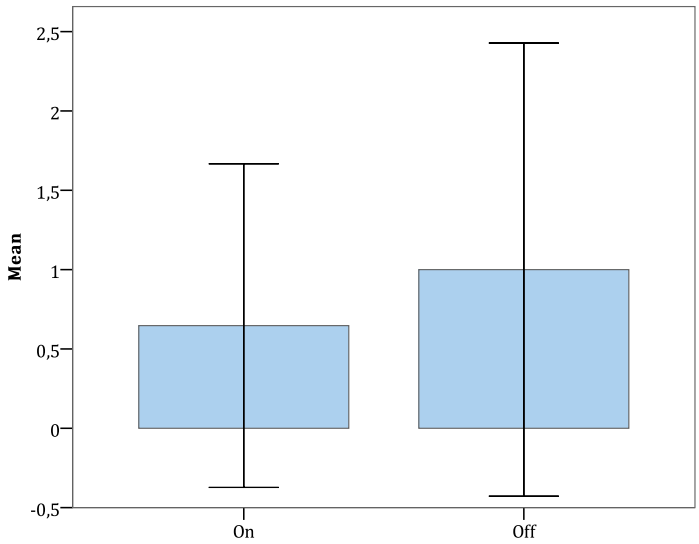


Figure 112: Mean of target errors for gestures with and without visual feedback. Error bars represent standard error of mean

8.2.3 DWELL SELECTION

TASK COMPLETION TIMES FOR DWELL SELECTION

There was no point in comparing *selection completion times* for the *dwell conditions* because the fixation durations were fixed. It was however of interest to see what affect fixation duration would have on the overall *task completion times*.

The data were explored. Mauchly's test of sphericity was significant and the data were skewed, so the assumptions which need to be fulfilled in order to conduct parametric statistics were not met.

Friedman's non-parametric ANOVA was used to compare the means and a post hoc analysis was used to determine were the differences were (Figure 113). Each mean was based on n=360 observations. The grand mean for *task completion time* for all *dwell conditions* was 1750ms.

The individual means were as follows (Table 10):

25ms	50ms	100ms	200ms	400ms
1550,39ms	1547,26ms	1538,45ms	1584,12ms	2532,30ms

Table 10: Mean task completion times for dwell selection conditions.

There was an overall significant effect between the *task completion times* $X^2_r = 24{,}141$ (4, n=360), $p < 0.01$. The 400ms dwell condition had significantly longer *task completion times* compared to all other dwell conditions, there was no significant difference between any of the other conditions.

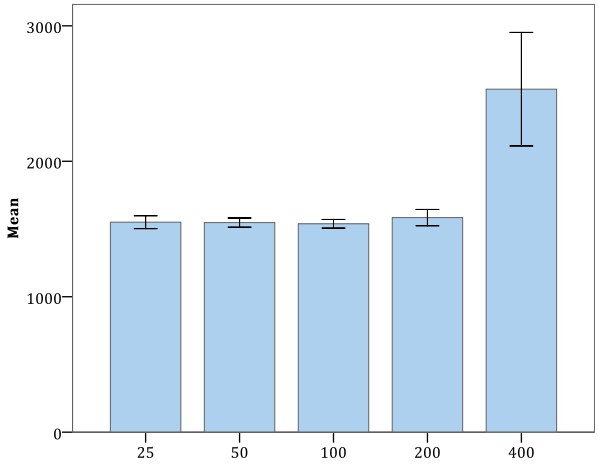


Figure 113: Mean of task completion time for all dwell conditions. Error bars represent standard error of mean

SELECTION ERRORS FOR DWELL SELECTION

Selection errors were the number of times a dwell-click was completed which did not correspond to the target being presented.

The data were explored. Sphericity could be assumed as Mauchly's test was significant p = 0,038. However, a normal distribution could not be achieved so non parametric methods were used.

Friedman's non parametric ANOVA was used to find an overall significant difference between the mean *selection error* values of the various dwell conditions (Figure 114). The mean were based on n=15 observations. The grand mean was 6,01.

The individual means for each *selection errors* for the dwell conditions were as follows (Table 11):

25ms	50ms	100ms	200ms	400ms
16,18	11,94	8,53	4,00	5,88

Table 11: Mean number of selection errors for the dwell selection conditions.

There was no overall significant effect between the number of *selection errors* for all of the *dwell conditions* $X^2_r = 6{,}349$ (4, n=18), $p < 0{,}175$.

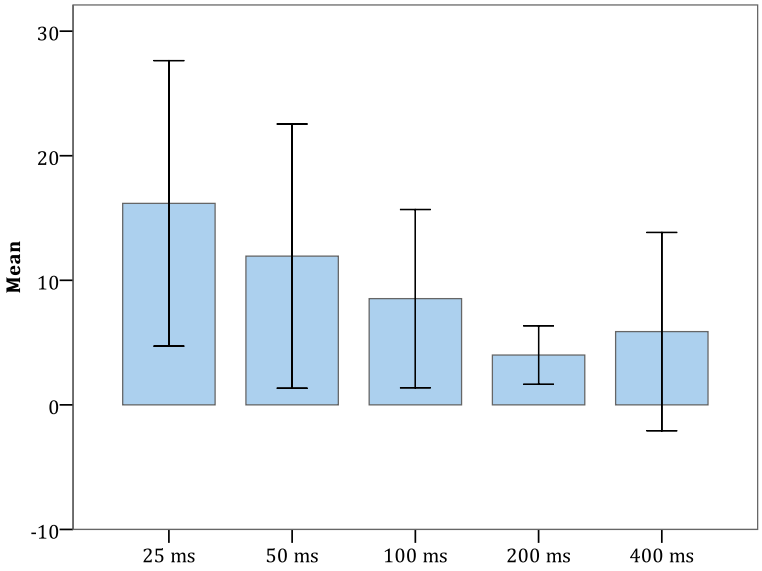


Figure 114: Mean number of selection errors time for dwell conditions. Error bars represent standard error of mean

TARGET ERRORS FOR DWELL SELECTION

This was an analysis of the number of targets which were allowed to cross the target area without being selected.

The data were explored and outliers removed (± 3 standard deviation) based on linear regression. This was done because a qualitative analysis of the data revealed that one subject fell greatly outside the norm. Sphericity and normal distribution could not be assumed so non parametric statistical methods were used.

Friedman's non-paramentric ANOVA was used to determine an overall significant difference (Figure 115). The mean values were based on n = 16 values. The grand mean for *target errors* was 0,18. The individual means are as follows (Table 12);

25ms	50ms	100ms	200ms	400ms
0,0	0,0	0,0	0,0	0,94

Table 12: Mean number of target errors for the dwell selection conditions.

There was an overall significant effect between the number of *target errors* for all of the dwell conditions $X^2_r = 12{,}000$ (4, n=16), $p < 0{,}01$. There were a significantly higher number of *target errors* in the 400ms *dwell condition*. There were no target errors in any of the other *dwell conditions*.

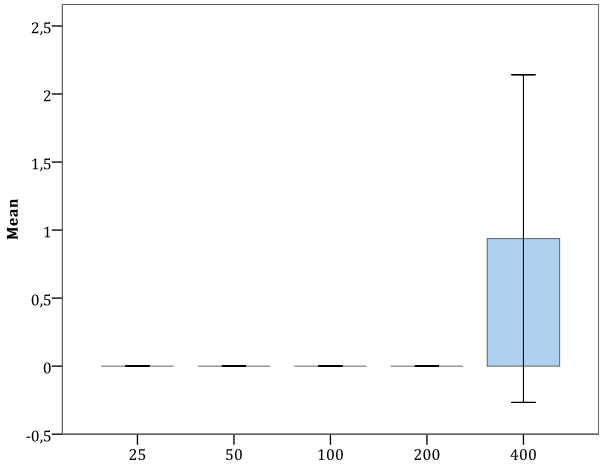


Figure 115: Mean number of target errors time for dwell conditions. Error bars represent standard error of mean

8.2.4 COMBINED MEASURE

OVERALL COMPARISON OF TASK TIMES

Seeing as the task was the same for all conditions, comparing *task completion times* between all conditions was of interest.

The data were explored. Sphericity could not be assumed so non parametric statistics were used. Friedman's non parametric ANOVA was used to compare the means, where *selection method* was the independent variable with nine levels (*short* SSGG *with visualization, short* SSGG *without visualization*, *long* SSGG *with visualization*, *long* SSGG *without visualization, 25ms dwell, 50ms dwell, 100ms dwell, 200ms dwell, 400ms dwell*). *Task completion time* was the dependent variable measured in ms (Figure 116). Post hoc analysis was used to determine where potential differences lay.

The means are based on n=360 observations per result. The grand mean for task completion time was 1838ms. The means for the individual conditions were as follows (Table 13):

Short ***With Vis***	**Short** ***Without Vis***	**Long** ***With Vis***	**Long** ***Without Vis***	**25ms**	**50ms**	**100ms**	**200ms**	**400ms**
1739ms	1754ms	2141ms	2161ms	1550ms	1547ms	1538ms	1584ms	2532ms

Table 13: Mean number of task completion times for all conditions.

There was an overall significant effect between the *task completion times* for all the conditions X^2_r= 43,568 (8, n=360), p < 0,01. 400ms dwell was significantly slower than all conditions, except both the *long* SSGG *condition*. The *long* SSGG conditions had significantly longer task times than all conditions except the 400ms *dwell* and the *short* SSGG *without visual feedback*. None of the other conditions were significantly different to one another.

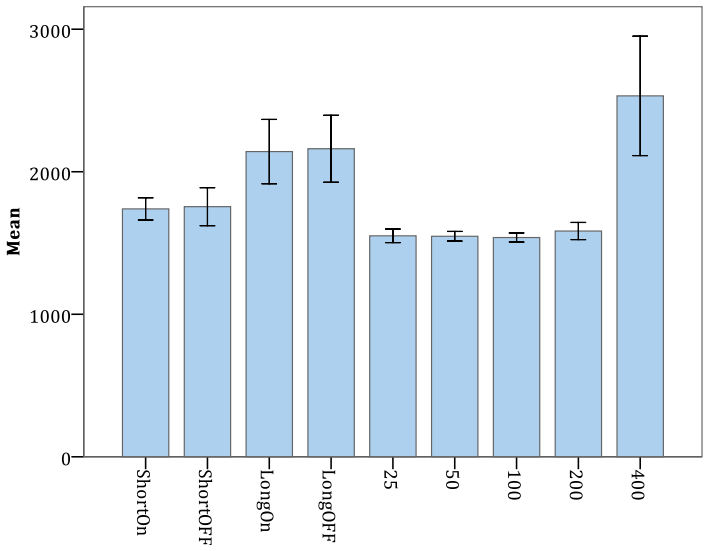


Figure 116: Mean task completion times for all conditions. Error bars represent standard error of mean

COMPARISON OF COMBINED MEASURES FOR ALL CONDITIONS

As with the *dwell, long* and *short* SSGG experiment it was necessary to use the combined measure from earlier in order to better compare the different methods of selections.

The *augmented selection completion time,* which was introduced earlier (Chapter 6, p. 125), was also applied here. The two main parameters of this equation are average *selection completion time* and average *selection error*.

The average *selection completion times* for the different conditions were (Table 14):

Short ***With Vis***	**Short** ***Without Vis***	**Long** ***With Vis***	**Long** ***Without Vis***	**25ms**	**50ms**	**100ms**	**200ms**	**400ms**
154ms	83ms	225ms	147ms	25ms	50ms	100ms	200ms	400ms

Table 14: Mean selection completion times for all conditions.

Friedman's non-parametric ANOVA was used to compare *selection errors* for all of the conditions. The average *selection errors* for each of the nine conditions were (Table 15):

Short ***With Vis***	**Short** ***Without Vis***	**Long** ***With Vis***	**Long** ***Without Vis***	**25ms**	**50ms**	**100ms**	**200ms**	**400ms**
3,82	4,06	4,18	5,12	16,18	11,94	8,53	4,00	5,88

Table 15: Mean number of selection errors for all conditions.

There was no overall significant effect between the number of *selection error* for any of the conditions X^2_r= 11,712 (8, n=17), p = 0,165.

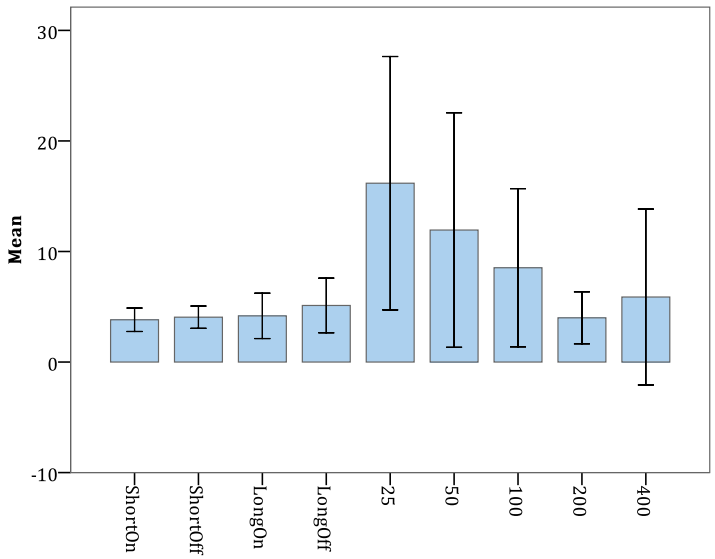


Figure 117: Mean selection error for all conditions. Error bars represent standard error of mean

After recalculating into the *augmented completion time* the data were explored. Levene's test of homogeneity of variance was significant and the data were skewed so parametric statistics could not be used.

Friedman's non-parametric ANOVA was used to compare the means (Figure 118). Each mean result was based on n=360 observations. The grand mean for *augmented selection completion time* was 842.43ms

The individual *augmented selection completion times* were (Table 16):

Short ***With Vis***	**Short** ***Without Vis***	**Long** ***With Vis***	**Long** ***Without Vis***	**25ms**	**50ms**	**100ms**	**200ms**	**400ms**
588,6ms	336,7ms	943,7ms	752,8ms	404,0ms	597,0ms	858,7ms	800,0ms	2352,0ms

Table 16: Mean number of selection errors for all conditions.

There was an overall significant effect between the *augmented selection completion times* for all the conditions $X\ {}^{2}_{r}= 1769{,}164$ (8, n=360), $p < 0, 01$.

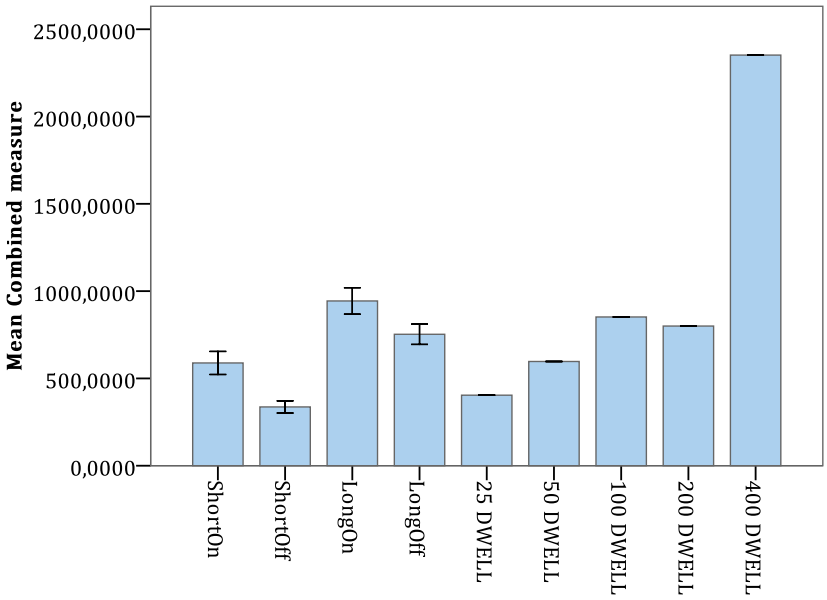


Figure 118: Mean augmented selection completion times for all conditions. Error bars represent standard error of mean

The fastest augmented selection completion time was for *short* SSGG *NO-VIS*, the only condition this was not significantly faster than was the *25ms dwell*. These two conditions were significantly faster than all other conditions.

The second fastest pair of conditions were *short* SSGG *VIS* and 50ms *dwell*. These were faster than all other conditions (except *short* SSGG *no vis* and 25ms dwell). There was no significant difference between long SSGG *vis*, *long* SSGG *no-vis*, *100ms dwell* and *200ms dwell*. 400ms dwell had a significantly higher *augmented selection completion time* than all other conditions.

The α-level calculation (Chapter 6, page 126) is an indication of how many errors are made per selection; thereby indicating whether or not a selection strategy is usable for practical purposes. For the nine conditions the α-levels were (Table 17), so all of the selection methods were usable:

Short ***With Vis***	**Short** ***Without Vis***	**Long** ***With Vis***	**Long** ***Without Vis***	**25ms**	**50ms**	**100ms**	**200ms**	**400ms**
0,191	0,203	0,209	0,256	0,809	0,597	0,426	0,2	0,294

Table 17: α-levels for all nine conditions.

8.3 DISCUSSION

In the following the results just presented will be discussed in regard to their initial null-hypotheses.

8.3.1 FACTOR 1, LONG AND SHORT SSGG:

NULL 1: THERE WAS NO DIFFERENCE BETWEEN THE *SELECTION COMPLETION TIME* FOR *LONG* AND *SHORT* SSGG.

This hypothesis was rejected as there was a significant difference between *selection completion times* for *long* and *short* SSGG. The selection completion time for *Short* SSGG was significantly faster than for *long* SSGG.

This result substantiated the observations which were made in the original *dwell*, *long* and *short* SSGG experiment. And it therefore also substantiates that potentially SSGG of different length can be applied to different tasks.

Further research should be done into what the ideal lengths of SSGG would be if the goal is both to avoid *accidental gesture completion* and make a distinction between *long* and *short* SSGG in the same interface.

Overall this observation holds interesting promise not only for SSGG, but also for future implementations of both *complex* and *continuous gaze gestures*. This point will be revisited in the final discussion chapter of this thesis.

NULL2: THERE WAS NO SIGNIFICANT DIFFERENCE BETWEEN *TASK COMPLETION TIMES* OF *LONG* AND *SHORT* SSGG.

This null hypothesis was rejected as there was a significant difference between the task completion times for *long* and *short* SSGG. *Short* SSGG had significantly faster *selection completion times*.

Even though there was not a significant difference between *task completion times* in the *dwell, long* and *short* SSGG experiment. There was still a substantial difference in favour of *short* SSGG, so some difference was expected; however, not necessarily a significant difference.

This might not seem surprising seeing as there was already a significant difference in *selection completion times* in the two conditions. However, the difference between the two mean *selection completion times* was 67ms and the difference between the two *task completion times* was 404. This means that there are 336,34ms of mean extra task time which cannot be accounted for by the faster *selection completion time*.

This could be a consequence of the navigation process being less time consuming for more centralized gestures. In other words the *short* SSGG activation fields were closer to the area in which the targets were being presented at the centre of the screens.

These findings regarding *long* and *short gaze gestures* on the screen require further investigation, because seemingly small changes in distances on the screen have a relatively large affect on the registration time of these eye movements.

NULL 3: THERE WAS NO DIFFERENCE IN THE NUMBER OF *SELECTION ERRORS* BETWEEN *LONG* AND *SHORT* SSGG.

This null hypothesis was retained as there was no significant difference in the number of *selection errors* between *long* and *short* SSGG.

This result again substantiated the observations found in the *dwell, long* and *short* SSGG experiment, where there also was no significant difference and the result was still surprising. It was expected that there might be a higher degree of accidental gesture completion in the *short* SSGG condition, especially in this experiment where the distance between the selection completion fields was even shorter (only covering 30 % of the screen).

However, the target area was also more centralized which meant that there was little reason to make navigational search patterns which would cause faulty selections. There seems to be an intricate relationship between *gaze gesture completion*, *accidental gesture completion* and *information visualization*.

NULL 4: THERE WAS NO DIFFERENCE IN THE NUMBER OF *TARGET ERRORS* BETWEEN *LONG* AND *SHORT* SSGG.

This null hypothesis was also retained as there was no significant difference in the number of *target errors* between *long* and *short* SSGG. This was the opposite of what was found in the initial *dwell*, *long* and *short* SSGG experiment. Where there were a significantly higher number

of *target errors* in the *long* SSGG condition. In the discussion regarding those results it was speculated that the main reason for the high number of *target errors* had to do with the delay in system response time on the QuickGlance, which was experienced by participants. In other words the system sometimes required users to repeat a selection process several times before a successful selection was made. This would cause a higher number of *selection errors* and create a backlog of targets which had already begun descending the screen. In turn this would cause a number of *target errors* to occur before the participant could 'catch up'.

The current experiment was completed on a system with very fast response time. Consequently it was shown that there was no significant difference in the number of *target errors*. This supports the earlier stated conclusion that the previous high number of *target errors* was a consequence mainly of system set-up.

8.3.2 FACTOR 2, VISUALIZATION:

NULL 5: THERE WAS NO OVERALL SIGNIFICANT DIFFERENCE BETWEEN *SELECTION COMPLETION TIMES* OF SSGG *WITH* OR *WITH VISUAL FEEDBACK*.

This null hypothesis was rejected as there was a significant difference in *selection completion time* between SSGG *with* and *without visualization*. SSGG *without visualization* was significantly faster than *with visualization*.

This result was not expected. It was expected that the lack of visual feedback to the user would cause selections to be made from one edge of the screen to the other, in other words using the edges of the screen as anchor points; thereby making longer and subsequently more time consuming gestures. This was not the case. Surprisingly, the *gestures without visual feedback* were faster.

The fact that SSGG *without visual feedback* can be completed and that *selection completion times* were even more efficient than with visualization indicates that it could be possible to implement SSGG as a transparent overlay in certain contexts. This will be explored further in the overall discussion.

NULL 6: THERE WAS NO OVERALL SIGNIFICANT DIFFERENCE BETWEEN *TASK COMPLETION TIMES* OF SSGG *WITH* OR *WITHOUT VISUAL FEEDBACK*.

This null hypothesis was retained as there was no difference in *task completion times* for SSGG with or without visualization.

Seeing as the *selection completion times* for SSGG *without visualizations* were faster, it might have been expected that this would have affected *task completion times.* The fact that they did not seem to affect each other can have a cause and effect as follows:

The cause could be that more navigation was required by the participants in the *without visual feedback condition*, because 'imaginary' fix-points had to be set.

The consequence of this and therefore the effect of this were longer *task completion times*. However, not longer than they can be comparable with the *task completion times* required for SSGG *with visualization*. In other words this observation was not enough to dismiss the use of SSGG *without visual feedback*.

NULL 7: THERE WAS NO OVERALL SIGNIFICANT DIFFERENCE BETWEEN NUMBER OF *SELECTION ERRORS* FOR SSGG *WITH* OR *WITHOUT VISUAL FEEDBACK*.

This null hypothesis was retained as there was no significant difference between the number of *selection errors with* or *without visualization*.

Again this went against the expectation that there would have been a higher number of *selection errors* in the condition which did not have visualization. The fact that this was not the case supports the potential use of SSGG *without visual feedback*.

NULL 8: THERE WAS NO OVERALL SIGNIFICANT DIFFERENCE IN THE NUMBER OF *TARGET ERRORS* FOR SSGG *WITH* OR *WITHOUT VISUAL FEEDBACK*.

This null hypothesis was also retained as there was no significant difference between the numbers of *target errors* in either condition. This metric is mainly relevant in regard to this specific task and therefore does not count that much either way in regard to how and when single SSGGs *with* and *without visualizations* should be implemented,

8.3.3 FACTOR 3, DWELL SELECTION:

NULL 9: THERE WAS NO OVERALL SIGNIFICANT DIFFERENCE BETWEEN THE *TASK COMPLETION TIMES* BETWEEN THE DIFFERENT *DWELL CONDITIONS*.

This null hypothesis was rejected as there was an overall difference in *task completion times* for the different *dwell conditions*. The *task completion time* for the *400ms dwell* condition was significantly longer.

What was interesting about this result was not the fact that the 400ms dwell condition produced longer *task completion times,* as this was expected. What was really interesting was the fact that there was no significant difference between the task completion times of any of the *dwell conditions* from 200ms and downwards. Not only that, there was no more than 50ms between the highest and lowest mean task completion time in that group, which was very little.

The main reason that this was interesting was that it indicates that there would potentially be no real benefit in implementing a *dwell duration* lower than 200ms. This number could even be higher, such as 250 or 275ms. However, these *fixation duration* times were not tested, which could therefore be the subject of a future study. The reason this was particularly interesting should be seen in correlation to the findings of Majaranta (2009) where trained users would adjust dwell fixation durations down towards an average of 282ms. In other words there might be a innate lower boundary for dwell fixation completion time. And users might implement this boundary if they were always able to adjust their activation speed as they gain experience.

NULL 10: THERE WAS NO OVERALL DIFFERENCE IN NUMBER OF *SELECTION ERRORS* FOR THE DIFFERENT *DWELL CONDITIONS*.

This null hypothesis was retained as there was no significant difference in the number of *selection errors* depending on dwell conditions.

This result was very surprising. In the initial *dwell, long* and *short* experiment the 100ms dwell condition had a much higher *selection error rate* than the other *dwell conditions*. From a practical perspective it was known that no-one generally uses a *dwell completion time* lower than 250ms because it causes too many *selection errors*. So why is there so little difference in

the number of *selection errors* for this experiment? – There are two reasons, which could be quite interesting for future implementations of dwell selection.

The first has to do with placement of the dwell-buttons in regard to the task-area on the screen. The second has to do with the task-targets and their visualization in regard to the dwell-buttons.

The fact that the dwell-buttons were placed around the field in which targets were presented meant that there was no overlap between *task navigation* (where the task was being perceived) and *selection navigation* (where the correct button was being selected).

The second and most important fact was that the targets themselves pointed to the dwell-button which was required in order to select them. This further emphasized the difference between navigation and selection; as the participants actually navigated by looking at the targets themselves and then subsequently completing a selection. In other words, the participants did not need to look at the dwell-buttons in order to know which selection to complete; they already knew before a selection was initiated what the result of that particular dwell action would be.

This could be implemented as a design feature that navigation and perception occur in separate visual spaces compared to selection. This will be explored further in the final discussion chapter.

NULL 11: THERE WAS NO OVERALL DIFFERENCE IN THE NUMBER OF *TARGET ERRORS* FOR THE DIFFERENT *DWELL CONDITIONS*.

This null hypothesis was rejected as there were a significantly different number of *target errors* depending on *dwell condition*. The only condition which experienced *target errors* was the 400ms *dwell condition*. The main reason for this was to be found in the balance of the experiment. Some participants were started with 25ms as their first selection condition and some started with 400ms dwell. The three people who ended up having *target errors* all started with the 25ms conditions. Indicating that in the context of this experiment it was more difficult when selection speed was increased compared to when it was decreased. It also indicates that there might be a tipping point for dwell activation time around the 200ms (or a bit higher) mark.

8.3.4 FACTOR 4, OVERALL AND COMBINED MEASURE:

NULL 12: THERE WAS NO OVERALL DIFFERENCE IN *TASK COMPLETION TIME* BETWEEN ANY OF THE *DWELL CONDITIONS*.

This null hypothesis was rejected as there was a significant overall difference in the *task completion times* between the *dwell conditions*. 400ms *dwell* and both the *long* SSGG conditions were in one group that had significantly slower *task completion times*. *Short* SSGG and the 25-200ms *dwell conditions* were in another group which had significantly faster *task completion times.*

The issues regarding *task completion times* have already been discussed for the SSGG and *dwell conditions* separately. This overall comparison was simply done in order to examine whether one of the overall *selection strategies*, *dwell* or SSGG, stood out, which they did not. This was a further indication that SSGG and *dwell* are not that far apart in terms of being effective strategies for solving similar tasks.

NULL 13: WHEN USING THE COMBINED MEASURE ON ALL OF THE NINE *SELECTION METHOD* CONDITIONS THERE WILL BE NO SIGNIFICANT DIFFERENCE BETWEEN CALCULATIONS.

This null hypothesis was rejected as there was significant difference between the *augmented selection completion times*.

Surprisingly the α-levels showed that all of the selection strategies presented in this context could potentially be usable.

The two fastest conditions for the augmented selection completion times was the *short* SSGG *without visualization* and the 25ms *dwell*. The main reason for the *short* SSGG *without visualization* being so effective in this calculation was the combination of having a low *selection completion time* and having a relatively low *selection error rate* as well. The 25ms *dwell condition* was, however, mainly benefitting from the low *selection completion time* as this meant that the 'penalty' for erroneous selection was also very low. So even though this condition had the highest number of *selection errors* it still fared relatively well. This was probably also mainly due to the previously mentioned visual layout, which was very conducive with short dwell time activations.

The second fastest pair of conditions was *short* SSGG *with visualization* and 50ms *dwell*. The same reasons as presented above can explain their relative efficiency within this augmented measure.

There was no significant difference between long SSGG *with visualization*, *long* SSGG *without visualization*, *100ms dwell* and *200ms dwell*. The main conclusion which can be drawn from this observation was that the overall SSGG and relatively low *selection completion times* are quite comparable; even in an interface layout and task which attempts to use the strengths of either selection method. These strengths being easy navigation in both designs, as the task targets provided the navigational information required to efficiently complete selections.

400ms *dwell* had a significantly higher *augmented selection completion time* than all other conditions. This indicates that the cost of a higher dwell completion time might not be outweighed by the presumed lower number of *selection errors* which should occur.

The goal of this research was not to compare *dwell selection* and SSGG to a point where one was better than the other, which also was not supported by the empirical evidence. The point was to justify whether or not SSGG *with* or *with-out visualization* could be an equally valid gaze selection strategy, which could supplement existing input. This *was* supported by the empirical evidence. The goal now becomes to determine when to use what selection strategy in order to create flexible gaze contingent applications.

To further explore the potential future toolbox of gaze selection strategies a final experiment was developed which was quite different to the previously presented experiments in both its intent and design.

9 COMPLEXITY THRESHHOLD OF GAZE GESTURES

The research presented in the previous chapters has inspired further investigation into the overall area of gaze gestures. This chapter describes a pilot study which is intended to inspire future research.

9.1 COMPLEXITY THRESHOLD

The research previously presented in this thesis resulted in the interest of pursuing certain aspects of gaze gesture interaction, specifically the affordances and constraints regarding single stroke gaze gestures (SSGG). The concept of complex finite gestures was defined in chapter 3, as finite shape based gestures.

An experiment regarding these types of gaze gestures was designed, implemented and a small study was conducted. The data collected presented challenges, which have helped establish some of the issues regarding further exploration of complexity in gaze gestures. The idea was that if SSGG could be completed *without* visual fix-points, perhaps more complex gaze gestures could also be completed *without* fix-points. Not only that, but complex gaze gestures could potentially be recognized purely based on shape completion, rather than relying on initiation and completion fields. This concept is called *free gaze gestures* and could, after further exploration, be added to the gaze gesture taxonomy presented in chapter 3. To explore the upper and lower boundaries of free gaze gestures a new experiment was designed. Initial observations are presented, along with considerations in regard to design as a full analysis of this research lies beyond the scope of this thesis.

There were three main areas which this experiment was designed to explore:

(1) The first was discovering whether or not there was a threshold for the potential complexity level of finite gestures while still retaining the ability of being sustainably repeated. There are two reasons for wanting to be able to implement complex finite gestures. The first is the increase in the *vocabulary* of gaze selection strategies. The rationale was that the more selection methods are available the better applications could potentially become, because diversity of selections strategies means that the *mono-modal* nature of gaze as sole input can still achieve *multi-modal* interaction capabilities. Another issue related to exploring

complex finite gestures, was accidental gesture completion. In certain situations for certain users SSGG will potentially be subject to high levels of accidental gesture completion, due to nystagmus, making complex finite gestures a potential alternative.

(2) The second focus which was explored in this experiment was inspired by the SSGG experiment with and without visualization presented in chapter 8. One thing was to be able to complete SSGG without on-screen fix-points; another would be to be able to complete complex finite gestures without feedback. This has not been explored in the field of gaze interaction. The idea is that a *free gaze gesture* could be completed anywhere on the screen as long as the shape being completed would be systematically recognizable. Figure 119 shows the difference between completing complex finite gestures based on fix-points and completing them freely.

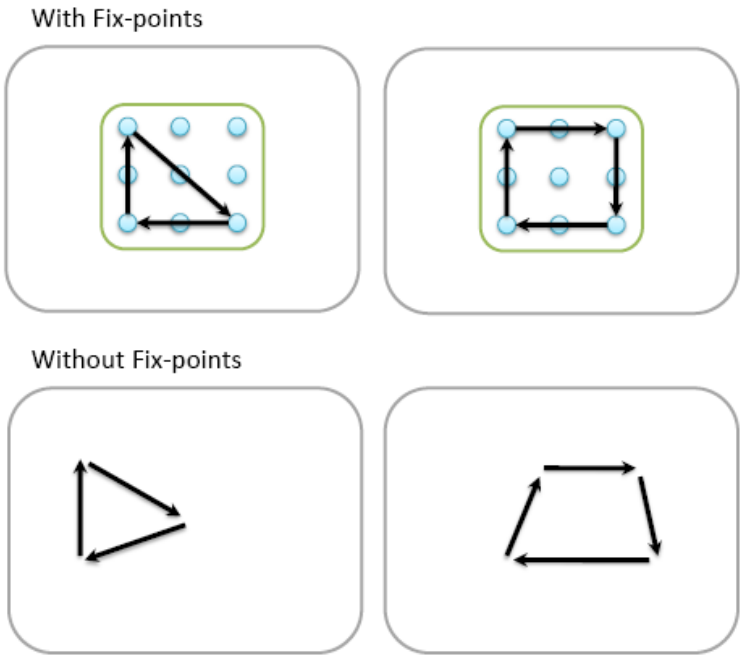


Figure 119: Complex finite gaze gestures with and without fix-points

All of the experiments on complex finite gestures (Wobbrock, 2003, Porta & Turina, 2008a, Drewes & Schmidt, 2007, Istance et al., 2009; Vickers et al., 2008; Istance et al., 2010; Istance et al., 2009) have been done using fix-points of some kind. Being able to implement complex finite gestures freely and without visualization relies on whether or not users are capable of completing recognizable patterns when they do not have fix-points on the screen. It was in that regard of interest to establish whether there would be an upper threshold for how

complex such gaze gestures could be, while still retaining the ability to be remembered and completed by the user. As well as a lower threshold for how simple such gaze gestures can be without causing too many accidental gesture completions.

(3) Finally, an experimental interest was exploring how on-screen visual interference would affect the completion of *free* single and complex finite gaze gestures; in order to develop guidelines for how these should be implemented appropriately in applications. One implementation of particular interest was to further develop gaze controlled driving, for instance to enable a disabled user control their wheelchair, as well as looking at direct object interaction as the foundation of environmental control. Mock ups of design implementations of these functions will be presented in the next chapter. However, this was the reason why it was deemed necessary to explore how *visual noise* (i.e., objects and people entering and exiting the screen beyond the user's control) affects the completion of *free* single and complex finite gaze gestures.

9.1.1 EXPERIMENTAL DESIGN

An alphabet of gaze gestures was created. The main consideration was whether or not the complex finite gaze gestures should have semantic meaning or not. In some of the experiments where these types of gaze gestures have been used in typing interfaces, the shapes which need to be completed were designed to resemble characters in the alphabet (Wobbrock et al., 2008). The gaze gestures in the experiment presented here were based on three basic geometric shapes: lines, triangles and squares; without any semantic meaning.

Another consideration was how to increase complexity gradually, in order to discover tipping points for the upper and lower boundaries. The gaze gestures were therefore implemented with one, two, three, four and five strokes respectively. Finally, the shape, size, and whether or not the figure was skewed was considered; in order to uncover whether these factors affected the completion process. The final gesture alphabet used can be seen in figure 120.

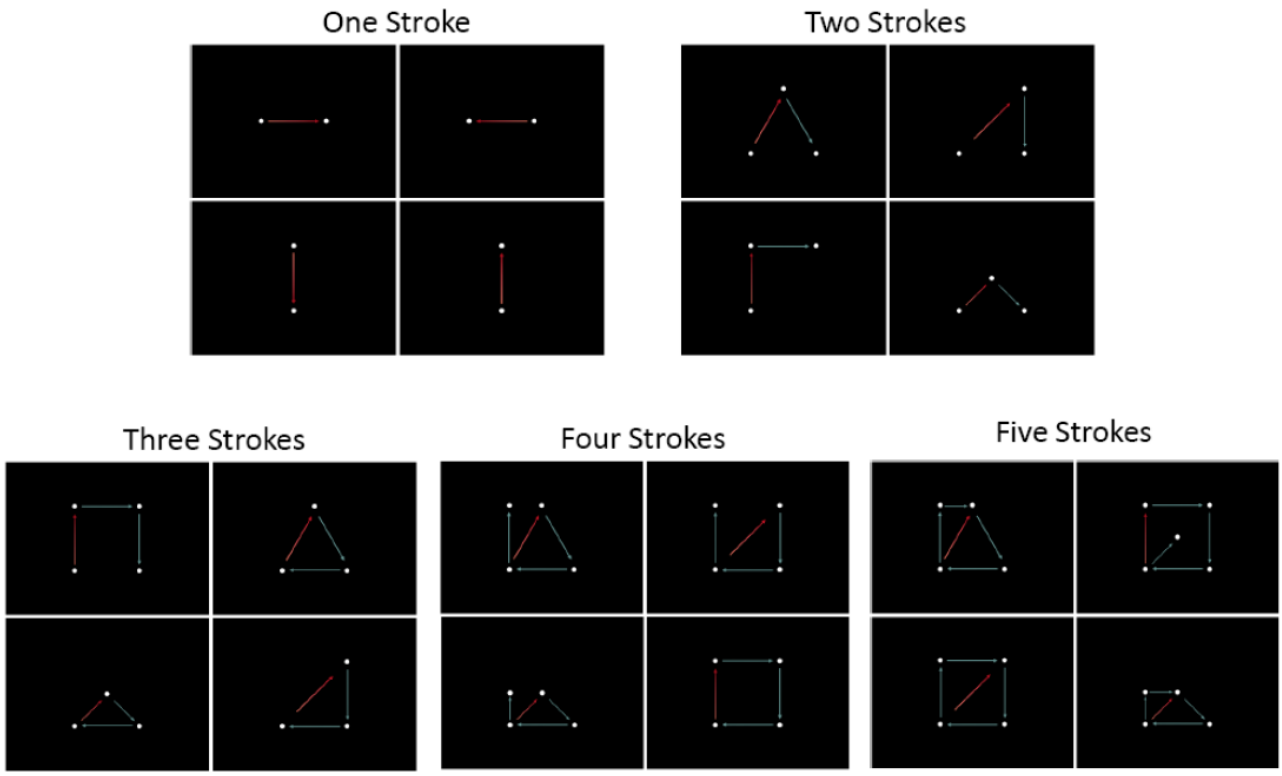


Figure 120: The gaze gesture alphabet as designed for this experiment. The red arrow symbolizes the first stroke of the gesture.

In order to determine how visual noise affected the completion of different types of gaze gestures; four different types of visual information were implemented. Gaze gestures were to be completed on (1) a black background as seen above (Figure 120). The other types of visual information were (2) a still image background, (3) a moving background without people and (4) a moving background with people. The intention with the moving backgrounds was to simulate a wheelchair control situation where an individual was moving down a hall. The reason for having images with and without people in them was to explore whether this type of interference was more or less likely to affect the successful completion of gaze gestures.

To build on the previous experiment where gaze gestures were completed *with* and *without visualizations*; a sequence of gaze gestures would be completed first with fix-points and then repeated without fix-points. This approach is also inspired by the research done in the context of other path based gestures that have attempted to define the boundaries of how this type of interaction could be implemented (Heikkilä & Räihä, 2009).

9.1.2 STUDY

4 participants (1 female) completed the study. All had previous experience with eye trackers.

Participants were asked to complete the 20 gaze gestures of varying complexity with and without fix-points and four different backgrounds: *Black background*; *still image background*; *moving image background without people* and *moving image background with people*.

Both gaze gesture type and background were presented randomly. However, the sequence of presentation was the same. First the gaze gesture was presented *with* fix-points and then *without*.

Analysis has been done on the gaze gestures with and without fix-points on black backgrounds, in order to develop a baseline method of analysis for this type of gaze experiment. Preliminary observations are reported which indicate that this is an interesting but complex area of research.

Figure 121, is an example of an overview of the preliminary recordings completed on a LC Technology Eye Follower system and plotted by the Nyan gaze diagnostics tool[13]. Nyan is a piece of software which allows experimental setups regarding analyzing eye movement patterns in various contexts such as scene viewing of dynamic content. It also provides direct access to the raw tracking data, as well as offering a full fixation and saccade analysis of the data based on the system's gaze analysis algorithm. This full analysis can then be plotted visually. What it does not provide is an ability to conduct a full analysis of eye movements in interactive applications. This is an area which needs further development.

The gaze gestures in this experiment were not interactive, but simply displayed on the screen. Participants were then asked to trace the gesture presented with fix-points and then repeat the pattern without fix-points. Figure 121 consists of 20 gesture sequences (the full vocabulary of this experiment). First the gesture is shown in each sequence; the red arrow indicating the first stroke. After this is the recording of the gaze gesture being completed with fix-points, and then the gaze gesture being completed without fix-points.

[13] http://www.interactive-minds.com/

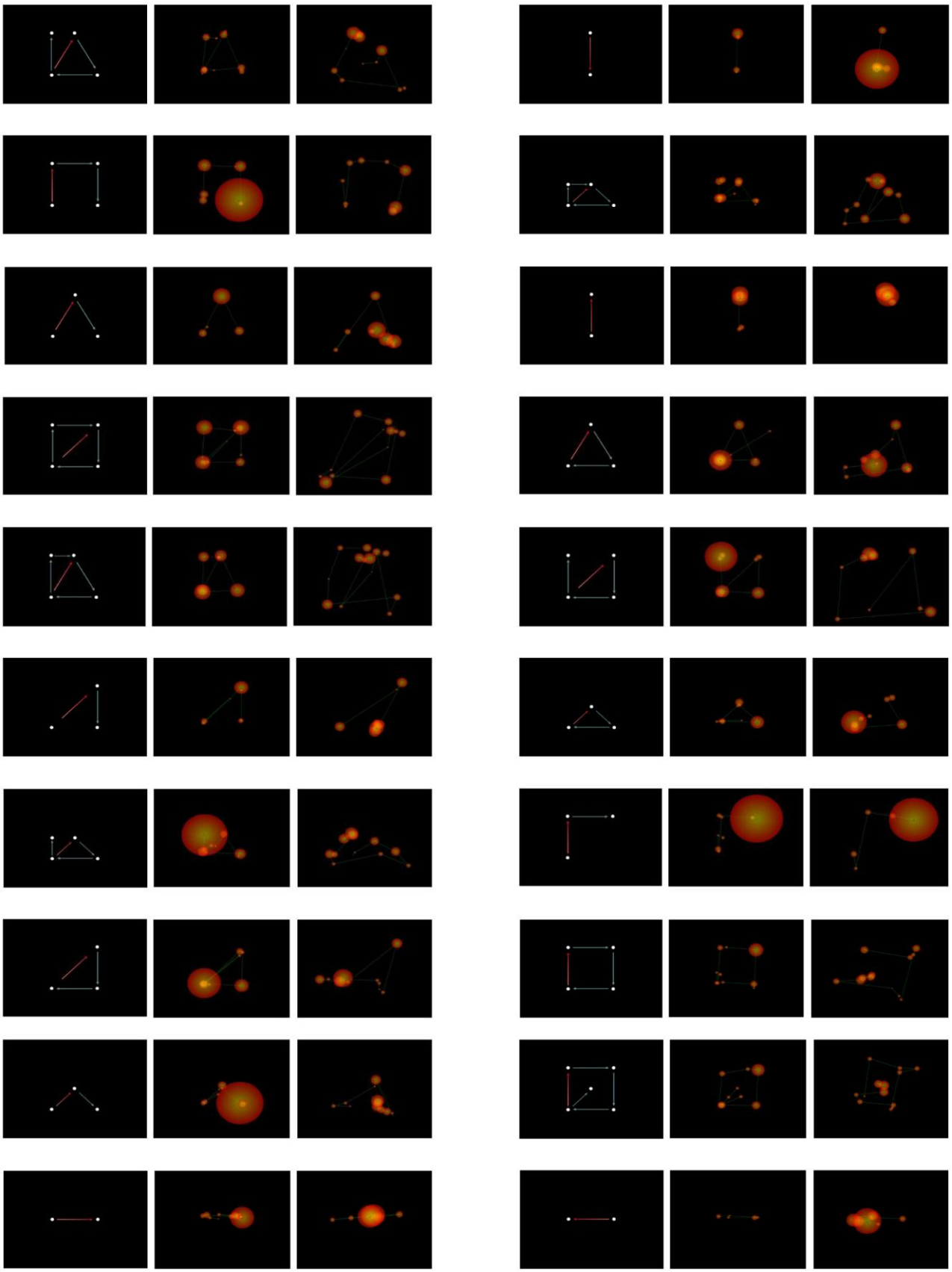


Figure 121: Preliminary recordings of eye movements while completing gestures with and without visualization.

There are four observations which can be derived from these preliminary recordings. First of all completing complex gaze gestures with fix-points create easily recognizable patterns (Figure 122).

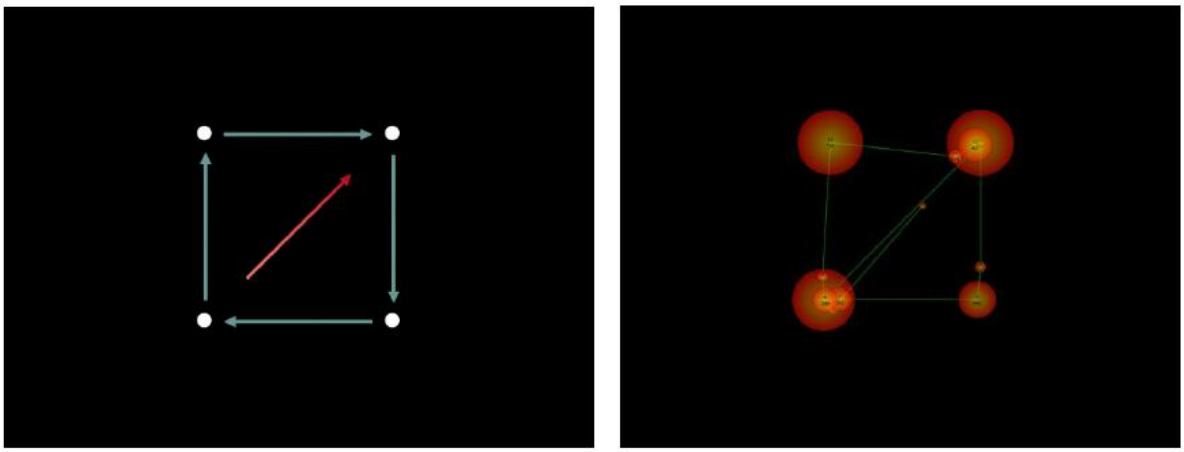


Figure 122: A five stroke gesture repeated with fixation points.

There does not seem to be an upper physiological threshold for how *complex* a gaze gesture can be. However the question still remains in regard to a potential cognitive threshold. Overall the findings indicate that complex finite gaze gestures are a valid solution if on-screen visualization is incorporated in the design. A design proposition, which incorporates the ability to complete complex gaze gestures, while removing the cognitive difficulties of remembering them and their consequences, has come from this and could be defined as gaze tracing. In this experimental design the gaze pattern was shown in its entirety, the subjects were therefore essentially *tracing* the pattern with the eyes. Tracing could become a selection strategy on its own; for instance, in situations where the user is to make a discrete choice where the outcome is of great consequence. In other words, for situations where the potential consequence of a selection error caused by a prolonged fixation (for a dwell- button) or a stray fixation (for a SSGG) – would be too great. An example of implementing *gaze tracing* as a selection strategy is shown in figure 123.

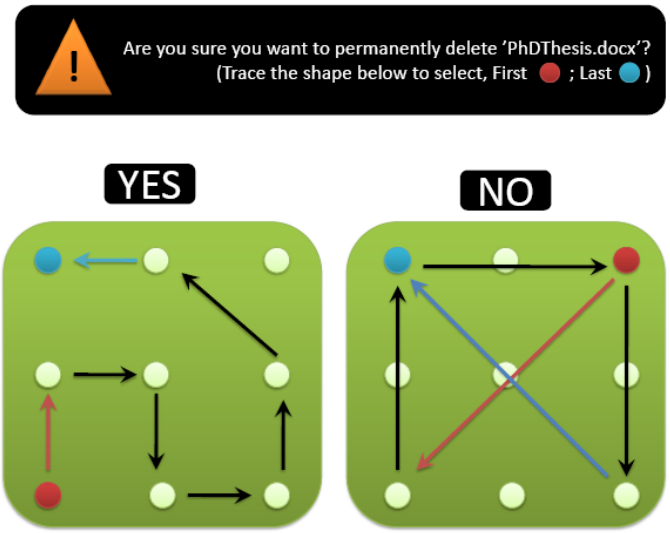


Figure 123: An example of implementing tracing as a selection strategy

The second observation which can be made was that there was a vast difference between completing free SSGG and free complex finite gaze gestures without fix-points; even though the gaze gesture shape had been shown right before the free complex finite gaze gesture was completed (Figure 124).

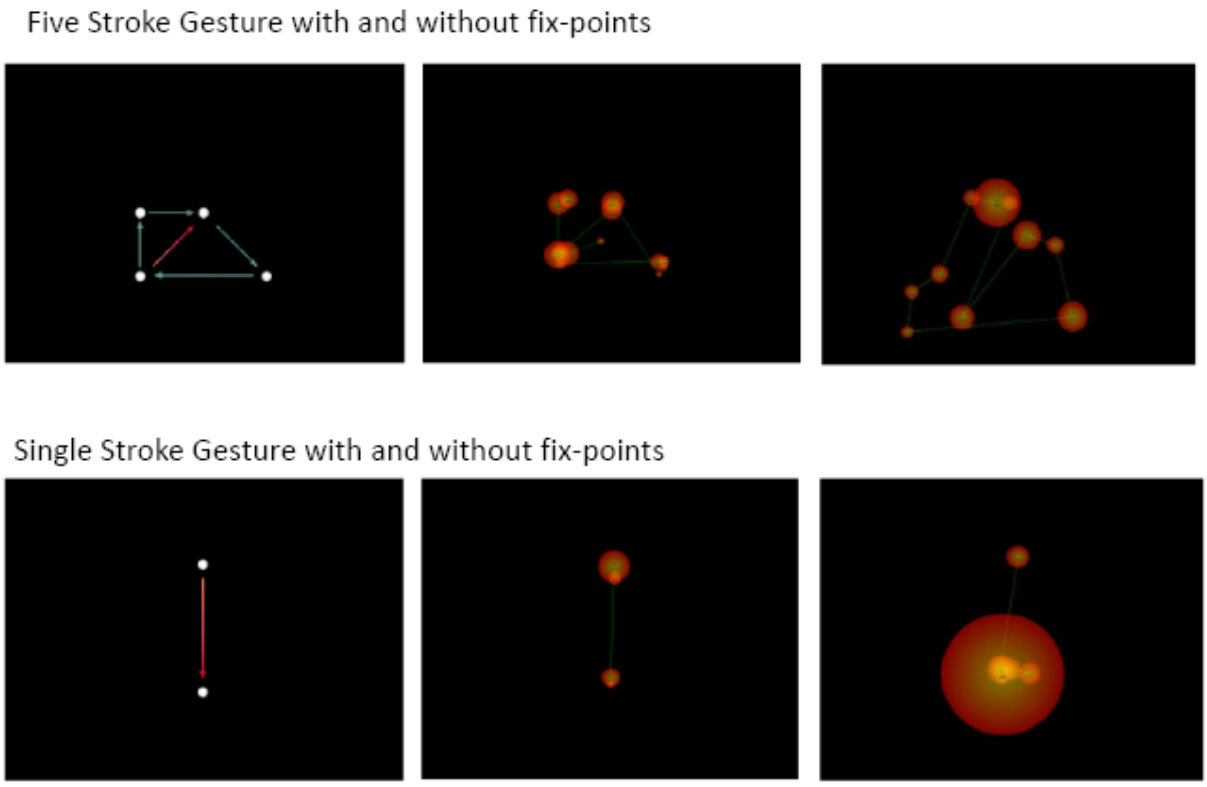


Figure 124: On the left the gesture image, the middle image is the gesture completed with fix-points and on the right the gestures completed without fix points.

This further substantiates that SSGG can indeed be completed successfully without visual fix-points. This infers that this type of gaze selection strategy would be useful in situations where it was of interest not to have gaze-related visual feedback presented on the screen.

The two final observations have to do with analyzing the results of this experiment quantitatively, but also have implications for future implementations of gaze gestures. If complex finite gaze gestures were to be implemented without visual feedback it would most likely require that there was an indication as to when a gaze gesture was started. In this design, it was difficult to distinguish which fixations were part of the gaze gesture and which were part of the visual inspection and this was on a black background with no noise. Potentially, start and stop fields could be implemented which informed the system when a gaze gesture was initialized and considered completed by the user (Figure 125). The system would then know what data to analyze for pattern recognition.

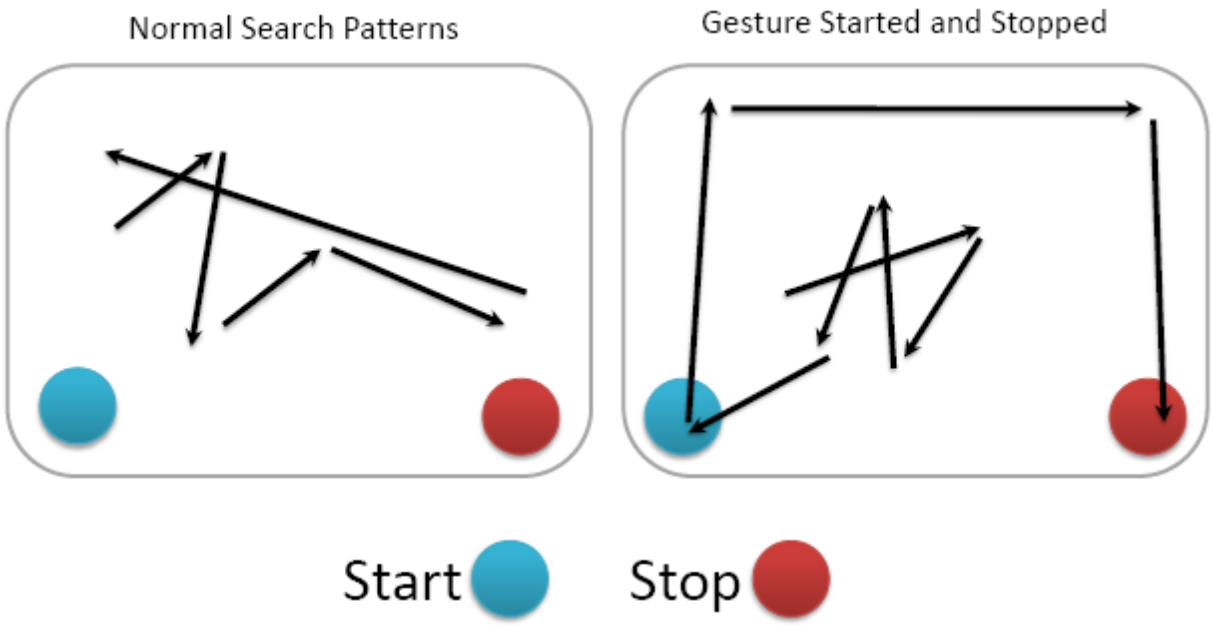


Figure 125: Start and stop fields regarding the initiation and completion of complex finite gaze gestures.

This leads to the final observation in regard to this experiment. In order to successfully analyze the data acquired in this experiment, an algorithm for *fix-point-free shape recognition* needed to be developed, which could recognize gaze gesture shapes even though they were scaled, skewed and angled slightly differently from the ideal gaze gesture which they were

intended to resemble. Only with such a system in place can free complex finite gaze gestures of varying degrees be tested and potentially implemented.

This research has uncovered that there are four sub-categories of complex finite gaze gestures (Figure 126).

(1)They can be based on the *gaze tracing* principle, where the shape which needs to be completed is presented on the screen. This method can cope with very complex gaze gestures as the cognitive load is very low, because the user does not have to remember the right order of the strokes or the potential outcome of a gaze gesture. However, this method is limited by the need for a relatively large visualization on screen. Therefore it would most likely be most appropriate in situations where discrete selections need to be made.

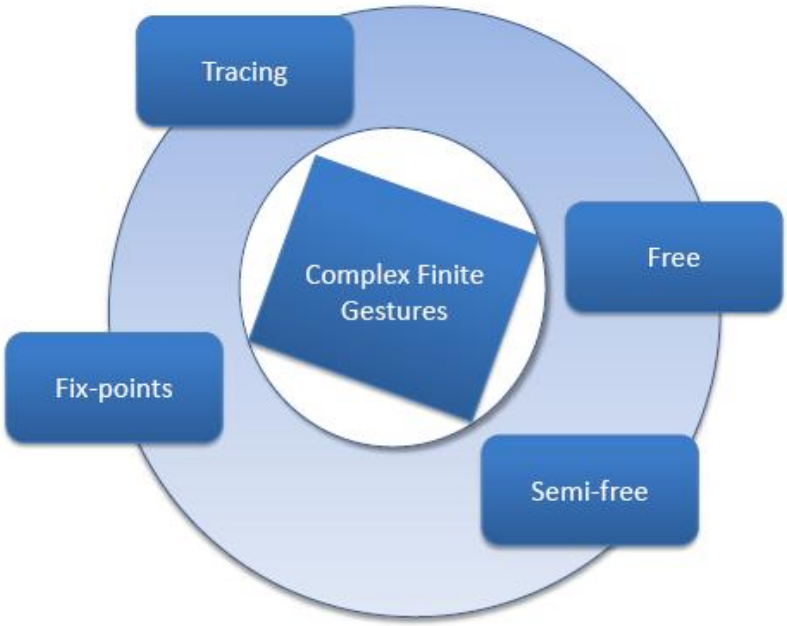


Figure 126: Sub-categorization of complex finite gaze gestures

(2)Complex finite gestures can also be implemented as gaze gestures which can be completed based on fix-points on the screen. This has been successfully examined on several occasions (Istance et al., 2009; Istance et al., 1996; Vickers et al., 2008; Istance et al., 2010; Istance et al., 2009b; Wobbrock et al., 2008). The limitation of this is that there needs to be fix-points on the screen, either opaque or semi-transparent. Also the cognitive load is potentially high as the

user has to remember both the combination of strokes and the consequence of each combination.

(3, 4) Finally, gaze gestures can be implemented as *free* or *semi-free gestures* (i.e., a start and stop feature implemented). Both of these gaze gesture types have a potential high cognitive load as the user has to remember stroke combinations and consequences. However, the lack of on-screen visualization might make them appropriate in applications which require the full use of the on-screen real-estate. There might possibly be a lower boundary when it comes to free gaze gestures. In other words single stroke gaze gestures might cause too many accidental gesture completions, as each individual saccade could be erroneously registered.

Further exploration of this area of research is considered, in the context of this research, to be an essential next step in the future development of gaze interaction. However, fully exploring the above mentioned aspects was beyond the scope and timeframe of this research. In the following chapter the main contributions will be reflected upon. There will also be a presentation of some designs which have been derived from this work.

PART IV

10 DISCUSSION

In this chapter the main contributions of this research will be presented, along with design suggestions regarding potential implementations of the various aspects presented in this thesis and suggestions for future research.

The overall approach taken in this research has dealt with four main aspects regarding the exploration of gaze selection strategies. Firstly, the fundamentals of eye movements have been explored with the intent of using these as the foundation for further development and understanding of gaze selection strategies. The second aspect dealt with identifying existing gaze selection strategies, so that application designers have a tool to asses and choose between the different gaze selection strategies and apply them appropriately. Thirdly, the focus has been on the end-users who will benefit from this type of interaction and have specific needs both physiologically and in terms of application requirements. The exploration of gaze selection strategies has attempted to take both into account. And finally, the simplest form of gaze gesture selection was explored to uncover the potential benefits and limitations.

10.1 RECAP IN THE CONTEXT OF THE ORIGINAL RESEARCH QUESTIONS

10.1.1 AFFORDANCES AND CONSTRAINTS OF THE EYE AS INPUT

The first original hypothesis and research questions regarding the understanding of eye movements stated:

Gaze as input has specific interaction design features.

What are these features?

What considerations are important when designing gaze selection strategies?

Understanding the nature of eye movements is essential when designing gaze selection strategies. The features of eye movements can be split into affordances and constraints.

The main affordances of gaze as sole input are:

1. The two main temporal states, saccades and fixations. The reason that these should be viewed as an affordance is the fact, that being able to systematically make a distinction between the two, is the foundation for interactive gaze.
2. The physiology of the eyes; this is both durable and sustainable, and is therefore an affordance which makes gaze well suited as an input modality.

The main constraints of eye movements as sole input are:

The eyes are always 'on', which means that the distinction between perception and interaction eye movements is the constraint.

1. For fixation based interaction the issue is the Midas touch problem, where everything can potentially be selected by viewing.
2. For saccade based interaction the issue is accidental gesture completion, where inspection and navigation patterns might accidentally complete unintended selections.

10.1.2 TAXONOMY OF GAZE INTERACTION PRINCIPLES

The second original hypothesis was regarding existing selection strategies in gaze interaction and stated:

There are many different types of gaze selection strategies.

What are these strategies?

When and how are they best used?

The main contribution, in regard to the above presented research questions, was the gaze selection strategy framework created and presented in chapter 3. Here an overview of existing selection strategies was given. The main distinction was again made between fixation based (dwell) and saccade based (gaze gestures) selection strategies.

Next implementation principles were defined and presented for each gaze selection strategy. These are intended to be used as tools when considering what and how gaze selection strategies should be implemented. For dwell the implementation principles were *selection complexity, visual feedback* and *fixation duration.* For gaze gestures the implementation

principles were *stroke complexity, finite* and *continuous gestures.* These principles were subcategorized and can and should be further developed. *Free gaze gestures* were introduced as a concept which could further expand the existing taxonomy. However, these need to be explored further to establish how this type of gesture relates to the taxonomy. The reason for defining these principles was in order to contribute to the overall understanding of gaze selection strategies. If the goal is to develop flexible gaze contingent applications for mono-modal input; multi-modal interaction strategies should be employed.

The 'when and how' selection strategies should be implemented is dependent on both system specifications (*application*, *task* and *visualization*) and the end-user (*capabilities*, *requirements* and *wants*).

10.1.3 A DECLARATION OF END-USER NEEDS

The third set of research questions that dealt with the end-user, their *capabilities*, *requirements*, and *wants* stated:
Users with severe motor impairments have special needs.

What are these needs?

How can these be assessed from a design perspective?

How do these affect design of gaze based applications and selection strategy?

The main contribution to these questions can be found in chapter 4. The overall *needs* of users with severe motor impairments were to gain as much autonomy as possible in the areas of *communication*, *environmental control*, and *mobility*. This should be done by developing gaze contingent applications which take into account the physiological capabilities of the user. The most important aspect is to make sure that applications can adapt to changes in the user's circumstances, which can occur either as a consequence of degenerative conditions or in the process of rehabilitation.
From a design perspective the main approach is the design applications and application control which can cater to the following areas:

1) The ability to point at objects of interest in the user's environment.
2) The ability to interact with objects in the environment.
3) The desire to control a multitude of applications.
4) Taking special considerations when design peer-to-peer communication systems.
5) Facilitating easy digital communication.
6) Creating gaze controlled wheelchair control.

The design of these applications should be approached from three perspectives: *Task analysis*, *information visualization* and *selection strategy*. These perspectives should in turn be viewed and assessed in regard to the *user context*, which entails cognitive and physiological affordances. Suggestions on how some of the end-user needs can be catered for by using a selection strategy such as the single stroke gaze gestures will be presented in the general discussion section.

10.1.4 FUNDAMENTAL OBSERVATIONS REGARDING SINGLE STROKE GAZE GESTURES

The final research questions dealt with the selection strategy of gaze gestures and stated:
Gaze Gestures constitute a selection strategy.

What is the simplest form of gaze gesture?

What are some of the basic characteristics of simple gaze gestures?

How could these characteristics affect the way gaze gestures are implemented?

The simplest form of gaze gesture was deemed the single stroke gaze gesture, which consisted of one stroke. The immediate assumption regarding this selection strategy was that it would potentially be too error prone, by causing a high number of accidental gesture completions.

However, the research presented in the thesis has shown that by using the edges of the screen, this single stroke gaze gestures can be successfully compared to dwell time selection; exceeding this in terms of *selection completion times*, and overall not having a substantially higher number of *selection errors* or significantly prolonged *task completion times.*

The main affordances of the single stroke gaze gesture are a low physiological and cognitive load, low *selection completion time* and the possibility of being implemented *without visual feedback*.

The main constraints are that single stroke gaze gestures only constitute a limited number of interactions and that there is a risk of accidental gesture completion, which accompanies all gaze gesture selection strategies.

The final research question deals with how single stroke gaze gestures could be implemented. Suggestions to this extent will be presented in the subsequent section.

10.2 GENERAL DISCUSSION

During the course of this research experiments have been made using approximately 60 participants and some observations go beyond the quantifiable nature of the experiments.

One such observation is that people approach eye tracking very differently. Some people would concentrate so hard on the screen that they would get a headache after a very short period of time– while others find it to be the easiest thing in the world. The *intention* and *preconception* which participants have to the task at hand greatly affects the outcome which they produce.

This is meaningful as seen in light of the work which Yarbus (Yarbus et al., 1973) conducted on eye movements where he concluded that the pattern of eye movements differed depending on the intention and cognitive ability of the person viewing the scene (Chapter 2). The same could very well be true for interactive eye tracking – the implications of this are, however, very hard to identify. This constitutes one of the challenges in the field of gaze research, as gaze can function as a direct reflection of what is going on cognitively. How to interpret and ultimately design for this is a subject for future research.

The second observation is that there is a need to move beyond the idea of *mono-modal interaction* for *mono-modal input*. Initially, the approach taken in this research was to show that single stroke gaze gestures were a better concept for gaze interaction than dwell-time selection and while *selection completion times* for SSGG might be significantly faster, the point that there is a need to develop selection strategies which supplement each other has become clear. The goal should therefore be to gain a sense of when to use *dwell*, *single stroke gaze gestures*, *complex gestures*, *zoom selection* etc. in a meaningful and appropriate way.

The main challenge facing the field of gaze interaction is not the technology, as most people believe, or the lack of a 'main-stream' application, such as computer games – although these would both be nice to have. The main challenge is still to design *sustainably* and *appropriately* for the *affordances* of gaze, along with an understanding of the task and subsequent information visualization of a given application. If meaningful ways of using gaze as input are developed; meaningful applications will follow. Most of the gaze selection strategies which have been developed so far have focussed entirely on the limitations of eye movements rather than the possibilities, because the limitations are so obvious. The most important research result in regard to gaze interaction that has come from this work is to break through the barrier of assuming that a *mono-modal input* can only lead to a *monotypic selection strategy*.

However, empirically a monotypic gaze selection strategy has been developed and explored. The main findings of the empirical research can be split into two overall interrelated areas of discovery, the first being findings regarding gaze selection strategies and the second regarding visual feedback.

10.2.1 FINDINGS REGARDING GAZE SELECTION STRATEGIES

THE GRAND MEAN

The grand mean for single stroke gaze gestures on the LC technologies system was 152ms and on the Tobii 1750 it was 105ms; these are substantially faster than a standard dwell selection time of approximately 400ms, which shows the potential of this gaze selection strategy as a very fast option; while still being comfortable, because it can be completed in the user's own time.

However, it is also a substantially faster finding compared with *stroke completion times* in other experiments. Istance et al. (Istance et al., 2010) showed an average stroke time of 247ms/stroke for 2 stroke gestures and 293ms/stroke for the 3 stroke gestures and Wobbrock (Wobbrock et al., 2003) showed an average stroke completion time of between 550-600ms.

There are two potential reasons for this. The first has to do with potential different approaches to measuring *stroke completion times*. The second has to do with the potential nature of complex gaze gesture versus single gaze gestures. In the previous discussion it was

stated that very little is known about how eye movements behave during interactive tasks. The difference in the *stroke completion times* could indicate that different processes occur during the completion of complex gaze gestures, compared with single gaze gestures. What these are is a subject for future research.

SINGLE STROKE GAZE GESTURES ARE NOT MORE ERROR PRONE THAN A STANDARD DWELL-TIME ACTIVATION.

In the gaze gesture and dwell experiment presented in chapter 6, there was no significant difference between the levels of accidental gesture completion in the single stroke gaze gesture conditions and the dwell conditions between 300ms and 1000ms. As mentioned in the discussion this was a surprise. The same was true in the single stroke gaze gesture experiment with and without visualization, Chapter 8.

This indicates that single stroke gaze gesture can indeed be used as a gaze selection method. The question is how they can be used appropriately in applications. Suggestions to this effect will be presented. But future research is also required to fully understand what the potential implications of this gaze selection strategy are.

THE AUGMENTED SELECTION COMPLETION TIME

In the augmented selection completion time the single stroke gaze gesture selections did very well, both in the gaze gesture and dwell experiment (chapter 6) and in the single stroke gaze gestures with and without visual feedback experiment (chapter 8). This further substantiates single stroke gaze gestures as a valid selection strategy.

However, there were a few surprises in this calculation. In chapter 5, the augmented calculation showed that the 500ms dwell condition was the least effective. This was very surprising seeing as this is generally considered a standard dwell time. The question is whether this standard has mainly been set based on the requirements of novice users, as could be suggested by the findings in (Majaranta et al., 2009). If this finding is combined with those presented in this research, it might suggest that the standard dwell completion time suggested by Jacob of 150- 250ms (Jacob, 1993) might ultimately be more appropriate. The key is to find a balance between the *costs* of making an error with the likelihood of making one.

It would be very interesting to explore what the results of this augmented completion time would be for other applications. It is proposed as a measure which can potentially allow for a comparison between different gaze selection strategies.

The α-level calculation was intended to show whether or not a selection strategy would even make sense. As mentioned the idea was that if there was a higher likelihood of making an error than making a successful selection, the method should be discarded. This was the case for the 100ms and 300ms dwell conditions in the dwell and gaze gesture experiment in chapter 6. Surprisingly all of the dwell conditions in the single stroke gaze gesture experiment with and without visualization in chapter 8 were usable. This was explained as being the consequence of the layout of information in this experiment being so conducive to dwell selection.

The question is if the threshold of requiring that there is less than one error per successful selection is low enough. A potential future investigation should explore what the appropriate α-level should be for practical purposes.

DWELL SELECTION TIMES BELOW 200MS AND TASK COMPLETION TIMES.

Even though the main focus of this research was not dwell-time selection a few observations of interest were inadvertently made regarding this gaze selection strategy.

One of these was the fact that dwell *completion times* below 200ms did not have any effect on the overall *task completion time.* This indicates that there might be a lower boundary for dwell at which point it might not be relevant, in terms of task completion times, to lower dwell-time any further. Whether this is the case for dwell in other applications could also be an issue for future studies.

THERE IS A DIFFERENCE BETWEEN LONG AND SHORT *GAZE GESTURES*

Some of the findings regarding gaze selection strategies can have direct design implications regarding how future applications and interface layouts should be developed. The potential of a real difference in *long* SSGG and *short* SSGG is one of these findings.

One way of using this in an interface is by defining different 'zones' of activation, where *long* SSGG and *short* SSGG could be applied depending on frequency of selection an example of a possible implementation is shown in figure 127.

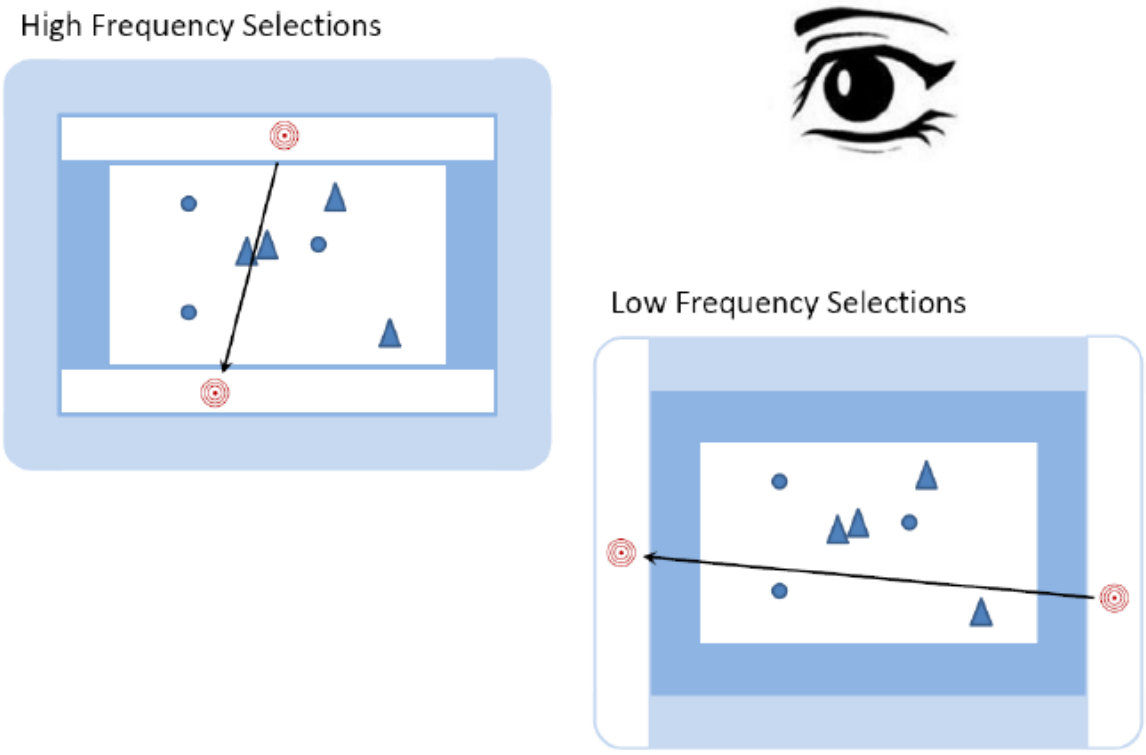


Figure 127: Potential implementation of high and low frequency gaze gestures. High frequency gestures are completed with short SSGG (left) and low frequency selections are completed with long SSGG (right).

Future research should deal with what should be the perfect balance between *gesture length* and *selection error*, to establish a threshold for accidental gesture completion based on single stroke gesture completion of different lengths.

HORIZONTAL SINGLE STROKE GAZE GESTURES ARE FASTER THAN VERTICAL SINGLE STROKE GAZE GESTURES.

The fact that there seems to be a systematic difference between *horizontal* and *vertical* single stroke gaze gesture can also have implications for how information on the screen should be presented.

An example of this could be to explore and further develop a system such as the one presented by Morimoto and Amir (2010). Here they presented a typing application which had two 'qwerty' keyboards placed on top of each other; selections were made by switching between the two with *vertical* eye movements. The findings in the present research indicate that there might be an advantage to be found, by placing the two keyboards next to each

other, causing selections to occur by *horizontal* eye movements instead. This could be another topic for future research.

SINGLE GAZE GESTURES CAN BE COMPLETED WITHOUT VISUALIZATION.

The fact that single stroke gaze gestures could be completed without visual feedback on the screen makes it a very useful and yet unheard of type of gaze selection strategy. A context in which this principle could be useful is when implemented gaze controls for watching movies on a PC or television, where it is of interest for the user to have control without there being large dwell-buttons or semi-transparent selection fields on the screen (Figure 128). This type of selection strategy could also be used in general for flicking through channels on the TV.

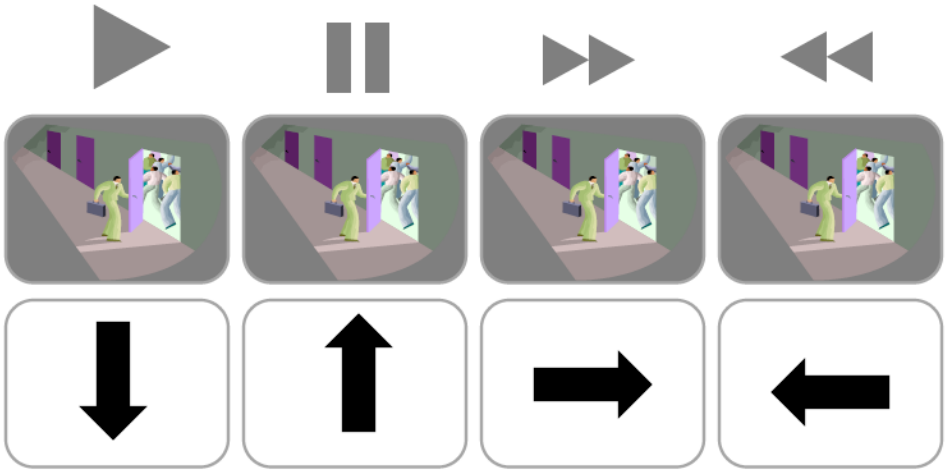


Figure 128: A potential implementation of movie controls based on SSGG

In that regard the completion of the final experiment on gaze gestures with different levels of noise in the background would have been useful. It needs to be uncovered what the chances are for accidental gesture completion to occur while watching a movie or other types of noisy visual input.

COMPLEX GAZE GESTURES

Investigating complex gaze gestures, based on the three principles presented in chapter 9 (i.e., *tracing*, *free*, *semi-free* and *fix-point based*) would be of great value in terms of understanding how *multi-modal interactions* for gaze as *mono-modal input* can be implemented. As mentioned in chapter 2, one of the main challenges of gaze detection is being able to threshold when fixations begin and end (Salvucci & Anderson, 2000). The same is especially true for the implementation of *free complex gaze gestures,* where a neural-network would need to be

developed which had the ability to adapt so as to better interpret a user's selection patterns. Tracing would also be a useful supplement for gaze interaction, because it affords a low cognitive load and the high complexity with which they can be created decreases the risk for accidental gesture completion.

10.2.2 VISUAL FEEDBACK

SEPARATING INSPECTION AND NAVIGATION FROM THE SELECTION PROCESS

The results in chapter 8, showed that even dwell-times of 25ms could be used without a significantly higher number of *selection errors*; most likely due to the layout and visualization of the information on the screen.

Very short dwell selection times can as a consequence potentially be applied if the visualization supports dwell free *inspection* and *navigation*. The same is probably true for very short single gaze gestures. These considerations can also be supported theoretically. As presented in chapter 2, there is evidence that we do not look at multiple objects simultaneously (Yarbus et al., 1973; Noton & Stark, 1971a) this is therefore probably also the case in screen-viewing. If *inspection*/*navigation* occurs separately from selection in terms of visual representation on the screen, the problems stemming from the interference between these types of eye movements can be avoided. In other words, by separating *inspection* /*navigation* from selection, the user will already know where to make a selection and what the consequence of that selection will be before the process has been initiated.

An example of this is given in figure 129. This is an illustration of how a typing interface could be implemented either by using very short dwell-time selections or very short SSGG.

The navigational space in the centre of the screen is not affected by gaze, so the user can navigate without urgency. When the letter of interest is located a selection can be made by looking at the outer rim. Whether it would be most appropriate to complete selections based on short *dwell-times* or simply short single stroke gestures would be a subject for future research. One concern with this design if it were implemented with SSGG would be whether accidental selections would occur if the user looked at the screen for other purposes or simply wanted to ponder while writing, rendering eye movements of no purpose. Addressing these types of design issues is a field of research for the future.

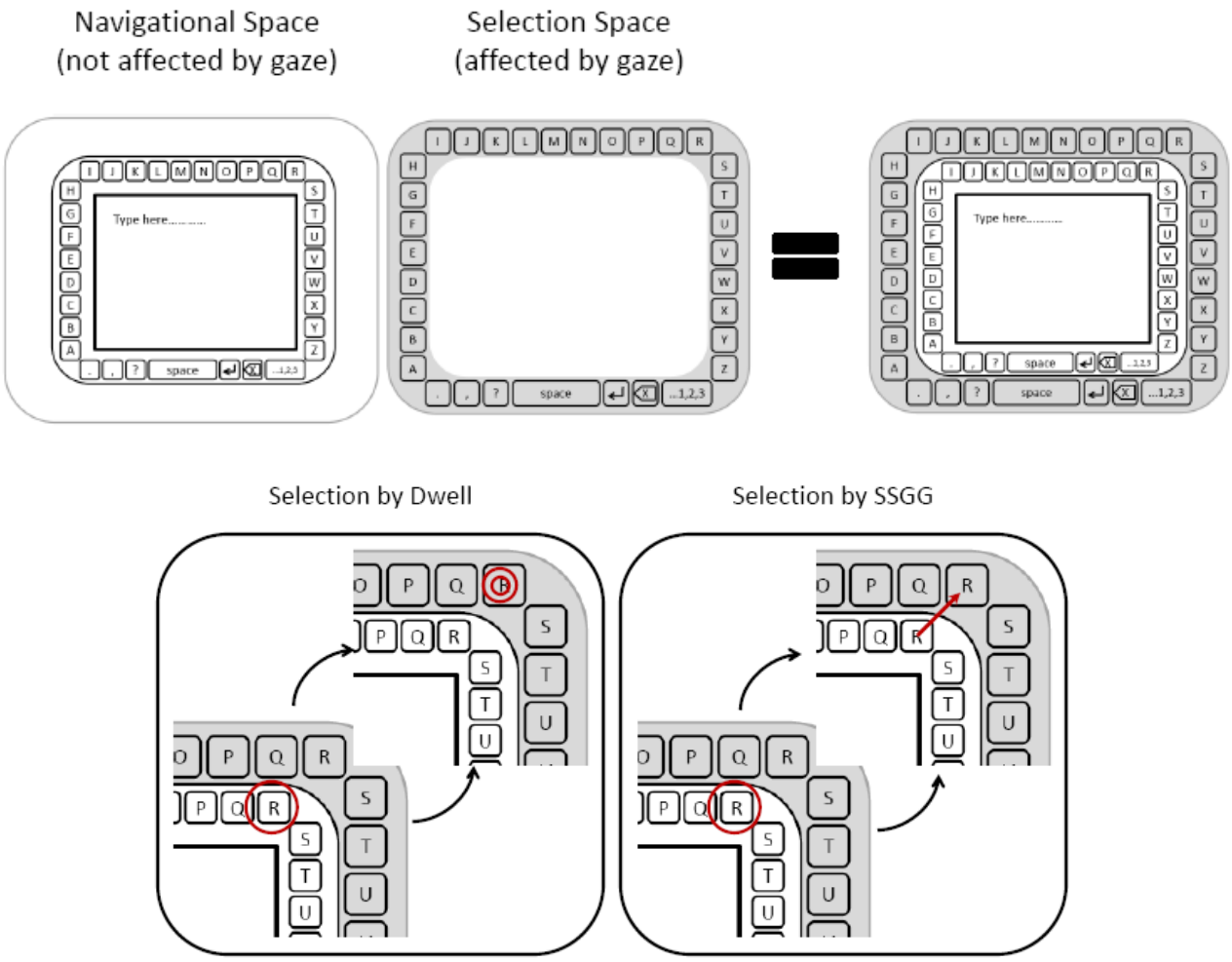


Figure 129: Inspection/navigation separated from selection. Selections occurring based on short dwell-time or short SSGG

This illustration is merely intended to visualize the concept. What the consequences of separating *inspection/navigation* from the selection process are, is also a subject for future research.

PRESENTING INFORMATION ON- AND OFF-TARGETS

Overall the experiments presented in this thesis have dealt with feedback on three different levels *multiple source feedback*, *single source feedback* and *no feedback*. In the initial pilot studies the user feedback was overcomplicated (*multiple source feedback*) and many of the design choices were redundant. In the gaze gesture and dwell experiment in chapter 6 and in parts of the experiment in chapter 8, *single source feedback* was used; in other words the feedback given to the user only came from the initiation and completion fields and only when viewed. Finally, *no feedback* was used as a condition in Chapter 8.

Other observations in regard to feedback have also been made. The first was in regard to how to implement target and action feedback. The distinction made here is between a *target*, which has an on-screen visualization and an *action*, which does not. An example of a *target* would be a *dwell-button*. An action could be initiating panning or scrolling based on dwell, but without a visual representation.

Interaction with computers is generally intended to cause a reliable relationship between a user-action and the subsequent system-response. The whole notion of feedback is to ensure that the consequences of a user-action are understood before the action is completed. This relationship can be implemented in gaze contingent interfaces based on three visualization principles: *on-target*, *off-target* (Figure 130) and *no-target* (Figure 131).

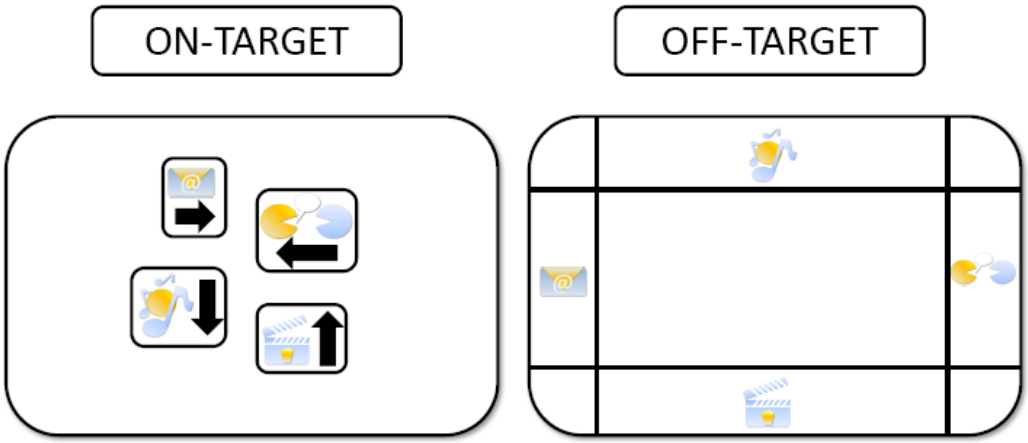


Figure 130: Illustration of the on-(left) and off- target (right) visualization principles.

Both *on* and *off – target* visualizations have *on-screen representation*. *On-target* visualization means that the user-action (in this case the SSGG) and system-response (activating any one of the applications) are displayed *on* the same target. *Off-target* visualizations mean that the user-action is implied in the design layout and therefore off the target, however, the system response is still visualized. Both of these principles can be applied to targets where the system-response is devoid of familiar mapping. In other words, there is no innate familiar mapping between a bottom/top SSGG and opening an email client.

No-target feedback requires an element of familiar mapping or easily learnt consequences. In the previously presented example panning and scrolling build on familiar mapping. An example of this could be reading a text on screen, and the text scrolling up when the eyes reach the bottom of the page. An example of easily learnt consequences could be the

previously presented example of controlling a film viewing session by using single stroke gaze gesture. Here there is a mix of familiar mapping with easily learnt consequences. The fast-forward and backward actions of looking from left/right and right/left respectively hold some familiar mapping; because these actions mirror the visual direction of the icons for these *actions* which are familiar to most people. However, looking from top/bottom to start a film and from bottom/top to pause it, holds no immediate familiar mapping, but might possibly be quite easily learnt.

Another example of the *no-target* visualization principle is shown in figure 131. While navigating in a 3D environment the diagonal SSGG from top-left/bottom-right has been applied to the '*zoom in*' action and the SSGG from bottom-left/top-right has been applied to the '*zoom out*' action.

This idea is the foundation for a potential environmental control system, which could be used by the end-user. The environment of the end-user should be displayed on a screen, by integrating dwell and gaze gestures camera control could be achieved. Dwell could be used for panning actions and gaze gestures for zoom. This then allows the user 360° of visual freedom, in their home environment. By zooming and panning the user can then 'point' to objects of interest in their surroundings. These objects of interest can either be inanimate, in which case they would be of most relevance to carers or other individuals in the user's environment. This could, for instance, aid in situations where the user wants to watch a specific DVD and selects it by 'pointing' to it or pointing to a book, or any other choice which involves inanimate objects. However, the user could potentially also zoom in on objects which could be subsequently controlled by themselves. An example of this could a TV, by zooming in to a certain level on the TV the system will recognize that this is an object which the user is interested in interacting with, and the controls for the TV will then be displayed on the user's monitor. Rather than having to search through elaborate menu structures, the user can essentially point to the object in their environment which they want to control (figure 131).

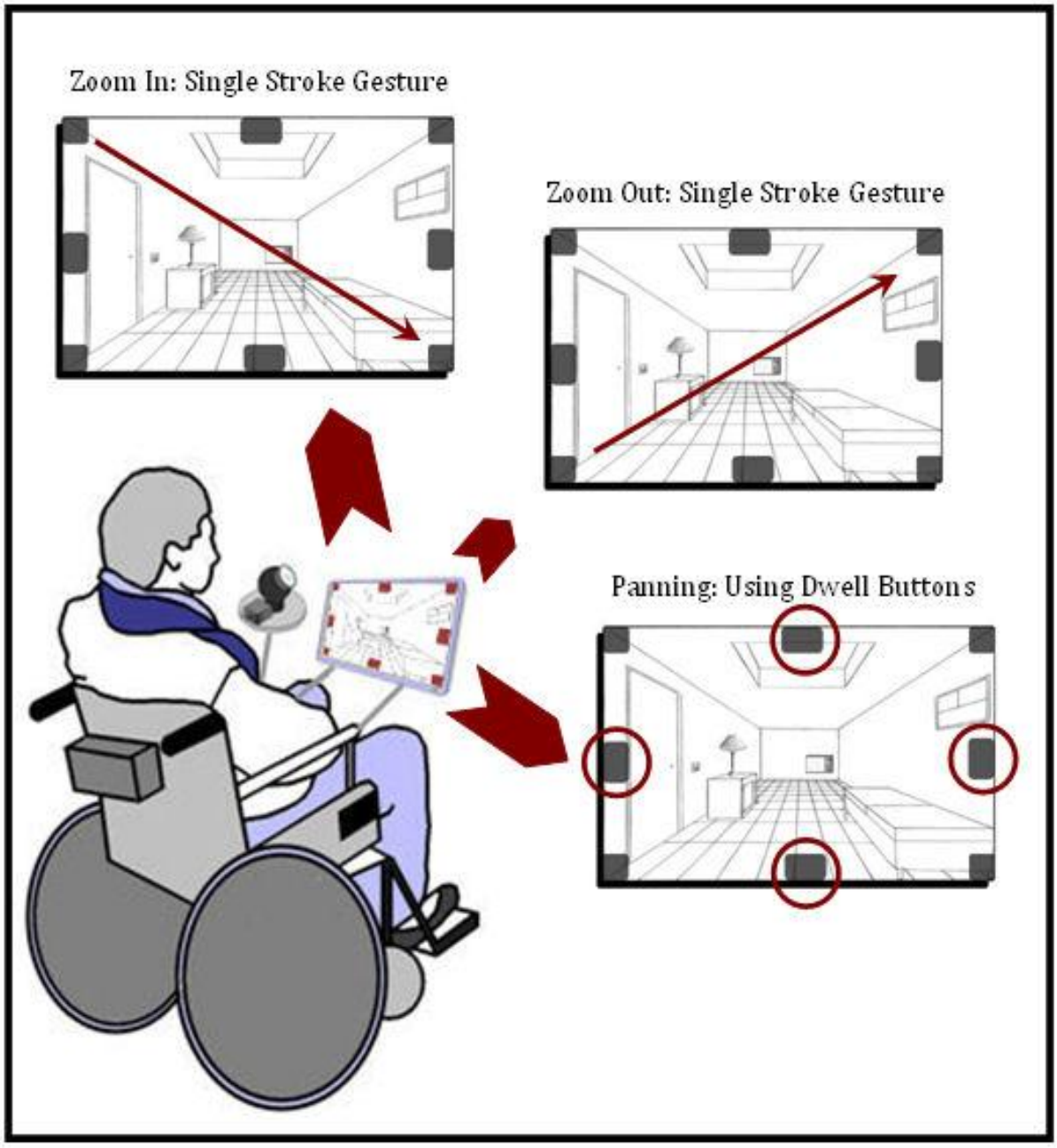


Figure 131: An example of an integrated version of dwell and gaze gesture selection

The key element of implementing gaze selection strategies is to ensure that they build on either familiar mapping or are easily learnt; this should be the overall goal for future gaze application design and development.

The main conclusion of this research, in regard to gaze selection strategies, is that large single stroke gaze gestures would be fundamentally well suited for simple change-of-context tasks and that they are a useful addition to the toolbox being developed for gaze interaction.

The path taken through this research has been exploratory and can be seen in figure 132.

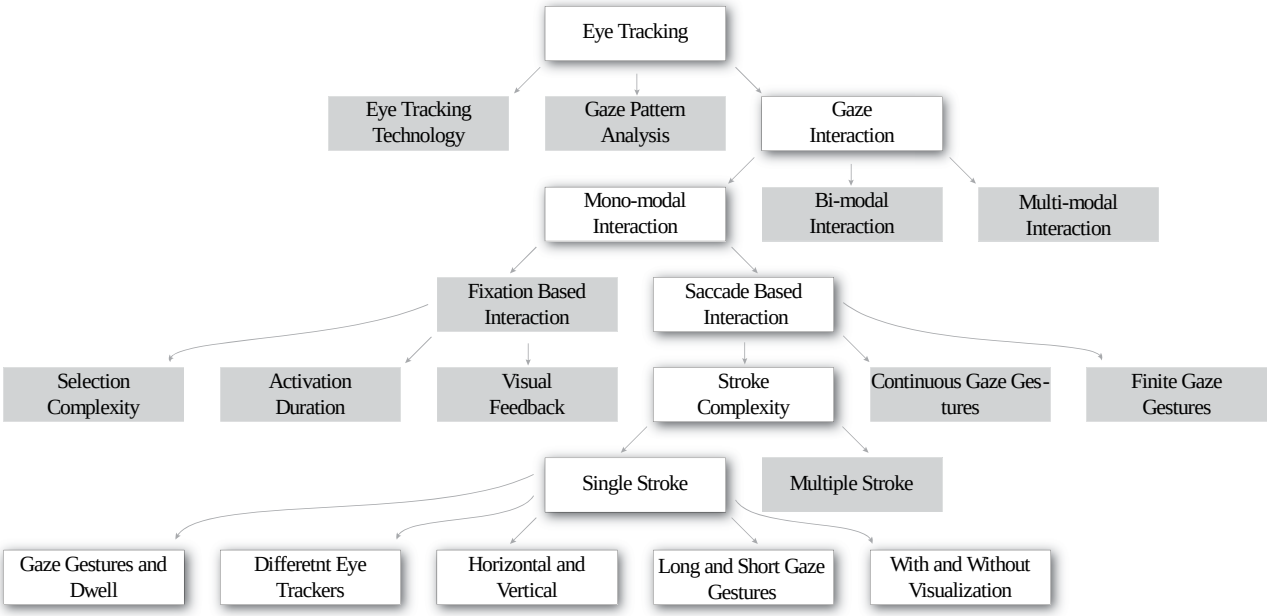


Figure 132: The path of research taken.

Through this work I have attempted to contribute to the understanding of gaze interaction in two ways; through exploring the application of a previously untested selection strategy and by conceptualizing a toolbox of selection strategies. Even though the work has been conducted from the top to the bottom of the diagram in figure 132, understanding the contribution of this work can be done by viewing it from the bottom up. At the bottom level (figure 132) are the five key findings which the experimental research conducted in this PhD has yielded. The first main finding is that single stroke gaze gestures are not significantly more error prone than dwell time selection. The second key finding is that different eye tracking equipment greatly affects the completion of gestures, both in terms of completion time and error. The third surprising finding is that horizontal single stroke gestures were completed faster than vertical ones. The fourth finding was that there was a significant difference in the completion time between short and long single stroke gaze gestures, short being significantly faster. The final finding was that single stroke gaze gestures can be completed without requiring visual fix-points on the screen. It is this final finding that in particular represents the strength of single stroke gaze gestures, as no other examples of this have currently been investigated to the author's knowledge. This fundamental research has

contributed to the understanding of single stroke gaze gestures, which in turn sheds light on the overall concept of stroke complexity in gaze gestures. Moving up through the diagram the conceptual contribution is next.

Conceptually the contribution has been made through the identification of a toolbox of principles that should enable designers of gaze contingent application to understand the choice of selection strategy which is available to them (figure 133).

Figure 133: Toolbox of principles which should be used when designing for gaze interaction.

The principles of selection complexity, activation and visual feedback should be considered when designing dwell time activation (chapter 3, page 44) and deciding between finite gaze gesture, continuous gaze gestures and level of stroke complexity should be considered when implementing gaze gestures (chapter 3, page 57). These concepts are intended to affect the understanding of *mono-modal input*.

Deciding which is the right selection strategy or combination of selection strategies is dependent on the context of use. The intention of the research presented here has been, as previously stated, to break the idea that *mono-modal input,* must result in *mono-modal interaction*. The future research direction in field of gaze interaction should therefore be exploring *multi-modal* gaze interaction*s* for *mono-modal input.* It is through these considerations that this work seeks to contribute to the overall field of *human computer interaction*.

More than anything else the end-users in this field are the inspiration and motivation for continued research, and they know what it means to need to break free from assumed constraints. On that note, the final thought in this thesis will be given by Arne Lykke Larsen, who was first presented in the introduction:

'*If by a miracle I should find an Aladdin's lamp (such a thing naturally doesn't exist, at least not scientifically!) and the genie inside (which also doesn't exist!) would grant me one wish; what would I choose? […]Would I choose to be able to eat and drink again? I love exotic food, especially from the Far East: Chinese soups, an Indonesian banquet with 26 little dishes, Indian*

Tandoori Chicken etc. Or would I choose to walk again? – I love walking in nature: Hiking in the Rocky Mountains, wandering on the Great Plains in Alaska or just a little trip to the local forest or marsh. Would I choose to get the use of my arms and hands back? – I love doing mathematical computations on paper, and I would be able to do it 100 times faster than on the computer, and 10 times faster than my secretary does them on the black board. Or would I choose to get my speech back? – Then I could start lecturing at the University again, which I really miss. Maybe I would choose to get my ability to breathe back? – wouldn't that be wonderful just to be able to breathe freely and effortlessly again! No, if I really could have one single physical wish granted, then I would wish that I could fly like the birds…. '

AUTHOR'S PUBLICATIONS

1. Hansen, D.W., Skovsgaard, H.H., Hansen, J.P. & Mollenbach, E. Noise tolerant selection by gaze-controlled pan and zoom in 3D. *Proceedings of the 2008 symposium on Eye tracking research & applications* 205–212 (2008).
2. Moellenbach, E., Gale, A. & Hansen, J.P. THE ASSISTIVE EYE: AN INCLUSIVE DESIGN APPROACH. *Contemporary ergonomics 2008* 285 (2008).
3. Mollenbach, E., Stefansson, T. & Hansen, J.P. All eyes on the monitor: gaze based interaction in zoomable, multi-scaled information-spaces. *Proceedings of the 13th international conference on Intelligent user interfaces* 373–376 (2008).
4. Mollenbach, E., Hansen, J.P., Lillholm, M. & Gale, A.G. Single stroke gaze gestures. *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems* 4555–4560 (2009).
5. Mollenbach, E., Lillholm, M., Gail, A. & Hansen, J.P. Single gaze gestures. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* 177–180 (2010).
6. Mollenbach, E., Hansen, J.P. & Gale, A.G. From On-screen Navigation to Real World Interaction, *The Scandinavian Workshop on Applied Eye-Tracking* **2008**,
7. San Agustin, J. et al. Evaluation of a low-cost open-source gaze tracker. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* 77–80 (2010).
8. Shi, F., Gale, A. & Mollenbach, E. Eye, Me and the Environment. *Computers Helping People with Special Needs* 1030–1033 (2008).
9. Tall, M. et al. Gaze-controlled driving. *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems* 4387–4392 (2009).

REFERENCES

1. Ahmadi, M. et al. Human Extraocular Muscles in ALS. *Investigative Ophthalmology & Visual Science* **51**, 3494 -3501 (2010).
2. Asari, S. & Ohmoto, T. Natural history and risk factors of unruptured cerebral aneurysms. *Clinical Neurology and Neurosurgery* **95**, 205-214 (1993).
3. Bates, R. Have patience with your eye mouse! Eye-gaze interaction with computers can work. *Proceedings of the 1st Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT)* (2002).
4. Bates, R. & Istance, H. Zooming interfaces!: enhancing the performance of eye controlled pointing devices. *Proceedings of the fifth international ACM conference on Assistive technologies* 119–126 (2002).
5. Bates, R., Istance, H., Donegan, M. & Oosthuizen, L. Fly where you look: Enhancing gaze based interaction in 3d environments. *Proc. COGAIN-05* 30–32 (2005).
6. Bederson, B.B., Meyer, J. & Good, L. Jazz: an extensible zoomable user interface graphics toolkit in Java. *Proceedings of the 13th annual ACM symposium on User interface software and technology* 171–180 (2000).
7. Bee, N. & André, E. Writing with your eye: A dwell time free writing system adapted to the nature of human eye gaze. *Perception in Multimodal Dialogue Systems* 111–122 (2008).
8. Blickenstorfer, C.H. Graffiti: wow. *Pen Computing Magazine* **1**, 30–31 (1995).
9. Bolt, R.A. Gaze-orchestrated dynamic windows. *ACM SIGGRAPH Computer Graphics* **15**, 109–119 (1981).
10. Bolt, R.A. Eyes at the interface. *Proceedings of the 1982 conference on Human factors in computing systems* 360–362 (1982).
11. Bracken, M.B. et al. Administration of methylprednisolone for 24 or 48 hours or tirilazad mesylate for 48 hours in the treatment of acute spinal cord injury: results of the Third National Acute Spinal Cord Injury Randomized Controlled Trial. *Jama* **277**, 1597 (1997).
12. Brown, C.M. *Human-computer interface design guidelines*. (Intellect Books: 1999).
13. Buchholz, M. & Holmqvist, E. Eye gaze assessment with a person having complex needs– the benefits of using an assessment method. *The 5th Conference on Communication by Gaze Interaction (COGAIN 2009)* (2009).
14. Buzan, T. & Buzan, B. *The mind map book*. (Pearson Education: 2006).
15. Carpenter, R.H. *Movements of the Eyes*. **158**, (Pion London: 1988).
16. Carpenter, R.H.S. The visual origins of ocular motility. *Vision and visual function* **8**, 1–10 (1991).
17. Chandler, S. et al. Matrix metalloproteinases, tumor necrosis factor and multiple sclerosis: an overview. *Journal of neuroimmunology* **72**, 155–161 (1997).
18. Chase, W.G. & Simon, H.A. Perception in chess* 1. *Cognitive psychology* **4**, 55–81 (1973).
19. Daniel, P.M. & Whitteridge, D. The representation of the visual field on the cerebral cortex in monkeys. *The Journal of Physiology* **159**, 203 (1961).
20. Distal, H.M.N. Spinal muscular atrophy.
21. Dodge, R. Visual perception during eye movement. *Psychological Review* **7**, 454–465 (1900).
22. Dodge, R. An improved exposure apparatus. *Psychological Bulletin* **4**, 10–13 (1907).
23. Drewes, H. & Schmidt, A. Interacting with the computer using gaze gestures. *Proceedings of*

the 11th IFIP TC 13 international conference on Human-computer interaction-Volume Part II 475–488 (2007).
24. Duchowski, A.T. *Eye tracking methodology: Theory and practice*. (Springer-Verlag New York Inc: 2007).
25. Ebisawa, Y. & others Unconstrained pupil detection technique using two light sources and the image difference method. *Visualization and Intelligent Design in Engineering* **79**, 89 (1989).
26. Franconeri, S.L. & Simons, D.J. Moving and looming stimuli capture attention. *Perception & Psychophysics* **65**, 999 (2003).
27. Goldberg, D. & Richardson, C. Touch-typing with a stylus. *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems* 80–87 (1993).
28. Hansen, D.W. & Hansen, J.P. Eye typing with common cameras. *Proceedings of the 2006 symposium on Eye tracking research & applications* 55 (2006).
29. Hansen, D.W. & Ji, Q. In the eye of the beholder: A survey of models for eyes and gaze. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **32**, 478–500 (2010).
30. Hansen, D.W., MacKay, D.J., Hansen, J.P. & Nielsen, M. Eye tracking off the shelf. *Proceedings of the 2004 symposium on Eye tracking research & applications* 58 (2004).
31. Hansen, J.P., Lund, H., Aoki, H. & Itoh, K. Gaze communication systems for people with ALS. *ALS Workshop, in conjunction with the 17th International Symposium on ALS/MND, Yokohama, Japan* (2006).
32. Hansen, D.W., Skovsgaard, H.H., Hansen, J.P. & M\ollenbach, E. Noise tolerant selection by gaze-controlled pan and zoom in 3D. *Proceedings of the 2008 symposium on Eye tracking research & applications* 205–212 (2008).
33. Hatfield, G. & Wright, R. Attention in early scientific psychology. *Visual attention* **8**, 3–25 (1998).
34. Hayhoe, M. & Ballard, D. Eye movements in natural behavior. *Trends in Cognitive Sciences* **9**, 188–194 (2005).
35. Heikkilä, H. & Räihä, K. Speed and Acuracy of Gaze Gestures. *Journal of Eye Movement Research* **2009**, 1-14
36. Helmholtz, H. Treatise on physiological optics (Vol. 3). *Menasha, WI: Optical Society of America* (1924).
37. Hubel, D.H. The Visual Cortex of the Brain. *Scientific American* **1963**, 2-10 (1963).
38. Isokoski, P. Text input methods for eye trackers using off-screen targets. *Proceedings of the 2000 symposium on Eye tracking research & applications* 21 (2000).
39. Istance, H.O., Spinner, C. & Howarth, P.A. Providing motor impaired users with access to standard Graphical User Interface (GUI) software via eye-based interaction. *Proceedings of the 1st European Conference on Disability, Virtual Reality and Associated Technologies (ECDVRAT'96)* 109–116 (1996).
40. Istance, H., Bates, R., Hyrskykari, A. & Vickers, S. Snap clutch, a moded approach to solving the Midas touch problem. *Proceedings of the 2008 symposium on Eye tracking research & applications* 221–228 (2008).
41. Istance, E.H., Štěpánková, O. & Bates, R. Communication, Environment and Mobility Control by Gaze Welcome to COGAIN 2008! (2008).
42. Istance, H., Hyrskykari, A., Vickers, S. & Chaves, T. For Your Eyes Only: Controlling 3D Online Games by Eye-Gaze. *Human-Computer Interaction–INTERACT 2009* 314–327 (2009).
43. Istance, H., Vickers, S. & Hyrskykari, A. Gaze-based interaction with massively multiplayer

on-line games. *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems* 4381–4386 (2009).
44. Istance, H., Vickers, S. & Hyrskykari, A. Gaze-based interaction with massively multiplayer on-line games. *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems* 4381–4386 (2009).
45. Istance, H., Hyrskykari, A., Immonen, L., Mansikkamaa, S. & Vickers, S. Designing gaze gestures for gaming: an investigation of performance. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* 323–330 (2010).
46. Jacob, R.J. The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Transactions on Information Systems (TOIS)* **9**, 152–169 (1991).
47. Jacob, R.J. The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Transactions on Information Systems (TOIS)* **9**, 152–169 (1991).
48. Jacob, R.J. Eye movement-based human-computer interaction techniques: Toward non-command interfaces. *Advances in human-computer interaction* **4**, 151–190 (1993).
49. Jacob, R.J. Human-computer interaction: input devices. *ACM Computing Surveys (CSUR)* **28**, 177–179 (1996).
50. Jacob, R.J. & Karn, K.S. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind* **2**, 4 (2003).
51. Jeppesen, B.B. *Er der mon bedre i himlen?* **1999**, (Dafolo:).
52. Kanizsa, G. *Organization in vision: Essays on Gestalt perception*. (Praeger Publishers: 1979).
53. Keates, S. & Robinson, P. The use of gestures in multimodal input. *Proceedings of the third international ACM conference on Assistive technologies* 35–42 (1998).
54. Kowler, E. & Anton, S. Reading twisted text: Implications for the role of saccades. *Vision Research* **27**, 45–60 (1987).
55. Kozma, L., Klami, A. & Kaski, S. GaZIR: Gaze-based zooming interface for image retrieval. *Proceedings of the 2009 international conference on Multimodal interfaces* 305–312 (2009).
56. Krauzlis, R.J. Recasting the smooth pursuit eye movement system. *Journal of neurophysiology* **91**, 591 (2004).
57. Krauzlis, R.J. & Lisberger, S.G. Temporal properties of visual motion signals for the initiation of smooth pursuit eye movements in monkeys. *Journal of Neurophysiology* **72**, 150 (1994).
58. Krigger, K.W. Cerebral palsy: an overview. *Am Fam Physician* **73**, 91–100 (2006).
59. Kristjansson, A., Vandenbroucke, M.W. & Driver, J. When pros become cons for anti-versus prosaccades: factors with opposite or common effects on different saccade types. *Experimental Brain Research* **155**, 231–244 (2004).
60. Land, M. & Tatler, B. *Looking and acting: vision and eye movements in natural behaviour*. (Oxford: Oxford University Press: 2009).
61. Land, M., Mennie, N. & Rusted, J. The roles of vision and eye movements in the control of activities of daily living. *PERCEPTION-LONDON-* **28**, 1311–1328 (1999).
62. Larsen, A.L. *Hellere dø af grin end af ALS – 99 sandfærdige historier om at leve med Amyotrofisk Lateral Sclerose*. **2009**, (Skriverforlaget:).
63. Laureys, S. et al. The locked-in syndrome: what is it like to be conscious but paralyzed and voiceless? *Progress in brain research* **150**, 495–511 (2005).
64. Laurutis, V.P. & Robinson, D.A. The vestibulo-ocular reflex during human saccadic eye

movements. *The Journal of Physiology* **373**, 209 (1986).
65. Level, D. D7. 2. Report on a market study and demographics of user population. (2004).
66. MacKenzie, I.S., Nonnecke, R.B., McQueen, J.C., Riddersma, S. & Meltz, M. A comparison of three methods of character entry on pen-based computers. *Human Factors and Ergonomics Society Annual Meeting Proceedings* **38**, 330–334 (1994).
67. Majaranta, P. & Räihä, K.J. Twenty years of eye typing: systems and design issues. *Proceedings of the 2002 symposium on Eye tracking research & applications* 22 (2002).
68. Majaranta, P., Ahola, U.K. & Špakov, O. Fast gaze typing with an adjustable dwell time. *Proceedings of the 27th international conference on Human factors in computing systems* 357–360 (2009).
69. Mankoff, J. & Abowd, G.D. Cirrin: a word-level unistroke keyboard for pen input. *Proceedings of the 11th annual ACM symposium on User interface software and technology* 213–214 (1998).
70. Matin, E. Saccadic suppression: A review and an analysis. *Psychological Bulletin* **81**, 899–917 (1974).
71. Moellenbach, E., Gale, A. & Hansen, J.P. THE ASSISTIVE EYE: AN INCLUSIVE DESIGN APPROACH. *Contemporary ergonomics 2008* 285 (2008).
72. Mollenbach, E., Stefansson, T. & Hansen, J.P. All eyes on the monitor: gaze based interaction in zoomable, multi-scaled information-spaces. *Proceedings of the 13th international conference on Intelligent user interfaces* 373–376 (2008).
73. Mollenbach, E., Hansen, J.P., Lillholm, M. & Gale, A.G. Single stroke gaze gestures. *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems* 4555–4560 (2009).
74. Mollenbach, E., Lillholm, M., Gail, A. & Hansen, J.P. Single gaze gestures. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* 177–180 (2010).
75. Mollenbach, E., Hansen, J.P. & Gale, A.G. From On-screen Navigation to Real World Interaction, *The Scandinavian Workshop on Applied Eye-Tracking* **2008**,
76. Morimoto, C.H. & Amir, A. Context switching for fast key selection in text entry applications. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* 271–274 (2010).
77. Morimoto, C.H., Koons, D., Amir, A. & Flickner, M. Pupil detection and tracking using multiple light sources. *Image and Vision Computing* **18**, 331–335 (2000).
78. Newhall, S.M. Instrument for observing ocular movements. *The American Journal of Psychology* **40**, 628–629 (1928).
79. Norman, D.A. Affordance, conventions, and design. *interactions* **6**, 38–43 (1999).
80. Noton, D. & Stark, L. Scanpaths in eye movements during pattern perception. *Science (New York, NY)* **171**, 308 (1971).
81. Noton, D. & Stark, L. Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research* **11**, 929–942 (1971).
82. Perlin, K. Quikwriting: continuous stylus-based text entry. *Proceedings of the 11th annual ACM symposium on User interface software and technology* 215–216 (1998).
83. Plaisant, C., Carr, D. & Shneiderman, B. Image-browser taxonomy and guidelines for designers. *Software, IEEE* **12**, 21–32 (2002).
84. Porta, M. & Turina, M. Eye-S: a full-screen input modality for pure eye-based communication. *Proceedings of the 2008 symposium on Eye tracking research & applications* 27–34 (2008).

85. Qvarfordt, P. & Zhai, S. Conversing with the user based on eye-gaze patterns. *Proceedings of the SIGCHI conference on Human factors in computing systems* 221–230 (2005).
86. Rashbass, C. The relationship between saccadic and smooth tracking eye movements. *The Journal of Physiology* **159**, 326 (1961).
87. Raskin, J. *The humane interface: new directions for designing interactive systems*. (ACM Press/Addison-Wesley Publishing Co. New York, NY, USA: 2000).
88. Reilly, R.G. & O'Regan, J.K. Eye movement control during reading: A simulation of some word-targeting strategies. *Vision Research* **38**, 303–317 (1998).
89. Roberts, D., Shelhamer, M. & Wong, A. A new wireless search-coil system. *Proceedings of the 2008 symposium on Eye tracking research & applications* 197–204 (2008).
90. Robinson, D.A. The mechanics of human saccadic eye movement. *The Journal of Physiology* **174**, 245 (1964).
91. Sadovnick, A.D. & Ebers, G.C. Epidemiology of multiple sclerosis: a critical overview. *The Canadian journal of neurological sciences. Le journal canadien des sciences neurologiques* **20**, 17 (1993).
92. Salvucci, D.D. & Anderson, J.R. Intelligent gaze-added interfaces. *Proceedings of the SIGCHI conference on Human factors in computing systems* 273–280 (2000).
93. Salvucci, D.D. & Goldberg, J.H. Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the 2000 symposium on Eye tracking research & applications* 71–78 (2000).
94. San Agustin, J. Off-the-Shelf Gaze Interaction. (2009).
95. San Agustin, J. et al. Evaluation of a low-cost open-source gaze tracker. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* 77–80 (2010).
96. Sekuler, R. & Blake, R. *Perception (3rd edn)*. (New York: McGraw-Hill.(4th edn, 2000): 1994).
97. Shein, F. et al. WiViK: A visual keyboard for Windows 3.0. *Proceedings of the 14th Annual Conference of the Rehabilitation Engineering and Assistive Technology Society of North America (RESNA '91)* 21–26 (1991).
98. Shi, F., Gale, A. & Purdy, K. Helping people with ICT device control by eye gaze. *Computers Helping People with Special Needs* 480–487 (2006).
99. Shi, F., Gale, A.G. & Purdy, K.J. Eye-centric ICT control. *Contemporary Ergonomics* 215–218 (2006).
100. Skovsgaard, H., Hansen, J.P. & Mateo, J.C. How can tiny buttons be hit using gaze only. *COGAIN 2008* (2008).
101. Starr, M.S. & Rayner, K. EYE MOVEMENTS DURING READING. *Psycholinguistics: critical concepts in psychology* 405 (2002).
102. Tall, M. NeoVisus:Gaze Driven Interface Components. *The 4th Conference on Communication by Gaze Interaction (COGAIN 2008)* (2008).
103. Tatler, B.W. The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision* **7**, (2007).
104. Tatler, B.W. & Wade, N.J. On nystagmus, saccades, and fixations. *PERCEPTION-LONDON-* **32**, 167–184 (2003).
105. Tian, Y., Kanade, T. & Cohn, J.F. Dual-state parametric eye tracking. *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on* 110–115 (2002).
106. Tinker, M.A. Eye movements in reading. *The Journal of Educational Research* 241–277 (1936).

107. Tscherning, M.H. & Weiland, C. *Physiologic optics: dioptrics of the eye, functions of the retina, ocular movements and binocular vision*. (Keystone: 1904).
108. Urbina, M.H. & Huckauf, A. Dwell time free eye typing approaches. *Proceedings of the 3rd Conference on Communication by Gaze Interaction (COGAIN 2007)* 3–4 (2007).
109. Urbina, M.H. & Huckauf, A. Alternatives to single character entry and dwell time selection on eye typing. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* 315–322 (2010).
110. Verstappen, S.M.M. et al. Overview of work disability in rheumatoid arthritis patients as observed in cross-sectional and longitudinal surveys. *Arthritis Care & Research* **51**, 488–497 (2004).
111. Vertegaal, R. The GAZE groupware system: mediating joint attention in multiparty communication and collaboration. *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit* 294–301 (1999).
112. Vickers, S., Istance, H.O., Hyrskykari, A., Ali, N. & Bates, R. Keeping an eye on the game: eye gaze interaction with Massively Multiplayer Online Games and virtual communities for motor impaired users. *Proc. 7th ICDVRAT with Art Abilitation, Maia, Portugal* (2008).
113. Ward, D.J., Blackwell, A.F. & MacKay, D.J. Dasher—a data entry interface using continuous gestures and language models. *Proceedings of the 13th annual ACM symposium on User interface software and technology* 129–137 (2000).
114. Ware, C. & Mikaelian, H.H. An evaluation of an eye tracker as a device for computer input2. *ACM SIGCHI Bulletin* **17**, 183–188 (1986).
115. Wobbrock, J.O., Myers, B.A. & Kembel, J.A. EdgeWrite: a stylus-based text entry method designed for high accuracy and stability of motion. *Proceedings of the 16th annual ACM symposium on User interface software and technology* 61–70 (2003).
116. Wobbrock, J.O., Rubinstein, J., Sawyer, M.W. & Duchowski, A.T. Longitudinal evaluation of discrete consecutive gaze gestures for text entry. *Proceedings of the 2008 symposium on Eye tracking research & applications* 11–18 (2008).
117. Yarbus, A.L. Eye movements during perception of complex objects. *Eye movements and vision* **7**, 171–196 (1967).
118. Zeki, S. & Marg, E. A Vision of the Brain. (1993).
119. Zhai, S., Morimoto, C. & Ihde, S. Manual and gaze input cascaded (MAGIC) pointing. *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit* 246–253 (1999).