Computational Analysis of Eye-Strain for Digital Screens based on Eye Tracking Studies

Mohsen Parisay

A Thesis
in
The Department
of
Computer Science and Software Engineering

Presented in Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy (Computer Science) at
Concordia University
Montréal, Québec, Canada

January 2022

Concordia University

School of Graduate Studies

This is to certify that the thesis prepared

By: **Mohsen Parisay**

Entitled: **Computational Analysis of Eye-Strain for Digital Screens based on Eye Tracking Studies**

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Computer Science)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Dr. Lata Narayanan	Chair
Dr. Christian Hansen	External Examiner
Dr. Najmeh Khalili-Mahani	External to Program
Dr. Tristan Glatard	Examiner
Dr. Adam Krzyzak	Examiner
Dr. Marta Kersten-Oertel	Thesis Supervisor
Dr. Charalambos Poullis	Co-supervisor

Approved by ______________________________
Dr. Leila Kosseim, Graduate Program Director

December 10, 2021 ______________________________
Dr. Mourad Debbabi, Dean
Gina Cody School of Engineering and Computer Science

Abstract

Computational Analysis of Eye-Strain for Digital Screens based on Eye Tracking Studies

Mohsen Parisay, Ph.D.

Concordia University, 2022

Computer vision syndrome (CVS) is composed of multiple eye vision problems due to the prolonged use of digital displays, including tablets and smartphones. These problems were shown to affect visual comfort as well as work productivity in both adults and teenagers. CVS causes eye and vision symptoms such as *eye-strain*, *eye burn*, *dry eyes*, *double vision*, and *blurred vision*. CVS, which causes severe vision and muscular problems due to repeated eye movements and excessive eye focus on computer screens, is a cause of work-related stress. In this thesis, we address this problem and present three general-purpose mathematical compound models for assessing eye-strain in eye-tracking applications, namely (1) Fixation-based Eye fatigue Load Index (FELiX), (2) Index of Difficulty for Eye-tracking Applications (IDEA), and (3) Eye-Strain Probation Model (ESPiM) based on eye-tracking parameters and subjective ratings to measure, predict, and compare the amount of fatigue or cognitive workload during target selection tasks for different user groups or interaction techniques. The ESPiM model is the outcome of both FELiX and IDEA, which benefit from direct subjective rating and, therefore, can be applied to assess the ESPiM model's efficacy. We present experiments and user studies that show that these models can measure potential eye-strain levels on individuals based on physical circumstances such as screen resolution and target positions per time.

Acknowledgments

This is the final station of my long journey of curiosity. I would like to thank my wonderful parents for their support of my education with their best efforts.

Special thanks to my supervisors Prof. Charalambos Poullis, and Prof. Marta Kersten-Oertel.

Contributions of Authors

I am the first author of four papers (three published) presented in this thesis. My role in this work was vital. I formed the ideas, designed the algorithms and implemented the systems, conducted the user studies, and led the writing of scientific publications. The contributions of co-authors include manuscript review, supervision, software development, and technical support. The following list contains the contributions of each author in detail:

Chapter 3: EyeTAP: Introducing a multimodal gaze-based technique using voice inputs with a comparative analysis of selection techniques. International Journal of Human-Computer Studies, Volume 154, 102676, 2021.

- Authors: Mohsen Parisay, Charalambos Poullis, Marta Kersten-Oertel
- Contributions: Conceptualization, Methodology, Software, Formal analysis, Writing - original draft, Writing - review & editing: Mohsen Parisay; Conceptualization, Methodology, Writing - review & editing: Charalambos Poullis, Marta Kersten-Oertel

Chapter 4: FELiX: Fixation-based Eye Fatigue Load Index A Multi-factor Measure for Gaze-based Interactions. 13th International Conference on Human System Interaction (HSI), 74-81, 2020. Recipient of the best paper finalist award.

- Authors: Mohsen Parisay, Charalambos Poullis, Marta Kersten-Oertel

- Contributions: Study design: Mohsen Parisay; software development: Mohsen Parisay; analysis of the results: all authors; data collection: Mohsen Parisay; paper write-up: all authors; supervision: Charalambos Poullis, Marta Kersten-Oertel; review and revision: all authors.

Chapter 5: IDEA: Index of Difficulty for Eye tracking Applications An Analysis Model for Target Selection Tasks. Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 1: HUCAPP, 135-144, 2021.

- Authors: Mohsen Parisay, Charalambos Poullis, Marta Kersten-Oertel
- Contributions: Study design: Mohsen Parisay; software development: Mohsen Parisay; analysis of the results: all authors; data collection: Mohsen Parisay; paper write-up: all authors; supervision: Charalambos Poullis, Marta Kersten-Oertel; review and revision: all authors.

Chapter 6: ESPiM: Eye-Strain Probation Model - An Eye Tracking Analysis Measure for Digital Displays. Prepared for the International Journal of Human-Computer Studies.

- Authors: Mohsen Parisay, Negar Haghbin, Charalambos Poullis, Marta Kersten-Oertel
- Contributions: Study design: Mohsen Parisay; software development: Mohsen Parisay and Negar Haghbin; analysis of the results: all authors; data collection: Mohsen Parisay; paper write-up: all authors; supervision: Charalambos Poullis, Marta Kersten-Oertel; review and revision: all authors.

Contents

List of Figures

List of Tables

List of Abbreviations

ANOVA Analysis of Variance.

CVS Computer Vision Syndrome.

ESPiM Eye-Strain Probation Model, a model to predict and measure eye-strain.

EyeTAP Eye tracking point-and-select by Targeted Acoustic Pulse, a contact-free multimodal interaction method for point-and-select tasks.

FELiX Fixation-based Eye fatigue Load Index, a model to measure eye-strain.

FQlS Fixation Qualitative Score, a measure that can reveal physical eye fatigue based on distance drift of fixation points.

IDEA Index of Difficulty for Eye tracking Applications, a model to measure eye-strain.

model A set of mathematical equations and logical expressions.

target A graphical user interface element to be selected.

Preface

The evolution of work-related cultures has shaped society and the lifestyles of the working force. The first industrial revolution (1760 to 1840) [123] started a new culture of regular work in the West. Before the industrial revolution, working hours and workplaces were mostly based on the seasonal agricultural environments, with the majority of people working on farmlands or in family-operated businesses[50]. The industrial revolution caused significant lifestyle changes in the greater part of Europe [50], leading to an advancement in innovation and technology, which transformed cities and standards of living.

This transformation not only exploited natural resources and polluted environments irreversibly [49], but it also caused new challenges for workers, including long working days (typically 12 hours a day [71]) in closed, unhealthy spaces dealing with chemicals and extremely loud and dangerous heavy machinery. Furthermore, this was the beginning of women and children entering the workforce [71]. The notable English writer Charles Dickens [135] described in detail child labour and cruel treatment of young workers during this time in his masterpiece Oliver Twist [136].

The second industrial revolution, which started in the 20th century, created more sophisticated materials and led to the creation of more advanced machinery and automation [123]. After the rise of electronic circuits and digital computers around 1960 [134], productivity increased in many areas, and more working skills and qualifications were needed to accomplish tasks in the workplace. These new advancements led to new challenges at work, and finally, in 1938 in the United States [122] a work-week structure declaring a 40-hour work-week, five days a week with eight hours a day was created. The concept of 9-to-5 [124, 85] was the beginning of the standard eight-hour working culture and is now accepted in many countries as the norm [184].

Today, most employees work in offices sitting in front of computers in many parts of the world for eight hours a day. Despite the luxury of modern working environments compared to factories, today's typical workplace has some negative aspects. The routine work atmosphere commonly in closed spaces, i.e. cubicles, dealing with multiple types of devices (PC, smart-phone, telephone) simultaneously (multitasking), causes an increased workload. These new working habits have also changed the social behaviour of workers [48]. Decades ago, work ended at the time the employees left their offices, but nowadays, due to high connectivity, many employees are expected to be accessible by their supervisors, colleagues or clients even on weekends and during vacations [19]. There is an ongoing debate, if the concept of 9-to-5 is still relevant to current working situations, [184, 20].

These circumstances have increased the stress of employees [125]. Although the theoretical working day is still eight hours, in practice, this is not the case and is causing *work-related stress*. In major cities worldwide, long commutes are also inevitable for many workers and most often do not count as working time. However, commuting can also contribute to work stress because this often is wasted time and reduces the spare time for private affairs. According to the statistics published by the University of Oxford [150, 149], the working week was reduced to 40 hours in industrial countries by 1980, but productivity has been increasing rapidly. This inconsistency in the labour-productivity ratio may indicate more workload and stress due to the higher complexity of tasks and demanded skills to accomplish routine tasks.

Chapter 1

Introduction

Today in many parts of the world, most employees work sitting in front of computers for eight hours a day. Despite the luxury of modern working environments compared to factories, today's typical workplace has some negative aspects. The typical work atmosphere commonly in enclosed spaces, i.e. cubicles, dealing with multiple devices (e.g. PC, smartphone, telephone) simultaneously (i.e. multitasking) can lead to an increased workload. Although the so-called open-plan[1] office concept was introduced to reduce the negative impacts of cubicles, this concept has multiple challenges such as higher blood pressure due to excessive ambient noise and a higher flow of airborne germs than closed-space offices [21]. In addition, open workspaces showed lower productivity, and employee morale [185]. According to a poll in the UK, employees spend 1700 hours working in front of a computer display yearly, and 37% of employees believe that these amounts of screen usage cause them headaches[9]. Moreover, scientists have found a correlation between excessive screen time and dry eyes due to a reduction in blink rates, for example, in a large study of office workers in Japan [155]. Computer vision syndrome (CVS), also known as digital eye-strain, is an umbrella term for multiple eye vision problems due to the prolonged use of digital displays,

[1] A large room with no divisions into smaller areas [45].

including tablets and smartphones [148]. These problems have been shown to affect visual comfort as well as work productivity in both adults and teenagers [148]. CVS causes eye and vision symptoms such as *eye-strain*, *eye burn*, *dry eyes*, *double vision*, and *blurred vision* [117]. Furthermore, CVS is a major cause of work-related stress due to repeated eye movements, and excessive eye focus on computer screens [119]. Therefore, in this research, we focus on eye-strain as one of the major CVS symptoms [117] and address this specific component of work-related stress among computer users.

1.1 Work-related Stress

Work-related stress, i.e., the stress caused by one's job, is one of the main challenges of the workforce in the 21st-century [163]. According to the World Health Organization (WHO), [127], work-related stress occurs when the workload demand is higher than the knowledge and abilities of workers to handle. Employees may feel overwhelmed with the amount of work to be accomplished in a limited amount of time and may feel no support to handle their tasks, and thus they feel stressed. According to the American Institute of Stress (AIS), 46% of stress relates to workload [125] and according to Statistics Canada 62% of Canadian workers claimed that work was the main source of stress in their lives [36]. Although stress is a natural phenomenon in human physiology that helps one to cope in dangerous situations [187], permanent stress has several harmful effects such as increasing chance of heart disease [90, 28], hypertension [166], diabetes [53, 98], Alzheimer's [108] and various psychological disorders such as depression [114, 162], anxiety [114, 118], changes in mood [171] and violence at the workplace [40]. Figure 1 illustrates the relation between stress increase and physical/mental activities based on the human functional curve proposed by Peter Nixon in 1979 [137]. The curve shows the struggle towards unhealthy conditions, which ends in a breakdown. The fatigue border (illustrated in a vertical red dashed line) represents the critical state towards distress.

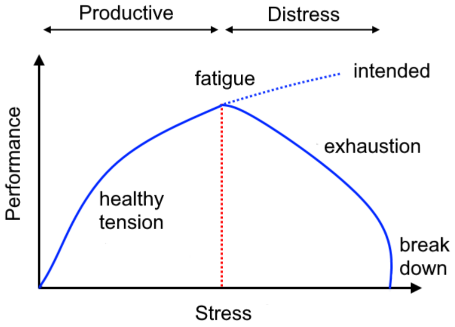


Figure 1: Relationship between stress growth and physical/mental activities. The dashed red line shows the critical boundary between good stress and distress based on the human functional curve proposed by Peter Nixon [137].

In addition, a poor diet such as that of cholesterol-rich processed food and fast food meals contributes to the side effects of work-related stress [67, 147]. Smoking or the consumption of mood and energy-boosting stimulant substances such as alcohol or caffeinated beverages, which have become part of the working culture in many countries, and the lack of physical activities due to automation and digitization of work tasks, are also associated with high work-related stress. However, the extent of these damages varies between male and female workers and depends on the type of work [28]. Due to these large and varied impacts that stress can have on a person and the people around them, stress has been labelled as a silent killer [38]. The severe health risks caused by work-related stress mentioned earlier might be complicated and expensive to treat, if not impossible. Human physiology and psychology are very complex, and the mentioned side effects are only a few examples among unknown issues that can arise. According to current evidence, permanent work-related stress is harmful with many complicated symptoms. Thus, it should be detected and handled in the early phases. Since exposure to work-related stress is inevitable and has become part of our daily lives, there need to be mechanisms to reduce its destructive impacts on our health.

1.2 Eye-strain

According to Vasiljevas *et al.*, fatigue is the increase of tiredness of a subject under load [178] and can be grouped into physical, e.g. lack of sleep, and mental related causes such as stress [66]. According to Marcora *et al.*, mental fatigue is the result of high cognitive activity [109]. Visual fatigue defined as "eyestrain or asthenopia, which can be caused by both two-dimensional and stereoscopic moving images" [78] and which can cause motion sickness [94], occurs when focusing on near objects. The visual function of the eyes may cause visual fatigue, especially in long-time periods.

1.3 Motivation

Given the increased use of digital displays in everyday life, CVS and eye-strain are becoming more and more common among computer users that spend prolonged periods working in front of monitors[178]. My research focuses on developing methods to measure visual fatigue or eye-strain and to provide models that can be used to evaluate user interfaces based on eye-strain in user studies. Thus in the following chapters, we: (1) review the aspects of eye-strain and its correlation with work-related stress, (2) investigate the potentials of eye-tracking as an effective and low-budget eye-strain analysis device, and (3) propose mathematical models to predict and measure eye-strain with and without subjective feedback.

1.4 Objectives and Contributions

The goals of this dissertation are twofold: first, we aim to enable researchers and designers of user interfaces to assess the amount of eye-strain the visual parameters of the interface may cause. Second, we aim to develop eye fatigue models that can be applied to different case scenarios, display types or compare different interaction techniques in user studies to identify and reduce the side effects of visual interactions with digital displays. Specifically, through the course of our research, we have developed three mathematical models, FELiX, IDEA and ESPiM, for predicting and measuring eye-strain (see Table 1).

Model	**Objective**	**Approach**
FELiX	Eye-strain measurement	Fixation points
IDEA	Task difficulty measurement	Fitts' law properties
ESPiM	Eye-strain measurement	Based on Shannon code

Table 1: Summary of trilateral models. FELiX and IDEA are more general measurement models, ESPiM relies solely on objective measures specifically for eye-strain prediction and measurements.

FELiX and IDEA are compound models comprised of subjective ratings integrated into objective measures as descriptive and measurement models with multi-purpose applications. However, the ESPiM model is a stand-alone objective measure applicable for both *prediction* and *measurement* purposes, specifically introduced for eye-strain on digital displays. Together, all three models provide unique tools to predict and measure eye-strain using only a simple eye tracker.

The main contributions of this dissertation can be summarized as follows:

1. Design, development and testing of a contact-free interaction technique for gaze-based interactions (EyeTAP [131]) which addresses the Midas touch problem [83].

2. Design and development of a novel compound multi-factor model to measure eye-strain in user studies based on both subjective and objective measures [129].

3. Design and development of a novel compound multi-factor model to measure task difficulty of eye tracking applications in user studies based on both subjective and objective measures [132].

4. Design and development of a novel objective compound model to measure eye-strain based on spatial and temporal parameters of interaction techniques on digital displays.

1.5 Terminology

The word *target* refers to graphical user interface (GUI) elements that are designed to be selected by clicks or to carry out inputs and outputs that need the user's attention. These GUI elements include commonly-known entities such as buttons, text boxes, or prompt dialogues. In addition, *eye-fatigue* and *eye-strain* are referred to the same concept in this dissertation. The words *mental fatigue* and *mental stress* have semantically different clinical definitions; we refer to them to describe the negative impacts of arousal effort shown in figure 1 in this thesis. Moreover, the word *cognitive workload* refers to the amount of mental effort used to perform a task by a person. The word *model* refers to a set of mathematical equations and logical expressions to describe and measure physical phenomena such as eye-strain based on case scenarios and test circumstances. In addition, the words *display* and *screen* are used interchangeably and refer to digital devices capable of producing visual images in pixels shown to users.

1.6 Organization

The remainder of this thesis is organized as follows: in Chapter 2, we introduce eye-tracking which contains an in-depth overview of the field of eye-tracking containing definitions, methods, applications, challenges, benefits, and guidelines. This chapter can be regarded as a survey for the field of eye-tracking which is necessary for the rest of this dissertation.

Chapter 3 proposes an alternative interaction technique called EyeTAP [131] for the most common challenge of eye-tracking known as the 'Midas touch' problem [83] presented in the earlier chapter on eye-tracking. We encourage readers to read Chapters 2 and 3 together to obtain a better understanding of the proposed interaction technique EyeTAP [131] if interested.

Having the fundamental concepts of eye-tracking and the issues of computer vision syndrome (CVS) discussed earlier, we will focus on the trilateral proposed models FELiX [129], IDEA [132], and ESPiM to mitigate eye-strain in Chapters 4, 5, and 6. Each chapter contains the necessary definitions, methodologies, and conclusions of each eye-strain model and can be studied independently from each other.

We then summarize and conclude the entire work presented in this dissertation in Chapter 7 and propose perspectives for future work.

Chapter 2

Background & Related Work

In the following chapter, we provide background to the physiology and function of the eye, eye tracking technology and different eye-tracking methods that have been used in human-computer interaction.

2.1 Physiology of the Human Eye

First, we review the visual system to better understand its functions and how humans perceive information. The anatomy of the eye can be described as follows [91]:

- Cornea: a transparent layer protecting the *iris* and *pupil*.
- Pupil: a controllable opening to adjust incoming light.
- Lens: the structure responsible for focusing received light on the retina.
- Retina: the layer of tissue at the back of the eye that is sensitive to light and color, constructs the picture of the observed scene.

The act of seeing is regarded as the first step of interacting with objects in the world [96]. Eye tracking technology tracks the eyes using devices that can observe natural eye behaviours such as movements, blink rate, and pupil size to determine where a user is focused on a visual scene [106]. By studying gaze behaviour it is possible to begin to understand the intentions or thinking of users [106].

2.2 Eye Movements

The invention of eye trackers led to the study of eye movements. In general, the most common types of eye movements: *fixations*, *saccades*, and *smooth pursuit*.

2.2.1 Fixations

Fixations are short pauses, in the range of 200 - 600 milliseconds, when the eye is virtually still, and visual input takes place [106]. Due to the physiology of the eye, specifically the decrease of visual acuity away from the center of retina[1], only a tiny portion of the visual field is perceived with high accuracy during a fixation [106]. Therefore, to gain a broader perception of the visual world, the eye moves around a scene changing focus from one point to another. These rapid eye movements between fixations are termed saccades [106]. Fixation points, fixation duration, and saccades can be detected based on the locations of eye gaze per time on the screen.

2.2.2 Saccades

According to Jacob *et al.* [81], saccades take 30 - 120 milliseconds, and once the path of the eye movement begins, it cannot be altered. A rapid rise of tension causes a saccade [174]. Saccadic eye movements are due to both *voluntary* and *sensory* factors [56]. According to Zimmermann *et al.*, saccades reveal an accurate representation of target locations. Furthermore, saccade adaptation (displacement of visual targets)

[1] The primary instrument of vision that receives the image formed by the lens and converts it into chemical and nervous signals which reach the brain by way of the optic nerve [46].

modifies the perception of targets in the brain [192]. In addition, microsaccades (very short in duration and magnitude) also affect the visual system [60]. Van Beers *et al.* investigated the source of variability in saccades [177]. They found that saccades in horizontal-, or vertical-only directions are less variable than diagonal directions. Saccadic parameters such as (1) amplitude, (2) duration time, and (3) peak velocity are all inter-correlated. The source of variability in saccade endpoints is due to the uncertainty in assessing the correct target location [177]. Furthermore, saccades with a duration time higher than average reach a peak velocity value smaller than the average. Figure 2 illustrates eye movement on a screen. Fixations are shown as circles and saccades as direct lines between the fixations. This figure shows the entire movement of a user on the screen for a specific task. Eye movement analysis can thus reveal users' areas of interest (AOI) based on fixations and the sequence of their observations on a visual scene.

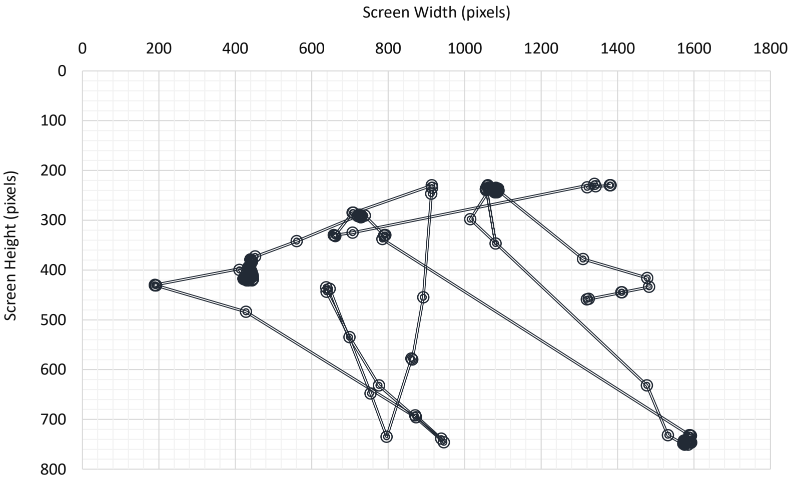


Figure 2: Eye movements on a screen with resolution of 1728 × 972 pixels. Circles represent fixations and lines illustrate the saccades between fixations. Focused areas of interests (i.e. longer fixations) are depicted as darker spots.

2.2.3 Smooth pursuit

A smooth pursuit is a form of eye movement that occurs when a moving stimulus (e.g. an object or animation) is followed with gaze [11] (see figure 3). The response to a moving visual stimulus initiates pursuit in which the *fovea* (responsible for sharp vision [44]) remains focused on the stimulus [99]. However, a moving target evokes both smooth and saccades. Saccades track the moving target, whereas pursuit corresponds to a process to initiate voluntary eye movements [10]. These two eye movements (saccades and smooth pursuit) are distinctive since (1) they are generated by separate neural systems, (2) have different latencies, and (3) react to different aspects of stimuli [99]. This is a major characteristic of primates that allows them to track small moving objects accurately, even across patterned backgrounds [99].

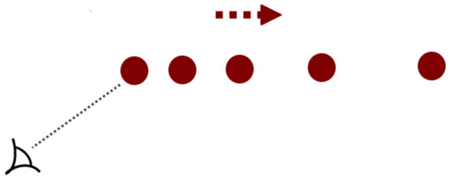


Figure 3: Overview of smooth pursuit.

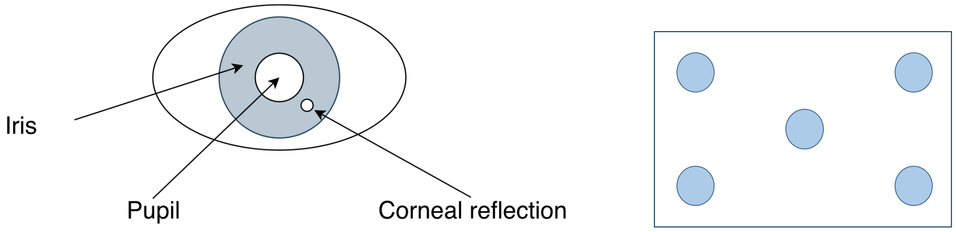
2.3 Eye Tracking

There are different technologies that exist for eye tracking including both head-worn sensors or remote (desktop) sensors. In general, there are three main techniques for eye tracking:

1. Videooculography (VOG): tracking using video cameras in visible light. Depending on the quality of camera the accuracy may vary in different cases. In addition, a dedicated system is required to process the recorded data and detect the gaze point.

2. Video-based infrared (IR): tracking using pupil-corneal reflection (PCR) detected by an infrared light beam. This small spot reflected on the iris offers a reference point to track the eye's gaze point (see figure 4a).

3. Electrooculography (EOG): the eye is an electric field, with a positive pole (cornea) and a negative pole (retina), which can be detected. Therefore, the eye movements can be measured using electrodes on the face. Although this method can work even when blinking, it is invasive and has lower accuracy.



(a) (b)

.

Figure 4: (a) Illustration of corneal reflection effect detected by LED light and infrared camera. The detected corneal reflection spot works as a reference point to track eye movements. (b) A five-point calibration screen. Each circle illustrates a reference point on the screen the user will focus on.

2.3.1 Tracking quality

Tracking quality depends on several properties, including the user's eye (i.e. the physiology of the eye and whether the user is wearing glasses/contact lenses), the tracking environment, and the tracking and calibration methods. The tracking environment refers to the lighting conditions of the environment (natural daylight vs. artificial lighting). The tracking method is the method by which the eye is tracked; examples include corneal reflection method vs. videooculography (tracking eye positions using video cameras in visible light) [31], or electrooculography (tracking eyes using electrodes worn on the face) [31], in addition, the resolution and focus of the tracker's camera are also important. The VOG method usually has lower quality (about 4°) than the IR method with 0.5°of visual angle. The EOG technique has the highest accuracy and therefore has medical applications; it is used to study diseases of the eye [106]. One visual degree corresponds to about 43 pixels on a 1920 × 1080 display with 70 cm distance to the display [167].

2.3.2 Calibration

Calibration is the process of mapping measurement points (reference points) to the user's eye positions (orientation) [145]. This procedure is essential for eye-tracking applications to improve accuracy. Many different methods exist, including five, seven or nine-point calibrations which record the user's eye position on some reference points on the screen. Figure 4b illustrates an overview of a five-point calibration procedure.

2.4 Application Domains

There are several research, industry and personal application areas with high potential use cases for eye tracking. The main benefits of eye-tracking over other interaction methods are related to its hands-free characteristics. Below we list some example domains where eye-tracking has been used or studied.

1. Basic interactions for disabled users: based on some physical disabilities, some users may not be able to interact with a mouse or keyboard, eye tracking may be used for interaction.

2. Medical environments: in surgical operation environments with high demand of sterility and low physical contacts to equipment.

3. Aviation and aerospace: in some environments, both hands may be occupied with different joysticks and controllers. In addition, high concentration and

interaction are needed quickly, and users may observe visual commands and trigger them without touching as they observe the scene.

4. Automotive: two areas in the automotive industry where eye tracking has been shown to be useful are:

    (a) Unobtrusive driver observation: useful for safety reasons by observing the level of consciousness of a driver in real-time to avoid accidents.

    (b) Interaction with car applications: the interface can be shown on the windshield glass where the commands can be triggered. In this way, the attention remains on the front screen and does not distract the driver from the road.

5. Remote collaboration: working on a shared visual object to improve coordination [35] or cooperation between a local worker and remote collaborator [68].

6. Augmented and virtual reality (AR/VR): optimizing the calibration of AR glasses and using eye gaze as an additional input modality [146].

7. Gaming: sharing eye gaze and visual attention in collaborative gameplay between several players [164].

8. Human identification (soft biometric): Cantoni *et al.* proposed an analysis technique (GANT) for human identification based on obtained fixation points during a user study [23]. The way of looking at an image can reveal a person's identification by comparing areas of interest (AOIs) to image landmarks. The comparison is based on created graphs based on fixation density[2], duration time and created graphs based on nodes connections on an image.

[2] Number of fixation points in a specific area [23].

2.5 Challenges of Eye Tracking

Eye tracking, like many emerging technologies, has its challenges. Before reviewing the benefits and the importance of eye-tracking, we first discuss its shortcomings.

1. Low accuracy: the accuracy of an eye-tracking cursor/pointer is typically lower than a mouse. It is also subjective to users and depends on the hardware and software of the eye-tracking sensor.

2. Midas touch problem: unintended activation of functions by eye gaze to hit a target accidentally.

3. Difficulty to control: it is difficult to move a cursor/pointer to a specific location as users tend to do with a mouse because of the natural micro-movements of the eye.

4. Device-dependency: the performance of eye-tracking applications is highly dependent on the quality of tracking sensors.

5. Application-dependency: the interaction with an eye tracker requires a specific user interface; therefore, it is not applicable for every application.

6. Eye fatigue: interacting with eye tracking applications leads to eye fatigue since the eye is responsible for observing information and sending commands simultaneously.

7. Configuration and calibration: an eye-tracking application needs prior setup to run. In addition, every single user of a system completes a short calibration process.

8. Training required: users need to learn how to interact with an eye-tracking application. These will be dependent on the type of interaction required (dwell-time or secondary input modality).

Midas Touch

One of the main challenges of gaze-based interactions is distinguishing normal eye function from a deliberate interaction with the computer system, commonly referred to as 'Midas touch'. The Midas touch problem occurs when a user accidentally activates a computer command when the intention is to look around and perceive the scene. According to Jacob [79] this problem occurs because eye movements are natural, i.e. the eyes are used to look around an object and to scan a scene, often without any intention to activate a command or function. This phenomenon is one of the significant challenges in eye interactions, and diverse methods have been proposed to reduce this effect.

2.6 Eye Tracking Methods Addressing Midas Touch

In the following section, we describe different techniques that have been used to address the Midas touch problem. These solutions can be categorized into four groups according to the interaction technique they employ: (a) dwell-time processing, (b) smooth pursuits, (c) gaze gestures, and (d) multimodal interaction. Below, we describe each of these solutions and provide example use-cases.

2.6.1 Dwell-time processing

Dwell-time is the amount of time that the eye gaze must remain on a specific target in order to trigger an event. Researchers have tried to detect specific thresholds to handle the Midas touch problem [142, 179]. For example, Pi *et al.* proposed a probabilistic model for text entry using eye gaze [142]. They reduced the Midas touch problem by assigning each letter a probability value based on the previously chosen letter such that a letter with lower probability requires a longer activation time to be activated and vice-versa. Velichkovsky *et al.* applied focal fixations to resolve the Midas touch problem by assigning the mean duration time (empirically set to 325 ms) of a visual search task to trigger a function [179]. Dwell time is even faster than the mouse in certain tasks, e.g. selecting a letter given an auditory cue [160]. The method of applying focal fixations may be very subjective since searching time varies across users when applying the dwell-time technique [12]. Moreover, increasing the threshold may increase the duration time of the entire interaction. Conversely, reducing the amount of dwell time may lead to more errors for some users [181]. Pfeuffer *et al.* investigated visual attention shifts in 3D environments for menu selection tasks [141]. They compared three interaction techniques for menu selection: (1) dwell-time (activation threshold of 1 sec.), (2) gaze button (applying eye gaze to point, selecting by a button press), and (3) cursor (applying eye gaze to point to a context, precise movement and selected by a manual controller). They found that the dwell-time technique was the fastest in terms of performance. In addition, the cursor technique was found to be the most physically demanding technique. They also found that dwell-time was considered to be the easiest method according to users. However, the gaze button and the dwell-time caused the highest eye fatigue.

2.6.2 Smooth pursuits

Smooth pursuits are a form of eye movement that occurs when a moving stimulus (e.g. an object or animation) is followed with gaze [11]. The method is typically implemented by using a visual point on the interface, then to activate the target, the user must fixate on one of these points. This technique has been used to select targets [182], control home appliances [180], activate functions such as mouse clicks [156] or to use the music player on a smartwatch (Orbits) [52]. Schenk *et al.* proposed a framework (GazeEverywhere) that enables users to replace mouse inputs [156]. This solution includes a computer to process gaze interactions (gaze PC), a computer to show the results (unmodified PC) connected via a micro-controller to trigger mouse click events, and a glass pane to project gaze targets a second screen. Vidal *et al.* introduced an interaction technique (Pursuits) for large screens using moving objects to be activated by eye gaze [182]. They used a desktop eye tracker and a public display to select targets on the screen. Velloso *et al.* presented a framework (AmbiGaze) to control ambient devices such as TVs and stereos (each assigned with an infrared (IR) beacon) with eye gaze using a head-mounted eye tracker [180]. The system employs a server to process gaze inputs and control the devices. Esteves *et al.* presented a framework for a multi-touch Android smartwatch to input commands using a head-mounted eye tracker [52]. They developed three use-cases: a music player, a notifications panel with six coloured points on the smartwatch screen representing six applications (e.g. social media apps), a missed call menu with four commands, call back, reply text, save the number and clear the notification.

Smooth pursuit gaze-based interaction has several drawbacks. First, it requires a moving stimulus [80] and, therefore, it requires implementing an additional graphical user interface (GUI) to handle the events. Second, this kind of point-and-select may slow down the interaction due to the pursuit time, adding latency to target selection completion time. In addition, the presence of moving paths on a limited screen size may limit users to a restricted set of functions. Third, this type of interface may lead to visual distraction on the screen and may not be suitable for long working sessions or users with disabilities. Moving objects require free space on a screen, therefore, dependent on the screen size. Thus, although smooth pursuits is a promising method for public and large digital displays, it is not ideal for everyday interaction.

Schenk *et al.* proposed a novel interaction technique, Smooth Pursuit Oculomotor Control Kit (SPOCK), to resolve the Midas touch problem [158]. This prototype can be regarded as an improved update of the *antisaccades* technique which applies a similar mechanism to activate a stimulus by focusing on a target for a predefined amount of time [72]. SPOCK employs smooth pursuit eye movements for button-based interfaces activated by eye gaze. Two small disc-shaped objects appear at the center of a target (button), and when a user focuses on the target, the objects slowly and simultaneously move towards the top and bottom of the target. This mechanism enables the user to follow one of the discs to select the target if the selection of the target was intended. If the user does not react to the discs, the cycle is repeated from the center of the target, as long as the user's gaze point is on the target (see figure 5).

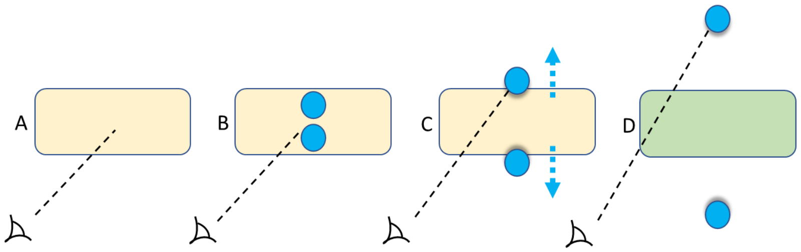


Figure 5: Illustration of the SPOCK interaction technique by smooth pursuits. User looks at a target (A), two similar discs appear after a predefined time (B), discs start to move smoothly in opposite directions (C), following each disc by gaze activates the target after a specific time (D).

Schenk *et al.* [158] claim that the application of two stimuli (symmetric design), compared to the antisaccades method, reduces involuntary eye movements. They conducted a two-part user study with 18 participants to compare SPOCK with the dwell-time processing method. The first part of the study contained a 3 × 3 square of targets to be selected, and the second part was based on a series of multiple-choice questions. Both interaction techniques were compared based on their performance on (1) failed attempts and (2) completion time. The SPOCK method showed lower failures and relatively higher completion time due to a slow selection mechanism.

2.6.3 Gaze gestures

Gaze gestures are sequences of eye movements that follow a predefined pattern in a specific order [47]. Researchers have proposed techniques that can be applied to analyze eye movements to detect unique gestures (e.g. [7, 47, 73, 77]). Drewes *et al.* assigned up, down, left, right and diagonal directions to different characters on the keyboard, thereby allowing a user to select a letter by moving the eye gaze in any direction [47]. In addition, they tried to distinguish between natural and intentional eye movements by using short fixation times during gesture detection and long fixation times to reset the gesture recognition. Istance *et al.* developed two-legged and three-legged gaze gestures (up, down and diagonal patterns) for command selection to play World of Warcraft for users with motor impairment disabilities [77]. In a similar work, Hyrskykari *et al.* studied both dwell-time and gaze gesture interactions in the context of video games and found that gaze gestures had better performance for command activation [73]. Moreover, gaze gestures produced fewer errors than dwell time and led to fewer visual distractions. Bâce *et al.* proposed an AR prototype, containing a head-mounted eye tracker and a smartwatch, to embed virtual messages to real-world objects to be shared with peer users [7]. The authors integrated eye gaze gestures as a pattern to encode and decode messages attached to a specific object previously tagged by another peer user, thus using gaze gestures as an authentication mechanism for secure communication. In general, gaze gestures have shown promising performance to address the Midas touch problem.

As gaze gesture techniques rely only on performing specific eye movements, they may lead to eye fatigue in a long working session as longer eye inputs are correlated with eye fatigue [141]. In addition, the detection algorithms may reduce the speed of interaction, and the limited amount of possible eye gestures may reduce the number of functions available to users. Further, applying gaze gesture commands requires a guiding system since users need to map commands with their corresponding gestures [37]. Learning the correct gestures may also be challenging and requires training for novice users [37]. Therefore, this kind of interaction solution may not be appropriate for users who must use a system over a long period or for users with disabilities. Figure 6 shows an example of a gaze gesture to trigger an action.

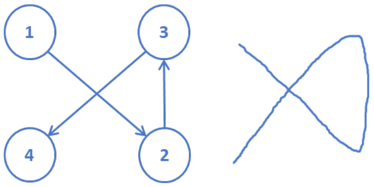


Figure 6: Illustration of a gaze gesture to initiate a command.

2.6.4 Multimodal Interaction

Multimodal techniques apply extra inputs from another modality (e.g. touch, audio, etc.) as the trigger of a function in addition to eye-tracking. They can be divided into the following sub-categories: mechanical switches, touch interaction, or facial gestures.

Applying a specific (mechanical) switch

Researchers have applied specific switches to activate an event or function for specific domains, such as rehabilitation and user groups (i.e. users with motor impairments or severe disabilities). For instance, Rajanna *et al.* [144] proposed a combined framework for users with disabilities that applies a foot pedal device to click on objects and to enter text. Meena *et al.* [112] applied a soft button on a wheelchair to control the movements of the wheelchair in different directions (horizontal, vertical and diagonal). Sidorakis *et al.* [161] applied a switch for a gazed-controlled multimedia framework on virtual reality head-mounted displays (Occulus Rift) to resolve the Midas touch problem. Biswas *et al.* [16] proposed a joystick to control point-and-select tasks for combat aviation platforms to address the Midas touch problem.

Touch interaction

Some researchers have proposed using touch interaction for a limited number of functions to increase the accuracy of target selection. Pfeuffer *et al.* [139] applied a cursor at the gaze point to be controlled by a finger holding a tablet where a finger tap on the screen leads to a click on the current location of the pointer (CursorShift method). In a similar study by Pfeuffer *et al.* [138], the authors investigated the integration of finger touch and pen inputs on a tablet for zooming or annotating tasks on images. Although this technique was not introduced as a solution to the Midas touch problem, it can increase the accuracy of selection, which reduces the Midas touch problem. However, this technique is not hands-free, and the application scenario is limited to tablet devices only.

Facial gestures recognition

Rozado *et al.* studied the potential of using live video monitoring to detect facial gestures to enhance eye tracking interaction [151]. Their work (FaceSwitch) associated facial gestures (opening mouth, raising eyebrows, smiling and twitching the nose up and down) with simulating left and right mouse clicks and customized some keyboard functions such as page down keypress. They found that increasing the number of gestures leads to lower recognition accuracy when monitored simultaneously.

Facial gesture recognition has several drawbacks. First, real-time video monitoring to detect the correct face gesture is very challenging beyond controlled lab conditions to address the real-life scenarios [110]. In addition, any emotional change or unwanted facial behaviour may lead to false activation of functions since modelling the human behaviour is challenging [110]. Another drawback is the latency between pointing using the eye tracker and selecting the facial gesture algorithm; precise timing is required for smooth interactions. Moreover, modelling of facial expressions requires a wide range of visual signal processing methods [110].

Eye gaze and head movements Stellmach *et al.* proposed multimodal techniques to interact with distant targets in which they studied combinations of gaze and head movements joint with a smartphone touch modality for precise selection, and manipulations [170]. Kytö *et al.* proposed similar techniques for AR headsets. They investigated head movements and eye gaze movements with a variety of combinations, including selection on device and hand gesture commands, and found the highest error rates and lowest completion time for the eye only selection technique [95].

Gaze and speech interaction Besides the above-related works aimed at addressing the Midas touch problem, multimodal interaction has also considered gaze and voice commands. Mayer *et al.* proposed an interaction technique (WorldGaze) to track users' fields of view and gaze point to refine the voice command engines on smartphones for more precise results [111]. Beelders *et al.* studied word processing tasks using voice commands and eye gaze compared with mouse and keyboard interactions in their work [13]. However, although they showed that speech interaction is feasible for word applications, the gaze and speech interaction technique could not reach the effectiveness and performance of keyboard interaction. Acartürk *et al.* reviewed the challenges and possibilities of gaze and speech modalities for elderly users in their work [2]. Esteves *et al.* conducted comparative studies using head-mounted displays (HMDs) to investigate the performance of hands-on and hands-free (including gaze and speech) interaction techniques and found that applying a clicker and dwell-time were the most favourable interaction techniques [51].

Miniotas *et al.* proposed a technique for selecting closely spaced targets based on speech commands [116]. They applied a grid of 5 × 5 squares as a stimulus to test two interaction techniques: (a) gaze and speech, and (b) gaze only. They suggested a dwell-time of 1500 ms for targets of the size of 30 × 30 pixels with a distance of 10 pixels for the best performing setup for target selections based on their results. However, they reported a slow performance in the case of selection speed when the activation threshold for the dwell-time increased.

Beelders *et al.* conducted an experiment to study eye gaze, and speech commands compared to the mouse for target selection tasks [14]. They applied a stimulus as the shape of a circle with 800 pixels diameter containing 16 squares on its edge to be selected in all directions. They found that the mouse had a significantly higher performance in the case of throughput and completion time and stated that using the dwell-time technique should be more efficient than speech commands. Sengupta *et al.* integrated gaze and voice inputs for web browsing tasks such as search, navigation, and bookmark of pages [159]. They found that the multimodal approach had a higher performance than each modality alone.

Zhao *et al.* proposed a multimodal technique of eye gaze by smooth pursuits and speech commands and found promising results when compared to mouse clicks [190]. They found that the selection of a word for confirmation should match the task for better performance. Further, participants who chose the activation word scored higher compared to those who used a pre-determined word. Similar to the EyeTAP method, the authors also suggested applications of other sound inputs such as pseudowords or exclamation for users with severe disabilities.

Gaze and hand gesture interaction Gaze has also been combined with hand gesture inputs. For example, Chatterjee *et al.* proposed an interaction technique that uses gaze and hand gestures to select targets at the most desired location on the screen [29]. They found that the combination of gaze and hand gesture outperformed each interaction modality alone. Pfeuffer *et al.* proposed a similar approach of applying eye gaze and a hand pinch to select and manipulate targets in a 3D space for virtual reality (VR) platforms [140]. Hand-gesture interactions are prone to muscular fatigue [70] and therefore may challenge users in certain circumstances.

Gaze and button press Hild *et al.* investigated multimodal gaze-based interactions: gaze and button press by hand, gaze and button press by foot, and the mouse input [69]. They found overall faster performance for gaze-based techniques than the mouse for task completion time. Kumar *et al.* proposed a technique (EyePoint) comprised of eye gaze and button press on the keyboard to improve the accuracy of gaze-based pointing in a Look-Press-Look-Release pattern of commands [93]. The EyePoint technique was designed in four steps to select a target accurately. The user looks at the desired target (Look), then presses and holds a hotkey on the keyboard, magnifying the specific spot on the screen (Press). A second look at the magnified scene is then done to refine the target's exact location (Look), then the key is released to select that target (Release). Gaze and button techniques have shown promising results in improving selection accuracy.

Gaze gesture recognition Istance *et al.* proposed a technique (Snap Clutch) to resolve the Midas touch problem [76]. They applied a disengagement technique to turn off gaze selections when not needed by defining four modes provided in the screen's up, left, right, and down directions. These modes are activated when looking at different directions (eye gesture), and visual feedback appears on the screen to confirm the intention.

2.7 Benefits of Eye Tracking

Although eye-tracking has several limitations, as described above, it offers high potential in both research fields and commercial applications. Furthermore, remote interaction offers the great potential of collaboration on shared objects when users are at a distance. There are several different areas where eye tracking has been applied for remote interaction. For example, collaborative virtual medical environments have been used in surgery and medical training[32]. For example, Black *et al.* studied the potentials of eye-tracking for a sterile operating room by offering hands-free interaction using eye-tracking, and audio feedback [17]. Human-robot interaction has also become a field of interest for enterprises to improve productivity, safety, and quality; for instance, eye tracking can replace conventional mouse or joysticks to send commands to a remote device in some dangerous and difficult situations. Yu *et al.* [188] studied the potentials of eye tracking in such cases. Interaction with large displays has also been studied and showed the benefits of eye-tracking. For example, a research study [88] showed that pedestrians could interact with a large display without touching any controller. The system was composed of a body tracker (Microsoft Kinect One), a rail system, a remote eye tracker (installed on the rail system) and a large display. The system could be activated when a person is detected, and their eyes are in the range of the eye tracker. The eye tracker could be moved horizontally as the user moves across the X axis. This kind of interaction (walk then interact) is useful in various scenarios, such as interacting with smart displays to obtain information for tourists and visitors to a public place without physical contact. In addition, eye tracking can be applied for analytic observations. As mentioned earlier, due to the raw data from an eye trackers' software, it is possible to detect users' areas of interest by analyzing the fixation points. Further, the saccade path reveals the user behaviour in terms of observing and perceiving information on a screen (see figure 2). Perhaps the most interesting aspect in human-computer interaction is the ability to avoid physical interaction with the computer, i.e. *hands-free* or *touchless* interaction. Eye-tracking can be applied to move a cursor on the screen; therefore, it provides an easy technique to interact with a computer with or without second modalities for selection or click actions. Furthermore, the reviewed prototypes for remote interaction [32, 188, 88] belong to the hands-free group as well.

2.8 Guidelines for Eye Tracking Applications

Feit *et al.* conducted a comprehensive study to investigate appropriate user interfaces, in the case of screen regions and target size, based on the accuracy and precision of eye-tracking sensors [54]. They measured the tracking quality in a user study with 80 participants in two lighting conditions (daylight, artificial light) and used two different eye trackers. The user study was designed such that a subject had to select 30 targets that were randomly distributed across the screen; the task was to look at each target for two seconds. The study measured both accuracies, i.e. the estimated distance to the actual gaze point and precision, and the standard deviation over all target fixations.

In addition, five filters were tested to improve the precision of tracking and to reduce errors. The filters include *stampe filter* [168], *weighted average (WA)* [86], *saccade detection*, *outlier correction* and *1€ filter* [25]. The results showed that the weighted average and saccade detection filters could improve tracking quality. Feit *et al.* conclude their study by proposing important factors regarding (1) target size and position on the screen, (2) applying menus, and (3) tracking quality. It is important to enlarge the size regarding the target size; however, this option is restricted to limited screen space. In addition, accuracy is worse in the Y axis, and the implication is that targets should have a larger height than width. Moreover, the precision of each target is worse at the right bottom border of the screen. Applying hierarchical menus for interaction may decrease performance and increase interaction time.

It is possible to zoom on targets by activation. However, this technique may increase visual distraction and use gaze gestures for selection that are hard to learn and may not be appropriate for all users. Moreover, it requires large saccades and may lead to eye fatigue. Using smooth pursuit (following moving objects on the screen by gaze) may lead to eye fatigue or increased visual distraction. No correlation was found between tracking quality and duration of recording for the same participant. Furthermore, accuracy and precision may decrease over time because of movement or eye fatigue. Thus applying filters can improve tracking quality. In general, it is important to consider (1) targets should have a greater height than width, (2) applying filters to improve tracking quality, (3) avoid hierarchical menus and (4) avoid placing targets in the right and bottom regions of the screen.

Chapter 3

EyeTAP: Introducing a Multimodal Gaze-based Technique using Voice Inputs with a Comparative Analysis of Selection Techniques

Preface

In the previous chapter, we reviewed a wide range of techniques that can be applied with good accuracy and are suitable for specific domains with specific peripherals or extra user interface designs. The need for contact-free gaze-based interactions is necessary to deal with the emerging requirements regarding hygiene interactions from a safe distance. Building on the promising results found for multimodal techniques, and specifically exploring the use of non-speech sounds to allow for a more diverse population of users, we developed EyeTAP [131]. EyeTAP can be applied to fill the gap for both able-bodied and disabled users with or without physical contact (to the microphone), with no need for specific user interface design or peripherals and using the simplicity of the Morse code to encode/decode input signals. EyeTAP uses a multimodal solution that combines eye-gaze with acoustic inputs (audio or speech detection) can be regarded as an alternative to the reviewed literature on multimodal interaction methods and has the advantage of not requiring additional hardware (in comparison to other gaze-based techniques) other than an eye tracker or a specialized user interface design. The EyeTAP paper was published in the International Journal of Human-Computer Studies.

Abstract

One of the main challenges of gaze-based interactions is the ability to distinguish normal eye function from a deliberate interaction with the computer system, commonly referred to as 'Midas touch'. In this paper we propose EyeTAP (Eye tracking point-and-select by Targeted Acoustic Pulse) a contact-free multimodal interaction method for point-and-select tasks. We evaluated the prototype in four user studies with 33 participants and found that EyeTAP is applicable in the presence of ambient noise, results in a faster movement time, and faster task completion time, and has a lower cognitive workload than voice recognition. In addition, although EyeTAP did not generally outperform the dwell-time method, it did have a lower error rate than the dwell-time in one of our experiments. Our study shows that EyeTAP would be useful for users for whom physical movements are restricted or not possible due to a disability or in scenarios where contact-free interactions are necessary. Furthermore, EyeTAP has no specific requirements in terms of user interface design and therefore it can be easily integrated into existing systems.

3.1 Introduction

In gaze-based interaction eye tracking sensors measure a user's gaze position on a computer screen and differing methods (e.g. dwell time, multimodal interaction, etc.) are employed to allow the user to interact with the system. Gaze-based interaction offers a suitable alternative to conventional input devices (i.e. keyboard and mouse) in several different scenarios including for users for whom manual interaction might be difficult or impossible, or in situations where contact-free interaction is required. However, gaze-based interaction has well-known challenges among which is *Midas touch*, where a system cannot distinguish the basic function of the eye (i.e. looking and perceiving) from deliberate interaction with the system. In this paper, we propose EyeTAP (Eye tracking point-and-select by Targeted Acoustic Pulse), a multimodal gaze-based interaction approach that addresses the Midas touch problem by integrating the user's gaze to control the mouse with audio input captured using a microphone to trigger button-press events for real-time interaction.

Traditionally, pointing and clicking is done with a mouse; a user uses a mouse to move a cursor to a target (pointing phase), and clicks on the mouse to select or trigger a function (selection phase). We designed EyeTAP as a multimodal method point and click interaction method that uses eye gaze for pointing and auditory input for selection. Specifically, with EyeTAP the mouse pointer position is captured using an eye tracker and selection is done by generating an acoustic signal (e.g. a tongue click, microphone tap, verbal command), which in our studies was captured by a headset microphone. Our solution thus provides a contact-free interaction method for users (including those with special needs) and addresses the Midas touch problem. EyeTAP provides contact free interactions in case scenarios where the use of speech commands are not possible, e.g. due to reasons such as difficulty of word detection by user's language, accent, or pronunciations of words; or for users with severe disabilities not capable of speaking or interacting with keyboard and mouse. Figure 7 illustrates the overview of EyeTAP.

In comparison to gaze-based multimodal interactions which use gestures, foot pedals, or buttons, using speech/sound enables contact-free interactions and supports users to point and select a target based on two separate modalities by simply using a microphone. This allows for a smooth and simple-to-use interaction technique that does not require extensive equipment or training. In addition, using sound input does not require users to shift their gaze focus (e.g. to a button or other hardware device) to trigger a function.

EyeTAP's ability to use different modes of interaction for selection, such as a mouth click or a microphone tap, overcomes the limitations of natural language processing methods and is applicable when speech commands are not feasible (e.g. due to disabilities or due to the surrounding environment). We showed that EyeTAP can be an alternative to using speech with no need for voice recognition engines independent from users' language or accent.

We performed four extensive user studies comparing EyeTAP to dwell-time, eye tracking with voice recognition, and mouse interaction for point-and-click tasks. The analysis of the results showed that although EyeTAP had comparable performance with other gaze-based interaction techniques, it did not outperform the dwell-time method on most criteria. At the same time, EyeTAP generally performed better than gaze-based interaction with voice recognition selection and thus might be suitable in cases where users cannot use voice commands, have restricted physical movement, or where manual interaction with an input device is not possible, e.g. medical practitioner having both hands busy or in a situation where physical contact with equipment should be avoided.

The contributions of this paper are twofold. First, we have designed and developed a simple-to-use, multimodal gaze-based interaction technique. The proposed approach allows for a completely hands-free interaction solution between the user and the computer system using only an eye-tracker and an audio input device. Second, we present four user studies comparing EyeTAP with two other widely-used gaze-based interaction techniques and the mouse.

3.2 Related Work

In this section, we provide an extensive literature review of gaze-based interaction techniques addressing the Midas touch problem. Although, some studies are not directly related to our proposed method, we were inspired by their intuitions and the approaches provided a broad view of both hands-on and hands-free multimodal gaze-based interaction techniques.

In eye-based interaction, the Midas touch problem occurs when a user accidentally activates a computer command using gaze when the intention was simply to look around and perceive the scene. According to Jacob [83], this problem occurs because eye movements are natural, i.e. the eyes are used to look around an object or to scan a scene, often without any intention to activate a command or function. This phenomenon is one of the major challenges in eye interaction techniques [82, 76], and diverse methods have been proposed to address the Midas touch problem. The solutions can be categorized into four groups according to the interaction technique they employ: (a) dwell-time processing, (b) smooth pursuits, (c) gaze gestures, and (d) multimodal interaction. Below, we describe each of these solutions and provide example use-cases, as well as describe their shortcomings or relationship to our work.

3.2.1 Dwell-time processing

Dwell-time is the amount of time that the eye gaze must remain on a specific target in order to trigger an event. Researchers have tried to detect specific thresholds to handle the Midas touch problem [142, 179]. For example, Pi *et al.* proposed a probabilistic model for text entry using eye gaze [142]. They reduced the Midas touch problem by assigning each letter a probability value based on the previously chosen letter such that a letter with lower probability requires a longer activation time to be activated and vice-versa. Velichkovsky *et al.* applied focal fixations to resolve the Midas touch problem by assigning the mean duration time (empirically set to 325 ms) of a visual search task to trigger a function [179]. Dwell time has been shown to be even faster than the mouse in certain tasks, e.g. selecting a letter given an auditory cue [160].

The method of applying focal fixations may be very subjective since searching time varies across users when applying the dwell-time technique [12]. Moreover, increasing the threshold may increase the duration time of the entire interaction. Conversely, reducing the amount of dwell-time may lead to more errors for some users [181]. Pfeuffer *et al.* investigated visual attention shifts in 3D environments for menu selection tasks [141]. They compared three interaction techniques for menu selection: (1) dwell-time (activation threshold of 1 sec.), (2) gaze button (applying eye gaze to point, selecting by a button press), and (3) cursor (applying eye gaze to point to a context, precise movement and selecting by a manual controller). They found that the dwell-time technique was the fastest in case of performance. In addition, the cursor technique was found to be the most physically demanding technique. They also found that dwell-time was considered to be the easiest method according to users. However, the gaze button and the dwell-time caused the highest eye fatigue.

Although dwell-time has been found to be the fastest technique among eye tracking techniques, some studies [181, 183, 107] show that it is error prone particularly in situations when a lower dwell-time is used. However, longer dwell times may cause eye discomfort or fatigue [141]. For this reason, we decided to turn towards multimodal techniques to address the Midas touch problem.

3.2.2 Smooth pursuits

Smooth pursuits are a form of eye movement that occurs when a moving stimulus (e.g. an object or animation) is followed with gaze [11]. The method is typically implemented by using a visual point on the interface, then to activate the target the user must fixate on one of these points. This technique has been used to select targets [182], control home appliances [180], to activate functions such as mouse clicks [156] or to use the music player on a smartwatch (Orbits) [52]. Schenk *et* *al.* proposed a framework (GazeEverywhere) which enables users to replace mouse inputs [156]. This solution includes a computer to process gaze interactions (gaze PC), a computer to show the results (unmodified PC) which are connected via a microcontroller to trigger mouse click events, and a glass pane to project gaze targets on a second screen. Vidal *et al.* introduced an interaction technique (Pursuits) for large screens using moving objects to be activated by eye gaze [182]. They used a desktop eye tracker and a public display to select targets on the screen. Velloso *et al.* presented a framework (AmbiGaze) to control ambient devices such as TVs and stereos (each assigned with an infrared (IR) beacon) with eye gaze using a head-mounted eye tracker [180]. The system employs a server to process gaze inputs and control the devices. Esteves *et al.* presented a framework for a multi-touch Android smartwatch to input commands using a head-mounted eye tracker [52]. They developed three use-cases: a music player, a notifications panel with six colored points on the smartwatch screen representing six applications (e.g. social media apps), and a missed call menu with four commands, call back, reply text, save number and clear the notification.

Smooth pursuit gaze-based interaction has several drawbacks. First, it requires a moving stimulus [80] and therefore, it requires implementing an additional graphical user interface (GUI) to handle the events. Second, this kind of point-and-select may slow down the interaction due to the pursuit time which can add latency to target selection completion time. In addition, the presence of moving paths on a limited screen size may limit users to a restricted set of functions. Third, this type of interface may lead to visual distraction on the screen and may not be suitable for long working sessions or for users with disabilities; in fact, moving objects require free space on a screen which is therefore dependent on the screen size. Thus, although smooth pursuits is a promising method for public and large digital displays, it is not an ideal method for everyday interaction.

3.2.3 Gaze gestures

Gaze gestures are sequences of eye movements that follow a predefined pattern in a specific order [47]. Researchers have proposed techniques which can be applied to analyze eye movements to detect unique gestures (e.g. [7, 47, 73, 77]). Drewes *et al.* assigned up, down, left, right and diagonal directions to different characters on the keyboard thereby allowing a user to select a letter by moving the eye gaze in any direction [47]. In addition, they tried to distinguish between natural and intentional eye movements by using short fixation times during gesture detection and long fixation times to reset the gesture recognition. Istance *et al.* developed two-legged and three-legged gaze gestures (up, down and diagonal patterns) for command selection to play World of Warcraft for users with motor impairment disabilities [77]. In a similar work, Hyrskykari *et al.* studied both dwell-time and gaze gesture interactions in the context of video games and found that gaze gestures had better performance for command activation [73]. Moreover, gaze gestures produced fewer errors than the dwell-time and led to less visual distractions. Bâce *et al.* proposed an AR prototype, containing a head-mounted eye tracker and a smartwatch, to embed virtual messages to real-world objects to be shared with peer users [7]. The authors integrated eye gaze gestures as a pattern to encode and decode messages attached to a specific object previously tagged by another peer user, thus using gaze gestures as an authentication mechanism for secure communication. In general, gaze gestures have shown promising performance to address the Midas touch problem.

As gaze gesture techniques rely only on performing specific sequences of eye movements, they may lead to eye fatigue in a long working session as longer eye inputs are correlated with eye fatigue [141]. In addition, the detection algorithms may reduce the speed of interaction and the limited amount of possible eye gestures may reduce the number of functions available to users. Further, applying gaze gesture commands requires a guiding system since users need to map commands with their corresponding gestures [37]. Learning the correct gestures may also be challenging and requires training for novice users [37]. This kind of interaction solution, therefore, may not be appropriate for users who must use a system over a long period of time or for users with disabilities.

3.2.4 Multimodal Interaction

Multimodal techniques apply extra inputs from another modality (e.g. touch, audio, etc.) as the trigger of a function in addition to eye tracking. They can be divided into the following sub-categories: using mechanical switches, touch interaction, head movements, facial gestures, hand gestures, and gaze gestures.

Applying a specific (mechanical) switch

For certain specific domains, such as rehabilitation, and user groups (i.e. users with motor impairments or severe disabilities), researchers have used mechanical switches to activate an event or function. For instance, Rajanna *et al.* proposed a combined framework for users with disabilities which applies a foot pedal device to click on objects and to enter text [144]. Meena *et al.* applied a soft button on a wheelchair to control the movements of the wheelchair in different directions (horizontal, vertical and diagonal) [112]. Sidorakis *et al.* applied a switch for a gazed-controlled multimedia framework on virtual reality head-mounted displays (Oculus Rift) to resolve the Midas touch problem [161]. Biswas *et al.* proposed a joystick to control point-and-select tasks for combat aviation platforms to address the Midas touch problem [16].

Touch interaction

Some researchers have proposed the integration of using touch interaction, for a limited number of functions, to increase the accuracy of target selection. Pfeuffer *et al.* applied a cursor at the gaze point to be controlled by a finger holding a tablet where a finger tap on the screen leads to a click on the current location of the pointer (CursorShift method) [139]. In a similar study by Pfeuffer *et al.*, the authors investigated the integration of finger touch and pen inputs on a tablet for zooming or annotating tasks on images [138]. Although this technique was not introduced as a solution to the Midas touch problem, it can increase the accuracy of selection which leads to reducing Midas touch. Stellmach *et al.* proposed an interaction technique to select targets on a remote screen via eye gaze and a handheld touchscreen device [169].

Eye gaze and head movements

Stellmach *et al.* proposed multimodal techniques to interact with distant targets in which they studied combinations of gaze and head movements joint with a smartphone touch modality for precise selection and manipulations [170]. Kytö *et al.* proposed similar techniques for AR headsets. They investigated head movements and eye gaze movements with a variety of combinations including selection on device and hand gesture commands and found the highest error rates and lowest completion time for the eye only selection technique [95].

Facial gestures recognition

Rozado *et al.* studied the potential of using live video monitoring to detect facial gestures to enhance eye tracking interaction [151]. In their work (FaceSwitch), they associated facial gestures (opening mouth, raising eyebrows, smiling and twitching the nose up and down) to simulate left and right mouse clicks and customized some keyboard functions such as page down key press. They found that increasing the number of gestures leads to lower recognition accuracy when monitored simultaneously.

Facial gesture recognition has several drawbacks. First, real-time video monitoring to detect the correct face gesture is very challenging beyond controlled lab conditions to address the real-life scenarios [110]. In addition, any emotional change or unwanted facial behavior may lead to false activation of functions, since modeling the human behavior is challenging [110]. Another drawback is the latency between pointing using the eye tracker and selecting using the facial gesture algorithm; precise timing is required for smooth interactions. Moreover, modeling of facial expressions requires a wide range of visual signal processing [110].

Gaze and speech interaction

Besides the above related works which were aimed at addressing the Midas touch problem, multimodal interaction have also considered gaze and voice commands. Mayer *et al.* proposed an interaction technique (WorldGaze) to track user's fields of view and gaze point to refine the voice command engines on smartphones for more precise results [111]. Beelders *et al.* studied word processing tasks using voice commands and eye gaze compared with mouse and keyboard interactions in their work [13]. However, although they showed the application of speech interaction is feasible for word applications, the gaze and speech interaction technique could not reach the effectiveness and performance of keyboard interaction. Acartürk *et al.* reviewed the challenges and possibilities of gaze and speech modalities for elderly users in their work [2]. Esteves *et al.* conducted comparative studies using head mounted displays (HMDs) to investigate the performance of hands-on and hands-free (including gaze and speech) interaction techniques and found that applying a clicker and dwell-time were the most favorable interaction techniques [51].

Miniotas *et al.* proposed a technique for selecting closely spaced targets based on speech commands [116]. They applied a grid of 5 × 5 squares as stimulus to test two interaction techniques: (a) gaze and speech, and (b) gaze only. They suggested a dwell-time of 1500 ms for targets of size of 30 × 30 pixels with distance of 10 pixels for the best performing setup for target selections based on their results. However, they reported a slow performance in case of selection speed when activation threshold for the dwell-time increased.

Beelders *et al.* conducted an experiment to study eye gaze and speech commands comparing to the mouse for target selection tasks [14]. They applied a stimulus as shape of a circle with 800 pixels diameter containing 16 squares on its edge to be selected in all directions. They found that the mouse had a significantly higher performance in case of throughput and completion time and also stated that using dwell-time technique should be more efficient than speech commands. Sengupta *et al.* integrated gaze and voice inputs for web browsing tasks such as search, navigation, and bookmark of pages [159]. They found the multimodal approach had a higher performance than each modality alone.

Zhao *et al.* proposed a multimodal technique of eye gaze by smooth pursuits, and speech commands and found promising results when compared to mouse clicks [190]. They found that the selection of a word for confirmation should match the task for a better performance. Further, participants who chose the activation word scored higher compared to those who used a pre-determined word. Similar to the EyeTAP method, the authors also suggested applications of other sound inputs such as pseudowords or exclamation for users with severe disabilities.

Gaze and hand gesture interaction

Gaze has also been combined with hand gesture inputs, for example, Chatterjee *et al.* proposed an interaction technique that uses gaze and hand gestures to select targets at the most desired location on screen [29]. They found that the combination of gaze and hand gesture outperformed each interaction modality alone. Pfeuffer *et al.* proposed a similar approach of applying eye gaze and a hand pinch to select and manipulate targets in a 3D space for virtual reality (VR) platforms [140]. Hand-gesture interactions are prone to muscular fatigue [70] and therefore may challenge users in certain circumstances.

Gaze and button press

Hild *et al.* investigated multimodal gaze-based interactions: gaze and button press by hand, gaze and button press by foot, and the mouse input [69]. They found overall faster performance for gaze-based techniques than the mouse for task completion time. Kumar *et al.* proposed a technique (EyePoint) comprised of eye gaze and button press on keyboard to improve the accuracy of gaze-based pointing in a Look-Press-Look-Release pattern of commands [93]. The EyePoint technique was designed in four steps to select a target accurately. The user looks at a desired target (Look), then presses and holds a hotkey on the keyboard which magnifies the specific spot on the screen (Press). A second look at the magnified scene is then done to refine the exact location of target to be selected (Look), then the key is released to select that target (Release). Gaze and button techniques have shown promising results in improving the selection accuracy.

Gaze gesture recognition

Istance *et al.* proposed a technique (Snap Clutch) to resolve the Midas touch problem [76]. They applied a disengagement technique to turn off gaze selections when not needed by defining four modes provided in up, left, right, and down directions on the screen. These modes are activated when looking at different directions (eye gesture) and visual feedback appear on the screen to confirm the intention.

3.2.5 Summary

We reviewed a wide range of techniques that can be applied with good accuracy and are suitable for specific domains with specific peripherals or extra user interface designs. The need for contact-free gaze-based interactions is necessary to deal with the emerging requirements regarding hygiene interactions from a safe distance. Building on the promising results found for multimodal techniques, and specifically exploring the use of non-speech sounds to allow for a more diverse population of users as suggested by Zhao *et al.* [190], we developed EyeTAP. EyeTAP can be applied to fill the gap for both able-bodied and disabled users with or without physical contact (to the microphone), with no need for specific user interface design or peripherals and using the simplicity of the Morse code [121] to encode/decode input signals.

3.3 EyeTAP Prototype

Using a multimodal solution that combines eye-gaze with acoustic inputs (audio or speech detection) can be regarded as an alternative to the reviewed literature on multimodal interaction methods and has the advantage of not requiring additional hardware (in comparison to other gaze-based techniques) other than an eye tracker or a specialized user interface design. Although there has been some work done on audio detection to simulate system events for computer interactions (e.g. [133, 39, 65]) on signal processing for complex interactions. Conversely, in our work we applied acoustic inputs only as a way of sending commands.

A simple mouse interaction consists of moving the pointer to a target (pointing phase), and clicking on it to trigger a function (selection phase). In the EyeTAP prototype the mouse pointer position is captured using an eye-tracker (in our case the Tobii 4C) and selection is done by generating an acoustic pulse by mouth (e.g. a mouth click) which is captured by a headset microphone (Logitech H370). The experiments using the EyeTAP prototype were run on a commodity computer system: 64-bit Windows 10 PC with Intel i7 2.67GHz CPU, 12 GB RAM, 1 TB hard disk and NVIDIA GeForce GTX 770 graphics card. Thus, EyeTAP is a cost-effective system that can be applied at almost any work space. Figure 7a gives an overview of the the EyeTAP system.

3.3.1 Eye Tracking: Pointing Phase

The Tobii SDK (TobiiEyeXSdk−Cpp−1.8.498) supports different events related to eye tracking activities such as providing the location of the current eye gaze, positions of both eyes, fixation points and user presence in front of the eye tracker. We employed the eye gaze library (API) to obtain users' gaze locations. These locations show the current gaze position on the screen as pixels. The SDK supports eye movements in a 3D coordinate system (horizontal, vertical, depth) but we applied a 2D coordinate system (x, y) such that the mouse cursor was synchronized with the gaze positions to control the mouse pointer on the screen. Eye tracking for the EyeTAP prototype was developed in C++ and integrated as a new plug-in into the Tobii SDK.

3.3.2 Auditory Processing: Selection Phase

To select a target the user makes a sound which is captured by a headset microphone. The intensity of the noise and distance of the microphone are adjusted by the user prior to using the system. A detected pulse in the real-time audio signal (amplitudes larger than a predefined threshold) is regarded as a click. The threshold's value can be adjusted based on the environment to reduce background ambient noise. When a significant increase in the signal (greater than the threshold) is detected a mouse click event is triggered as shown in Figure 7b. In general, recording is categorized into two phases: audible and silent periods. Any audible period with an intensity (amplitude) greater than the predefined threshold triggers an input signal to the system; on the other hand, values smaller than the threshold value are suppressed. Thus, any spoken sound e.g. speaking into the microphone or clicking the tongue, can trigger a click-event. Signal detection is continuous and works in real-time. The selection time-point is the moment the input pulse goes over the specified threshold at which point the click-event is triggered. This is purposedly designed to reduce possible synchronization issues resulting from eye gaze drifting away from the initial selection point. Thus our method initiates the selection phase as soon as it detects a trigger signal while the gaze pointer is still on the target.

Specifically, click detection is implemented as follow. First we capture the analog sound wave stream received from the microphone via the *AudioFormat* class provided in the Java Platform Standard Edition. 7 API [126] and digitize it using the sampling rate of 44100 Hz in a fixed buffer size of 256 bytes at a time. The buffer size is regarded as a *detection window* which is a queue for further processing. We set an empirical amplitude as threshold for pulse detection based on the available noise in the environment. Any receiving signal with an amplitude higher than the threshold is regarded as a 'click candidate' if it remains above the threshold for a minimum of

3 consecutive time-steps in which case it is considered a physical click and a mouse event is triggered. This step is necessary to enable a smooth flow of clicks in the case of noise or random vocal inputs by users and to reduce the effects of sudden noise inputs to the auditory detection API to avoid 'over clicking' events. The output of the auditory processing module is a series of 0s and 1s which are coupled with a mouse interaction event handler to trigger a left click based on 1 values. The entire workflow of the auditory module operates in real-time.

The intuition behind the auditory processing was inspired from the simplicity of the Morse code [121], which consists of a series of ON/OFF signals triggered by tone or light. In this case, information is interpreted using dots and dashes and therefore can be used to represent transmitted signals through a sequence of True/False variables. Figure 7b illustrates the step-wise operation of target selection phase by the EyeTAP technique.

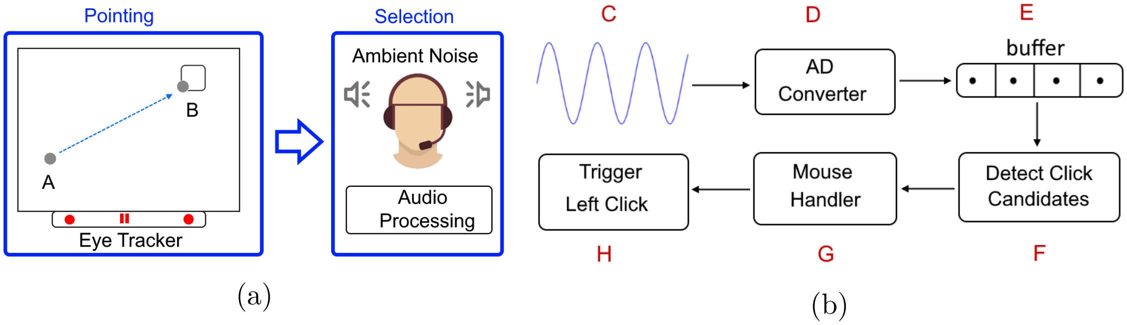


Figure 7: (a) The EyeTAP system: the eye tracker is used to move the pointer from A to B. The user makes an acoustic pulse and the signal processing module interprets the signal as an input and triggers a click event to select B. (b) The pipeline of the audio processing module. Analog audio waves are received from the microphone (C), and converted to a digital using an analog to digital converter (AD Converter) (D). The converted signal is stored in a fixed-sized buffer for further processing (E). A function detects the amplitudes higher than threshold as click candidates from the buffer (F). More than three click candidates in buffer are recognized as a click signal to be sent to a mouse event handler (G). Mouse handler triggers a left click (H).

3.3.3 Hypotheses

We hypothesize that a multimodal gaze-based interaction technique based on sound inputs can be applied to (a) enable a high accuracy contact-free interaction and (b) provide an alternative to mitigate the Midas touch problem. Furthermore, we hypothesize that our proposed technique will be easier to use compared to dwell-time and gaze with voice recognition and will be faster than the voice recognition technique.

3.4 Evaluation

To evaluate the effectiveness of the developed EyeTAP method, we ran four user studies with 33 participants (13 female, from 22 to 35 years old, $mean = 26.06$). Prior to running the experiments, subjects were informed about the purpose of the study, trained on each of the methods to be tested, and participated in a pre-test questionnaire probing them on their background in the fields of eye tracking, voice recognition technologies and their preferred kind of interaction in the case of contact-free alternatives. The Tobii calibration software was used to calibrate the system for each participant before starting the study. At the end of the user studies subjects filled out a post-test questionnaire, which consisted of the NASA TLX questionnaire [59] followed by specific questions about the subjects' perceptions of the different interaction methods. The order of interaction method was randomly selected for each participant.

We played an artificial ambient noise through stereo desktop speakers of 50 dB to simulate a typical work environment since EyeTAP and voice recognition rely on audio inputs. Participants were asked to produce a tongue click type sound ('tick') which lasted for 2 seconds on average.

To determine the effectiveness of the EyeTAP method, we analyzed the results of our experiments using an analysis of variance (ANOVA) followed by Bonferroni posthoc tests with the IBM SPSS software, and applied descriptive statistics based on dispersion with the JASP 0.11.1 software [84].

3.4.1 Interaction Techniques

We applied two eye tracking techniques to be compared with the performance of EyeTAP and included mouse as the baseline technique for point-and-select tasks. In other words, for all tests our independent variable is the interaction technique: (a) the mouse, (b) dwell-time, (c) eye tracking with voice-recognition, and (d) EyeTAP.

Mouse

For the mouse method (our baseline method for comparison), subjects simply used a mouse to move to targets and select them in numerical order.

Dwell-time

For the dwell-time method an internal timer was used to determine if a target was selected. Given the range of dwell-time is typically 300-1100 milliseconds for target selection [165], we defined the target activation threshold to 500 milliseconds, since this showed the best performance in [103, 165]. In other words, a target was selected when a subject focused on a target for 0.5 seconds, and if the subject moved their gaze away from the target prior to 0.5 seconds the target selection process would restart.

Eye Tracking with Voice recognition

For voice recognition, eye tracking was used for pointing and voice for selection. The method was developed using the built-in Windows 10 speech recognition capabilities available in the .NET framework. We implemented a C# application to respond to the activation keyword ‘select’ to trigger a mouse click. The same microphone was used as for the EyeTAP test.

3.4.2 User Study 1: Matrix-based Test

In the first user study, the EyeTAP interaction method was compared with: (a) the mouse, (b) dwell-time, and (c) eye tracking with voice-recognition. In this test, a matrix of buttons (targets), were randomly distributed across the screen. The task of the subjects was to point and click on buttons shown on the screen in increasing numerical order for various levels of difficulty from 1 (easy) to 5 (hard), described in detail below. The order of interaction methods seen by each subject was randomly selected for each participant however, the level of difficultly was presented in ascending order.

We were inspired by Miniotas *et al.*’s work that applied a stimulus composed of a grid of 5 × 5 squares [116]. The matrix grid was designed to cover a large area of the screen and to have equally-sized targets in close adjacent proximity. This enabled the analysis of errors that are most important for the Midas touch problem. Furthermore, since different areas of a screen have different accuracy in target selection for eye tracking applications [54], this test allowed us to study target selection accuracy on different areas of the screen.

Stimulus

The stimulus consisted of 77 buttons (11 columns × 7 rows) some labeled with numbers and others not, which covered the entire screen at a resolution of 1920 × 1080 pixels on a Dell P2411Hb monitor. Two marginal columns (far left, far right) and two rows (top, bottom) were removed from the active selection due to the high difficulty to be selected by users during the pilot-test. Buttons that were not labeled are considered as *barriers* or *distractions*. To provide feedback to the subject, labeled buttons change color after the user has successfully pointed and selected on the correct button. Wrongly selected barriers (buttons with no label) are highlighted in red. The level of difficulty of the stimulus was also increased across subject trials. This was done by increasing the number of targets that had to be selected by the subject. Five levels of difficulty were used for each interaction method: level 1 (4 targets), level 2 (6 targets), level 3 (8 targets), level 4 (10 targets) and level 5 (12 targets). Targets were randomly distributed over the entire screen for each level. Figure 8 shows the matrix-based test during difficulty level 5. The cursor that was used was a black circle because it was easier for users to keep it on the target's boundary rather than a pointer. The rationale of 'difficulty' for a higher number of targets lies in the experience that the selection of more targets caused eye fatigue for some users during the test, especially for the dwell-time method.

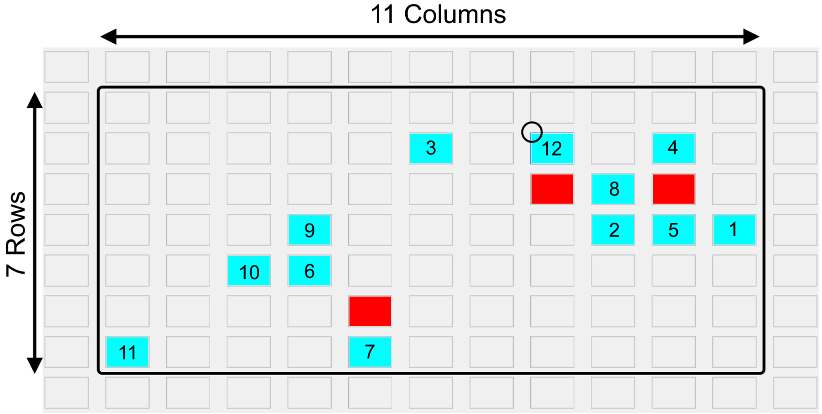
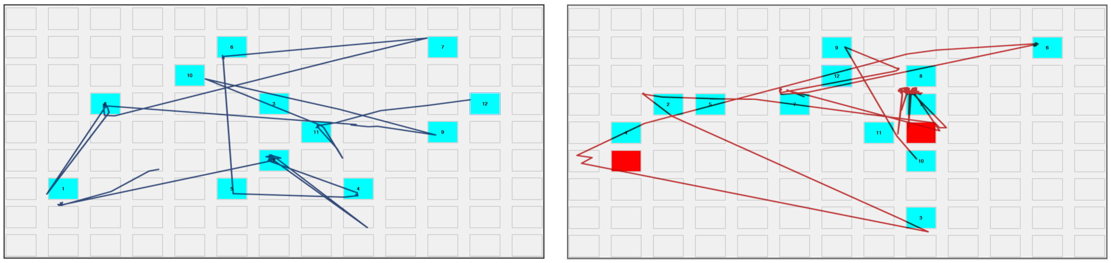


Figure 8: The matrix-based test for difficulty level 5. Target buttons are distributed randomly across the screen. The red buttons illustrate errors. The black circle on number 12 shows the current eye gaze location. Labels were enlarged for higher visibility.

Measures

The following dependent variables were recorded: *completion time*, *path cost of selecting targets*, *error locations*, and *cognitive load* (based on the NASA TLX scores). An internal logging module recorded subjects' actions, selection times, as well as the number of correct and wrong selections.

For the path cost measure the shortest path between targets and the produced path by each interaction method was processed. The intuition behind this measure was to analyze the trajectory of pointer movements (footprints) of each interaction technique. In other words, since the pointer was mapped with eye gaze, we could detect which interaction technique would select targets with less eye movements (see Figure 9). This measure was specifically designed to test the hypothesis whether dwell-time requires less eye movements than multimodal techniques due to pointer drift caused by synchronization between *pointing* and *selection* phases. To compare the shapes of the generated paths, we used the dynamic time warping (DTW) algorithm [15, 120, 153]. Since DTW works on a time-value domain the paths produced by the eye tracker were decomposed into their horizontal and vertical values and compared with their associated shortest path models' X and Y values. We applied the built-in DTW function in the Python DTW 1.3.3 module [143] to measure the deviations of each path from the shortest path model.



(a)

(b)

Figure 9: The path cost overview of (a) dwell-time, and (b) EyeTAP on the screen.

Results

A two-way repeated measures ANOVA (methods × difficulty levels) was performed to examine the effect of interaction type on: (1) *completion time* and (2) *path costs of target selection* for each method and difficulty levels. We also analysed the distribution of each measure since it indicates the consistency of each interaction technique on most users.

Completion time: We found a significant effect of interaction method on completion time (F(12,384)=8.51, $p < .001$). A posthoc Bonferroni comparison test showed a significant difference between mouse ($M = 1.04$ *sec*, $SE = 0.02$ *sec*) and all other eye tracking methods (see Figure 10a). In addition, EyeTAP ($M = 2.57$ *sec*, $SE = 0.12$ *sec*), dwell-time ($M = 1.40$ *sec*, $SE = 0.06$ *sec*) and voice recognition ($M = 3.20$ *sec*, $SE = 0.25$ *sec*) are significantly different ($p < .05$). Figure 10a illustrates the overall completion time per method for each target.

We also looked at the distribution of values for completion time, and found a large range for both EyeTAP ($range = 8.69$ *sec*, $IQR = 0.90$ *sec*) and voice recognition ($range = 7.71$ *sec*, $IQR = 1.39$ *sec*) comparing to the mouse (0.70 *sec*, $IQR = 0.14$ *sec*) and dwell-time (1.80 *sec*, $IQR = 0.84$). The interquartile range comparison was the narrowest for mouse and highest for voice recognition, but there was a similar variability between EyeTAP and dwell-time.

Path costs of target selections: To examine the paths produced by selecting targets we compared the original locations of the targets and the shortest path (ideal path model), as described earlier. For each method, we had a $\frac{distance}{cost}$ measure to the shortest path. This metric can be regarded as the *footprint* of each interaction technique on the display. A two-way repeated measures ANOVA (methods × difficulty levels) showed that there was a significant effect of interaction type on path cost (F(12,384)=2.57, $p < .05$). A Bonferroni posthoc test showed that dwell-time ($M = 76.73$ $pixels$, $SE = 5.09$ $pixels$) produced the shortest path among all other interaction techniques, even better than the mouse interaction ($M = 109.25$ $pixels, SE = 3.82$ $pixels$) with $p < .05$. There were no significant differences between dwell-time ($M = 76.73$ $pixels$, $SE = 5.09$ $pixels$), EyeTAP ($M = 84.80$ $pixels$, $SE = 3.59$ $pixels$) and voice recognition ($M = 82.03$ $pixels$, $SE = 4.41$ $pixels$). Figure 10b, which shows the path costs for all interaction methods, reveals that eye tracking movements produce significantly lower movements than mouse on a large screen. We found the highest variability in paths for dwell-time ($range = 126.81$ $pixels$, $IQR = 43.13$ $pixels$) and the lowest for mouse ($range = 79.21\, pixels$, $IQR = 33.26\ \ pixels$). Voice recognition ($range = 111.11\, pixels$, $IQR = 29.91$ $pixels$) showed a larger range compared to EyeTAP ($range = 88.88$ $pixels$, $IQR = 22.76\, pixels$). All eye tracking techniques reached a significantly lower median than the mouse which reflects a shorter path for eye gaze pointing on the screen than mouse pointing. EyeTAP reached the narrowest interquartile range for gaze path on screen among all interaction techniques which represents similar performance for most users comparing to other interaction techniques. The dwell-time method showed the highest variability and voice recognition reached the second highest variability based on the interquartile range measure.

Errors in target selections: To measure the effectiveness of each Midas touch solution we need to consider a penalty for wrongly selected neighboring targets. These targets are shown in red on the screen (see Figure 8). We projected the locations of errors per each interaction method, since difficulty level 5 has the highest number of targets (12 targets) on the screen, we illustrate the locations for this difficulty level in Figure 11. EyeTAP has the highest number of errors, however the figure reveals the potential regions of the screen which are more error prone. As shown in the figure, most errors occurred from the center towards the right side of the screen. In fact, the right side of the screen produces more errors than the left side. Moreover, the lower side produces more errors than the top side. This is similar to Feit *et al.*'s finding showing that the bottom and right regions of the screen have lower accuracy [54]. We confirm their results and also demonstrate that the same regions are also more error prone.

3.4.3 User Study 2: Dart-based Test

The purpose of this user study was to measure the accuracy of EyeTAP in comparison to the previously proposed eye-based interaction methods. Specifically, we wanted to focus on target selection accuracy. The task of the subject was to select, as accurately as possible, the bull's-eye of a dart target using each interaction method. In this test, the eye tracker was used for the pointing phase for each of the interaction methods, however selection of the target was triggered by different methods, i.e. dwell-time, voice command or EyeTAP acoustic signal. In order to take into consideration the fact that eye tracking has different accuracy in different regions of the monitor, we computed an average value based on five trials for each interaction method where the stimulus was shown at different areas of the screen near the center of the screen randomly. Each new randomly chosen trial began two seconds after selection of the previous target, allowing users time to change their gaze and to focus on the new target. For the dwell-time method, a countdown (from 5 to 0) representing the time left in milliseconds until the target selection was displayed and after each selection visual feedback was given to the user by showing the achieved distance to target.

Stimulus

The stimulus for this test consisted of a dart-like target with three circles, green (0 to 30 pixels radius), blue (30 to 60 pixels radius) and red (60 to 90 pixels radius) as shown in Figure 13a. Points within the center area i.e. green have the lowest range of distances to the bulls-eye; each other co-centric circle has a larger range of distance values. Any point lying outside the three co-centric circular areas is considered as having a fixed maximum distance of 90 pixels. For this test, a cross-hair icon was used.

Measures

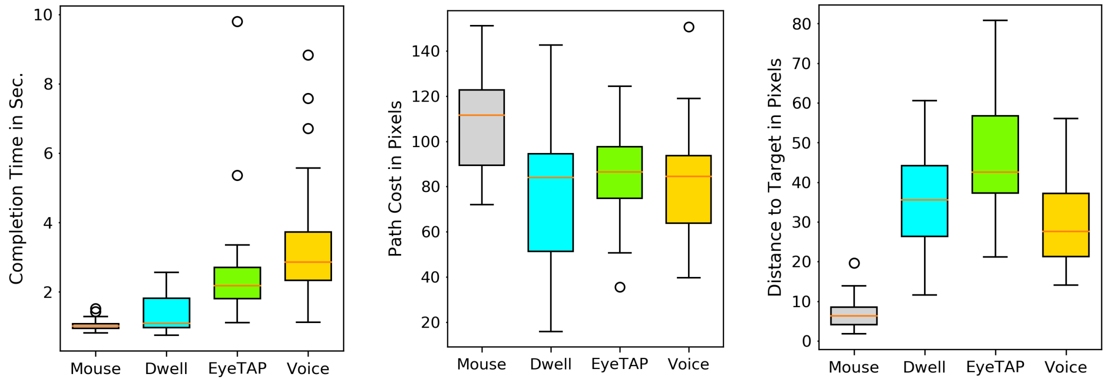
The purpose of this test was to measure the selected point's distance on the dart target to the center of the core circle (in green), thus the accuracy (i.e. dependent variable) is measured in pixels. Since the measured trials are chosen randomly, the average is calculated to compare different methods based on accurate selection.

Results

We performed a one-way repeated measures ANOVA to compare the effect of the different interaction methods on accuracy. The results of the ANOVA showed all eye tracking methods have statistical difference (F(3,96)=104.92, $p < .001$) on selection accuracy. In fact, the mouse interaction has the lowest distance to target (highest accuracy) compared to eye tracking techniques. EyeTAP ($M = 45.11\ pixels$, $SE = 2.28\ pixels$) achieved the highest mean pixel accuracy compared to dwell-time ($M = 35.30\ pixels$, $SE = 2.11\ pixels$) and voice recognition ($M = 29.27\ pixels$, $SE = 2.07\ pixels$). Figure 10c depicts the results of the accuracy test.

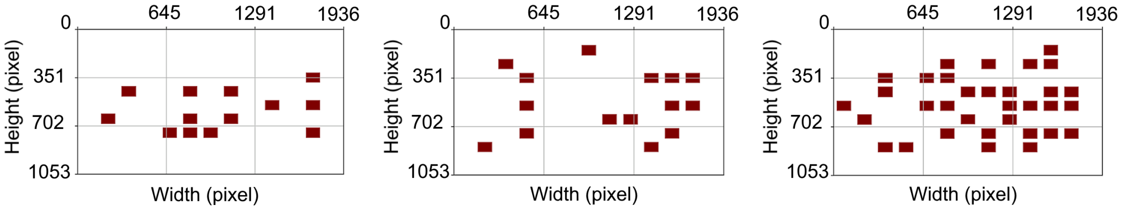
We found the highest variability for EyeTAP on both measures ($range = 59.62\ pixels$, $IQR = 19.42$) among eye tracking techniques whereas the voice recognition technique reached the lowest distribution ($range = 41.05\ pixels$, $IQR = 15.87$) and lowest distance to the target, and dwell-time ($range = 48.96\ pixels$,

$IQR = 17.91$) showed a higher distribution than mouse ($range = 17.76\ pixels$, $IQR = 4.39$).



(a) (b) (c)

Figure 10: (a) Completion time of point-and-select tasks for each target ($p < .001$). (b) Path cost comparison calculated using the dynamic time warping (DTW) algorithm. All eye tracking techniques have shorter path lengths than mouse interaction for traversing items on a screen for matrix-based user study ($p < .05$). (c) The distance to target in pixels for dart-based test ($p < .001$).



(a) Dwell-Time (b) Voice Recognition (c) EyeTAP

Figure 11: The locations of errors on the screen during the matrix-based user study (see Figure 8) for difficulty level 5. The right side of the screen as well as bottom side are more error prone than the left and top sides.

3.4.4 User Study 3: Ribbon-shaped Test

In order to compare our method to other gaze-based techniques, we measured the performance target selection based on the Fitts' law [57]. This study is used to analyze pointing interaction methods in accordance to well-established academic standards.

As part of this study, we measured three metrics to compare the performance of all interaction techniques for point-and-select tasks, (1) *throughput* (how good a selection technique operates), (2) *movement time* and (3) *error rates* for ribbon-shaped targets (see Figure 13b).

The intuition of this test was to test interaction techniques based on the Fitts' law with rectangular buttons ('FittsStudy' application [186]).

Stimulus

The stimulus for this test consisted of two ribbon-shaped buttons to be selected on the left and right sides of the screen with random widths and distances as shown in Figure 13b. The test sessions includes three distances (256, 384, 512) pixels, and two widths (96, 128) pixels.

Measures

The following dependent variables were recorded: *movement time*, *throughput*, and *error rates* for this test. We applied the 'FittsStudy' application by Wobbrock *et al.* [186] for this test.

Results

A one-way repeated measures ANOVA was performed to examine the effect of interaction type on: (1) *movement time*, (2) *throughput* and (3) *error rates* for each interaction method. We also analysed the distribution of each measure since it indicates the consistency of each interaction technique on most users.

Movement time: We found a significant effect of the interaction method on movement time (F(3,96)=69.42, $p < .001$). A posthoc Bonferroni comparison test showed a significant difference between mouse ($M = 684.15\ ms$, $SE = 16.80\ ms$)

and all other eye tracking methods (Figure 12a). In addition, among all eye tracking methods, dwell-time ($M = 599.39$ ms, $SE = 18.76$ ms) achieved significantly lower movement time than EyeTAP ($M = 1794.89$ ms, $SE = 170.90$ ms) and voice recognition ($M = 2014.20$ ms, $SE = 89.28$ ms) techniques. However, there is no statistical significance between EyeTAP and voice recognition. The lower movement time of dwell-time method compared to mouse interaction is associated with the low activation time (500 ms).

We found the highest variability for EyeTAP ($range = 5.67$ sec, $IQR = 0.69$ sec) among all interaction techniques, whereas dwell-time ($range = 0.42$ sec, $IQR = 0.09$ sec) and voice recognition ($range = 2.03$ sec, $IQR = 0.37$ sec) reached lower distributions among eye tracking techniques. The mouse reached the narrowest range ($range = 0.34$ sec) but larger interquartile range ($IQR = 0.11$ sec) than dwell-time. We found the dwell-time as the best interaction technique based on the movement time measure for the ribbon-shaped test as illustrated in Figure 12a.

Throughput: We found a significant effect of the interaction method on throughput ($F(3, 96) = 75.13$, $p < .001$). A posthoc Bonferroni comparison test showed a significant difference between dwell-time ($M = 3.30$ $bits/sec$, $SE = 0.36$ $bits/sec$) and all eye tracking methods (Figure 12b). The mouse ($M = 4.81$ $bits/sec$, $SE = 0.11$ $bits/sec$) achieved higher throughput than the eye tracking methods. However, there is no statistical difference between voice recognition ($M = 1.15$ $bits/sec$, $SE = 0.09$ $bits/sec$) and EyeTAP ($M = 1.34$ $bits/sec$, $SE = 0.12$ $bits/sec$).

We found that EyeTAP ($range = 2.73$ $bits/sec$, $IQR = 0.78$ $bits/sec$) had the narrowest range of values for throughput, and dwell-time ($range = 7.64$ $bits/sec$, $IQR = 2.86$ $bits/sec$) the highest variability based on both measures among all interaction techniques. The voice recognition ($range = 2.043$ $bits/sec$, $IQR =$

0.63 $bits/sec$) reached lower variability than mouse ($range = 2.83$ $bits/sec$, $IQR = 0.95$ $bits/sec$) on both measures. However, both EyeTAP and voice recognition reached lower throughput than dwell-time on average, dwell-time reached the highest variability due to having a sparse distribution compared to the other interaction techniques.

Error rates: We found a significant effect of interaction method on error rates ($F(3,96) = 27.15$, $p < .001$). A posthoc Bonferroni comparison test showed a significant difference between mouse ($M = 0.01$ $errors$, $SE = 0.005$ $errors$) and all eye tracking interactions (see Figure 12c). In addition, dwell-time ($M = 0.28$ $errors$, $SE = 0.03$ $errors$) reached a higher error rate than EyeTAP ($M = 0.18$ $errors$, $SE = 0.02$ $errors$) and voice recognition ($M = 0.10$ $errors$, $SE = 0.02$ $errors$).

We also analysed the distribution of errors among users and found that EyeTAP ($range = 0.66$ $errors$, $IQR = 0.16$ $errors$) had a similar range compared to dwell-time ($range = 0.66$ $errors$, $IQR = 0.25$ $errors$) but lower variability based on the interquartile range measure. The voice recognition technique ($range = 0.58$ $errors$, $IQR = 0.16$ $errors$) showed a narrower range than EyeTAP but similar variability based on the interquartile range measure. The mouse ($range = 0.08$ $errors$, $IQR = 0.00$ $errors$) reached the lowest variability based on both measures among all interaction techniques. The voice recognition technique reached the lowest distribution of errors among eye tracking techniques based on error rates as illustrated in Figure 12c.

3.4.5 User Study 4: Circle-shaped Test

This test is similar to the Ribbon-shaped test, however, contains different target shapes. Figure 13c illustrates the screenshots of this test which contains uni-variate endpoint deviation (SD_x) through X axis and bi-variate endpoint deviation ($SD_{x,y}$)

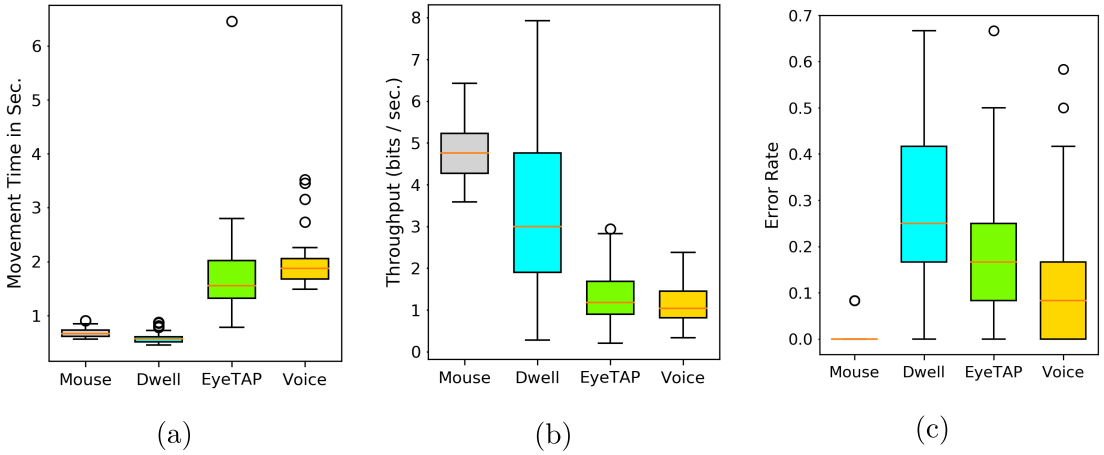


Figure 12: (a) Calculated movement time, (b) throughput, and (c) the error rates per method for the ribbon-shaped test. For all measures $p < .001$.

through both X, Y axes for throughput calculations which results in better Fitts' law model [186]. The 'FittsStudy' application by Wobbrock *et al.* [186] was used for this test.

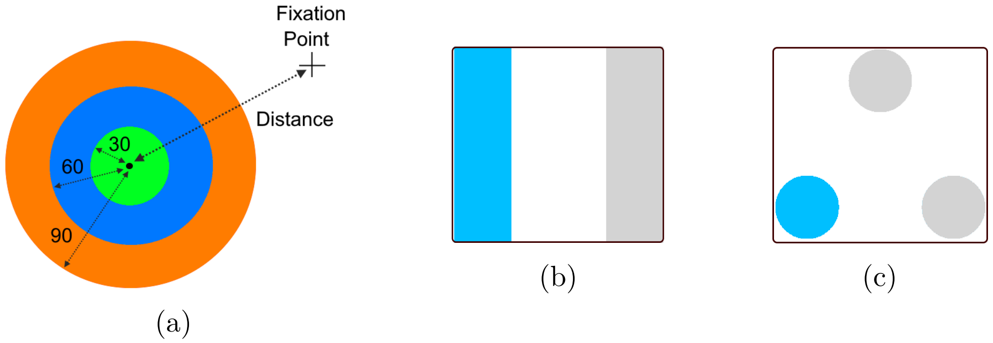
The intuition of this test was to test the interaction techniques based on the Fitts' law with circular buttons provided by the 'FittsStudy' application [186].

Stimulus

The stimulus for this test consisted of three circle-shaped buttons to be selected located in the middle of the screen with random widths and distances as shown in Figure 13c. The test sessions includes three distances (256, 384, 512) pixels, and two widths (96, 128) pixels.

Measures

The following dependent variables were recorded: *movement time*, *throughput* (with two variations), and *error rates* for this test.



(a) (b) (c)

Figure 13: (a) Shows the Dart-based test stimuli: the accuracy is highest in the green area. The cross-hair icon indicates the correct eye gaze location, (b) Illustrates the ribbon-shaped stimuli, and (c) shows the circle-shaped stimuli of the 'FittsStudy' application [186]. Targets highlighted in blue represent active targets to be selected.

Results

A one-way repeated measures ANOVA was performed to examine the effect of interaction type on: (1) *movement time*, (2) *throughput* and (3) *error rates* for each interaction method. This test is similar to ribbon-shaped test but contains an extra metric to measure throughput of each method.

Movement time: We found a significant effect of the interaction method on movement time (F(3,96)=67.48, $p < .001$). A posthoc Bonferroni comparison test showed a significant difference between EyeTAP ($M = 1578.95\ ms$, $SE = 95.34\ ms$), dwell-time ($M = 638.80\ ms$, $SE = 24.35\ ms$), voice recognition ($M = 2123.35\ ms$, $SE = 132.42\ ms$) and mouse ($M = 727.91\ ms$, $SE = 46.12\ ms$). However, there is no statistical difference between mouse ($M = 727.91\ ms$, $SE = 46.12\ ms$) and dwell-time ($M = 638.80\ ms$, $SE = 24.35\ ms$). Figure 14a illustrates the mean movement time per method for the circle-shaped test.

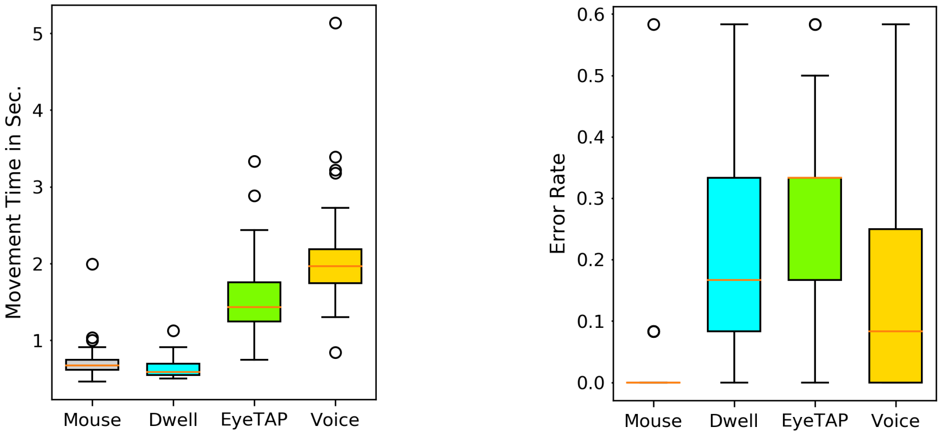
We found that dwell-time ($range = 0.62\ sec$, $IQR = 0.15\ sec$) has the narrowest, and voice recognition ($range = 4.29\ sec$,$IQR = 0.44\ sec$) the largest range. EyeTAP ($range = 2.58\ sec$, $IQR = 0.51\ sec$) showed a narrower range than voice recognition but larger interquartile range than voice recognition, dwell-time and mouse ($range =$

1.53 *sec*, $IQR = 0.12$ *sec*). This analysis shows higher consistency for dwell-time compare to the other interaction techniques.

Error rates: We found a significant effect of the interaction method on error rates ($F(3, 96) = 18.25$, $p < .001$). A posthoc Bonferroni comparison test showed a significant difference between mouse ($M = 0.02$ *errors*, $SE = 0.01$ *errors*), dwell-time ($M = 0.23$ *errors*, $SE = 0.03$ *errors*), voice recognition ($M = 0.13$ *errors*, $SE = 0.02$ *errors*) and EyeTAP ($M = 0.28$ *errors*, $SE = 0.02$ *errors*). Voice recognition ($M = 0.13$ *errors*, $SE = 0.02$ *errors*) reached the lowest error rate among eye tracking methods, however, there is no statistical difference between dwell-time ($M = 0.23$ *errors*, $SE = 0.03$ *errors*) and EyeTAP ($M = 0.28$ *errors*, $SE = 0.02$ *errors*). Figure 14b illustrates the calculated error rates for the circle-shaped test.

We found that mouse ($range = 0.58$ *errors*, $IQR = 0.0$ *errors*), dwell-time ($range = 0.58$ *errors*, $IQR = 0.25$ *errors*), voice recognition ($range = 0.58$ *errors*, $IQR = 0.25$ *errors*), and EyeTAP ($range = 0.58$ *errors*, $IQR = 0.16$ *errors*) showed the same variability based on range measure, but EyeTAP reached a lower distribution based on the interquartile range among eye tracking techniques.

Throughput: Since the circle-shaped test contains two variations (uni-variate, bi-variate) to measure throughput [186], we ran a two-way repeated measures ANOVA (throughput × variation) and found a significant effect of the interaction method on throughput (F(3,96)=19.75, $p < .001$). A posthoc Bonferroni comparison test showed a significant difference between mouse ($M = 4.16$ *bits/sec*, $SE = 0.18$ *bits/sec*), dwell-time ($M = 3.20$ *bits/sec*, $SE = 0.25$ *bits/sec*), voice-recognition ($M = 1.24$ *bits/sec*, $SE = 0.07$ *bits/sec*) and EyeTAP ($M = 1.04$ *bits/sec*, $SE = 0.13$ *bits/sec*). However, there is no statistical difference between voice-recognition ($M = 1.24$ *bits/sec*, $SE = 0.07$ *bits/sec*) and EyeTAP ($M = 1.04$ *bits/sec*,



(a)

(b)

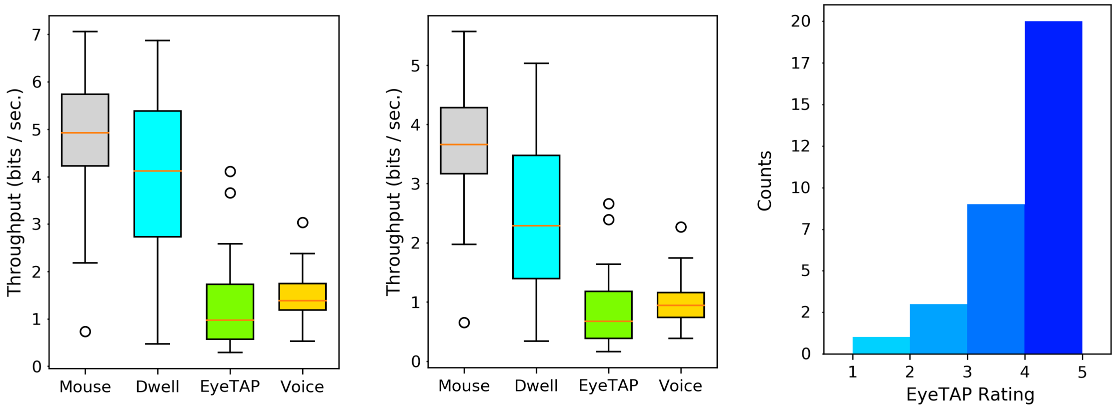
Figure 14: (a) Calculated movement time, and (b) error rates per method for the circle-shaped test. For all measures ($p < .001$).

$SE = 0.13$ *bits/sec*). Figure 15a shows uni-variations of throughput, and Figure 15b shows the bi-variations of throughput per interaction method.

We found that dwell-time ($range = 6.40$ *bits/sec*, $IQR = 2.66$ *bits/sec*) showed the highest variability among all interaction techniques based on both measures, range and interquartile range for uni-variation throughput measure. Whereas, voice recognition ($range = 2.50$ *bits/sec*, $IQR = 0.55$ *bits/sec*) showed the lowest variability. EyeTAP ($range = 3.81$ *bits/sec*, $IQR = 1.16$ *bits/sec*) showed lower variability than mouse ($range = 6.32$ *bits/sec*, $IQR = 1.51$ *bits/sec*) on both measures as illustrated in Figure 15a.

We found that dwell-time ($range = 4.69$ *bits/sec*, $IQR = 2.08$ *bits/sec*) and mouse ($range = 4.91$ *bits/sec*, $IQR = 1.11$ *bits/sec*) showed the highest variability on both range and interquartile range measures. Whereas voice recognition ($range = 1.88$ *bits/sec*, $IQR = 0.42$ *bits/sec*) and EyeTAP ($range = 2.49$ *bits/sec*, $IQR = 0.79$ *bits/sec*) showed lower variability for the bi-variate throughput measure as illustrated in Figure 15b.

This analysis confirms that EyeTAP has the lowest throughput based on mean value, and voice recognition has the lowest distribution (higher consistency) among all interaction techniques for throughput measure based on both uni-variation and bi-variation of the circle-shaped user study (see Figure 15).



(a) (b) (c)

Figure 15: (a) Calculated throughput for uni-variate, (b) throughput for bi-variate per method for the circle-shaped test, and (c) shows the ratings of EyeTAP from 1 (worst) to 5 (best) for 33 participants. For all measures in (a) and (b) ($p < .001$).

3.5 Results

3.5.1 EyeTAP Rating by Users

We asked participants to evaluate the overall performance of EyeTAP in the post-test questionnaire on a scale from 1 (worst) to 5 (best). EyeTAP reached the average rate of 3.64 ($SD = 0.99$) by 33 users. Figure 15c illustrates the subjective ratings obtained from the post-test questionnaire.

3.5.2 NASA TLX Scores

Figure 16 shows the NASA TLX scores for all interaction methods obtained during the user study. The overall workload is the average of scale values since we assume all scales equally important and therefore eliminated the weighting calculation to apply a simplified version [62] of the basic NASA TLX ratings [59]. According to our findings, the dwell-time method has the lowest workload among other eye tracking techniques. However, EyeTAP shows relatively lower workload compared to the voice recognition technique.

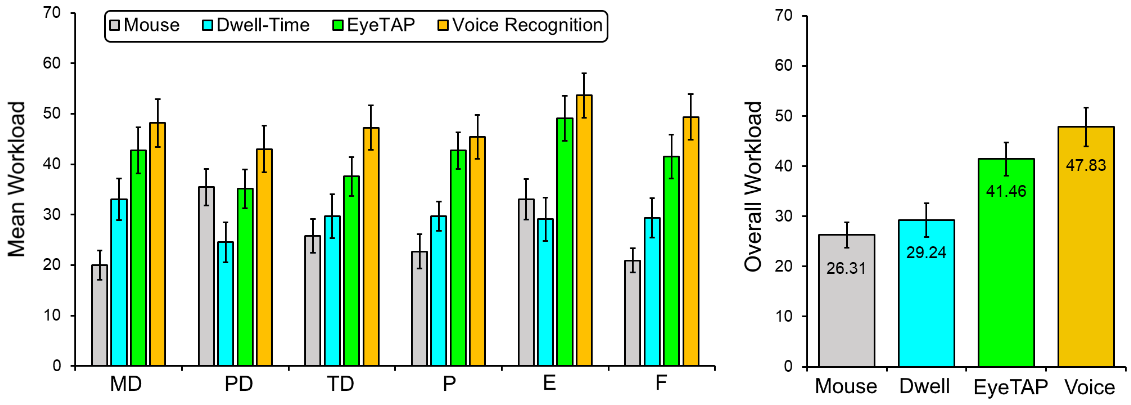
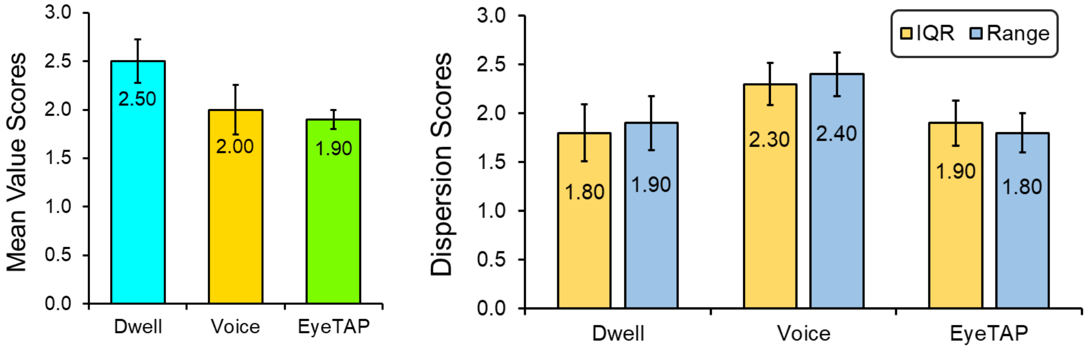


Figure 16: The NASA TLX scores for the interaction methods. (Left) Comparison of each method based on different scales. (Right) The overall mean workload of tested interaction methods. Error bars represent standard error.

3.5.3 Comparative Scores

We analyzed the results of the eye tracking techniques based on (1) the analysis of variance (ANOVA), and (2) the descriptive statics based on dispersion of data, as illustrated earlier in this section. Since we measured the interaction techniques based on various criteria, we need to obtain a single measure comprised of all reviewed measures for comparison. Therefore, we applied a simple scoring technique and assigned an integer value in the set of {1 (worst), 2 (medium), 3 (best)} to eye tracking techniques based on their performance and calculated the arithmetic average for each interaction techniques of the entire criteria. Furthermore, we assigned the value of 2 (medium) to interaction techniques when they showed statistically similar or very close performance. Table 2 shows the details of this scoring technique for the ANOVA-based measures, and Table 3 contains the details of dispersion analysis scoring. The higher the calculated average score shows the better performance of the entire measures.

Figure 17a illustrates the results of Table 2 and Figure 17b shows the calculated average of both measures (range and IQR measures) of Table 3. The dwell-time reached the highest score (the best performance) based on the average value of objective measures of our user studies, although the difference between voice recognition and EyeTAP is not significant. However, EyeTAP and voice recognition reached relatively higher scores (higher consistency) than the dwell-time method based on dispersion analysis, however, the differences are not statistically significant. We showed that dwell-time performs very well for some participants, but shows sparse distribution on some criteria. Furthermore, EyeTAP may be considered as an interaction technique that has potential for improvement and can be adapted for most participants with sufficient training.



(a)

(b)

Figure 17: (a) Calculated scores from 1 (worst) to 3 (best) on all objective measures for eye tracking techniques shown in Table 2. The dwell-time method shows the highest scores based on ANOVA analysis results. (b) The calculated scores of average of both dispersion analysis results (range and IQR measures) shown in Table 3. Higher scores are better in both figures.

	Mean Values		
Criteria	Dwell	Voice	EyeTAP
Comp. Time	3	1	2
	(1.40)	(3.20)	(2.57)
Path Costs	2	2	2
	(76.73)	(82.03)	(84.80)
Distance	2	3	1
	(35.30)	(29.27)	(45.11)
MT_{Ribbon}	3	1	2
	(0.59)	(2.01)	(1.79)
TP_{Ribbon}	3	2	2
	(3.30)	(1.15)	(1.15)
ER_{Ribbon}	1	3	2
	(0.28)	(0.10)	(0.18)
MT_{Circle}	3	1	2
	(0.63)	(2.12)	(1.57)
$\text{TP}_{Circle-uni}$	3	2	2
	(3.90)	(1.48)	(1.24)
$\text{TP}_{Circle-bi}$	3	2	2
	(2.50)	(1.00)	(0.84)
ER_{Circle}	2	3	2
	(0.23)	(0.13)	(0.28)
Average	2.50	2.00	1.90

Table 2: Summary of scores per interaction techniques based on comparison of their mean values. Scores are integer values from 1 (worst) to 3 (best). Statistically similar mean values ($p > .05$) were assigned the value of 2. Values represented in parenthesis denote the mean values of each measure. MT, TP, and ER represent movement time, throughput, and error rates.

		R			**IQR**	
Criteria	Dwell	Voice	EyeTAP	Dwell	Voice	EyeTAP
Comp. Time	3	2	1	3	1	2
	(1.80)	(7.71)	(8.69)	(0.84)	(1.39)	(0.90)
Path Costs	1	2	3	1	2	3
	(126.81)	(111.11)	(88.88)	(43.13)	(29.91)	(22.76)
Distance	2	3	1	2	3	1
	(48.96)	(42.05)	(59.62)	(17.91)	(15.87)	(19.42)
MT$_{Ribbon}$	3	2	1	3	2	1
	(0.42)	(2.03)	(5.67)	(0.09)	(0.37)	(0.69)
TP$_{Ribbon}$	1	3	2	1	3	2
	(7.64)	(2.04)	(2.73)	(2.86)	(0.63)	(0.78)
ER$_{Ribbon}$	2	3	2	1	2	2
	(0.66)	(0.58)	(0.66)	(0.25)	(0.16)	(0.16)
MT$_{Circle}$	3	1	2	3	2	1
	(0.62)	(4.29)	(2.58)	(0.15)	(0.44)	(0.51)
TP$_{Circle-uni}$	1	3	2	1	3	2
	(6.40)	(2.50)	(3.81)	(2.66)	(0.55)	(1.16)
TP$_{Circle-bi}$	1	3	2	1	3	2
	(4.69)	(1.88)	(2.49)	(2.08)	(0.42)	(0.79)
ER$_{Circle}$	2	2	2	2	2	3
	(0.58)	(0.58)	(0.58)	(0.25)	(0.25)	(0.16)
Average	1.90	2.40	1.80	1.80	2.30	1.90

Table 3: Summary of scores per interaction techniques based on comparison of dispersion on both measures (1) range (R), and (2) interquartile range (IQR) values. Scores are integer values from 1 (worst) to 3 (best). We assigned value of 2 for similar mean values. Values represented in parenthesis denote the actual values of each measure. MT, TP, and ER represent movement time, throughput, and error rates.

3.6 Discussion

Regarding the experiments with the reviewed Midas touch solutions, we found several benefits and disadvantages of each method. We discuss each method individually.

3.6.1 EyeTAP

We found several benefits of using EyeTAP in comparison to the other interaction techniques. First of all, it has no dependent features, rather it requires only an acoustic pulse (making a sound) near a microphone to send a signal. In fact, the output of EyeTAP in a noisy environment can appear deterministic after a number of repetitions. According to the results of our study, it achieved faster completion time in the matrix-based test, and faster movement time in the circle-shaped test than voice recognition. In addition, it showed a similar path cost (pointer footprint on display) with the other eye tracking techniques. It also achieved lower cognitive workload in comparison to the voice recognition technique. Furthermore, EyeTAP was a popular choice of interaction (36.4%) compared to voice recognition (9.1%). However, EyeTAP showed relatively lower accuracy and higher error rates than voice recognition, perhaps due to the fact most users had no prior experiences with this kind of interaction. Suggesting that with more training the performance of EyeTAP could be improved.

EyeTAP achieved the lowest variability for path cost of pointer movements on screen for the matrix-based test. In addition, it showed lower variability than dwell-time and mouse on throughput measures of both ribbon-shaped and circle-shaped test. The low variability of EyeTAP reflects the predictability of its performance on subjects, thus this method can be adopted for different users or different case scenarios. In general, EyeTAP allows for point-and-select interaction because it separates the actions of *pointing* and *selecting* to two different modalities while relaxing the requirement for accurate voice recognition. The results of our user study demonstrate that EyeTAP is a feasible alternative interaction technique. Moreover, it is a viable and effective solution to the Midas touch problem for eye tracking platforms and can be regarded as an alternative to voice recognition technique. EyeTAP showed the similar dispersion on average based on both measures *range*, and *interquartile range* (IQR) with dwell-time as shown in Table 3 and Figure 17b.

However, the range of activation threshold for the dwell-time method is reported in the range of (300-1100 ms) in the literature [165]. Compared to a 500 ms dwell-time, EyeTAP showed acceptable results. To our surprise however, EyeTAP did not generally outperform dwell-time in terms of either time or errors. This may suggest that a well-tuned dwell-time method even on commercial hardware components does not suffer greatly from the Midas touch problem.

EyeTAP showed a lower error rate than the dwell-time in the ribbon-shaped test (see Figure 12c) with relatively large targets. We posit that with larger size targets, the eyes to move around the target causing the dwell-time method to have more errors. Conversely, target size should not impact EyeTAP as much, as the selection is multimodal so as soon as the eye is on target the user can confirm the selection with a sound. These features caused the reduction of wrong selections by users to select relatively large targets in a left-right shift of movements applying the EyeTAP technique. In contrast, selecting smaller-sized targets in different orientations on the screen (360 degrees) of the circle-shaped test (see Figure 13c) caused a larger number of errors for EyeTAP compared to dwell-time and voice recognition. These show that EyeTAP is more suitable to select larger targets with eye movements in opposite directions (left-right, up-down) based on error rates.

EyeTAP is an effective and robust alternative to previous gaze-based interaction techniques. It may be more robust than voice-based techniques and cause less fatigue than the dwell-time method. Based on our study results, we believe it would be particularly useful when there is ambient noise, or users feel uncomfortable speaking out loud, such as the case in a communal workplace.EyeTAP showed a lower variability than the voice recognition technique, and a comparable variability to the dwell-time technique based on dispersion analysis (see Figure 17b) when applied on participants which is beneficial to apply EyeTAP on different users.

Another advantage of EyeTAP relies on its dual-purpose applications for able-bodied and severely disabled users who may not use a voice recognition engine to send their commands and has also difficulties using a dwell-time technique for their basic interaction needs.

Finally, the interesting advantage of EyeTAP lies in its fundamental auditory technique which is based on the Morse code [121] which enables a series of commands based on binary input variables. This feature provides an extension of new commands from simple to complex functionalities which offers a design flexibility for future applications and case scenarios. Although currently, EyeTAP is designed for selection tasks only, its functionalities can be extended. EyeTAP can be considered as a competitive alternative to speech recognition techniques for selection tasks. Furthermore, when users are uncomfortable using a mouth sound (and having the physical capacity to do so), they can tap the microphone to initiate the required acoustic pulse for selection.

3.6.2 Voice Recognition

This interaction method showed relatively acceptable results but suffers from some limitations. In general, a voice recognition engine depends on the user's voice, gender, language, and accent. Additionally, it is not applicable to users with speech impediments. Another drawback is the need of prior training samples to detect words correctly. Furthermore, similar words may lead to false recognition as we experienced during our user study. The quality of the microphone and its distance to the user is also another factor to be considered for this kind of interaction. Regarding the accuracy of recognition, the choice of recognition software plays an important role. Finally, speaking commands out loud may not be suitable in certain working environments.

In general, voice recognition presented some challenges for the users in terms of wrongly recognized words, need for action word repetition, and delay between input and feedback. The subjects' rating of this technique was very low (9.1%) in our user study. Voice recognition showed the highest completion time in the matrix-based test and highest movement time in the circle-shaped test and reached the highest cognitive workload among all interaction techniques.

The lowest error rates in both Fitts' studies reflect that the voice recognition technique is easier to control than EyeTAP and dwell-time to select targets (see Figures 12c and 14b). Voice recognition had the highest selection accuracy measured by the dart-based test. This suggests that it may be a well-suited interaction technique when on small screens and/or with small-sized targets. In addition, the voice recognition technique reached the lowest variability based on our dispersion analysis on distance to target (as shown in Figure 10c and Table 3), and throughput measures (shown in Figures 12b, 15a, and 15b and Table 3) among all eye tracking techniques. The voice recognition technique achieved the highest score based on dispersion analysis as shown in Table 3, and Figure 17b. These show its adaptability on different users which is a useful feature to apply it on a larger population with a predictable performance for suitable case scenarios.

Beelders *et al.* stated that using the dwell-time technique should be more efficient than speech commands [14]. However, we have shown that speech commands have better performance for error rates (see Figures 12c, 14b), selection accuracy (see Figure 10c), and higher consistency on users based on dispersion analysis (see Figure 17b). Zhao *et al.* experienced issues with their voice recognition engine such as speaking words loudly [190], we also had the same difficulties in our experiments. This is one of the challenges of voice recognition engines.

3.6.3 Dwell-Time

The dwell-time method showed the fastest completion time in the matrix-based test, and fastest movement time and highest throughput in both Fitts' experiments due to the low amount of activation time (500 ms). In addition, it reached the lowest amount of cognitive workload. However, it showed the highest error rates in the ribbon-shaped test and with EyeTAP in the circle-shaped test. Moreover, some users complained about eye fatigue after a while during test sessions. Since the dwell-time method relies on the activation time, any changes may produce different results.

We believe that the reason for faster completion time for dwell-time relates to the fact that it has a singular activation function which demands significantly lower cognitive workload (see Figure 16) to select targets at different locations, whereas the multimodal technique relies on mental coordination between both modalities to point and select a target. We posit that the synchronization of these modalities was a major factor in dwell-time outperforming the EyeTAP technique on most measures.

The dwell-time technique showed the lowest variability on task completion time and movement time measures among all eye tracking techniques, but the highest variability on path cost of target selection, throughput of both ribbon-shaped and circle-shaped tests and the highest variability on error rates of the ribbon-shaped test. This method reached similar variability as EyeTAP based on both measures *range*, and *interquartile range* (IQR) as shown in Table 3 and Figure 17b. Except the high error rates for the dwell-time method, it has been shown to be comparable with the mouse interaction for target selections in our studies which makes it still a superb eye tracking interaction technique. However, the EyeTAP technique showed competitive performance compared to the voice recognition technique with promising results. Pfeuffer *et al.* found the dwell-time the fastest technique in their study [141]. We confirm their findings regarding the completion time in our user studies for the dwell-time technique. However, they found dwell-time eye tiring and the least favorable technique by users due to relatively high activation time (1 sec). In contrast, we found the lowest workload for the dwell-time based on the NASA TLX scores (see Figure 16) but had similar feedback about eye fatigue. Since we employed half of the activation threshold used in Pfeuffer *et al.*'s experiment, dwell-time was found to be the easiest and fastest technique among eye tracking techniques in our user studies. In another work for head mounted displays (HMDs), Esteves *et al.* found a dwell-time of 400 ms a faster interaction technique than applying a clicker and speech commands [51]. We confirm their findings based on our user studies' results. Moreover, they found the dwell-time and clicker the most popular interaction techniques by users. We found relatively high error rates for dwell-time in our studies. Esteves *et al.* showed that increasing the activation threshold for dwell-time (400 ms to 1 sec) can decrease error rates to zero. These confirm that the choice of activation threshold is a key factor in applying the dwell-time method which is a trade-off between performance and error rates.

Miniotas *et al.* applied a dwell-time of 1500 ms in their experiments and showed the lowest error rate for that threshold [116]. However, although increasing the dwell-time may reduce error rates, it may also cause eye fatigue as we experienced in our user studies, especially during long-time sessions. The dwell-time method with 500 ms threshold is regarded as the best performing version of dwell-time [103].

3.6.4 The Mouse

We applied the mouse interaction as a baseline technique for comparison with the gaze-based techniques. Overall, we found higher performance for mouse interaction, however, it showed higher pointer movements on the screen (see Figure 10b) than eye tracking techniques. Beelders *et al.* found that mouse interaction has significantly higher performance than eye tracking techniques in the case of throughput and completion time. We confirm these findings, however we also found that in the case of completion time, the dwell-time technique reached similar performance (see Figures 12a, 14a). These show the potentials of a fine-tuned dwell-time technique as an alternative for the mouse.

3.7 Conclusion and Future Work

In this paper, we proposed EyeTAP (Eye tracking point-and-select by Targeted Acoustic Pulse), an eye tracking interface that addresses the Midas touch problem with acoustic input detection capabilities. The performance of the prototype was measured in four user studies with 33 participants based on eight criteria: (1) *completion time*, (2) *path cost of target selection*, (3) *error rate*, (4) *error locations on screen*, (5) *accuracy of target selection*, (6) *movement time*, (7) *throughput*, and (8) *cognitive workload.*

In addition, we performed a statistical analysis based on (1) variance, and (2) dispersion of data. The results of our user studies showed that the dwell-time method outperformed other eye tracking techniques, including EyeTAP on most criteria based on an analysis of variance (ANOVA), but suffers from a high level of distribution on some criteria. At the same time we found that EyeTAP, in comparison to the other tested methods provides a faster task completion time, faster movement time and lower workload than voice recognition. In addition, EyeTAP showed similar performance compared to the dwell-time method and a lower error rate in the ribbon-shaped test.

Moreover, our study showed that eye tracking has a lower footprint (eye gaze mapped with mouse pointer) on the screen compared to a mouse pointer in time scale. Additionally, we confirmed that center regions towards the right and bottom side of the screen are more error prone than the left and top sides. Finally, we developed two user tests (Matrix-based, and Dart-based tests) that would be effective in studying different target selection in gaze-based interaction techniques.

Although we only developed the left mouse click event, EyeTAP demonstrates a completely contact-free alternative to mouse interaction for users with disabilities and users who need to avoid physical contact with input devices considering their workplace or situation. Thus, we believe EyeTAP can be regarded as a competitive technique to both dwell-time, specifically in cases where users may experience physical disabilities or restrictions, and voice recognition, particularly when dealing in workplaces, accents or speech disabilities. EyeTAP showed a higher consistency (lower variability) based on the dispersion analysis, thus it may be more easily accessible to a larger diverse population (e.g. children, users with disabilities, and elderly users).

The global outbreak of COVID-19 showed the importance of contact-free interactions, specifically in public places and for healthcare personnel. The potential of EyeTAP can be considered on public devices such as ATM machines and self check-in platforms at airports. We hope, that EyeTAP inspires researchers into developing contact-free interaction techniques for emerging case scenarios and equipment. In future work, we will apply the EyeTAP technique on AR/VR headsets to measure its usability in different case scenarios for able-bodied and participants with motor disabilities.

Chapter 4

FELiX: Fixation-based Eye Fatigue Load Index A Multi-factor Measure for Gaze-based Interactions

Preface

As we spend more time on digital displays on a daily basis, it is very important to measure the amount of workload or eye-strain on our eyes. Such a measurement would enable researchers to compare different interaction techniques in user studies. Previous metrics used to determine for eye-strain [1, 92, 8, 113, 41], are not always appropriate or effective and some require expensive eye tracking sensors. Thus, here we propose and evaluate a model (FELiX) comprised of both *objective* and *subjective* criteria to be applied in user studies with two variations *performance-based* and *accuracy-based* for different test conditions. FELiX with its variations takes users' feedback into account which is useful for consumers of digital contents to select a software product (fully-functioning interaction technique) with the lowest amount of eye-strain. FELiX can be applied using budget-friendly eye tracking sensors which is suitable for many use case scenarios in the research community. In addition, FELiX can be applied as an alternative measurement technique for biological sensors to measure eye-strain and work-related stress. In testing the developed model, we compared the dwell-time technique with a voice-recognition technique.

This Chapter is based on a paper was presented at the 2020 13th International Conference on Human System Interaction (HSI), was a *Best Paper Finalist, and was published in the proceedings of the conference. The results of this chapter were the fundamental motivation of our next models regarding eye-strain on digital media presented in Chapters 5 and 6.

Abstract

Eye fatigue is a common challenge in eye tracking applications caused by physical and/or mental triggers. Its impact should be analyzed in eye tracking applications, especially for the dwell-time method. As emerging interaction techniques become more sophisticated, their impacts should be analyzed based on various aspects. We propose a novel compound measure for gaze-based interaction techniques that integrates subjective NASA TLX scores with objective measurements of eye movement fixation points. The measure includes two variations depending on the importance of (a) performance, and (b) accuracy, for measuring potential eye fatigue for eye tracking interactions. These variations enable researchers to compare eye tracking techniques on different criteria. We evaluated our measure in two user studies with 33 participants and report on the results of comparing dwell-time and gaze-based selection using voice recognition techniques.

4.1 Introduction

4.1.1 Cognitive Workload

Cognitive workload is defined as the amount of mental effort of a person performing a task or in the process of problem-solving. It is related to a person's working memory which has a limited capacity [30, 173]. It is important to measure the amount of cognitive workload related to performing a task given a specific interface in order to compare the usability of different systems. The NASA Task Load Index (TLX) questionnaire is a well-known multidimensional method used to measure subjects' perceived workload in user studies [64, 63]. The TLX questionnaire, which has been shown to be a valid tool to measure workload [152], comprises of six scales: (1) physical demand, (2) mental demand, (3) temporal demand, (4) effort, (5) performance and

(6) frustration, each on a 100-point range with 5-point steps [59]. Each scale can be weighted based on its importance and used to calculate the average value - known as the *overall workload*. The overall workload serves as a measure of the efficacy of the interaction technique and can be used for comparing different methods based on their workload. However, the results of the NASA TLX are subjective and suffer from several limitations. One such limitation is that subjects often confound task performance with the perceived mental effort. Furthermore, as the results are obtained after a task is completed so as not to interrupt the task, the NASA TLX is not ideally suited for real-time scenarios [189]. For these reasons, more robust and accurate methods should be applied for measuring cognitive load, such as the use of physiological data [6]. Researchers are thus beginning to investigate the use of physiological signals, for example by measuring brain activity. Techniques for measuring brain activity include: (1) Electroencephalography (EEG) which detects brain waves [42], (2) Magnetoencephalography (MEG) that records magnetic fields of electrical activities in brain [43], and (3) Near-infrared spectroscopy (NIRS) which is a spectroscopic method that uses wavelengths in the near-infrared range to measure blood flow changes in the frontal cortex [74]. These methods although accurate in detecting brain activity require specialized and sometimes cumbersome equipment. In addition, these techniques are intrusive for users and therefore are restricted to controlled environments such as laboratories [189].

4.1.2 Eye Fatigue

According to Vasiljevas *et al.*, fatigue is the increase of tiredness of a subject under load [178] and can be grouped into physical, e.g. lack of sleep, and mental related causes such as stress [66]. According to Marcora *et al.*, mental fatigue is the result of high cognitive activity [109]. Visual fatigue defined as “eyestrain or asthenopia, which can be caused by both two-dimensional and stereoscopic moving images" [78] and which can cause motion sickness [94], occurs when focusing on near objects. The visual function of the eyes may cause visual fatigue, especially in long-time periods. Other symptoms of visual fatigue include: tiredness, headaches, and irritation of the eyes [176]. In this paper, we propose an integrated measure to detect task load and visual fatigue during gaze-based interactions. Our focus is on visual fatigue as it is a common issue among computer users due to the prolonged periods of time they spend working in front of a monitor[178]. We believe *that a comprehensive measure that combines the quantitative aspects of eye tracking fixation points with the qualitative aspects of the NASA TLX scores could provide an effective means to distinguish task load and fatigue in different gaze-based interactions.* The developed measure is an alternative to using sensory devices in situations where the application of biological sensors are either not possible, or cumbersome to participants for user studies.

The contribution of this paper is twofold. Firstly, we introduce FELiX: Fixation-based Eye Fatigue Load Index, an integrated measure of task workload and visual fatigue. The term *eye fatigue load* is defined as a combined measure of task workload and visual fatigue. The FELiX measure combines the accuracy of the objective eye tracking data (quantitative inputs) with the subjectivity of user's experience as calculated by the NASA TLX scores (qualitative inputs) during gaze-based interactions. Secondly, we investigate the ability of FELiX to measure eye fatigue load by conducting two user studies comparing two gaze-based interaction techniques: dwell time and voice recognition. The results of our studies show that FELiX is able to distinguish between different gaze-based interaction methods.

4.2 Related Work

Researchers proposed various measures to measure eye fatigue based on either eye movement analysis or biological sensor inputs. Zheng *et al.* investigated the correlation between eye blinks and mental workload among surgeons. They found that shorter blink duration and frequency indicate an increase of the mental workload [191]. Additionally, Borghini *et al.* studied brain activity and heart rate of car drivers and found the same results regarding the eye blink rates with mental workload [18]. Lanthier *et al.* studied the correlation between fixations and eye fatigue during visual search tasks and found that fixation duration increases with fatigue [97]. Abdulin *et al.* showed that the distance drift of fixation points in response to a stimuli can reveal physical eye fatigue [1] and calculated this using the fixation qualitative score (FQlS) [92]. Vasiljevas *et al.* examined an analytical model of muscle fatigue proposed to measure athletes fatigue [22] and adopted it to assess eye fatigue in gaze-based tasks [178]. In studying the impact of learning on fatigue, they found that the required break time for gaze-based interactions can be measured. Researchers have also applied self-evaluation questionnaires to evaluate eye fatigue in user studies for gaze-based applications [105]. There are saccades-based approaches to measure eye fatigue [8, 113, 41]. However, according to Abdulin *et al.*, analysis of saccades raw data requires expensive eye trackers, and these approaches are not applicable on budget-friendly devices [1]. Building on the previous works, we propose a fixation-based approach which can be applied on most eye trackers. Although, previous measures can be used to measure eye fatigue with high probability, they rely solely on eye movements or sensor inputs. To the best of our knowledge, there are currently no measures that integrate NASA TLX scores with the measurements of eye movements using eye tracking to assess eye fatigue. To take advantage of both physiological data and user perceptions, we integrate eye movements (fixation points) as an objective measure, and NASA TLX scores as a subjective measure in FELiX. By combining workload and eye fatigue in one measure, FELiX is ideally suited to compare different interaction techniques in gaze-based interaction user studies.

4.3 Eye Fatigue Load Index (FELiX)

We propose two variations of FELiX, both of which integrate the beneficial features (simplicity and direct ratings by users) of the NASA TLX with gaze fixations to measure eye fatigue load. Out of the six TLX questionnaire scales, we only employ scores for the following three scales: *physical demand* (PD), *mental demand* (MD), and *performance* (P). This choice is based on the fact that the *physical* and *mental* demands best describe the concept of workload to users, whereas *performance* is best interpreted by the users as the overall performance of the method. In contrast, the other three scales (e.g. temporal demand, frustration, effort) focus on usability and user satisfaction. The first variation of the proposed measure $FELiX_{per}$ incorporates fixations recorded (x, y, *timestamp*) during a gaze-based test as well as the error rates of target selections and can be used in experiments where performance is of high importance. On the other hand, if accuracy is of higher importance, the second variation $FELiX_{acc}$ can be used, which incorporates the Euclidean distance to the target as well as the number of fixations. The proposed measures, *performance-based* and *accuracy-based*, measure the eye fatigue load for any gaze-based interactions relying on eye movement measurements.

4.3.1 Cognitive and Eye-Tracking Coefficients

FELiX involves two coefficients, namely *cognitive* and *eye-tracking* coefficients. The *cognitive coefficient* is a qualitative factor which is calculated based on the users' rating scores of the NASA TLX questionnaire for the scales PD, MD, P (rated on a scale of 1-100). The *eye tracking coefficient* is a quantitative factor which is calculated from the eye-tracking data recorded during the test session. Since the recorded values used in calculating the eye-tracking coefficient can vary depending on the test conditions, we use the logarithmic function to scale down to a lower range the potentially large index values. Furthermore, to avoid cases where one coefficient diminishes the effect of the other (e.g. eye-tracking coefficient or cognitive coefficient is close to zero), we offset the coefficients by 1 and 9 respectively, such that the lowest value is ≥ 1 as explained below. We applied similar parameters introduced by previous research based on saccades [8, 113, 41], and fixation analysis such as *average fixation duration time* (AFD), and *average number of fixations* (ANF), as proposed by Komogortsev *et al.* [92].

4.3.2 Performance-based FELiX ($FELiX_{per}$)

Equation 2 shows the formula for the first variation of the measure, $FELiX_{per}$. This measure can be used for calculating the eye fatigue load index for interaction techniques and is dependent on the following parameters:

- average fixation duration time (AFD),
- error rate (ER) which is the total number of error selections divided by the total number of targets,
- average number of fixations (ANF), and
- NASA TLX questionnaire (3 scores: PD, MD, P)

The conditions and range of each of the parameters are given by,

1. $PD = \{a \mid a \in \mathbb{Z} \land 1 \leq a \leq 100\}$,

   $MD = \{b \mid b \in \mathbb{Z} \land 1 \leq b \leq 100\}$,

   $P = \{c \mid c \in \mathbb{Z} \land 1 \leq c \leq 100\}$

   TLX scores are integers in range of 1 to 100.

2. $ER \in \mathbb{R} \land 0 \leq ER \leq 1$

   The error rate is a real number from 0 to 1.

3. $(ER \times P) \in \mathbb{R} \land 0 \leq (ER \times P) \leq 100$

   The product of error rate and performance score is a real number from 0 to 100.

4. $CC_{per} \in \mathbb{R} \land 1 \leq CC_{per} \leq 200$

   The cognitive coefficient CC_{per} (equation 1) is the average of TLX scores (PD, MD) added to the product of error rate and performance score (P) which results in a real number from 1 to 200. This coefficient reflects the increase of task workload by multiplying the error rate factor. In the case of an error-free condition, the performance factor is removed to lower the cognitive coefficient.

$$CC_{per} = (\frac{PD + MD}{2}) + (ER \times P) \tag{1}$$

5. $\left(\frac{ANF}{AFD}\right) \in \mathbb{R}_{>0}$

   The eye tracking coefficient is comprised of the average number of fixations (ANF) divided by average fixation duration time (AFD) which results in a positive real number greater than 0. This measure reflects the duration of fixation points on average.

6. $FELiX_{per} \in \mathbb{R} \land FELiX_{per} \geq 1$

   $FELiX_{per}$ (equation 2) is the product of (a) logarithm of cognitive coefficient CC_{per} with the fixed constant value 9 in base 10, and (b) the eye tracking coefficient $\frac{ANF}{AFD}$ with the fixed constant value 1 which results in a real number greater or equal than 1.

$$\boxed{FELiX_{per} = \log_{10}(9 + \underbrace{CC_{per}}_{\text{cog. coeff.}}) \times (1 + \underbrace{\frac{ANF}{AFD}}_{\text{eye-track. coeff.}})} \tag{2}$$

4.3.3 Accuracy-based FELiX ($FELiX_{acc}$)

Equation 4 shows the formula for the second variation of the measure, $FELiX_{acc}$. The measure can be used to calculate the eye fatigue load index for interaction techniques where accuracy is of utmost importance i.e. distance to target, such as in target selection tasks. $FELiX_{acc}$ is dependent on the parameters:

- average number of fixations (ANF),
- average Euclidean distance to the target (ADT), and
- NASA TLX questionnaire (2 scores: PD, MD) as described above.

The distance (ADT) is measured as the difference between the 2D coordinates of the center of a target and the coordinates of the corresponding fixation point. The conditions and ranges of each of the parameters are defined as,

1. $PD = \{a \mid a \in \mathbb{Z} \land 1 \leq a \leq 100\}$,
   $MD = \{b \mid b \in \mathbb{Z} \land 1 \leq b \leq 100\}$
   TLX scores are integers in range of 1 to 100.

2. $\left(\frac{ANF}{ADT}\right) \in \mathbb{R}_{>0}$
   The eye tracking coefficient is comprised of the average number of fixations (ANF) divided by average Euclidean distance to the target (ADT) which results in a positive real number greater than 0. This measure reflects the distance of fixation points to the target on average.

3. $CC_{acc} \in \mathbb{R} \wedge 1 \leq CC_{acc} \leq 100$

   The cognitive coefficient CC_{acc} (equation 3) is the average of TLX scores PD and MD which results in a positive real number between 1 and 100.

$$CC_{acc} = \frac{PD + MD}{2} \tag{3}$$

4. $FELiX_{acc} \in \mathbb{R} \wedge FELiX_{acc} \geq 1$

   $FELiX_{acc}$ (equation 4) is the product of (a) logarithm of cognitive coefficient CC_{acc} with the fixed constant value 9 in base 10, and (b) the eye tracking coefficient $\frac{ANF}{ADT}$ with the fixed constant value 1 which results in a real number greater or equal than 1.

$$\boxed{FELiX_{acc} = \log_{10}(9 + \underbrace{CC_{acc}}_{\text{cog. coeff.}}) \times (1 + \underbrace{\frac{ANF}{ADT}}_{\text{eye-track. coeff.}})} \tag{4}$$

4.3.4 Discussion: Rational of FELiX

We employed quantitative parameters typically recorded in eye tracking applications in our measure since they reflect technical workflow of an interaction technique. These technical parameters are bound to test applications and equipment. Additionally, we applied workload parameters obtained from the NASA TLX scores to include direct ratings of participants who were involved in the practical aspects of an interaction technique. The proposed measure should result in a single value based on both technical and empirical parameters regarding the available measures. The purpose of multiplication of both coefficients (quantitative and qualitative) is to control the influence of both coefficients. In fact, the proposed measure should be balanced in the way that no aspects of an interaction technique (technical or empirical) can undermine the impact of the other.

4.4 Methodology

To evaluate the effectiveness of the proposed measures we calculated the FELiX measure based on two gaze-based interaction studies with 33 participants (13 female, from 22 to 35 years old, $SD = 2.96$). All subjects partook in both experiments. The equipment is illustrated in Figure 18a.

4.4.1 Interaction Methods

Dwell-time

The dwell-time method integrates both pointing and selection phases using the eye tracker only. The range of dwell-time has been between 300-1100 milliseconds for target selection in the literature [165]. We defined the target activation threshold to 500 milliseconds, since it showed the best performance in [103] and participants prefer dwell-times of around 500 ms[165]. In other words, the target was considered as selected when a subject focused on it for 0.5 seconds; if the subject moved their gaze away from the target prior to the 0.5 seconds the selection process would restart.

Eye Tracking with Voice Recognition

For voice recognition, eye tracking was used for pointing and voice for selection. The selection phase for the voice recognition technique is triggered by a voice command which in our case was the word 'select' that was interpreted as a mouse click. The voice command is captured by a headset microphone (Logitech H370). An artificial ambient noise was introduced in the background through stereo desktop speakers at a volume of 50 dB to simulate a typical work environment. The method was developed using the built-in Windows 10 speech recognition capabilities available in the .NET framework. We implemented a C# application to respond to the activation keyword 'select' to trigger a mouse click.

4.4.2 Hypotheses

Based on the previous literature, which has demonstrated dwell-time to be one of the most effective gaze-based interaction techniques [172], but one which can suffer from issues related to Midas touch [76], we hypothesized that:

1. The accuracy-based FELiX ($FELiX_{acc}$) will be lower for dwell-time than voice recognition because dwell-time should have lower fixation distances to target ($\frac{ANF}{ADT}$), as well as, lower physical demand (PD) and mental demand (MD).
2. The performance-based FELiX ($FELiX_{per}$) will be higher for dwell-time than voice recognition because dwell-time tends to result in more errors due to Midas touch and should have higher duration of fixation points ($\frac{ANF}{AFD}$).
3. The analysis of both FELiX variations will allow us to distinguish dwell-time and a multi-modal interaction technique.

4.4.3 Apparatus

In our user study, the mouse pointer position is captured using the Tobii 4C eye tracker[1]. All test applications were developed and the user studies were run on a commodity computer system: 64-bit Windows 10 PC with Intel i7 2.67GHz CPU, 12 GB RAM, 1 TB hard disk and NVIDIA GeForce GTX 770 graphics card. Figure 18a shows the required equipment of both interaction techniques.

[1] https://tobiigaming.com/product/tobii-eye-tracker-4c/

Eye Tracking: Pointing Phase

The Tobii SDK (TobiiEyeXSdk-Cpp-1.8.498) supports different events related to eye tracking activities such as the location of the current eye gaze, positions of both eyes, fixation points, and user presence in front of the eye tracker. We employed the eye gaze library (API) to obtain users' gaze locations. These locations show the current gaze position on the screen in pixel coordinates. The SDK supports eye movements in a 3D coordinate system (horizontal, vertical, depth). However, we applied a 2D coordinate system (x,y) combined with a unique timestamp corresponding to the recorded location such that the mouse cursor was synchronized with the gaze positions to control the mouse pointer on the screen. Eye-tracking for both user studies was developed in C++ and integrated as a new plug-in into the Tobii SDK. The samples were recorded in distance of 60 cm (23.6 in) to the eye tracker with the sampling rate of 90 Hz.

Voice Processing: Selection Phase

To simulate a click on the item to be selected a headset microphone listens to the user while suppressing the background ambient sounds/noise in real-time. The Windows 10 Speech Recognition engine (available in the .NET framework) was selected to parse the received commands and a C# program was developed to trigger a left mouse click.

4.4.4 Experimental Design

Prior to running the studies, subjects were informed about the purpose of the study, trained on each of the methods to be tested, and participated in a pre-test questionnaire inquiring on their background in the fields of eye tracking, voice recognition technologies and their preferred kind of interaction. After the pre-test questionnaire the Tobii calibration software was used to calibrate the system for each participant before starting the study. During the study each user partook in two experiments with different stimuli: (1) matrix-based and (2) dart-based. Overall, the studies took 8 minutes on average for each participants, 6 minutes for the matrix-based, and 2 minutes for the dart-based test.

4.4.5 User Study 1: Matrix-based Test

In the first experiment, a matrix of buttons (targets), were randomly distributed across the screen. The task of the subjects was to point and click on buttons shown on the screen in increasing numerical order for various levels of difficulty from 1 (easy) to 5 (hard), described in detail below. The level of difficultly was presented in ascending order. Further, the transition from lower levels to higher levels was done automatically, thus the whole test session for each participant was continuous.

Stimulus

The stimulus consisted of 77 buttons (11 columns × 7 rows) in size of 110 × 80 pixels, some labeled with numbers and others not, which covered the entire screen at a resolution of 1920 × 1080 pixels on a Dell P2411Hb monitor. Two marginal columns (far left, far right) and two rows (top, bottom) were removed from the active selection due to the high difficulty to be selected by users during the pilot-test. Buttons that were not labeled are considered as *barriers* or *distractions*. To provide feedback to the subject, labeled buttons change color after the user has successfully pointed and selected on the correct button. Wrongly selected barriers (buttons with no label) are highlighted in red. The level of difficulty of the stimulus was also increased across subject trials. This was done by increasing the number of targets that had to be selected by the subject. Five levels of difficulty were used for each interaction method: level 1 (4 targets), level 2 (6 targets), level 3 (8 targets), level 4 (10 targets) and level 5 (12 targets). Targets were randomly distributed over the entire screen for each level. Figure 18b shows the matrix-based test during difficulty level 5.

Measures

The following variables were recorded: *fixation duration time*, *number of fixations*, *error rates*, and *subjective ratings* (based on the NASA TLX scores). An internal logging module recorded subjects' actions, fixation duration times, wrongly selected targets, as well as the number of fixations per each method.

4.4.6 User Study 2: Dart-based Test

In this experiment the subject was to select, as accurately as possible, the bull's-eye of a dart target using each interaction method. In order to take into consideration the fact that eye tracking has different accuracy in different regions of the monitor [54], we computed an average value based on five trials for each interaction method where the stimulus was shown at different areas of the screen near the center of the screen randomly. Each new randomly chosen trial began two seconds after selection of the previous target, allowing users time to change their gaze and to focus on the new target. For the dwell-time method, a countdown (5 to 0) representing remaining 100 milliseconds was displayed during the selection phase and users needed to focus on the dart shape before this time was up.

Stimulus

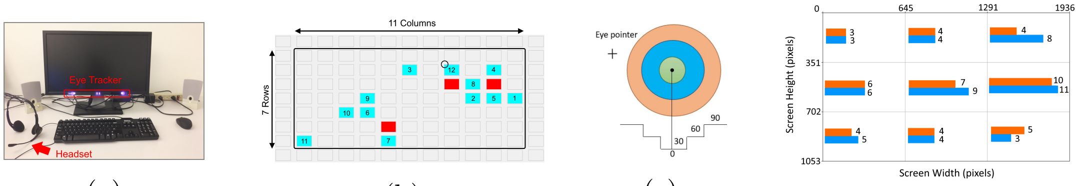
The stimulus for this experiment consisted of a dart-like target with three circles: green (0 to 30 pixels radius), blue (30 to 60 pixels radius) and red (60 to 90 pixels radius) as in Figure 18c. Points within the center area i.e. green have the lowest range of distances to the bulls-eye; each other co-centric circle has a larger range of distance values. Any point lying outside the three co-centric circular areas is considered as having a fixed maximum distance of 90 pixels. For this experiment, a cross-hair icon was used.

Measures

The purpose of this test was to measure the selected point's distance on the dart target to the center of the core circle (in green), thus the accuracy is measured in pixels. The distance between the selected location and the center of the stimulus is calculated based on the Euclidean distance. Since the measured trials are chosen randomly, the average is calculated to compare the two different methods based on accurate selection. In addition, the number of fixation points for each method was recorded.

4.4.7 Test Workflow

The order of interaction methods was randomly selected for each participant. At the end of the two studies subjects filled out a post-test questionnaire, which among other questions consisted of the NASA TLX questionnaire [59].



(a)

(b)

(c)

(d)

Figure 18: (a) shows test setting and equipment for both user studies. (b) shows the matrix-based test. The red button represents an error selection. The circle on number 12 represents the eye pointer. (c) shows the Dart-based test stimuli, and (d) shows error locations on screen. Orange bars represent total number of errors for voice recognition, and blue bars for dwell-time method.

4.5 Results

We analyzed the results of our experiments using an analysis of variance (ANOVA) followed by Bonferroni posthoc tests with the JASP 0.11.1 software[2].

4.5.1 User Study 1: Matrix-based Test

A one-way repeated measure ANOVA was performed to examine the effect of interaction type on (1) number of fixations, (2) fixation duration time, (3) error rate, and (4) eye fatigue load index. Since we calculate average values on the entire test session for each participant, we can ignore the difficulty level factor in the analysis and take the total number of targets (40) into account.

Number of fixations

We found a significant effect of interaction method on average number of fixations (F(1,32)=7.79, $p < .05$). A posthoc Bonferroni comparison test showed a significant difference between dwell-time ($M = 262.97\ fixations$, $SE = 34.06\ fixations$) and voice recognition ($M = 425.84\ fixations$, $SE = 68.75\ fixations$).

Fixation duration time

We found a significant effect of interaction method on average fixation duration (F(1,32)=32.93, $p < .001$). A posthoc Bonferroni comparison test showed a significant difference between dwell-time ($M = 16.52\ sec$, $SE = 1.32\ sec$) and voice recognition ($M = 39.77\ sec$, $SE = 3.97\ sec$).

[2] https://jasp-stats.org/

Error rate

We found a significant effect of interaction method on error rate (F(1,32)=5.26, $p < .05$). A posthoc Bonferroni comparison test showed a significant difference between dwell-time ($M = 0.12\ errors$, $SE = 0.03\ errors$) and voice recognition ($M = 0.05\ errors$, $SE = 0.01\ errors$). Table 4 summarizes test results of the Matrix-based test.

Error locations on screen

Previous research has shown that the right side of a monitor has lower precision for eye tracking applications [54]. We studied the regions of the screen in regard to errors. We divided the screen size into nine equally-sized squares and counted the number of errors occurring in each location. In our study, errors are defined as wrongly selected targets (depicted in red in Figure 18b). Errors on the borders were counted for all adjacent regions. For instance, errors which occur in two regions are counted as occurring in both regions. Figure 18d illustrates the total number of errors for all participants for both interaction techniques.

Eye fatigue load index (performance-based)

We found a significant effect of interaction method on our eye fatigue load index (F(1,32)=24.09, $p < .001$). A posthoc Bonferroni comparison test showed a significant difference between dwell-time ($M = 17.24$, $SE = 1.2$) and voice recognition ($M = 11.85$, $SE = 0.94$). Figure 20a illustrates the calculated performance-based eye fatigue load index for the Matrix test. This confirms our second hypothesis that $FELiX_{per}$ is higher for dwell-time than voice recognition.

	Dwell-Time	Voice Recog.	Sig.
Mean number of fixations	262.97	425.84	$p < .05$
Mean fixation duration (sec.)	16.52	39.77	$p < .001$
Error rate	0.12	0.05	$p < .05$

Table 4: Test results of the Matrix-based test. Dwell-Time caused significantly more errors as expected.

4.5.2 User Study 2: Dart-based Test

A one-way repeated measure ANOVA was performed to examine the effect of interaction type on (1) number of fixations, (2) average distance to target, and (3) eye fatigue load index.

Number of fixations

We found a significant effect of interaction method on average number of fixations (F(1,32)=26.38, $p < .001$). A posthoc Bonferroni comparison test showed a significant difference between dwell-time ($M = 455.52\ fixations$, $SE = 1.71\ fixations$) and voice recognition ($M = 1379.66\ fixations$, $SE = 179.17\ fixations$).

Average distance to target

We found a significant effect of interaction method on average distance to target (F(1,32)=8.33, $p < .05$). A posthoc Bonferroni comparison test showed a significant difference between dwell-time ($M = 35.30\ pixels$, $SE = 2.11\ pixels$) and voice recognition ($M = 29.27\ pixels$, $SE = 2.07\ pixels$). Since our accuracy-based FELiX (see equation 4) calculates the average distance to target in its eye tracking coefficient ($\frac{ANF}{ADT}$), it is similar with the FQlS measure [92] in measuring distance to target. In comparing the two measures, we found that $FELiX_{acc}$ decreases when the distance to target increases. In other words, higher distance to the target (lower accuracy) is associated with lower eye fatigue (Figure 19a). On the contrary, $FELiX_{per}$ increases with distance to target (Figure 19b). Table 5 summarizes the test results of the Dart-based test.

	Dwell-Time	Voice Recog.	Sig.
Mean number of fixations	455.52	1379.66	$p < .001$
Average distance to target	35.30	29.27	$p < .05$

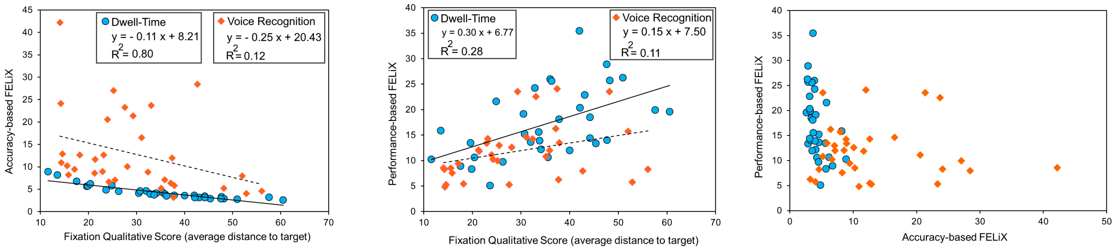
Table 5: Test results of the Dart-based test. Dwell-Time reached significantly lower number of fixations as expected.

Eye fatigue load index (accuracy-based)

We found a significant effect of interaction method on eye fatigue load index (F(1,32)=31.74, $p < .001$). A posthoc Bonferroni comparison test showed a significant difference between dwell-time ($M = 4.28$, $SE = 0.26$) and voice recognition ($M = 12.96$, $SE = 1.53$). Figure 20b illustrates the calculated accuracy-based eye fatigue load index for the Dart test. This result confirms our first hypothesis that dwell-time has a lower $FELiX_{accc}$ score than voice recognition.

4.5.3 Bi-variate Comparison

We proposed two variations on different criteria (performance and accuracy). Each interaction technique can be analyzed on both measures. A one-way repeated measure ANOVA was performed to examine the effect of interaction type on the mean of both FELiX variations. We found no significant effect of interaction method on bivariate eye fatigue load index (F(1,32)=3.77, $p > .05$). A posthoc Bonferroni comparison test showed no significant difference between dwell-time ($M = 10.76$, $SE = 0.53$) and voice recognition ($M = 12.40$, $SE = 0.86$). Figure 20c illustrates the calculated bivariate (performance-, and accuracy-based) average of eye fatigue load index. Table 6 summarizes calculated FELiX values on both criteria.



(a) (b) (c)

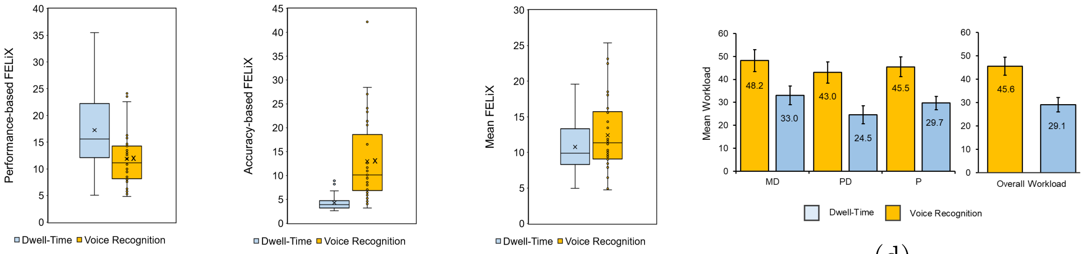
Figure 19: Correlations of the accuracy-based (a) and performance-based (b) FELiX with fixation qualitative score (FQlS), for 33 participants. Dashed lines represent regression through voice recognition and solid lines through dwell-time. (c) shows eye fatigue load index on both variations. Voice recognition technique shows sparse values on both variations.

	Dwell-Time	Voice Recog.	Sig.
$FELiX_{per}$	17.24	11.85	$p < .001$
$FELiX_{acc}$	4.28	12.96	$p < .001$
Bi-variate FELiX	10.76	12.40	$p > .05$

Table 6: Test results of FELiX calculations. Dwell-Time caused significantly higher eye fatigue based on performance and lower eye fatigue based on accuracy as expected.

4.5.4 NASA TLX Scores

Figure 20d shows the required NASA TLX scores by FELiX variations from the post-test questionnaire.



(a) (b) (c) (d)

Figure 20: (a) shows performance-based eye fatigue load index for the Matrix test ($p < .001$), and (b) shows accuracy-based eye fatigue load index for the Dart test ($p < .001$). (c) shows the calculated mean of both variations ($p > .05$). The cross symbols show mean, and the horizontal lines show median points. (d) shows NASA TLX scores. Error bars represent standard error.

4.6 Discussion

The results indicate that the developed multi-factor simple-to-calculate measure, which is solely dependent on the recorded data of a user study, can be used to accurately assess the amount of eye fatigue on participants based on available measures and NASA TLX scores. Further, we showed how to compare our measures with the available FQlS measure and illustrated the correlations between them (Figures 19a, 19b).

Although we only studied voice recognition as a multi-modal gaze-based interaction technique, the dwell-time results confirmed our assumptions that it results in a lower number of fixations and lower fixation duration time compared to a multi-modal interaction technique. Although dwell-time showed lower accuracy (higher distance to the target) than voice recognition (see Table 5), it reached significantly lower eye fatigue based on accuracy (see Table 6 and Figure 20b) confirming our first hypothesis that $FELiX_{acc}$ is lower for dwell-time. This is due to a significantly lower number of fixations (Table 5) and lower TLX scores (Figure 20d) for dwell-time. The higher distance to the target for dwell-time is due to the activation threshold which bounds a user's decision time into a limited time window to respond to target movements. The results of the performance-based FELiX depicted in Figure 20a shows higher eye fatigue for the dwell-time technique. This is due to higher error rate and higher duration of fixation points ($\frac{ANF}{AFD}$) of dwell-time as expected (see table 4). This confirms our second hypothesis that $FELiX_{per}$ is higher for dwell-time than voice recognition. Although the bivariate comparison of both FELiX variations (Figure 20c) shows relatively lower eye fatigue for dwell-time, the difference is statistically not significant (see Table 6).

Additionally, Figure 19c shows distinctive clusters of dwell-time and voice recognition techniques based on FELiX variations and reflects the potential of FELiX measure to analyze similar eye tracking techniques based on their eye fatigue values, and therefore our third hypothesis that dwell-time can be distinguished from a multi-modal interaction technique based on FELiX variations is confirmed. We believe that these results would generalize, and that FELiX is an effective means of determining eye fatigue load and can differentiate different gaze-based interaction methods based on their tendencies to cause the user more discomfort in terms of visual fatigue and task load. We also studied the role of target locations on screen and their relation with error rate and eye fatigue. As illustrated in Figure 18d, the middle row of the screen, towards the right side, has higher eye fatigue potential according to the performance-based FELiX as these regions produced higher errors. Since we applied no biological sensor devices in our user studies, we could not compare the results to study the correlations between our proposed measure and physiological data. We leave this for future work. We did, however, demonstrate that FELiX is an alternative measure to be used in user studies with no access to electronic sensors.

Although the eye tracking parameters involved in FELiX measure can be analyzed individually, the emerging interaction devices offer a variety of quantitative parameters. Therefore, the application of different parameters may be difficult to compare different techniques. The analysis of our results indicates the potential of our multi-aspect evaluation measure on two similar interaction techniques. This experiment provides new insight into the feasibility of multi-factor compound evaluation measures for gaze-based interactions.

4.7 Conclusion and Future Work

As emerging interaction techniques become more sophisticated and multi-dimensional, the need for more complex and multi-factor measures is necessary. Therefore, we propose *fixation-based eye fatigue load index* (FELiX), a compound evaluation measure for gaze-based interactions based on the NASA TLX scores and recorded eye tracking data. Our measure combines the quantitative (technical) and qualitative (empirical) aspects of interaction techniques in a simple-to-calculate measure. Since NASA TLX scores are very common in user studies, we can take benefit of its simplicity to assess cognitive workload of different interaction techniques on the same tasks.

FELiX includes two variations to measure visual eye fatigue based on (a) performance, and (b) accuracy. These measures enable researchers to compare different eye tracking techniques, specially dwell-time and multi-modal techniques, based on eye fatigue load index on different criteria. The *performance-based* measure can be applied when the duration of the entire fixation sequences and the error rates of target selection are recorded, and the *accuracy-based* measure is applicable for case scenarios where distance to target (selection accuracy) is available in the analysis process and can be measured in user studies. Both measures take benefit of three scores from the NASA TLX, (a) physical demand, (b) mental demand, and (c) performance. The application of the proposed measures can be regarded as a feasible alternative to biological sensor inputs or to adopt gaze-based applications for children, users with disabilities or elderly users to assess the amount of eye fatigue in user studies before final release of eye tracking applications.

In addition, we presented an in-depth analysis of the dwell-time method as the most common gaze-based interaction technique with different approaches. As well as developing measures for eye fatigue load, we proposed two test applications to analyze eye tracking applications. In future work, we plan on applying the proposed eye fatigue measures on VR headsets with integrated eye trackers to study motion sickness in VR applications.

Chapter 5

IDEA: Index of Difficulty for Eye tracking Applications An Analysis Model for Target Selection Tasks

Preface

In the following chapter, we propose a new analysis model to measure difficulty of task selection tasks in eye tracking applications, a next phase of our FELiX model presented in the previous chapter to measure eye-strain. Both models take users' feedback into account via the NASA Task Load Index (TLX) which is a valid tool for user studies. In the FELiX model we focus on eye-strain, here although, the IDEA model is specifically proposed to measure task difficulty, it can also be applied to measure eye-strain in user studies. In addition, it takes benefit from the well-known Fitts' law [57] for comparison between user interface concepts. Further, IDEA can be applied to study the Midas touch problem [83] which, as previously described, is a common issue in eye tracking interaction techniques. Although there are related works regarding applications of Fitts' law [89, 61, 75, 191] or the NASA TLX [89, 61, 75, 191], there is no suitable model to combine both measures into a single simple-to-calculate model for eye tracking applications. IDEA is specifically useful to adopt eye tracking applications on different target groups such as children or the elderly. IDEA can be regarded as an extension of the Fitts' law to be applied in emerging case scenarios with eye tracking selection techniques. Although the IDEA model was developed for eye tracking interactions, it can also be applied on various selection techniques to measure difficulty levels based on test conditions (target size and distance) and subjective ratings. The following Chapter, which describes the IDEA concept, is based on a paper that was presented at the International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP HUCAPP).

Abstract

Fitts' law is a prediction model to measure the difficulty level of target selection for pointing devices. However, emerging devices and interaction techniques require more flexible parameters to adopt the original Fitts' law to new circumstances and case scenarios. We propose *Index of Difficulty for Eye tracking Applications* (IDEA) which integrates Fitts' law with users' feedback from the NASA TLX to measure the difficulty of target selection. The COVID-19 pandemic has shown the necessity of contact-free interactions on public and shared devices, thus in this work, we aim to propose a model for evaluating contact-free interaction techniques, which can accurately measure the difficulty of eye tracking applications and can be adapted to children, users with disabilities, and elderly without requiring the acquisition of physiological sensory data. We tested the IDEA model using data from a three-part user study with 33 participants that compared two eye tracking selection techniques, dwell-time, and a multi-modal eye tracking technique using voice commands.

5.1 Introduction

In this paper we introduce IDEA: Index of Difficulty for Eye tracking Applications, an integrated prediction model of task workload and performance of target selection tasks. The IDEA model combines the effective contact-free target selection of eye tracking with direct feedback of user's experience obtained from the NASA TLX scores. The IDEA model calculates a prediction index value based on objective technical specifications such as the target's size and distance, and subjective measures from the NASA TLX questionnaire obtained from user studies. To demonstrate the efficacy of IDEA, we measured target selection performance with data from three user studies that compared two eye tracking interaction techniques (dwell-time, and selection by voice commands) and showed that our predictions correlate with *throughput* and *movement time* of the Fitts' prediction model.

5.1.1 Fitts' Law

Paul Morris Fitts introduced a mathematical prediction model to measure the difficulty level of target selection in 1954 [57]. This model, which has been extensively applied in user study interface evaluations [58], correlates the required movement time (MT) to activate a target with a specific size (W), at a certain distance (D). Fitts' Law is formulated as: $MT = a + b\ .\ ID$, and $ID = \log_2(\frac{2D}{W})$ where ID denotes the index of difficulty, and a and b are empirically defined constant values. In the field of HCI, the Shannon formulation is most commonly used to calculate the index of difficulty, $ID = \log_2(1 + \frac{D}{W})$ as described in [102]. Fitts' law has been applied effectively in numerous user studies to analyse the performance of selecting specific targets such as buttons (e.g. [34], [87]). One of the earliest applications of Fitts' law in HCI was to compare four devices (mouse, joystick, step keys and text keys) for text selection on a monitor [24]. Researchers have also proposed variations to extend the original Fitts' law, for example MacKenzie *et al.* [104] extended Fitts' law from a one-dimension to a 2D model for target acquisition tasks to improve the accuracy of the index of difficulty measure for interactive computer systems.

5.1.2 Cognitive Workload

Cognitive workload refers to the amount of mental effort used to perform a task by a person. The NASA Task Load Index (TLX) questionnaire is a well-known method to measure subjective workload in user studies [64] and has been shown to be an effective tool to measure cognitive workload [152]. The questionnaire includes: physical demand, mental demand, temporal demand, effort, performance, and frustration with the maximum range of 100 points [59]. Although there is physiological data (e.g. electroencephalogram or EEG) which can be used to measure subjects' workload, these methods although accurate in detecting brain activity require specialized and sometimes cumbersome equipment. In addition, these techniques are intrusive for users and therefore are restricted to controlled environments such as laboratories [189]. Thus in our model, we focus on the NASA TLX.

5.1.3 Midas Touch Problem

Eye tracking, like many emerging technologies, has its challenges. The Midas touch problem which refers to unintended activation of functions by eye gaze to select a target is one of the major challenges to be considered when dealing with eye tracking applications. According to Jacob (1990), this problem occurs since the eyes are used to look around an object or to scan a scene, often without any intention to activate a command or function. Thus, numerous research has focused on solving the Midas touch problem for gaze-based interactions (e.g. [142], [179], [180], and [157]).

5.2 Related Work

Both Fitts' law and the NASA TLX are popular tools for user studies. Felton *et al.* applied these tools to study mental workload during brain-computer interactions [55]. Kim *et al.* applied Fitts' law in a driving safety simulation to analyze the usability of touch-key sizes [89]. Hansen *et al.* made use of Fitts' law to analyze the performance of gaze and head tracking for point and selection tasks when using head-mounted displays (HMDs) [61]. In addition, Fitts' law was applied to reduce dwell-time for gaze-based selection techniques by considering the estimated target acquisition time and the actual eye movement time [75]. Researchers have investigated the relation between eye blinks and mental workload among surgeons [191], finding that shorter blink duration and frequency indicate an increase of mental workload [191]. Borghini *et al.* studied brain activity and heart rate of car drivers and also found shorter blink rates correlate with mental workload [18]. Lanthier *et al.* studied the correlation between fixations and eye fatigue during visual search tasks and found that fixation duration increases with fatigue [97]. Abdulin *et al.* showed that the distance drift of fixation points in response to a stimuli can reveal physical eye fatigue [1] and calculated this using the fixation qualitative score (FQlS) [92]. Another study looked at developing a metric based on fixation points and the NASA TLX to determine the possibility of eye fatigue in gaze-based interactions [130]. There are also approaches to measure eye fatigue based on saccades, however, analysis of saccades requires expensive eye trackers, and these approaches are not applicable on budget-friendly devices [1], such as the one used in our study.

Building on previous work, we propose a non-invasive approach which can be applied on any remote eye trackers without the need of raw data analysis of the specific eye tracking sensors. We apply eye tracking for target selection from a safe distance and assess the difficulty levels including subjects' ratings independent from device abilities or tracking techniques. The primary purposes of IDEA are (1) to compare different eye tracking applications, and (2) to enable adaptation of eye tracking applications on different user groups such as children, users with disabilities, and the elderly. Furthermore, IDEA has the potential to be applied for eye fatigue assessment, and stress level measures based on target selection tasks. To the best of our knowledge, there are no models that integrate the index of difficulty of the Fitts' law (ID) and the NASA TLX scores for eye tracking applications without the need of technical parameters such as *blink rates*, *fixation duration time*, *average number of fixations*, and *saccade duration.*

5.3 Index of Difficulty (IDEA)

Users' perceived rating is one of the most valuable sources of data in any user study and the NASA TLX questionnaire is a valid tool for this purpose. On the other hand, Fitts' law can reflect the difficulty and performance of target selection tasks based on test specifications. Therefore, we integrated users' feedback into the Fitts' law model to result in a combined value reflecting both technical and experimental aspects of target selection tasks for eye tracking applications. In addition, the entire workload of a task (subjective rating) can be modulated by a selection ratio parameter (selection distance divided by screen diameter) which is determined based on test conditions, users' ability to select targets, and interaction techniques. The purpose of modulating the technical factor with the experimental factor is to combine the importance of both into a single index value. In other words, the multiplication combines both, technical aspects which are bound to case scenarios, with subjective understanding of the actual functions. This results in a single value for comparison. Thus, the IDEA analysis model is a novel simple-to-calculate compound model for eye tracking techniques based on the Fitts' law [57] and the NASA TLX questionnaire [59] to measure the difficulty of target selection tasks. IDEA is device-independent and can be applied on any eye tracker, and depends on the following parameters:

- All scores from the NASA TLX questionnaire: *physical demand* (PD), *mental demand* (MD), *temporal demand* (TD), *effort* (E), *performance* (P), and *frustration* (F).
- Diameter of screen (D): represents the longest distance on screen $D = \sqrt{x^2 + y^2}$ where x and y represent screen width and height.
- Selection ratio (S): represents the difficulty of target selection (distance to target) in regards to the screen diameter (see Figure 21).

- DISTANCES: the set of target distances from each other.

- WIDTHS: the set of target sizes (widths).

The conditions and range of each of the parameters are given by:

1. $PD = \{x_1 \mid x_1 \in \mathbb{Z} \wedge 1 \leq x_1 \leq 100\}$,

   $MD = \{x_2 \mid x_2 \in \mathbb{Z} \wedge 1 \leq x_2 \leq 100\}$,

   $TD = \{x_3 \mid x_3 \in \mathbb{Z} \wedge 1 \leq x_3 \leq 100\}$,

   $E = \{x_4 \mid x_4 \in \mathbb{Z} \wedge 1 \leq x_4 \leq 100\}$,

   $P = \{x_5 \mid x_5 \in \mathbb{Z} \wedge 1 \leq x_5 \leq 100\}$,

   $F = \{x_6 \mid x_6 \in \mathbb{Z} \wedge 1 \leq x_6 \leq 100\}$

   All NASA TLX scores are integers in the range of 1 to 100.

2. $D \in \mathbb{Z} \wedge D > 0$

   Diameter of screen is an integer value greater than 0 in pixels.

3. $r \in \mathbb{R} \wedge 0 \leq r \leq D$

   The distance to target (r) is a real number between 0 and screen diameter in pixels (see Figure 21a).

4. $S = \frac{r+1}{D} \wedge S \in \mathbb{R} \wedge S > 0$

   Selection ratio (S) is the ratio of distance to target (r) over diameter of the screen (D). The constant value of 1 added to the equation to avoid the 0 case for distance to target (see Figure 21b).

5. $DISTANCES = \{d \mid d \in \mathbb{R} \wedge d > 0\}$

   DISTANCES is the set of real numbers containing distances of targets from each other greater than 0.

6. $WIDTHS = \{w \mid w \in \mathbb{R} \land w > 0\}$

   WIDTHS is the set of real numbers containing widths (sizes) of targets greater than 0.

7. $m = |WIDTHS| \land m \geq 1$

   m is the count of members in the $WIDTHS$ set greater than or equal to 1.

8. $n = |DISTANCES| \land n \geq 1$

   n is the count of members in the $DISTANCES$ set greater than or equal to 1.

9. $Technical\ Factor \in \mathbb{R} \land Technical\ Factor > 0$

   The technical factor (Equation 5) is the sum of all distances ($d \in DISTANCES$) doubled and divided by the width values ($w \in WIDTHS$) derived from the Fitts' law [57]. This results in a real number greater than 0 which resembles the index of difficulty of the Fitts' law $ID = \log_2 \frac{2D}{W}$. The technical factor represents the *precondition of target properties* (distances and widths).

$$Technical\ Factor = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{2d_i}{w_j} \tag{5}$$

10. $R \in \mathbb{R} \land 1 \leq R \leq 100$

    The subjective rating (R) is the mean of all TLX scores which is a real number between 1 and 100 shown in Equation 6.

$$R = \frac{PD + MD + TD + E + P + F}{6} \tag{6}$$

11. $Experimental\ Factor \in \mathbb{R} \land Experimental\ Factor > 1$

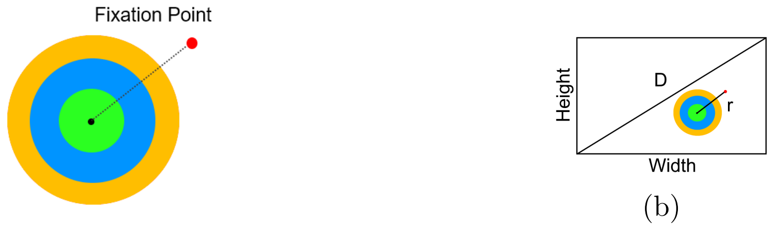
    The experimental factor (Equation 7) is defined as the product of the calculated selection ratio (S) depicted in Figure 21, and the subjective rating (R) which results in a real number greater than 1.

$$Experimental\ Factor = S \times R \tag{7}$$

12. $IDEA \in \mathbb{R} \wedge IDEA > 1$

    The proposed index of difficulty for eye tracking applications (IDEA) is calculated by multiplying (a) the technical factor, and (b) the experimental factor offset by a constant value of 2 which results in a real number greater than 1 (Equation 8). We offset both technical and experimental factors by the constant value of 2 in case these factors are close to zero, therefore the calculated IDEA value starts from $1.x$. Figure 22 shows the 3D visualization of IDEA and its factors.

$$IDEA = \log_2\left(\underbrace{(\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{2d_i}{w_j})}_{Tec.\ Fac.} \times \underbrace{S \times R}_{Exp.\ Fac.} +2 \right) \tag{8}$$



(a)

(b)

Figure 21: (a) overview of the dart-test to measure Euclidean distance, and (b) the concept of selection ratio regarding diameter of screen (D) and the selection distance (r).

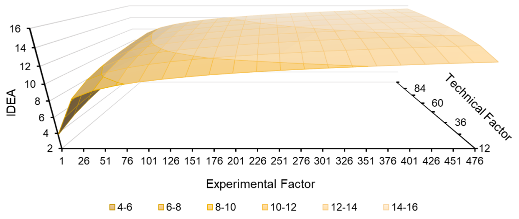


Figure 22: 3D illustration of the IDEA model.

5.4 Methodology

We conducted a three-part repeated measures user study to evaluate the efficacy of our proposed model with 33 participants (20 male, from 22 to 35 years old, $mean = 26.06$). Subjects were asked to navigate and select highlighted targets (see Figure 24) under two gaze-based interaction techniques: (1) dwell-time with 500 ms threshold, and (2) eye tracking using voice commands. Prior to running the experiments, participants were informed about the objectives of the user study, trained on each of the interaction techniques, and filled out a pre-test questionnaire. Before running the tests, the built-in eye tracking software was used to calibrate eye positions for each participant. The order of interaction techniques was randomly selected for each participant. Overall, the user studies took 8 minutes on average for a participant to finish. At the end of the two experiments measuring the Fitts' law parameters (Figure 24) participants were asked to fill out a post-test questionnaire consisting of the NASA TLX questionnaire.

5.4.1 Interaction Techniques

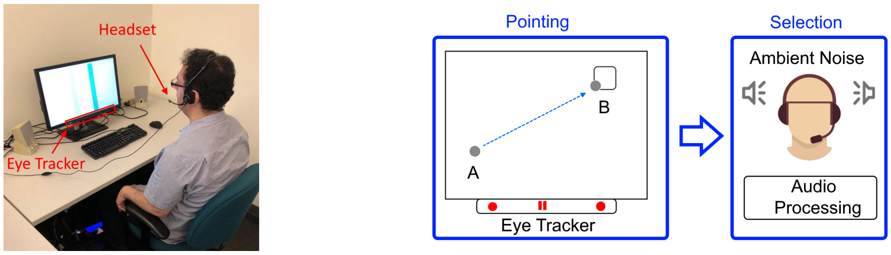
We applied two eye tracking techniques (single and multi-modal interactions) to evaluate the efficacy of our proposed model. We ran the mentioned interaction techniques on an Intel i7 PC with the 64-bit Windows operating system. Figure 23 illustrates the test setting and overview of the interaction techniques.

Dwell-time

The dwell-time method can select a target only by eye gaze fixations after a predefined threshold is reached. We defined the target selection threshold to 500 milliseconds which is in the typically accepted range of 300-1100 milliseconds [165], and has been shown to be the best-suited threshold for the dwell-time method [103], [165]. In other words, when a subject focuses for 0.5 seconds on a target it gets selected, and any gaze movement from the target boundaries prior to that threshold causes the restart of target selection process.

Eye Tracking with Voice recognition

The voice recognition method operates in two phases, (1) pointing phase using the eye tracker, and (2) selection phase using voice commands. Figure 23b illustrates the overview of these phases. The process of voice recognition was developed using the built-in Windows 10 speech recognition functionalities provided in the Microsoft .NET framework. We developed a C# application to capture user's activation command 'select' to activate a left mouse click.



(a)

(b)

Figure 23: (a) test setting and equipment, and (b) system overview and workflow of both interaction techniques.

5.4.2 Interaction Modules

Eye Tracking: We used the Tobii 4C eye tracker to capture the mouse pointer position to enable users to interact with the system with their gaze. Moreover, we employed the Tobii SDK to obtain users' gaze locations (2D coordinates) on the screen and synchronize the mouse pointer to these coordinates in pixel. The eye tracking module for both interaction techniques was developed in C++ and integrated into the Tobii SDK as a new plug-in. The samples were recorded at a distance of 60 cm (24 in) from the eye tracker with a sampling rate of 90 Hz on a 24 inch screen with the resolution of 1920 × 1080 pixels. The dwell-time technique relies solely on the eye tracking module.

Voice Processing: We used a headset microphone (Logitech H370) to capture the user's voice commands in the presence of an artificial ambient noise around 50 dB played by stereo speakers (Figure 23a) to simulate a typical working office. The voice recognition module received the commands in real-time to be activated by the keyword 'select' to trigger a left mouse click.

5.4.3 Hypotheses

Based on the previous literature, which has demonstrated the effectiveness of Fitts' law [34], [87], [24], and [104] and the NASA TLX questionnaire [64], [63], and [152], we propose a compound simple-to-calculate mathematical model to measure the difficulty level of eye tracking applications independent from device type and technical capabilities during user studies. This model enables analysis of eye tracking applications based on user groups and their abilities to interact with an eye tracking device or interaction technique. Specifically, we hypothesize that:

1. When IDEA is higher on average for an interaction technique, the calculated *throughput* based on the Fitts' law will be lower, and vice versa.

2. When IDEA is higher on average for an interaction technique, the calculated *movement time* based on the Fitts' law will be higher as well, and vice versa.

3. When IDEA is higher on average for an interaction technique, the registered *error rates* will be higher as well, and vice versa.

5.4.4 User Study

The user study described above was used to analyze the mentioned eye tracking interaction techniques to evaluate the proposed IDEA model, according to well-established academic standards. We measured four parameters in our 3 part study: (1) distance to target, (2) throughput, (3) movement time, and (4) error rates. We developed a dart-like application (Figure 21a) to measure distance to target and used the application developed by Wobbrock *et al.* [186] called the *FittsStudy* version 4.2.7 which includes two widths (96, 128), and three distances (256, 384, 512) pixels to record the rest of the measures.

Dart test

The stimulus consisted of three circles, green from 0 to 30 pixels, blue from 30 to 60 pixels, and red from 60 to 90 pixels in radius as illustrated in Figure 21a. Any selection outside of the dart colored circles is recorded as the fixed maximum range of 90 pixels for that selection. The purpose of this experiment was to measure the Euclidean distance to target to be applied in Equation 7 by calculating the fraction of distance (r) over diameter of screen (D) as shown earlier ($S = \frac{r+1}{D}$). Subjects were asked to select, as accurately as possible, the center of a dart target using both interaction methods. Since eye tracking has different accuracy in different regions of a screen [54], we calculated an average of five trials for each interaction techniques where the stimulus moved to different areas around the center of screen randomly. Each random trial started in two second intervals enabling subjects to change their gaze before recording the distance measures. A countdown timer with intervals of 100 ms was displayed from 5 to 0 to show the remaining time to subjects.

Ribbon-shaped test

The stimulus contains two vertical bars to be selected (clicked), each at a time shown in Figure 24a. The variation of distances and widths are chosen randomly by the FittsStudy [186] application and the order of each interaction method for each participant were also chosen randomly.

Circle-shaped test

This test is the same as the ribbon-shaped test with circular-shaped targets illustrated in Figure 24b. This experiment measures two variations for throughput, (1) univariate endpoint deviation (SD_x) through one axis, and (2) bi-variate endpoint deviation ($SD_{x,y}$) through both axes which results in a better Fitts' law model [186]. The stimulus contains equally-sized circles with different distances and widths to be selected (clicked), each at a time shown in Figure 24b. The variation of distances and widths are chosen randomly by the FittsStudy [186] application and the order of each interaction method for each participant was also chosen randomly.

Figure 24: The *FittsStudy* application [186]. (a) Ribbon-shaped, and (b) Circle-shaped targets.

Workflow and parameters

The user study was conducted on a screen with the resolution of 1920 × 1080 pixels which results in a *diameter* (D) of 2203 pixels (rounded up). *Distances* of 256, 384, and 512 pixels between targets were used with a target *width* of 96 and 128 pixels.

The *distance to target* (r) for both interaction techniques was measured by the dart test application (Figure 21) based on the Euclidean distance in pixels. Lastly, the selection ratio was calculated by measured distance to target over the screen diameter ($S = \frac{r+1}{D}$). The constant value of 1 is added to the measured distance for selecting the target exactly in the middle which results in a distance to target of 0.

5.5 Results

The results of our experiments were analyzed using paired-sample t-tests with the JASP[1] software. Figure 25b shows the NASA TLX scores and the calculated average workload based on Equation 6 for both interaction techniques from the post-test questionnaire.

As per Equation 5, the technical factor, which is 42, was the same for both interaction techniques as it depends on distances and widths which were constant in our user study. This is the case in our experiments as both interaction techniques were evaluated on the same device with the same screen resolution and the same target distances and widths. However, the technical factor can be different for varying case scenarios. A paired-sample t-test was applied to check the effectiveness of the interaction technique on the experimental factor based on Equation 7 with (t(32)=2.86, $p < .05$). A significant difference was found between dwell-time ($M = 0.48$, $SE = 0.06$) and voice recognition ($M = 0.65, SE = 0.06$). Specifically, dwell-time had a lower experimental factor than the voice recognition technique. This suggests that the multiplication of users' selection ratio on screen (S) and their rating scores (R) is significantly lower for the dwell-time method than the voice recognition technique.

[1] https://jasp-stats.org/

A paired-sample t-test was applied to check the effectiveness of interaction technique on the index of difficulty based on the Equation 8 shown in Figure 25a and Table 7. A significant difference (t(32)=3.19, $p < .05$) was found between dwell-time ($M = 4.17, SE = 0.15$) and voice recognition ($M = 4.66, SE = 0.15$). This suggests that the dwell-time method has a significantly lower IDEA value than the voice recognition technique. Dwell-time can thus be considered an easier eye tracking technique for our subjects comparing to the voice recognition.

	Dwell-Time	Voice Recog.
Distance	35.30	29.27
Selection ratio	0.016	0.014
Tech. factor	42	42
Exp. factor	0.48	0.65
IDEA	4.17	4.66

Table 7: Summary of IDEA calculations.

Dart Test: Paired-sample t-tests were performed to study the effect of interaction type on (1) distance to target, and (2) selection ratio. A significant difference (t(32)=2.88, $p < .05$) was found between dwell-time ($M = 35.30 \ \ pixels, SE = 2.11 \ pixels$) and voice recognition ($M = 29.27 \ pixels$, $SE = 2.07 \ pixels$) on distance to target (r) depicted in Figure 26a. This shows that the voice recognition technique has a higher target selection accuracy (lower distance to target) than the dwell-time method. This is likely the case because this method splits the pointing (eye tracking) and selecting (voice command) into different modalities.

A paired-sample t-test was also applied to check the effectiveness of interaction technique on selection ratio (S) depicted in Figure 26b. A significant difference (t(32)=2.88, $p < .05$) was found between dwell-time ($M = 0.016 \ pixels, SE = 9.620e-4 \ pixels$) and voice recognition ($M = 0.014 \ pixels, SE = 9.409e-4 \ pixels$). This means that users are more accurate to select targets using the voice recognition technique than the dwell-time.

Ribbon-shaped Test: Paired-sample t-tests were performed to study the effect of interaction type on (1) throughput, (2) movement time, and (3) error rate. There was a significant difference (t(32)=5.96, $p < .001$) of throughput for dwell-time ($M = 3.30 \ bits/sec, SE = 0.36 \ bits/sec$) and voice recognition ($M = 1.16 \ bits/sec$, $SE = 0.09 \ bits/sec$) as seen in Figure 26c. This confirms our hypothesis that a lower IDEA value for an interaction technique reflects a higher throughput.

A paired-sample t-test was applied to check the effectiveness of interaction technique on movement time depicted in Figure 26d. A significant difference (t(32)=15.13, $p < .001$) was found between dwell-time ($M = 0.60 \ sec$, $SE = 0.01 \ sec$) and voice recognition ($M = 2.01 \ sec$, $SE = 0.08 \ sec$). This confirms our hypothesis that a lower IDEA value for an interaction technique reflects a lower movement time.

A paired-sample t-test was applied to check the effectiveness of interaction technique on error rate depicted in Figure 26e. A significant difference (t(32)=4.84, $p < .001$) was found between dwell-time ($M = 0.28 \ errors, SE = 0.03 \ errors$) and voice recognition ($M = 0.11 \ errors, SE = 0.02 \ errors$). This rejects our hypothesis that an interaction technique with a lower IDEA value should cause lower error rate. The cause of errors in eye tracking applications as explained above are mostly due to the Midas touch problem [83]. Thus as the dwell-time method relies on eye tracking solely, and selection is done based on fixation time there were higher error rates in this method than in the multi-modal voice method where selection is done based on a voice command.

Circle-shaped Test: Paired-sample t-tests were performed to study the effect of interaction type on (1) throughput with two variations, (2) movement time, and (3) error rate. For univariate throughput (illustrated in Figure 27a) there was a significant difference (t(32)=7.98, $p <$ *.001*) between dwell-time ($M = 3.91\ bits/sec, SE = 0.31\ bits/sec$) and voice recognition ($M = 1.48\ bits/sec, SE = 0.09\ bits/sec$). This confirms our hypothesis that an interaction technique with a lower IDEA value should reach higher throughput.

A paired-sample t-test was applied to check the effectiveness of interaction technique on bivariate throughput illustrated in Figure 27b. A significant difference (t(32)=7.19, $p <$ *.001*) was found between dwell-time ($M = 2.51\ bits/sec$, $SE = 0.22\ bits/sec$) and voice recognition ($M = 1.01\ bits/sec, SE = 0.06\ bits/sec$). This confirms our hypothesis that an interaction technique with a lower IDEA value should reach higher throughput.

A paired-sample t-test was applied to check the effectiveness of interaction technique on movement time illustrated in Figure 27c. A significant difference (t(32)=11.31, $p <$ *.001*) was found between dwell-time ($M = 0.64\ sec$, $SE = 0.02\ sec$) and voice recognition ($M = 2.12\ sec, SE = 0.13\ sec$). This confirms our hypothesis that an interaction technique with a lower IDEA value should reach a lower movement time.

A paired-sample t-test was applied to check the effectiveness of interaction technique on error rate illustrated in Figure 27d. A significant difference (t(32)=2.26, $p <$ *.05*) was found between dwell-time ($M = 0.23\ errors$, $SE = 0.03\ errors$) and voice recognition ($M = 0.13\ errors$, $SE = 0.02\ errors$). This rejects our hypothesis that an interaction technique with a lower IDEA value should have a lower error rate. As described above, the cause of errors in eye tracking applications are mostly due to the Midas touch problem and thus the single mode method which requires gaze for both pointer movement and selection is more error prone.

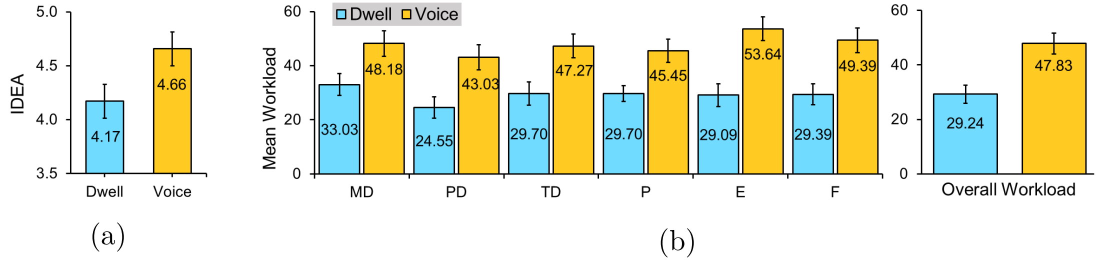


Figure 25: (a) shows index of difficulty for eye tracking applications (IDEA) based on Equation 8 for both interaction techniques ($p < .05$), and (b) illustrates the results of the NASA TLX scores. Error bars represent SE.

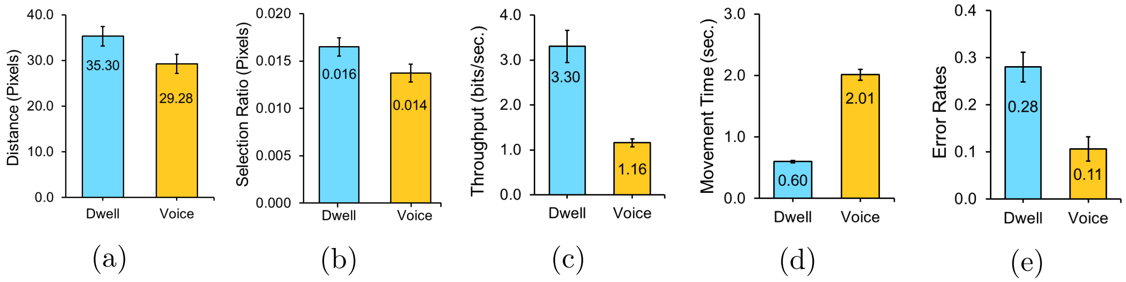


Figure 26: (a) Euclidean distance to target measure (r). (b) Calculated selection ratio ($S = \frac{r+1}{D}$) for both interaction techniques. (c) Throughput (TP), (d) Movement time (MT), and (e) Error rates (ER) for both interaction techniques of the ribbon-shaped test. Error bars represent SE. ($p < .05$ on (a) and (b) measures, $p < .001$ on (c), (d), and (e) measures).

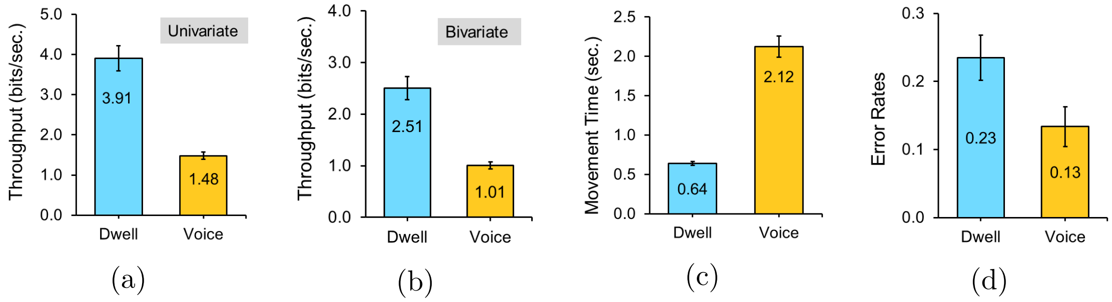


Figure 27: Calculated measures of the circle-shaped test. (a) Univariate throughput (TP) ($p < .001$), (b) Bivariate TP ($p < .001$), (c) Movement time (MT) ($p < .001$), and (d) Error rates for both interaction techniques ($p < .05$). Error bars represent SE.

5.6 Discussion

The results reflect the efficacy of our two-factor model to measure the performance of eye tracking applications independently of device type. We showed that our model can predict the difficulty of eye tracking applications solely based on Fitts' law and the NASA TLX scores. Further, we showed our model correlates with the standard measures (throughput and movement time) described by Fitts' law.

The global pandemic of COVID-19 showed the importance of computer interactions from a safe distance without physical contact. Eye tracking applications, specifically the dwell-time method, are suitable candidates to enable safe interactions on shared and public devices for selection tasks. Therefore, our proposed model can be applied in pilot studies to measure the usability and performance of selection techniques to address different user groups such as children, users with disabilities, or elderly based on the experimental factor which reflects (a) subjective ratings (NASA TLX scores), and (b) perceived difficulty levels of interaction techniques or user groups.

Although we only studied voice recognition as a multi-modal interaction technique, the results of the user studies confirm our first and second hypotheses regarding the correlation between *throughput* and *movement time* calculated by the Fitts' law and the predictions by our proposed model. However, eye tracking applications suffer from the Midas touch problem, and since the dwell-time method relies on eye gaze only, it reached higher error rates than the multi-modal selection technique using voice recognition with separate modalities for point and selection.

The analysis of our results emphasizes the potential of our two-factor prediction model on two similar eye tracking interaction techniques. We hope, this experiment leads to more innovations of multi-dimensional compound models for gaze-based interactions.

5.7 Conclusion and Future Work

In this paper we proposed the Index of Difficulty for Eye tracking Applications (IDEA) a compound two-factor model to measure the performance and usability of selection techniques based on calculations of Fitts' law and the results of a NASA TLX questionnaire. As emerging interaction techniques are required to cope with emerging users' demands, the need for more complex models to compare different techniques requires more attention. We present our model to asses the efficacy of eye tracking applications for pilot studies with different user groups such as children, users with disabilities, or elderly. Our configurable model can be applied for case scenarios as well as to discriminate specific interaction techniques.

In addition, we presented an in-depth analysis of the dwell-time method based on the Fitts' law measures. Although our model was developed to address eye tracking interactions, it can be applied on any selection technique to measure difficulty levels based on test specifications (target size and distance) and users' subjective ratings. Further, we showed eye tracking techniques can be compared without analysis of technical raw data such as fixation duration time and blink rates. These enable researchers to run pilot studies independently from device type.

We predict the transition from conventional interaction techniques, such as keyboard and mouse, to contact-free techniques from a safe distance caused by the latest global outbreak of viral infections, especially for equipment in healthcare sectors, and shared public devices. IDEA enables researchers to run user studies based on video eye tracking techniques via remote webcams to comply with restrictions caused by viral diseases which limit the physical presence of participants in laboratories or attaching sensory equipment to record users' feedback. We plan on applying our proposed model on AR and VR headsets with internal eye trackers to study usability of target selection in our future work.

Chapter 6

ESPiM: Eye-Strain Probation Model - An Eye-Tracking Analysis Measure for Digital Displays

Preface

We propose ESPiM, a mathematical model based on the previously proposed models FELiX (Chapter 4), and IDEA (Chapter 5) discussed earlier in this dissertation. Since we have already proposed models with integrated users' feedback, it is also necessary to run user studies without the presence of users to compare design concepts. ESPiM is a computational model based on spatial properties of user interfaces such as size, distance, area of targets, and area of screen. In addition, another feature of ESPiM is the integration of time in its equation. The goal of ESPiM is to measure eye-strain over a specific duration which is suitable to test commercial software applications before release for comparison and optimization. In contrast to previous models that address eye-strain (e.g. [1, 92, 8, 113, 41]), ESPiM expands the range of measurement techniques in the research community by providing an eye-tracking measure (eye fixation points) integrated into a compound model which is applicable on most eye-tracking sensors.

ESPiM has multi-purpose applications beyond the measurement of eye-strain on digital displays; it is capable of comparison between various interaction techniques/design concepts, and different display types. Another application of ESPiM would be to compare video games based on the amount of eye-strain on a specific play duration. We present two evaluations of ESPiM to show its usability in user studies with a remote eye-tracking sensor for in-person user studies, and a video-based eye-tracking technique that was used during an online user study. The following chapter is based on a paper that will be submitted to the International Journal of Human-Computer Studies.

Abstract

Eye-strain is a common issue among computer users due to the prolonged periods they spend working in front of a monitor, which can lead to vision problems, such as irritation and tiredness of the eyes, and headaches. Eye-strain is mainly caused by moving objects on the screen and occurs when focusing on close objects. We propose the Eye-Strain Probation Model (ESPiM), a computational model, based on eye-tracking data, that measures eye-strain on digital displays based on the spatial properties of the user interface and display area for a required period of time. As well as measuring eye-strain, ESPiM can be applied to compare (a) different user interface designs, (b) different display devices, and (c) different interaction techniques. Two user studies were conducted to evaluate the effectiveness of ESPiM. The first was conducted in form of an in-person study with an infrared eye-tracking sensor with 32 participants. The second was conducted in form of an online study with video-based eye-tracking technique via webcams on users' computers with 13 participants. Our analysis showed significantly different eye-strain patterns based on video gameplay frequency of participants. Further, we found distinctive patterns among users on a regular 9-to-5 routine versus those with more flexible work hours in terms of (a) error rates, and (b) reported eye-strain symptoms.

6.1 Introduction

Today, in many parts of the world, many employees work sitting in front of computers for eight hours a day. In fact, the typical work atmosphere which includes working in enclosed spaces, e.g. cubicles or offices, and dealing with multiple devices (e.g. PC, smartphone, telephone) at the same time (i.e. multitasking) can lead to an increased workload. Yet, scientists have found a correlation between excessive screen time and health risks such as cardiovascular diseases, impaired vision, and bone density [100] and a correlation between long screen times of elementary school students with dry eyes (the malfunctioning of tear production in the eyes [175]) and learning abilities [115]. Other symptoms of CVS include *eye-strain*, *eye burn*, *double vision*, and *blurred vision* [117].

These CVS symptoms not only affect visual comfort, but also are a major cause of work-related stress [119] and productivity in both adults and teenagers [148]. In this paper we address eye-strain, also known as visual fatigue, as one of the major CVS symptoms [117]. Eye-strain can be caused by moving images [78] and occurs when focusing on near objects. It is a common issue among computer users due to the prolonged periods they spend working in front of a monitor [178]. Eye-strain can cause motion sickness [94], vision problems such as irritation and tiredness of the eyes, and headaches [176]. We introduce the Eye-Strain Probation Model (ESPiM), an integrated measurement model for eye-strain based on target selection tasks relying on spatial targets' and screen properties via Fitts' law and eye-tracking fixation points.

To demonstrate the efficacy of ESPiM, we measured target selection performance with data from two eye-tracking user studies, one in-person laboratory study and one online web application study. We considered the distinctive patterns among participants based on (1) biological sex, and (2) video gameplay frequency and found that females and participants with lower frequency of gameplay experienced higher eye-strain base on our model. Moreover, we studied the eye-strain patterns including eye symptoms among typical 9-to-5 participants and flexible (anytime beyond 9-to-5) participants and found the flexible groups experienced higher number of eye symptoms than the 9-to-5 group. Moreover, we recorded significantly higher error rates for the 9-to-5 groups than the flexible group.

6.2 Related Work

In the following section we describe Fitts' Law which is related to task difficulty of target selection tasks and previous work in eye-strain and eye-tracking models.

6.2.1 Fitts' Law

Originally proposed to measure task difficulty, Fitts' law predicts the amount of movement time (MT) to activate a target based on specific size (W), at a certain distance (D). Fitts' Law is formulated as: $MT = a + b\ .\ \log_2(\frac{2D}{W})$, where a and b are empirically defined constant values. However, the Shannon formulation is most commonly used to calculate the index of difficulty in the field of HCI, $ID = \log_2(1+\frac{D}{W})$ [102]. Among the earliest applications in HCI, Fitts' law was used to compare input devices (mouse, joystick, step keys and text keys) for text selection on a display [24]. Fitts' law has also been applied effectively in various studies to analyze the performance of selecting targets (e.g. [34], [87]). A number of extensions of the original Fitts' law have been proposed for different case scenarios. For instance MacKenzie *et al.* proposed an extension from a one-dimension to a 2D model for target acquisition tasks enabling the improvement of index of difficulty for interactive computer systems with higher accuracy [104].

Fitts' law is a popular tool for user studies that has been extensively been applied in evaluations [58]. For example, Kim *et al.* analyzed the usability of touch-key sizes in a driving safety simulation [89]. Hansen *et al.* studied the performance of gaze and head tracking for point and selection tasks on head-mounted displays (HMDs) [61]. In addition, researchers applied Fitts' law to reduce dwell-time for gaze-based interactions by taking into account the estimated target acquisition time and eye movement time [75].

6.2.2 Eye-strain Models

Researchers have proposed various means to measure eye-strain based on eye movement analysis. Lanthier *et al.* showed that eye fixations and eye-strain increases with fatigue [97]. Komogortsev *et al.* proposed the Fixation Quantitative Score (FQnS) to consider the amount of fixation points in regards to a stimulus which may reveal physical eye-strain [92]. Furthermore, Vasiljevas *et al.* adopted an analytical model for muscle fatigue to assess eye-strain in gaze-based tasks [178]. Researchers have also applied self-evaluation rating questionnaires to measure eye-strain for gaze-based applications [105]. Further, saccades-based approaches were proposed to measure eye-strain [8, 113, 41]. However, analysis of saccades cannot be applied on budget-friendly devices, and therefore fixation-based approaches have been preferred [1].

Considering the previous works, we propose a dual-purpose approach which can be applied in user studies with eye trackers to measure eye-strain based on screen and target properties for a specific duration. In our previous works, we proposed one approach to measure eye-strain involving subjective ratings (FELiX) [129], and introduced an index of difficulty for eye tracking applications (IDEA) [132]. These approaches were compound models based on subjective and objective measures. The introduction of ESPiM is based on objective measures only which may improve and optimize user interface design concepts based on eye-strain criterion. Based on the results of our previous models, we propose a predictable objective model suitable for assessments of visual prototypes on digital displays being used for a specific period of time to reduce costs of productions by comparing design ideas in early steps via pilot studies.

6.3 Eye-Strain Probation Model (ESPiM)

The ESPiM model relies on properties (spatiotemporal parameters) that are related to screen size, target dimensions and distances (spatial) shown in Fig. 28a, task duration time (temporal), and eye tracking fixations as described below:

1. x: width of the screen in pixels.

   $x \in \mathbb{R} \land x > 0$

2. y: height of the screen in pixels.

   $y \in \mathbb{R} \land y > 0$

3. z: diameter of the screen in pixels.

   $z = \sqrt{x^2 + y^2} \land z \in \mathbb{R} \land z > 0$

4. Area of screen (AoS): arithmetic surface area of the screen in which test applications are executed on in pixels.

   $AoS = x \times y \land AoS \in \mathbb{R} \land AoS > 0$

5. Area of target (AoT): arithmetic surface area of the target (user interface element) in which user tries to focus on in pixels. Typically there are multiple targets on a user interface, we calculate the area of a single target since users focus on one target at selection time. In case of targets with various areas, the average of all targets will be considered.

   $AoT \in \mathbb{R} \land AoT > 0 \land AoT \leq AoS$

6. Distance of target (D): distance of the target centers from each other in pixels which must be smaller than or equal to screen diameter.

   $D \in \mathbb{R} \land D > 0 \land D \leq z$

7. Width of target (W): width of target on the screen in pixels, which must be smaller than or equal to screen width.
   $W \in \mathbb{R} \land W > 0 \land W \leq x$

8. Shannon code ($\log_2(1 + \frac{D}{W})$): index of difficulty for point-and-select tasks [102] based on the Fitts' law [57].
   $\log_2(1 + \frac{D}{W}) \in \mathbb{R} \land \log_2(1 + \frac{D}{W}) > 0$

9. Task duration (TD): time spent on a specific task.
   $TD \in \mathbb{R} \land TD > 0$

10. Average number of fixations (ANF): average number of fixations recorded by an eye-tracking sensor of an entire task.
    $ANF \in \mathbb{R} \land ANF > 0$

ESPiM: the calculated eye-strain value is based on Equation 9 which is greater than 0 based on the square root function growth: ($ESPiM \in \mathbb{R} \land ESPiM > 0$).

The ESPiM model is based on pure test conditions such as screen and target properties regarding the dimensions, and distances to be measured for a desired duration of time. This model provides an initial assessment to researchers about task difficulties. The ESPiM model can be used considering any 2D flat display type such as smartphones, tablets, and laptop/desktop monitors. This enables researchers to predict the difficulty level of target selection tasks on any device regardless of the applied interaction techniques.

The ESPiM model given in Equation 9 reflects *the level of difficulty given the size and distances of targets over the screen* ($\log_2(1 + \frac{D}{W})$), *to select a portion of the screen covered by a target* ($\frac{AoS}{AoT}$) *for the specific period of time* (TD). The equation is offset by 1 in case of very small values for the parameters described earlier. The purpose of this addition is to set the minimum threshold of the square root function to start from $\sqrt{\frac{\epsilon+1}{\lambda+1}}$, where ϵ and λ denote very small values. The average number of fixations (ANF) is one of the most used eye tracking variables which contributes to the accuracy of the calculations, the higher number of fixations should cause higher eye fatigue.

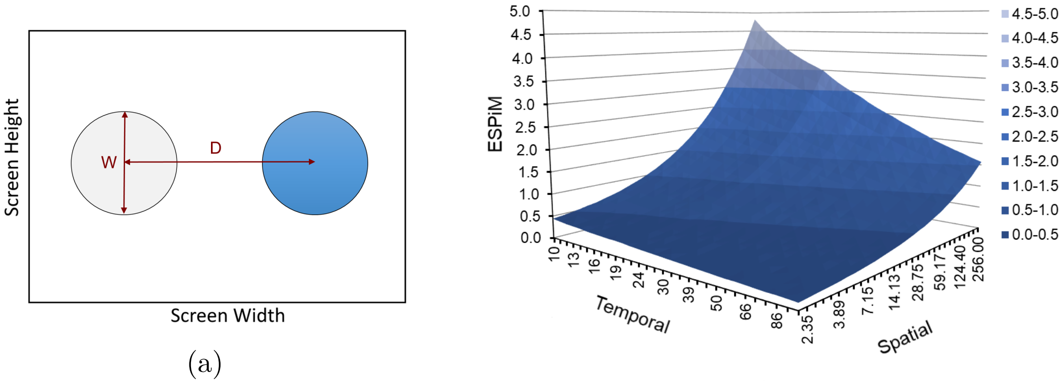
These parameters are bound into the square root function to shape a positive continuous predictable increase or decrease of values which are suitable for machine learning algorithms. We assign the unit of *bits* for the ESPiM model. Fig. 28b shows a 3D visualization of the ESPiM model calculated for sample generated values.

$$ESPiM = \sqrt{\frac{\left(\overbrace{(\frac{AoS}{AoT}) \times \log_2(1 + \frac{D}{W})}^{spatial} \quad \times \quad \overbrace{ANF}^{eye-tracking}\right) + 1}{\underbrace{TD}_{temporal} \; +1}} \tag{9}$$

6.3.1 Applications of ESPiM

Although we have designed the ESPiM model to measure eye-strain on digital displays primarily, it can be applied to (a) compare user interface designs, (b) compare different display devices, and (c) compare different interaction techniques based on eye-strain of users. As the spatial parameters are included in the ESPiM model, it can be used for estimation of eye-strain before testing in a user study which can be beneficial to both research communities and industrial producers of digital contents.

In fact, we incorporate spatial parameters of screen and targets including the application of Fitts' law and eye-tracking fixation points for a desired period of time in a single measure. Moreover, considering the fact that eye fixations are typically bound to a certain range (200-600 ms) [106] and therefore the average number of fixations for a specific period of time can be estimated. This property makes ESPiM a suitable choice for testing and evaluating interfaces even when there is no access to eye-tracking sensors.



(a)

(b)

Figure 28: (a) The overview and spatial parameters of ESPiM. The blue-colored target represents a selected target by eye gaze as provided in 'FittsStudy' application [186] and the area of active target (AoT) measure that a user focused to select. ESPiM considers targets' properties in regards to the screen dimensions. (b) A 3D visualization of the ESPiM model for sample generated values. The spatial parameter axis represents the product of the relative area ratio of the screen over target, and the Shannon code as described in Equation 9 with a constant average number of fixations (ANF).

6.4 User Study 1: Fitts' Study (in-person)

We conducted two user studies to evaluate the ESPiM model using (1) an infrared desktop eye tracker and (2) video-based eye-tracking for both an in-person and remote study. The purpose of the first evaluation was to study the ESPiM model. In this study, we used the unpublished dataset parts of our previous paper EyeTAP [131] in which we collected large amount of infrared eye-tracking data.

6.4.1 Methods

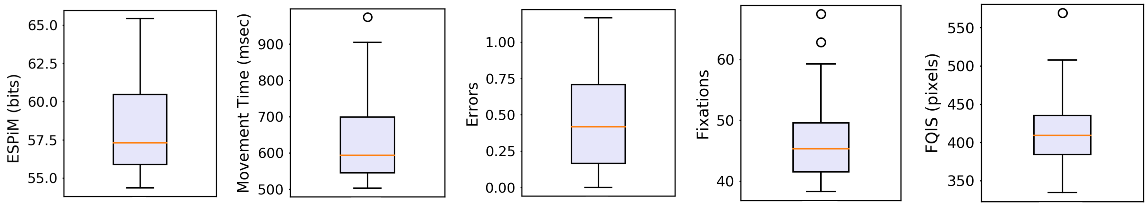
During this study, participants were asked to select circular targets using eye gaze, and specifically the dwell-time method. We used the 'FittsStudy' V4.2.7 application [186] for our user study which enabled us to run experiments based on Fitts' law. Our stimuli contains two target widths (96 and 128 pixels) at three distances apart (256, 384 and 512 pixels) to record the required measures. An activation of 500 milliseconds for dwell-time was used as this has been shown to be the best-suited threshold in previous studies [103, 165]. Thus, when a participant focuses for 500 milliseconds on a target it triggers a click event and the target gets selected, and any gaze movement from the target borders causes pointer movement and therefore restarts the target selection process.

The user study took 12 minutes on average for each participant, 10 minutes for preparation including description of the task, training and eye calibration, and 2 minutes for the actual target selection task in 6 trials (2 widths × 3 distances). Considering the differences between target selections using conventional input devices such as keyboard and mouse, and an eye-tracking sensor with a low activation threshold (500 ms) which challenges the control of pointer on screen, the relative short duration of target selection was sufficient to record required measures as well as not to overwhelm participants with heavy tasks.

To capture the eye-tracking data a remote eye-tracking sensor (Tobii 4C) with a sampling rate of 90 Hz on a monitor with the resolution of 1920 × 1080 pixels (24″) with a distance of 60 cm (≈23.5 in) to the eye tracker running on an Intel i7 Windows 10 PC was used. Specifically, the following data was collected during the study: (1) movement time based on the Fitts' law, (2) recorded errors of target selections, (3) average number of fixations (ANF), and (4) Fixation Qualitative Score (FQlS) (a measure that can reveal physical eye fatigue based on distance drift of fixation points) [92]. These measures accompanied by ESPiM calculations based on Equation 9 were needed for our analysis.

6.4.2 Results and Discussion

We analyzed our raw data based on three categories: (1) model analysis regarding the effectiveness of ESPiM, (2) gender of participants, and (3) participants' frequency of video game play. Fig. 29 illustrates the calculated ESPiM values and the recorded measures and Table 8 shows descriptive statics of 32 participants. ESPiM can be applied to compare results of different groups in user studies. We analyse these basic results among (1) gender groups, and (2) groups of video gameplay frequency.



(a) (b) (c) (d) (e)

Figure 29: Illustration of (a) ESPiM, (b) calculated movement time based on Fitts' law, (c) recorded errors, (d) eye fixations, and (e) Fixation Qualitative Score (FQlS) measure for 32 participants observed in our user study.

Table 8: Descriptive statistics of recorded measures for 32 participants.

	ESPiM	Movement Time	Errors	Fixations	FQlS
Mean	58.2	637.5	0.4	46.8	413.8
Median	57.2	593.6	0.4	45.3	409.3
SD	2.9	125.5	0.3	7.3	46.1
IQR	4.5	154.2	0.5	8.0	51.0
Range	11.0	471.5	1.1	29.1	234.8
Min	54.3	502.8	0.0	38.2	334.3
Max	65.4	974.3	1.1	67.4	569.1

Gender-based Analysis

Although the analysis of results based on gender was not in our hypotheses, we found distinctive patterns between male and female participants regarding the measures presented earlier. Since male subjects outnumbered the females, we selected 13 males from the total 19 participants randomly to analyze the subjects into two balanced groups. Fig. 30 illustrates the results of the gender-based analysis.

Eye-strain probation model (ESPiM): We applied a paired-samples t-test to look at the effect of gender on the calculated ESPiM and found a significant difference ($t(12) = 4.16, p < .001$) between male ($M = 56.97$ $bits$, $SE = 0.70$ $bits$) and female ($M = 60.01$ $bits$, $SE = 0.76$ $bits$) groups as illustrated in Fig. 30a. The results show that females experienced a higher eye-strain level in comparison to males based on our proposed model.

Movement time: We applied a paired-samples t-test to check the effect of gender on movement time based on the Fitts' law and found a significant difference ($t(12) = 2.65, p < .05$) between male ($M = 594.76$ $msec$, $SE = 28.95$ $msec$) and female ($M = 700.62$ $msec$, $SE = 39.09$ $msec$) groups as illustrated in Fig. 30b. In general males achieved quicker test run-times than females.

Error rates: We applied a paired-samples t-test to check the effect of gender on the recorded errors and found a significant difference ($t(12) = 2.93, p < .05$) between male ($M = 0.37$ $errors$, $SE = 0.09$ $errors$) and female ($M = 0.69$ $errors$, $SE = 0.09$ $errors$) groups as illustrated in Fig. 30c. This shows higher error rates for females than males in the test.

Eye fixations: We applied a paired-samples t-test to check the effect of gender on eye fixations and found no significant difference ($t(12) = 2.15, p > .05$) between male ($M = 44.85$ $fixations$, $SE = 2.00$ $fixations$) and female ($M = 50.03$ $fixations$, $SE = 2.16$ $fixations$) groups as illustrated in Fig. 30d. This shows both groups experienced a similar amount of eye fixations during the test.

Fixation qualitative score (FQlS): We applied a paired-samples t-test to check the effect of gender on the calculated FQlS measure (a measure that can reveal physical eye fatigue based on distance drift of fixation points) [92] and found no significant difference ($t(12) = 0.36, p > .05$) between male ($M = 411.05$ *pixels*, $SE = 11.87$ *pixels*) and female ($M = 417.83$ *pixels*, $SE = 14.84$ *pixels*) groups as illustrated in Fig. 30e.

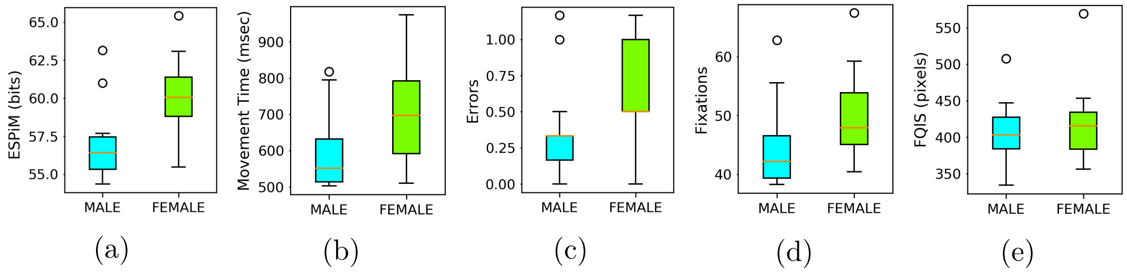


Figure 30: Illustration of gender-based analysis (a) ESPiM ($p < .001$), (b) movement time based on Fitts' law ($p < .05$), (c) error rates ($p < .05$), (d) eye fixations, and (e) Fixation Qualitative Score (FQlS) ($p > .05$) for 26 participants (13 male, 13 female).

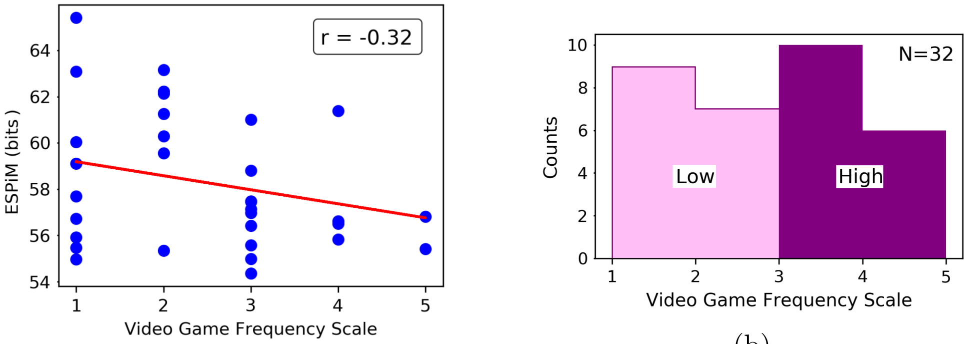
Table 9: Descriptive statistics of 26 participants based on gender (13 male, 13 female).

	ESPiM		Movement Time		Errors		Fixations		FQlS	
	M	F	M	F	M	F	M	F	M	F
Mean	56.9	60.0	594.7	700.6	0.3	0.6	44.8	50.0	411.0	417.8
Median	56.4	60.0	551.8	697.5	0.3	0.5	42.1	47.8	403.2	415.3
SD	2.5	2.7	104.4	140.9	0.3	0.3	7.2	7.7	42.8	53.5
IQR	2.1	2.5	118.8	200.8	0.1	0.5	7.2	8.8	43.0	50.5
Range	8.8	9.9	315.0	464.1	1.1	1.1	24.5	27.0	173.2	213.1
Min	54.3	55.4	502.8	510.1	0.0	0.0	38.2	40.4	334.3	356.0
Max	63.1	65.4	817.8	974.3	1.1	1.1	62.8	67.4	507.6	569.1

Studies have shown that there are gender-based differences in eye movements [27, 154]. Our results also suggest these differences from another perspective according to the proposed ESPiM model. Although there was statistically no significant differences on average number of fixations (ANF) measure among gender groups, the higher fixations for female group ($IQR_{female} > IQR_{male}$, $Range_{female} > Range_{male}$) and slower eye movements between targets (see Fig. 30b) led to a significantly higher ESPiM value for females.

Video-gameplay Analysis

In a pre-test questionnaire, we asked the participants to rate their video gameplay frequency on an integer scale of 1 (never) to 5 (every day). We found a negative correlation (Pearson's r = -0.32, and $p > .05$) which was not statistically significant, although it shows a trend for further discussion as shown in Fig. 31a. In order to further analyse the result, we divided our subjects into two groups based on their rate regarding the mean (2.5) of the scale. Participants with a rate smaller than the average value are labeled as *low* and those with a rate greater or equal than the average as *high* groups as illustrated in Fig. 31b. Each group was assigned exactly 16 participants. In general we found significantly higher eye-strain level, movement time, and error rates for the low group compared to the high group while no significantly different values for the eye fixations and the FQlS measures were recorded.



(a)

(b)

Figure 31: (a) Correlation of video game frequency and the calculated ESPiM values (Pearson's $r = -.32$), and (b) histogram of 32 participants based on their video game frequency on a scale of 1 (never) to 5 (every day) divided into Low and High groups.

Eye-strain probation model (ESPiM): We applied a paired-samples t-test to check the effect of video gameplay frequency on the calculated ESPiM ($t(15) = 2.14, p < .05$) and found a significant difference between low ($M = 59.52$ *bits*, $SE = 0.81$ *bits*) and high ($M = 57.04$ *bits*, $SE = 0.48$ *bits*) groups as illustrated in

Fig. 32a. Thus, those that frequently play video games experienced a significantly lower amount of eye-strain than the group of lower frequency of playing video games.

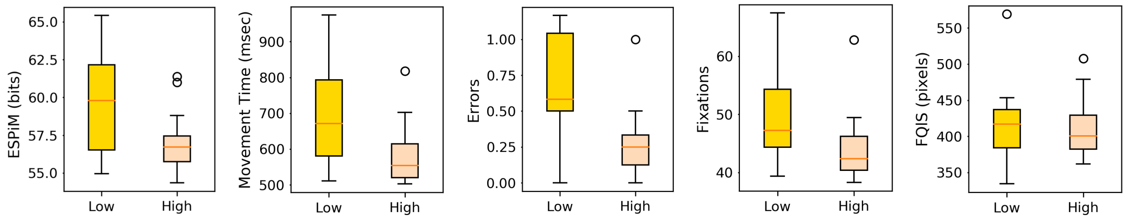
Movement time: We applied a paired-samples t-test to check the effect of video gameplay frequency on movement time based on the Fitts' law and found a significant difference ($t(15) = 2.36, p < .05$) between low ($M = 690.37$ $msec$, $SE = 34.63$ $msec$) and high ($M = 584.72$ $msec$, $SE = 21.53$ $msec$) groups as illustrated in Fig. 32b. Thus, those with a higher frequency of video gameplay were able to perform the task in a shorter amount of time than those who play video games less frequently.

Errors: We applied a paired-samples t-test to check the effect of video gameplay frequency on the recorded errors and found a significant difference ($t(15) = 3.62, p < .05$) between low ($M = 0.69$ $errors$, $SE = 0.09$ $errors$) and high ($M = 0.27$ $errors$, $SE = 0.06$ $errors$) groups as illustrated in Fig. 32c. This suggests that those with a higher frequency of video gameplay made less errors than those who do not play or play less frequently.

Eye fixations: We applied a paired-samples t-test to check the effect of video gameplay frequency on eye fixations with ($t(15) = 2.12, p > .05$) and found no significant difference between low ($M = 49.53$ $fixations$, $SE = 1.93$ $fixations$) and high ($M = 44.13$ $fixations$, $SE = 1.51$ $fixations$) groups as illustrated in Fig. 32d. Thus, gameplay frequency did not impact eye fixations during the test.

Fixation qualitative score (FQlS): We applied a paired-samples t-test to check the effect of video gameplay frequency on the calculated FQlS measure (a measure that can reveal physical eye fatigue based on distance drift of fixation points) with ($t(15) =

$0.28, p > .05$) and found no significant difference between low ($M = 416.00$ *pixels*, $SE = 13.13$ *pixels*) and high ($M = 411.67$ *pixels*, $SE = 10.12$ *pixels*) groups as illustrated in Fig. 32e. In other words, gameplay frequency does not impact fixation drift distances.



(a) (b) (c) (d) (e)

Figure 32: Illustration of (a) ESPiM ($p < .05$), (b) movement time based on Fitts' law ($p < .05$), (c) error rates ($p < .05$), (d) eye fixations, and (e) Fixation Qualitative Score (FQlS) ($p > .05$) based on video gameplay frequency of 32 participants divided into 2 equal groups (Low, High) with 16 participants in each group.

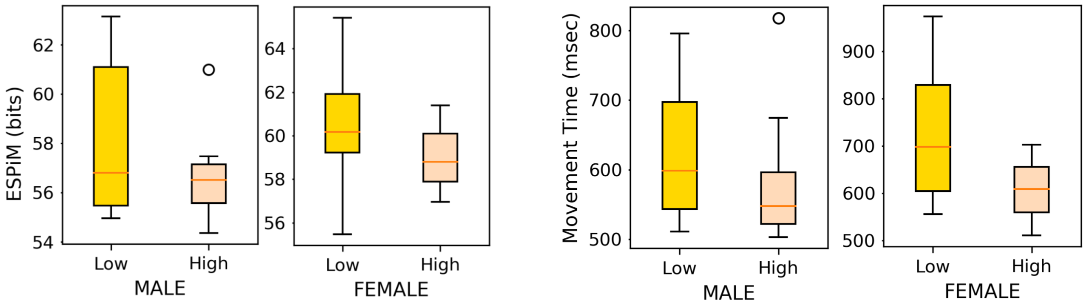
Table 10: Descriptive statistics based on video gameplay frequency divided into equal groups (Low, High) with 16 participants.

	ESPiM		Movement Time		Errors		Fixations		FQlS	
	Low	High	Low	High	Low	High	Low	High	Low	High
Mean	59.5	57.0	690.3	584.7	0.6	0.2	49.5	44.1	416.0	411.6
Median	59.8	56.7	671.6	554.3	0.5	0.2	47.2	42.3	416.8	400.3
SD	3.2	1.9	138.5	86.1	0.3	0.2	7.7	6.0	52.5	40.4
IQR	5.6	1.6	212.5	94.4	0.5	0.2	9.9	5.8	52.8	46.6
Range	10.4	7.0	463.3	315.0	1.1	1.0	28.1	24.5	234.8	145.8
Min	54.9	54.3	511.0	502.8	0.0	0.0	39.3	38.2	334.3	361.7
Max	65.4	61.3	974.3	817.8	1.1	1.0	67.4	62.8	569.1	507.6

We found a similar pattern on higher ANF (Average Number of Fixations) measure ($IQR_{low} > IQR_{high}$, and $Range_{low} > Range_{high}$). However, there was no significant differences among low and high groups, significantly higher movement times were the causes of the higher ESPiM values for the *low* group. This suggests that participants with higher frequency of gameplay were more experienced in moving their eyes in shorter time and therefore produced lower fixations and consequently lower eye-strain based on our model. This result may be due to the fact that video games can increase visual abilities. Previous studies have shown the relationship between video gameplay and eye movements, for instance shorter saccadic reaction time in video game players [101], the usage of video games to train visual skills [3], and to enhance visual search in players [26]. These characteristics of frequent game players enable them to experience lower eye-strain as indicated by our results.

Video-gameplay Among Gender Groups

We also analyzed the calculated ESPiM values among male and female participants based on their video-gameplay frequencies as shown in Fig. 33 and 34. Despite unequal number of participants based on their video-gameplay frequencies among male and female groups, participants with a lower frequency of gameplay (low group) show sparser distributions than participants with a higher frequency of gameplay (high group). This suggests that users with a higher frequency of video gameplay achieved similar results.

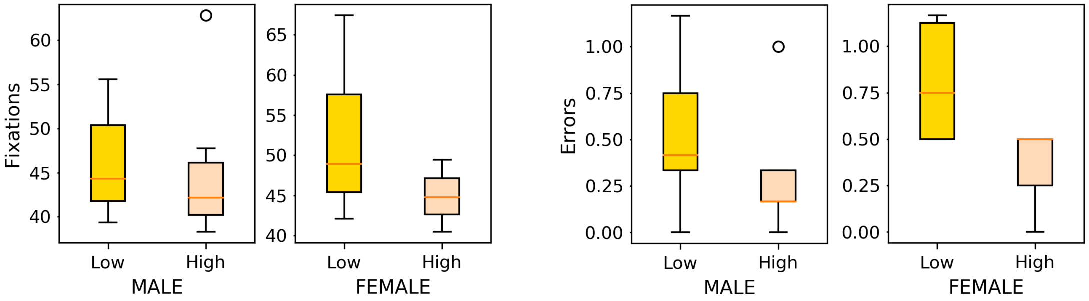


(a)

(b)

Figure 33: Illustration of video-gameplay frequency among gender groups on (a) ESPiM, and (b) movement time based on Fitts' law. The male group contains 19 (low=6, high=13), and female 13 (low=10, high=3) participants.

Further, we analyzed the differences among male and female participants based on their video-gameplay frequency and found participants in both gender groups show higher variability when their video-gameplay frequency is low as shown in Fig. 33



(a) (b)

Figure 34: Illustration of video-gameplay frequency among gender groups on (a) eye fixations, and (b) error rates. The male group contains 19 (low=6, high=13), and female 13 (low=10, high=3) participants.

Table 11: Descriptive statistics of 19 male participants (low=6, high=13) based on their video-gameplay frequencies.

	ESPiM		Movement Time		Errors		Fixations	
	Low	High	Low	High	Low	High	Low	High
Mean	58.2	56.5	626.4	579.6	0.5	0.2	46.1	43.9
Median	56.8	56.5	598.8	547.5	0.4	0.1	44.3	42.1
SD	3.6	1.6	112.2	87.0	0.4	0.2	6.3	6.4
IQR	5.6	1.5	153.8	73.8	0.4	0.1	8.6	5.8
Range	8.2	6.6	284.5	315.0	1.1	1.0	16.2	24.5
Min	54.9	54.3	511.0	502.8	0.0	0.0	39.3	38.2
Max	63.1	60.9	795.5	817.8	1.1	1.0	55.5	62.8

and 34 with details in Tables 11 and 12. In other words, for subjects that do not frequently play video games it is more difficult to predict eye-strain based on our proposed model.

Furthermore, we observed that male participants and participants with high frequency of video gameplay produced lower errors than their counterparts in target selections as shown in Fig. 30c, and 32c. The cause of low error rates in participants with high video game experience is due to their fast eye movements as shown in Fig. 32b. Similarly, fast eye movements in male participants was the cause of lower error rates shown in Fig. 30b. Thus those who have more "trained" visual search through video gameplay[3] tend to perform better with eye-tracking based selection.

Table 12: Descriptive statistics of 13 female participants (low=10, high=3) based on their video-gameplay frequencies.

	ESPiM		Movement Time		Errors		Fixations	
	Low	High	Low	High	Low	High	Low	High
Mean	60.3	59.0	728.7	606.9	0.8	0.3	51.5	44.8
Median	60.1	58.7	697.9	608.5	0.7	0.5	48.8	44.7
SD	2.9	2.2	143.6	96.0	0.3	0.2	8.0	4.5
IQR	2.7	2.2	224.4	96.0	0.6	0.2	12.1	4.5
Range	9.9	4.4	418.3	192.0	0.6	0.5	25.3	9.0
Min	55.4	56.9	556.0	510.1	0.5	0.0	42.1	40.4
Max	65.4	61.3	974.3	702.1	1.1	0.5	67.4	49.4

The mentioned observations based on gender reported in this study should not be interpreted to discriminate users towards a specific gender group. Additionally, the higher performance of experienced participant related to video gameplay cannot be interpreted as a promotion in favour of video games.

6.5 User Study 2: Focus Shift Simulator (online)

The second study, described here, was conducted to include a more in-depth analysis of eye-strain characteristics using a video-based eye-tracking technique applied via the WebGazer API [128], a recognized tool for remote eye-tracking user studies.

6.5.1 Methods

We developed a custom web application that runs on the client-side completely and requires no video footage to be sent to the server. The WebGazer API [128] uses the user's webcam to track eye movement and maps features of the eye and positions on the screen. It begins by obtaining the participant's consent to use the webcam followed by a calibration process. During this process, the user has to click on reference points with the mouse while looking at the cursor.

The application was developed in *Java* and the server interaction with the WebGazer API [128] was implemented in *JavaSctipt*. The AWS Elastic Beanstalk [5] provided by Amazon Web Services [4] was used for deploying our web application. Fig. 35 shows the workflow of the *focus shift simulator* user study.

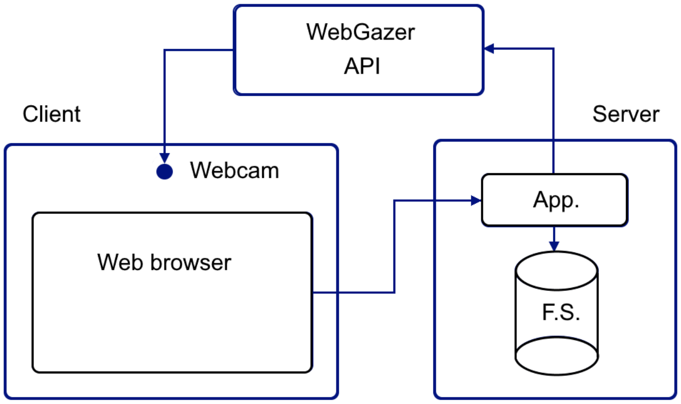
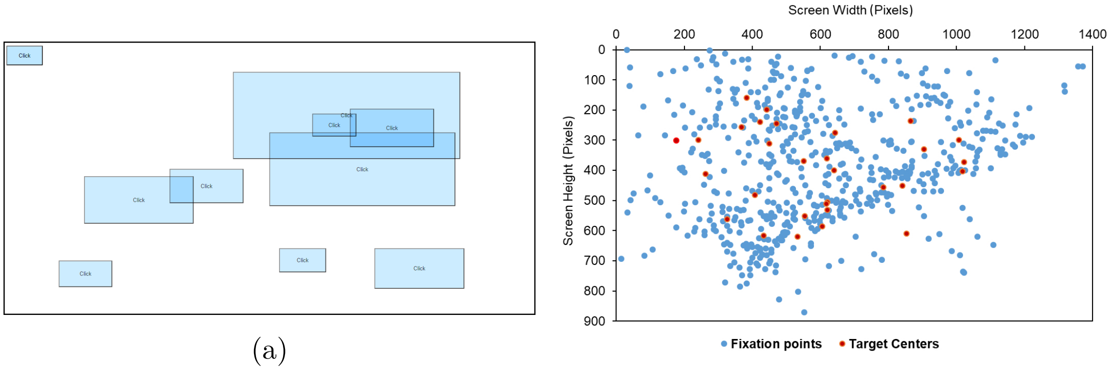


Figure 35: Overview of the focus shift simulator test application using the WebGazer API [128]. The user runs the application from a web browser, the client-server connection is established and the WebGazer API acquires control of the webcam to track user's eyes. The results of the entire test session is saved on the file system (F.S.) on the server.

In terms of the study, participants were asked to launch the developed web application at 3 moments of time in the day (from 08:00 to 12:00, 12:00 to 18:00 and after 18:00) over a period of 7 days (did not have to be consecutive). Prior to each trial in the study we asked the user how many hours they have been working or using a screen. Next, the participant was prompted to click on 30 buttons that randomly appear on the screen one at a time. Each button has a different position and size adjusted to the screen dimension as illustrated in Fig. 36a. By clicking on the last button, the recorded raw data is sent to the server. At the end of the trial we asked the participant to rate their current eye-strain level on a scale from 1 (none) to 5 (a lot).

The intuition behind the developed *focus shift* application is based on the multitasking efforts to interact with different interactions, applications, and events while working on a digital display. The user needs to move their focal point to react to situations. For instance, reading emails and constantly receiving messages on a messenger application and monitoring running applications in background which may interrupt user's attention with prompt dialog boxes to confirm/decline certain actions are typical tasks of an office employee. Fig. 36a illustrates the screenshot of the test application with the goal of simulating multi-tasking events on a computer and Fig. 36b shows the pattern of eye fixations when shifting to different locations on the screen to follow stimuli and illustrates the intuition of fixation points integration in our proposed model (see Equation 9).



(a)

(b)

Figure 36: (a) Shows the screenshot of the focus shift simulator application for 10 targets. We reduced the number of targets for this screenshot to reduce overlapping of targets for higher visibility. Targets contain 'click' label and appear one by one randomly across the screen with different sizes. As soon as user clicks on a button, the application removes the selected button and loads the next one. (b) Illustrates the recorded fixation points of an entire test session with 30 targets on a laptop display. The red circles represent center of targets, and the blue circles represent fixation points. This figure shows the intuition of fixation points to be integrated in our eye-strain model to be considered for a specific period of time.

This study was specifically designed to be applicable remotely to comply with physical restrictions caused by the COVID-19 pandemic [33]. Moreover, since the study collects data from 24 hours a day from users, it could be executed anytime and anywhere via the URL to the server to provide participants freedom and control over their interactions with the system rather to run the tests in a laboratory equipped with eye-tracking sensors. Therefore the application of video-based eye tracking was found to be the ideal choice of design.

After analysing the preliminary results we found that some participants failed to stick to the restricted time slots. This may be due to the fact that our testing sessions collided with personal tasks and that since the start of the COVID-19 pandemic [33], many workers began working from home with varying work routines rather than the traditional 9-to-5 workday. As we originally wanted to study our ESPiM model for classical 9-to-5 working hours, based on our data we defined two groups of participants:

- **9-to-5**: any time between 09:00 and 17:00 o'clock.
- **Flexible**: any times not in the 9-to-5 group.

These slots were chosen based on the concept of 9-to-5 working hours and there is no intersections between the groups (9-to-5 $\cap$ Flexible $= \emptyset$).

We had four hypotheses that we considered in this study:

- H1: Users perceive higher eye-strain level beyond the standard 9-to-5 working hours: ESPiM(9-to-5) $<$ ESPiM(Flexible).
- H2: The group of 9-to-5 participants may cause more errors than the flexible group. In simple words, we assume that the flexible group may choose their preferred hours of work or take more time in between work periods and thus make less errors.
- H3: The longer users spend on a digital display, the more eye-strain they have.
- H4: The increase of the calculated ESPiM value correlates with the increase of mouse pointer movements.

We collected 70 samples from 13 participants (10 Male, 3 Female) with an average age of 31.33 years (SE = 2.01) to be analysed based on their working times and ratings. Specifically, we collected (a) screen resolution, (b) test duration, (c) errors, (d) eye fixations, (e) display hours, (f) perceived eye-strain rating, and (g) eye-strain symptoms of participants.

6.5.2 Results and Discussion

Since the start of the COVID-19 pandemic [33] many knowledge workers and students were forced to work from home which typically deviates from the routine of 9-to-5 working hour schedules. This phenomenon motivated us to apply ESPiM to study the impacts of remote working via measuring eye-strain based on our proposed equation (see Equation 9). Although the calculated ESPiM values of both groups (9-to-5, and flexible) showed no significant difference, we observed significant patterns in the (a) subjective ratings, (b) time spent on a digital display, and (c) recorded error rates among groups.

To test our first hypothesis, we applied a paired-samples t-test to check the difference of the calculated ESPiM (t(34)=1.90, $p > .05$) among both groups and found no significant difference between 9-to-5 (M = 30.58 $bits$, SE = 1.22 $bits$) and flexible (M = 27.76 $bits$, SE = 1.34 $bits$) groups as illustrated in Fig. 37a. This suggests our first hypothesis (H1) concerning a lower eye-strain level for 9-to-5 participants than the flexible group is rejected.

To test our second hypothesis, we applied a paired-samples t-test to check the effect of errors on the calculated ESPiM (t(34)=2.47, $p < .05$) and found a significant difference between 9-to-5 (M = 0.31 $errors$, SE = 0.09 $errors$) and flexible (M = 0.08 $errors$, SE = 0.04 $errors$) groups as illustrated in Fig. 38b. As we had predicted a higher error rates of the 9-to-5 group than the flexible group in our second hypothesis

(H2) which is thus confirmed by these results. This shows the flexible group, who are more flexible to work on their digital devices, was able to finish the test with fewer mistakes.

We applied a paired-samples t-test to check the effect of subjective ratings on the calculated ESPiM (t(34)=3.58, $p < .001$) and found a significant difference between 9-to-5 ($M = 1.34$ *points*, $SE = 0.10$ *points*) and flexible ($M = 2.00$ *points*, $SE = 0.15$ *points*) groups as illustrated in Fig. 37b. Further, we looked at the effect of display hours on the calculated ESPiM (t(34)=6.52, $p < .001$) and found a significant difference between 9-to-5 ($M = 2.41$ *hours*, $SE = 0.27$ *hours*) and flexible ($M = 5.64$ *hours*, $SE = 0.37$ *hours*) groups as illustrated in Fig. 37c. Thus, although we predicted a higher eye-strain level based on ESPiM values for participants who spend more time on a digital display in our third hypothesis (H3), and we recorded significantly higher working hours for the flexible group than the 9-to-5 participants, the difference of ESPiM values was not statically significant and therefore we reject our third hypothesis. We posit that the reason for higher working hours but relatively similar eye-strain level of the flexible group lies in the fact that flexible participants could work anytime based on their convenience and therefore were benefited from breaks rather that those bound to a certain time window as the 9-to-5 group.

To test our fourth hypothesis, we studied the correlation of ESPiM and mouse pointer movements for each group and found positive correlations among those measures as predicted and therefore confirm the hypothesis (H4) as shown in Fig. 39a, and 39b. This result suggest that tired eyes may lead to an increase of mouse pointer movements among users for target selection tasks.

We also explored the effect of eye fixations on the calculated ESPiM (t(34)=1.46, $p > .05$) and found no significant difference between 9-to-5 ($M = 904.28$ $fixations$, $SE = 64.06$ $fixations$) and flexible ($M = 791.08$ $fixations$, $SE = 76.35$ $fixations$)

groups as illustrated in Fig. 38a. This shows both groups of users experienced similar amount of eye fixations during our test.

In addition, we studied the effect of mouse pointer movements (t(34)=1.81, $p > .05$) and found no significant difference between 9-to-5 ($M = 667.14$ *movements*, $SE = 47.87$ *movements*) and flexible ($M = 565.11$ *movements*, $SE = 45.04$ *movements*) groups as illustrated in Fig. 38c. However, the pointer movements correlate with the calculated ESPiM of each group. This shows both groups of users applied similar amount of mouse pointer movements during our test.

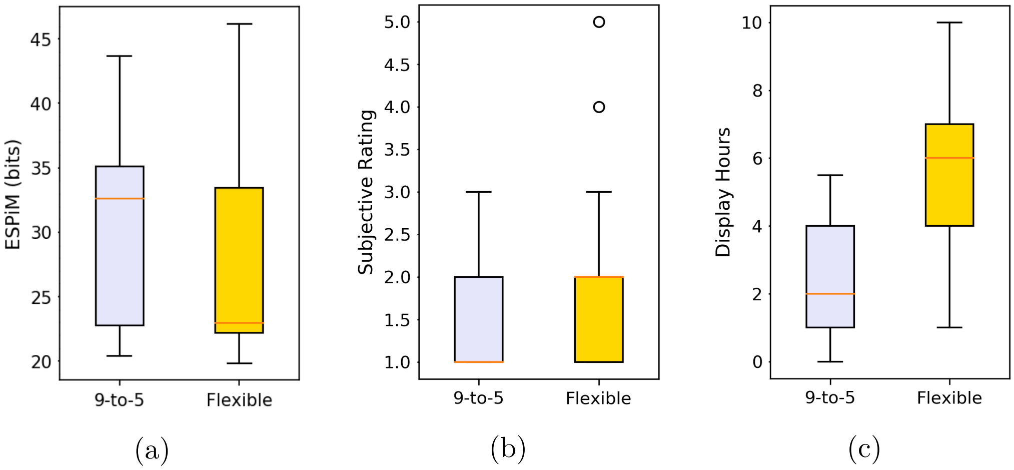


Figure 37: Illustration of (a) ESPiM (p > .05), (b) subjective ratings of perceived eye-strain level ($p < .001$), and (c) display hours before test ($p < .001$) of both testing groups. More details in Tables 14 and 15.

Furthermore, we recorded participants' symptoms of eye-strain after each session as illustrated in Fig. 40. It should be noted that these symptoms are based on subjective feedback of participants and do not reflect clinical definitions. We found that the flexible group reported more eye-strain symptoms (29 symptoms) than the 9-to-5 group (15 symptoms). The large gap between groups are related to *tired eyes*, *dry eyes*, and *blurred vision* which can be explained due to the higher working hours on a computer display of the flexible group as shown earlier in Fig. 37c. In simple

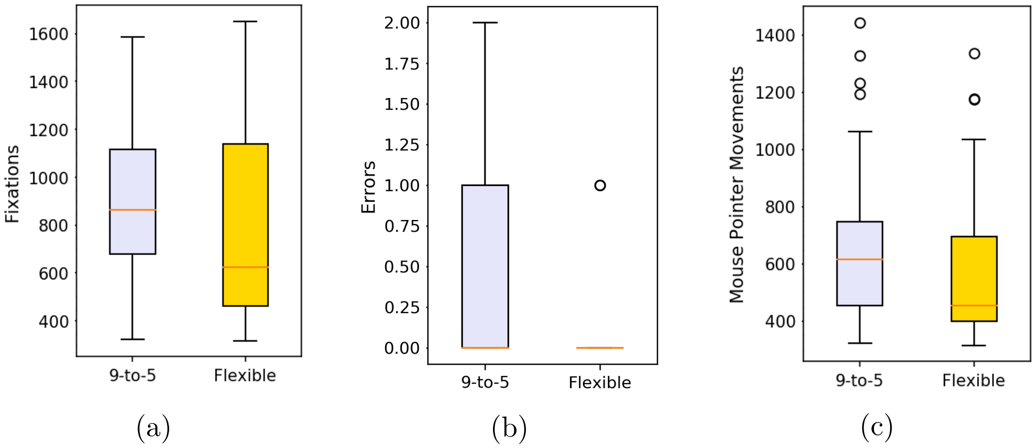


Figure 38: Illustration of (a) recorded eye fixation points ($p > .05$), (b) error rates ($p < .05$), and (c) mouse pointer movements of both testing groups ($p > .05$). More details in Tables 14 and 15.

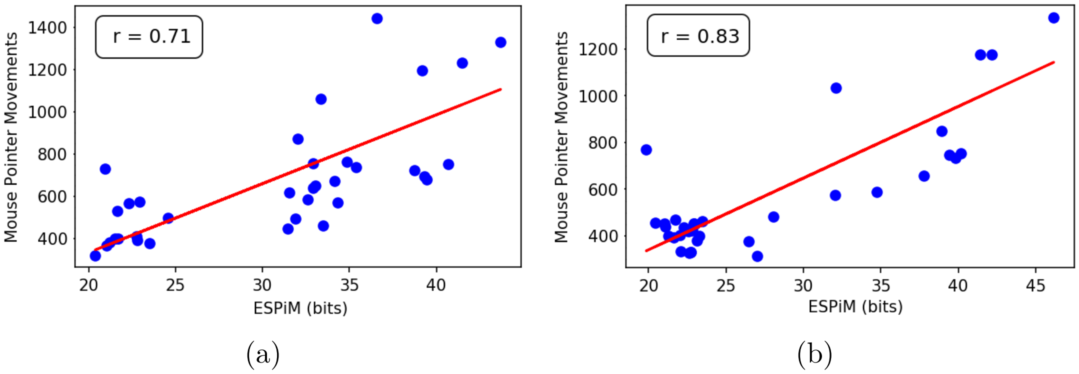


Figure 39: Correlations of ESPiM and mouse pointer movements for (a) the 9-to-5 group (Pearson's $r = .71$, $p < .001$), and (b) the flexible group (Pearson's $r = .83$, $p < .001$).

words, working in a 'flexible' routine seems to lead to higher amount of hours spent on a display, which in turn leads to more eye-strain symptoms.

Since participants ran the test on their personal computers at home, we recorded different screen resolutions and analysed the effectiveness of ESPiM on different screens as illustrated in Fig. 41. We found no significant difference between the 9-to-5 group ($M = 32.26\ bits, SE = 2.81\ bits$) and the flexible group ($M = 30.83\ bits, SE = 3.25\ bits$) based on screen resolution, although there is a slight decrease of ESPiM for screen dimension 1280 × 720 pixels for the 9-to-5 group.

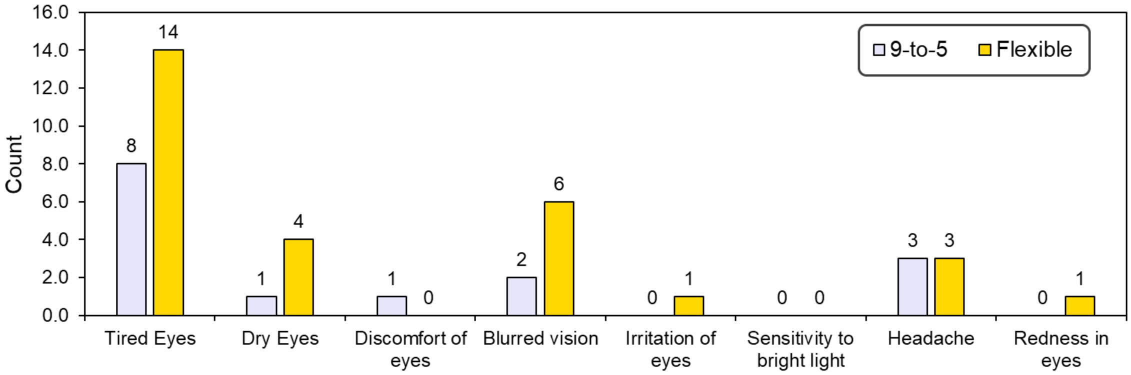


Figure 40: The recorded eye-strain symptoms of participants after running the test.

Additionally, we calculated the difference of ESPiM values for each resolution point (shown as P1-P7 on Fig. 41) to analyze the increase and decrease of ESPiM values per resolution increase. In fact, we calculated $P_n - P_{n-1}$ $(1 \leq n \wedge P_0 = 0)$ for each resolution as shown in Table 13. Negative values represent *decrease*, and positive values *increase* of ESPiM per bits. The summation of both groups shows a slightly higher increase for the flexible group with the value of 32.11 *bits* compared to that of the 9-to-5 group with the value of 32.01 *bits*. However, the difference lies only on a single resolution (1280 × 720 pixels), this might suggest that users with a flexible working time may experience higher eye-strain levels as screen dimensions increase than the group of standard 9-to-5 routine. However, the increase of eye-strain for both groups decreases for screen dimensions larger than 2048 × 1152 pixels. This may suggest that the choice of screen size is essential in how users experience eye-strain for any working time schedules. This result might be of interest for video game designers in finding suitable screen resolutions for their audience to recommend for the best game experience with lower eye-strain levels.

This previous analysis is important as ESPiM takes both target's and screen's area in to account to calculate eye-strain value for comparison as shown in Equation 9. Therefore, ESPiM is applicable on any screen resolution with no further adjustments in its equation. This feature enables designers of user interfaces, or producers of digital displays to compare different screen dimensions based on eye-strain on consumers.

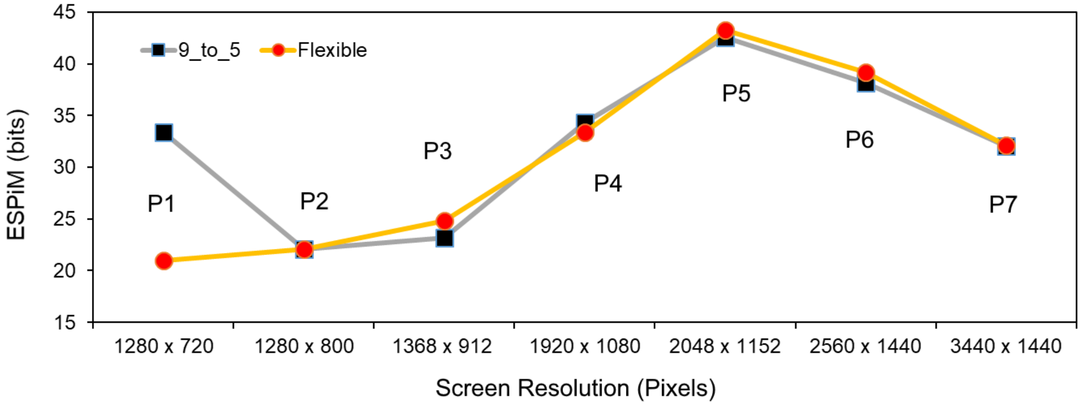


Figure 41: The recorded ESPiM values of both groups based on screen resolutions. The comparison points are shown as $P1$ to $P7$ which build comparison sections to study ESPiM on screen expansion. We analysed the increase and decrease of ESPiM in each section.

Table 13: ESPiM difference based on screen resolution for both 9-to-5, and flexible groups. Positive values represent increase and negative values decrease of ESPiM per bits respectively.

Resolution (Pixels)	Measure	ESPiM 9-to-5	ESPiM Flex.
1280 × 720	P1	33.38	20.98
1280 × 800	P2 - P1	-11.28	1.10
1368 × 912	P3 - P2	1.11	2.73
1920 × 1080	P4 - P3	11.13	8.58
2048 × 1152	P5 - P4	8.25	9.86
2560 × 1440	P6 - P5	-4.41	-4.03
3440 × 1440	P7 - P6	-6.17	-7.11
	Σ	**32.01**	**32.11**

We demonstrated the application of ESPiM in a remote user study with a video-based eye-tracking technique. We have shown the effectiveness of ESPiM for user studies to analyse eye-strain on digital displays. Although the results of ESPiM cannot be interpreted as clinical analysis, they can be applied in low-budged user studies with physical restrictions to participants.

Table 14: Descriptive statistics of the focus shift simulator study.

	ESPiM		Fixations		Display Hours		Errors	
	9-to-5	Flex.	9-to-5	Flex.	9-to-5	Flex.	9-to-5	Flex.
Mean	30.58	27.76	904.28	791.08	2.41	5.64	0.31	0.08
Median	32.61	22.96	865.00	625.00	2.00	6.00	0.00	0.00
SD	7.24	7.96	378.98	451.72	1.65	2.21	0.53	0.28
IQR	12.35	11.24	437.00	676.00	3.00	3.00	1.00	0.00
Range	23.31	26.32	1264.00	1334.00	5.50	9.00	2.00	1.00
Min	20.38	19.83	322.00	317.00	0.00	1.00	0.00	0.00
Max	43.69	46.16	1586.00	1651.00	5.50	10.00	2.00	1.00

Table 15: Descriptive statistics of mouse pointer movements.

	Mouse Pointer Movements		Subjective Rating	
	9-to-5	Flex.	9-to-5	Flex.
Mean	667.14	565.11	1.34	2.00
Median	617.00	454.00	1.00	2.00
SD	283.22	266.50	0.59	0.93
IQR	291.00	295.50	1.00	1.00
Range	1121.00	1021.00	2.00	4.00
Min	322.00	315.00	1.00	1.00
Max	1443.00	1336.00	3.00	5.00

6.6 Conclusion and Future Work

Eye-strain is a common issue among computer users due to prolonged periods spent working and using digital displays, which leads to vision problems such as irritation and tiredness of the eyes, and headaches. We proposed the Eye-Strain Probation Model (ESPiM), an easy-to-apply computational model to measure eye-strain on digital displays based on the spatial properties of the user interface and display area for a required period of time, using eye-tracking analysis integrated into a single measure. We conducted two user studies to evaluate the effectiveness of ESPiM, its functionalities and potentials and showed how to measure potential eye-strain levels of a specific user interface suitable for pilot studies to compare various design prototypes before the application by end users. We evaluated the effectiveness of ESPiM in an in-person user study with an infrared eye-tracking sensor and found interesting patterns among (a) gender, and (b) video gameplay play frequency groups. In addition, we showed the application of ESPiM in a remote user study that complied with the COVID-19 safety measures. Since eye fixations can be assessed which is typically bound to a certain range (200-600 ms), ESPiM can be applied in pilot studies with no access to eye-tracking devices for analysis estimations. Despite the relatively short test session for each participant of both user studies, we were able to evaluate our hypotheses and recognized distinctive patterns among participants. Furthermore, we showed that ESPiM has strong correlations with *error rates* of target selections. The correlation with error rates can be interpreted as an indicator to estimate and reduce the impacts of the Midas touch problem in gaze-based interactions by analysing the user interface properties. We also found significantly different eye-strain patterns based on video gameplay frequency of participants. The results showed that participants with frequent video gameplay reached significantly lower eye-strain, and lower error rates compared to their counterparts. This may suggest that users with higher training of eye focus shifts (e.g. video games) might experience a lower eye-strain with prolonged use of digital displays for singular (not multi-tasking) or enjoyable tasks. Furthermore, we found that mouse pointer movements increase as eye-strain levels increase. This would suggest users tend to move their mouse pointers more frequently to select targets in case of tired eyes.

Beyond the prediction and measurement of eye-strain, ESPiM can be applied to compare different gaze-based interaction techniques, and evaluate different user interface prototypes to reach a comfortable design based on eye-strain.

As more individuals get access to digital content provided on digital displays, especially children and students of a younger age, the consumption of digital media becomes more prevalent. Furthermore, since the start of the COVID-19 pandemic, many schools moved to online teaching which caused new challenges for elementary students and their parents. Many in-person events still occur on digital devices now to comply with the required safety measures. Thus, there is a need for compound models to cope with the emerging trends in order to measure and compare the impact of digital displays on our health.

Today, most smart-phones include high quality cameras accessible by younger generations, thus the application of video-based eye-tracking that we showed in this paper can be feasible to design and conduct large-scale user studies on smartphones to reach a larger population of digital media consumers. In future work, we plan to apply ESPiM on smart-phones to compare educational video games for school students. Additionally, the continuous and predictable form of ESPiM makes it suitable for machine learning algorithms which will be explored in the future. Finally, we hope our proposed model makes a step forward towards the reduction of eye-strain on digital displays, inspires researchers and user interface designers and leads to more discussions in research communities.

Chapter 7

Conclusion

Computer vision syndrome (CVS) is composed of multiple eye vision problems due to the prolonged use of digital displays, including tablets and smartphones. These problems have been shown to affect visual comfort and work productivity in both adults and teenagers. CVS causes eye and vision symptoms such as *eye-strain*, *eye burn*, *dry eyes*, *double vision*, and *blurred vision.* Furthermore, CVS causes severe vision and muscular problems and is a cause of work stress due to repeated eye movements and excessive eye focus on computer screens. Work-related stress is one of the main challenges of the workforce in the 21st century, and according to the World Health Organization (WHO), work-related stress occurs when the workload demand is higher than the knowledge and abilities of workers to handle. Employees may feel overwhelmed with the amount of work to be accomplished in a limited amount of time and may feel no support to handle their tasks; thus, they feel stressed.

The severe health risks caused by work-related stress mentioned in this dissertation might be complicated and expensive to cure. Human physiology and psychology are very complex, and the mentioned side effects are only a few examples among unknown issues that can arise. According to the current evidence, permanent work-related stress is harmful with many complicated symptoms. Thus, it should be detected and handled in the early phases. Since exposure to work-related stress is inevitable and has become part of our daily lives, there need to be mechanisms to reduce its destructive impacts on our health. In this thesis, we examined eye-strain as one of the significant CVS symptoms that could be used to address this aspect of work-related stress among computer users.

Eye-strain, also known as visual fatigue, is growing more common among digital device users. Symptoms do include not only irritation of the eyes but also tiredness and headaches. Eye-strain is a common issue among computer users due to the prolonged periods working in front of a monitor.

As emerging interaction techniques become more sophisticated and multi-dimensional, the need for more complex and multi-factor measures is necessary. We have developed multi-purpose mathematical models *Fixation-based Eye fatigue Load Index* (FELiX), *Index of Difficulty for Eye tracking Applications* (IDEA), and *Eye-Strain Probation Model* (ESPiM) based on eye-tracking parameters and subjective ratings to measure, predict, and compare the amount of eye-strain, fatigue or cognitive workload during target selection tasks for different user groups or interaction techniques. The ESPiM model is the outcome of both FELiX and IDEA relying on objective measures solely. These trilateral models enable researchers to predict and quantify potential eye-strain levels on individuals based on physical circumstances such as screen resolution and target positions per time and, consequently, could reduce work-related stress.

7.1 Summary of Findings

The contributions of our proposed models are twofold: firstly, they enable researchers and designers of user interfaces to assess the amount of eye-strain as a *proxy* to work stress based on the visual parameters. Secondly, they can be applied to different case scenarios, display types or compare different interaction techniques in user studies to identify and reduce the side effects of visual interactions on digital displays. Although there is always a gap between **prediction** and **reality**, our trilateral compound models contribute to assess the challenges of gaze-based software development and provide new insights into the feasibility of multi-factor prediction measures for gaze-based interactions. Table 16 summarizes all research questions explored in this thesis and their conclusions.

Although we have designed the ESPiM model to measure eye-strain on primarily digital displays, it can be applied to (a) compare user interface designs, (b) compare different display devices, and (c) compare different interaction techniques based on eye-strain on users. Since spatial parameters are included in the ESPiM model, it can be used to estimate eye-strain before testing in a user study, which can benefit both research communities and the industry of creating digital content. We incorporate the screen's and targets' spatial parameters, including applying Fitts' law and eye-tracking fixation points for the desired period in a single measure. Moreover, eye fixations are typically bound to a specific range (200-600 ms). Therefore, the average number of fixations for a specific period can be estimated, making ESPiM a suitable choice for testing and evaluating purposes with no access to eye-tracking sensors.

7.2 Limitations of the Eye-strain Models

However the proposed models can be applied in a variety of user studies, there are some limitations to be considered. A major issue that applies to all proposed models is the analysis of long-term aspects of these models on users in further studies. We review the limitations based on each model.

7.2.1 FELiX

The major limitation of FELiX is the absence of the entire scores of the NASA TLX. However, it includes the most important parameters of the subjective rating, the rating scores of *frustration*, *temporal demand*, and *effort* parameters may reveal some important aspects of a subjective rating coefficient.

7.2.2 IDEA

However the IDEA model takes the entire NASA TLX scores into account as the subjective rating criterion, it requires a separate test to measure the selection ratio parameter S (the ratio of distance to target over the diameter of the screen) which demands a test application and might be time-consuming for some user studies.

7.2.3 ESPiM

Although the ESPiM model is designed to be as flexible as possible, a precise measurement of test execution time is required for any user study. Thus an accurate synchronization between the start and end of a test is needed. Further, it does not take preparation activity times into account, for instance, spent time on the calibration process before running a test.

7.3 Future Works

There are a number of avenues for future work as follows. We review these research and development directions briefly.

7.3.1 Smartphones, Tablets, and AR/VR Devices

Applying our trilateral models on smartphones/tablets may reduce eye-strain levels on mobile applications and video games specifically designed for children. The ESPiM model enables researchers to optimize their user interface designs based on small-size screens of smartphones and tablets.

Our trilateral proposed models can be applied on head-mounted devices for AR/VR applications with integrated eye-tracking sensors to study and analyze motion sickness, which is a common challenge in these devices. All our models were originally designed for 2D screens. Thus, an exciting research direction is to evaluate and expand these models to 3D view volumes, typically associated with stereo vision, for example, using VR headsets.

7.3.2 Driver Awareness Analysis

Our models can also be suitable for analyzing Microsleep, a sudden loss of awareness or a short sleep in car drivers. Especially for truck drivers, this could reduce traffic collisions by measuring eye-strain levels for different times and conditions in user studies. Since our models contain mathematical equations, researchers may apply machine learning techniques to run or expand these models in real-time use case scenarios to detect driver's awareness based on eye-strain.

7.3.3 User Interface and Screen Design

Our models, especially the ESPiM model, are suitable for user interface designers, producers of display devices, and video game developers to optimize their output in user studies to reduce eye-strain before the final release of their works. Since ESPiM integrates both (1) target's area and (2) screen's area into account, the calculation of eye-strain can be adopted on various screen dimensions as we showed the results of our online user study.

7.3.4 Alternative Method to Biological Sensors

Measuring biological inputs from users is one of the most effective methods to record involuntary users' reactions during user studies. The application of these sensors is limited to certain domains, may need specific ethical approval and acquisition of expensive equipment. Moreover, these sensors may be intrusive for test participants, which may affect the results. Thus, our models can be applied as low-budget and easy-to-calculate alternatives to biological sensors since they include eye-tracking measures (e.g. fixations points, fixation duration, etc.), which are involuntary inputs from participant's eyes in response to visual stimuli for pilot studies or preliminary testing phases.

7.3.5 Comparison Measures

Finally, our trilateral models can be applied as simple *comparison* measures for any user study to find distinctive patterns based on different criteria (gender, age, experiences with eye-tracking, etc.) by applying eye-tracking fixation points recorded through a commodity webcam for preliminary analysis. These comparisons are helpful to distinguish different criteria, interaction techniques, or conditions for user studies as we showed some results of our user studies.

7.4 Perspective

In this thesis, we have reviewed eye-strain and work-related stress caused by the workplace and discussed detection techniques and equipment to measure stress.

Specifically, we focused on eye-tracking as a feasible and low-budget tool for the measurement of eye-strain. In the era of smart devices capable of visually representing data, we believe it is essential to determine stress levels by applying eye-tracking techniques and analyzing the eye-strain of particular targets and screen dimensions for a specific period. The results of the proposed work can lead to predict and prevent work overload and stressful situations by offering estimation reports on the difficulty of selection tasks for users who spend hours working in front of a computer. This prevention is not only economical but would help to improve work-life balance. We hope that our models make a step forward towards the reduction and control of work-related stress.

Table 16: Summary of the research questions and findings of this dissertation.

Research Question	Finding
Can we propose an alternative method to dwell-time method, and voice recognition selection technique for the Midas touch problem?	Yes. We designed, developed, and evaluated EyeTAP with a completely contact-free approach to address the Midas touch problem with comparable results.
Can we measure eye-strain based on the NASA TLX subjective feedback?	Yes. We designed, developed, and evaluated FELiX, and IDEA which integrate subjective scores in form of the NASA TLX scores into their equations.
Can we expand the Fitts' law for eye-tracking applications?	Yes. We designed, developed, and evaluated IDEA which measures task difficulty for selection techniques with inclusion of width and distances of targets. Additionally, we proposed ESPiM with the integration of the Fitts' law principles.
Can we measure eye-strain by a standalone model relying solely on objective measures to be applicable on any 2D display?	Yes. We designed, developed, and evaluated ESPiM in two user studies with a dedicated infrared eye-tracking sensor and an online study using video-based eye-tracking technique via commodity webcams on user's computer.
Can we apply eye-tracking as a simple alternative to biological sensor for user studies with low budget or no access to measuring sensors?	Yes. We proposed FELiX, IDEA, and ESPiM which can be applied in user studies to record user's inputs from a safe distance with no extra peripherals via a low-budget eye-tracking sensor, or a webcam.
Can we propose a model to measure eye-strain applicable on various screen dimensions?	Yes. We designed, developed, and evaluated ESPiM which integrates both targets' and screen dimensions into its equation.
Can we expand the original Fitts' law to include time?	Yes. We designed, developed, and evaluated ESPiM which takes benefit of the Fitts' law principles for a specific duration.

Bibliography

[1] E. Abdulin and O. Komogortsev. User eye fatigue detection via eye movement behavior. In *Proceedings of the 33rd annual ACM conference extended abstracts on human factors in computing systems*, pages 1265–1270. ACM, 2015.

[2] C. Acartürk, J. Freitas, M. Fal, and M. S. Dias. Elderly speech-gaze interaction. In M. Antona and C. Stephanidis, editors, *Universal Access in Human-Computer Interaction. Access to Today's Technologies*, pages 3–12, Cham, 2015. Springer International Publishing.

[3] R. L. Achtman, C. S. Green, and D. Bavelier. Video games as a tool to train visual skills. *Restorative neurology and neuroscience*, 26(4, 5):435–446, 2008.

[4] I. o. i. a. Amazon Web Services. `https://aws.amazon.com/`, 2021.

[5] I. o. i. a. Amazon Web Services. Aws elastic beanstalk. `https://aws.amazon.com/elasticbeanstalk/`, 2021.

[6] J. Annett. Subjective rating scales: science or art? *Ergonomics*, 45(14):966–987, 2002.

[7] M. Bâce, T. Leppänen, D. G. de Gomez, and A. R. Gomez. ubigaze: Ubiquitous augmented reality messaging using gaze gestures. In *SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications*, SA '16, pages 11:1–11:5, New York, NY, USA, 2016. ACM.

[8] A. T. Bahill and L. Stark. Overlapping saccades and glissades are produced by fatigue in the saccadic eye movement system. *Experimental neurology*, 48(1):95–106, 1975.

[9] G. Bailey. Office workers spend 1,700 hours a year in front of a computer screen. `https://www.independent.co.uk/news/uk/home-news/office-workers-screen-headaches-a8459896.html`, 2018.

[10] G. Barnes. Cognitive processes involved in smooth pursuit eye movements. *Brain and Cognition*, 68(3):309–326, 2008. A Hundred Years of Eye Movement Research in Psychiatry.

[11] G. R. Barnes. Rapid learning of pursuit target motion trajectories revealed by responses to randomized transient sinusoids. *Journal of Eye Movement Research*, 5(3), 2012.

[12] R. Bednarik, T. Gowases, and M. Tukiainen. Gaze interaction enhances problem solving: Effects of dwell-time based, gaze-augmented, and mouse interaction on problem-solving strategies and user experience. *Journal of Eye Movement Research*, 3(1), Aug. 2009.

[13] T. R. Beelders and P. J. Blignaut. The usability of speech and eye gaze as a multimodal interface for a word processor. *Speech Technologies*, pages 386–404, 2011.

[14] T. R. Beelders and P. J. Blignaut. Using eye gaze and speech to simulate a pointing device. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 349–352, 2012.

[15] R. Bellman and R. Kalaba. On adaptive control processes. *IRE Transactions on Automatic Control*, 4(2):1–9, November 1959.

[16] P. Biswas and P. Langdon. Multimodal intelligent eye-gaze tracking system. *International Journal of Human-Computer Interaction*, 31(4):277–294, 2015.

[17] D. Black, M. Unger, N. Fischer, R. Kikinis, H. Hahn, T. Neumuth, and B. Glaser. Auditory display as feedback for a novel eye-tracking system for sterile operating room interaction. *International journal of computer assisted radiology and surgery*, 13(1):37–45, 2018.

[18] G. Borghini, G. Vecchiato, J. Toppi, L. Astolfi, A. Maglione, R. Isabella, C. Caltagirone, W. Kong, D. Wei, Z. Zhou, et al. Assessment of mental fatigue during car driving by using high resolution eeg activity and neurophysiologic indices. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6442–6445. IEEE, 2012.

[19] A. Brownawell and K. Wiggins. Americans stay connected to work on weekends, vacation and even when out sick, September 2013. `https://www.apa.org/news/press/releases/2013/09/connected-work`.

[20] M. Byrom. The death of the workday: Is 9 to 5 working obsolete?, May 2019. `https://www.business.com/articles/the-death-of-the-workday-is-9-to-5-working-obsolete/`.

[21] D. Calleja. The problem with open-plan offices (and how to fix it). `https://www.azuremagazine.com/article/open-plan-office-problems/`, 2018.

[22] T. W. Calvert, E. W. Banister, M. V. Savage, and T. Bach. A systems model of the effects of training on physical performance. *IEEE Transactions on systems, man, and cybernetics*, (2):94–102, 1976.

[23] Cantoni, Virginio and Galdi, Chiara and Nappi, Michele and Porta, Marco and Riccio, Daniel. GANT: Gaze analysis technique for human identification. *Pattern Recognition*, 48(4):1027–1038, 2015.

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

[25] G. Casiez, N. Roussel, and D. Vogel. 1€ filter: a simple speed-based low-pass filter for noisy input in interactive systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2527–2530. ACM, 2012.

[26] A. D. Castel, J. Pratt, and E. Drummond. The effects of action video game experience on the time course of inhibition of return and the efficiency of visual search. *Acta Psychologica*, 119(2):217 – 230, 2005.

[27] V. Cazzato, D. Basso, S. Cutini, and P. Bisiacchi. Gender differences in visuospatial planning: An eye movements study. *Behavioural Brain Research*, 206(2):177 – 183, 2010.

[28] T. Chandola, A. Britton, E. Brunner, H. Hemingway, M. Malik, M. Kumari, E. Badrick, M. Kivimaki, and M. Marmot. Work stress and coronary heart disease: what are the mechanisms? *European heart journal*, 29(5):640–648, 2008.

[29] I. Chatterjee, R. Xiao, and C. Harrison. Gaze+ gesture: Expressive, precise and targeted free-space interactions. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pages 131–138, 2015.

[30] F. Chen, J. Zhou, Y. Wang, K. Yu, S. Z. Arshad, A. Khawaji, and D. Conway. *Robust multimodal cognitive load measurement*. Springer, 2016.

[31] H. Chennamma and X. Yuan. A survey on eye-gaze tracking techniques. *arXiv preprint arXiv:1312.6410*, 2013.

[32] A. S. A. Chetwood, K.-W. Kwok, L.-W. Sun, G. P. Mylonas, J. Clark, A. Darzi, and G.-Z. Yang. Collaborative eye tracking: a potential training tool in laparoscopic surgery. *Surgical Endoscopy*, 26(7):2003–2009, Jul 2012.

[33] M. Ciotti, M. Ciccozzi, A. Terrinoni, W.-C. Jiang, C.-B. Wang, and S. Bernardini. The covid-19 pandemic. *Critical reviews in clinical laboratory sciences*, 57(6):365–388, 2020.

[34] E. Crossman and P. Goodeve. Feedback control of hand-movement and fitts' law. *The Quarterly Journal of Experimental Psychology Section A*, 35(2):251–278, 1983.

[35] S. D'Angelo and D. Gergle. An eye for design: Gaze visualizations for remote collaborative work. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 349. ACM, 2018.

[36] K. Dankwa. Infographic: Work-related stress, May 2019. https://www150.statcan.gc.ca/n1/pub/11-627-m/contest/finalists-finalistes_2-eng.htm.

[37] W. Delamare, T. Han, and P. Irani. Designing a gaze gesture guiding system. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 1–13, 2017.

[38] S. Denning. How stress is the business world's silent killer, May 2019. https://www.forbes.com/sites/stephaniedenning/2018/05/04/what-is-the-cost-of-stress-how-stress-is-the-business-worlds-silent-killer/#324db2576e06.

[39] A. K. Dey, R. Hamid, C. Beckmann, I. Li, and D. Hsu. A cappella: Programming by demonstration of context-aware applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 33–40, New York, NY, USA, 2004. ACM.

[40] V. Di Martino. *Relationship between work stress and workplace violence in the health sector*. ILO Geneva, 2003.

[41] L. L. Di Stasi, M. Marchitto, A. Antolí, and J. J. Cañas. Saccadic peak velocity as an alternative index of operator attention: A short review. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, 63(6):335–343, 2013.

[42] M.-W. O. Dictionary, May 2019. https://www.merriam-webster.com/dictionary/electroencephalography.

[43] M.-W. O. Dictionary, May 2019. https://www.merriam-webster.com/dictionary/magnetoencephalography.

[44] M.-W. O. Dictionary. fovea, February 2020. https://www.merriam-webster.com/dictionary/fovea.

[45] M.-W. O. Dictionary. open-plan. https://www.merriam-webster.com/dictionary/open-plan, 2020.

[46] M.-W. O. Dictionary. retina, February 2020. https://www.merriam-webster.com/dictionary/retina.

[47] H. Drewes and A. Schmidt. Interacting with the computer using gaze gestures. In C. Baranauskas, P. Palanque, J. Abascal, and S. D. J. Barbosa, editors,

Human-Computer Interaction – INTERACT 2007, pages 475–488, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.

[48] P. F. Drucker. The age of social transformation. *The Atlantic Monthly*, 274(5):53–80, November 1994. https://www.theatlantic.com/past/docs/issues/95dec/chilearn/drucker.htm.

[49] H. Editors. Water and air pollution. *HISTORY*, August 2018. https://www.history.com/topics/natural-disasters-and-environment/water-and-air-pollution.

[50] H. Editors. Industrial revolution. *HISTORY*, January 2019. https://www.history.com/topics/industrial-revolution/industrial-revolution.

[51] A. Esteves, Y. Shin, and I. Oakley. Comparing selection mechanisms for gaze input techniques in head-mounted displays. *International Journal of Human-Computer Studies*, 139:102414, 2020.

[52] A. Esteves, E. Velloso, A. Bulling, and H. Gellersen. Orbits: Gaze interaction for smart watches using smooth pursuit eye movements. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software &#38; Technology*, UIST '15, pages 457–466, New York, NY, USA, 2015. ACM.

[53] G. Falco, P. Pirro, E. Castellano, M. Anfossi, G. Borretta, and L. Gianotti. The relationship between stress and diabetes mellitus. *Journal of Neurology and Psychology*, 3(1):1–7, 2015.

[54] A. M. Feit, S. Williams, A. Toledo, A. Paradiso, H. Kulkarni, S. Kane, and M. R. Morris. Toward everyday gaze input: Accuracy and precision of eye tracking and implications for design. In *Proceedings of the 2017 Chi conference on human factors in computing systems*, pages 1118–1130. ACM, 2017.

[55] E. A. Felton, J. C. Williams, G. C. Vanderheiden, and R. G. Radwin. Mental workload during brain–computer interface training. *Ergonomics*, 55(5):526–537, 2012.

[56] Findlay, John M. Properties of the Saccadic Eye Movement System Introduction. In *Advances in psychology*, volume 22, pages 51–53. Elsevier, 1984.

[57] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, 47(6):381, 1954.

[58] J. Gori, O. Rioul, Y. Guiard, and M. Beaudouin-Lafon. One fitts' law, two metrics. In *IFIP Conference on Human-Computer Interaction*, pages 525–533. Springer, 2017.

[59] N. H. P. R. Group. Nasa task load index (tlx) paper and pencil package, 1986.

[60] Hafed, Ziad M and Chen, Chih-Yang and Tian, Xiaoguang. Vision, perception, and attention through the lens of microsaccades: mechanisms and implications. *Frontiers in systems neuroscience*, 9:167, 2015.

[61] J. P. Hansen, V. Rajanna, I. S. MacKenzie, and P. Bækgaard. A fitts' law study of click and dwell interaction by gaze, head and mouse with a head-mounted display. In *Proceedings of the Workshop on Communication by Gaze Interaction*, pages 1–5, 2018.

[62] S. G. Hart. Nasa-task load index (nasa-tlx); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9):904–908, 2006.

[63] S. G. Hart. Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 50, pages 904–908. Sage publications Sage CA: Los Angeles, CA, 2006.

[64] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.

[65] B. Hartmann, L. Abdulla, M. Mittal, and S. R. Klemmer. Authoring sensor-based interactions by demonstration with direct manipulation and pattern recognition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 145–154, New York, NY, USA, 2007. ACM.

[66] J. Hawley and T. Reilly. Fatigue revisited. *Journal of sports sciences*, 15(3):245, 1997.

[67] S. Health. The risks of poor nutrition, July 2019. `https://www.sahealth.sa.gov.au/wps/wcm/connect/public+content/sa+health+internet/healthy+living/is+your+health+at+risk/the+risks+of+poor+nutrition`.

[68] K. Higuch, R. Yonetani, and Y. Sato. Can eye help you?: effects of visualizing eye fixations on remote collaboration scenarios for physical tasks. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 5180–5190. ACM, 2016.

[69] J. Hild, P. Petersen, and J. Beyerer. Moving target acquisition by gaze pointing and button press using hand or foot. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*, pages 257–260, 2016.

[70] J. D. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani. Consumed endurance: a metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1063–1072, 2014.

[71] E. Hopkins. Working hours and conditions during the industrial revolution: A re-appraisal. *The Economic History Review*, 35(1):52–66, 1982.

[72] A. Huckauf and M. H. Urbina. Object selection in gaze controlled systems: What you don't look at is what you get. *ACM Trans. Appl. Percept.*, 8(2):13:1–13:14, Feb. 2011.

[73] A. Hyrskykari, H. Istance, and S. Vickers. Gaze gestures or dwell-based interaction? In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '12, pages 229–232, New York, NY, USA, 2012. ACM.

[74] Y. Ishii, H. Ogata, H. Takano, H. Ohnishi, T. Mukai, and T. Yagi. Study on mental stress using near-infrared spectroscopy, electroencephalography, and peripheral arterial tonometry. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4992–4995. IEEE, 2008.

[75] T. Isomoto, T. Ando, B. Shizuki, and S. Takahashi. Dwell time reduction technique using fitts' law for gaze-based target acquisition. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, pages 1–7, 2018.

[76] H. Istance, R. Bates, A. Hyrskykari, and S. Vickers. Snap clutch, a moded approach to solving the midas touch problem. In *Proceedings of the 2008 symposium on Eye tracking research & applications*, pages 221–228, 2008.

[77] H. Istance, A. Hyrskykari, L. Immonen, S. Mansikkamaa, and S. Vickers. Designing gaze gestures for gaming: An investigation of performance. In *Proceedings of the 2010 Symposium on Eye-Tracking Research &#38; Applications*, ETRA '10, pages 323–330, New York, NY, USA, 2010. ACM.

[78] I. IWA. ISO, image safety: Reducing the incidence of undesirable biomedical effects caused by visual image sequence, 2005.

[79] R. J. Jacob. What you look at is what you get: eye movement-based interaction techniques. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 11–18. ACM, 1990.

[80] R. J. Jacob. Eye movement-based human-computer interaction techniques: Toward non-command interfaces. *Advances in human-computer interaction*, 4:151–190, 1993.

[81] R. J. Jacob. Eye tracking in advanced interface design. *Virtual environments and advanced interface design*, pages 258–288, 1995.

[82] R. J. Jacob and K. S. Karn. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In *The mind's eye*, pages 573–605. Elsevier, 2003.

[83] R. J. K. Jacob. What you look at is what you get: Eye movement-based interaction techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '90, pages 11–18, New York, NY, USA, 1990. ACM.

[84] JASP Team. JASP (Version 0.12.2)[Computer software], 2020.

[85] J. B. Jentz. Eight-hour movement, May 2019. `http://www.encyclopedia.chicagohistory.org/pages/417.html`.

[86] J. Jimenez, D. Gutierrez, and P. Latorre. Gaze-based interaction for virtual environments. *J. UCS*, 14(19):3085–3098, 2008.

[87] S. W. Keele and M. I. Posner. Processing of visual feedback in rapid movements. *Journal of experimental psychology*, 77(1):155, 1968.

[88] M. Khamis, A. Hoesl, A. Klimczak, M. Reiss, F. Alt, and A. Bulling. Eyescout: Active eye tracking for position and movement independent gaze interaction with large public displays. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pages 155–166. ACM, 2017.

[89] H. Kim, S. Kwon, J. Heo, H. Lee, and M. K. Chung. The effect of touch-key size on the usability of in-vehicle information systems and driving safety during simulated driving. *Applied ergonomics*, 45(3):379–388, 2014.

[90] M. Kivimäki, M. Virtanen, M. Elovainio, A. Kouvonen, A. Väänänen, and J. Vahtera. Work stress in the etiology of coronary heart disease–a meta-analysis., 2006.

[91] H. Kolb. Gross anatomy of the eye. In *Webvision: The Organization of the Retina and Visual System [Internet]*. University of Utah Health Sciences Center, 2007.

[92] O. V. Komogortsev, D. V. Gobert, S. Jayarathna, D. H. Koh, and S. M. Gowda. Standardization of automated analyses of oculomotor fixation and saccadic behaviors. *IEEE Transactions on Biomedical Engineering*, 57(11):2635–2645, 2010.

[93] M. Kumar, A. Paepcke, and T. Winograd. Eyepoint: practical pointing and selection using gaze and keyboard. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 421–430, 2007.

[94] J. Kuze and K. Ukai. Subjective evaluation of visual fatigue caused by motion images. *Displays*, 29(2):159–166, 2008.

[95] M. Kytö, B. Ens, T. Piumsomboon, G. A. Lee, and M. Billinghurst. Pinpointing: Precise head-and eye-based target selection for augmented reality.

In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2018.

[96] M. F. Land and S. Furneaux. The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 352(1358):1231–1239, 1997.

[97] S. Lanthier, E. Risko, D. Smilek, and A. Kingstone. Measuring the separate effects of practice and fatigue on eye movements during visual search. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 35, 2013.

[98] J. Li, M. N. Jarczok, A. Loerbroks, I. Schöllgen, J. Siegrist, J. A. Bosch, M. G. Wilson, D. Mauss, and J. E. Fischer. Work stress is associated with diabetes and prediabetes: cross-sectional results from the miph industrial cohort studies. *International journal of behavioral medicine*, 20(4):495–503, 2013.

[99] S. G. Lisberger, E. Morris, and L. Tychsen. Visual motion processing and sensory-motor integration for smooth pursuit eye movements. *Annual review of neuroscience*, 10(1):97–129, 1987.

[100] G. Lissak. Adverse physiological and psychological effects of screen time on children and adolescents: Literature review and case study. *Environmental Research*, 164:149–157, 2018.

[101] D. J. Mack and U. J. Ilg. The effects of video game play on the characteristics of saccadic eye movements. *Vision Research*, 102:26 – 32, 2014.

[102] I. S. MacKenzie. A note on the information-theoretic basis for fitts' law. *Journal of motor behavior*, 21(3):323–330, 1989.

[103] I. S. MacKenzie. Evaluating eye tracking systems for computer input. In *Gaze interaction and applications of eye tracking: Advances in assistive technologies*, pages 205–225. IGI Global, 2012.

[104] I. S. MacKenzie and W. Buxton. Extending fitts' law to two-dimensional tasks. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 219–226, 1992.

[105] P. Majaranta, U.-K. Ahola, and O. Špakov. Fast gaze typing with an adjustable dwell time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 357–360. ACM, 2009.

[106] P. Majaranta and A. Bulling. Eye tracking and eye-based human–computer interaction. In *Advances in physiological computing*, pages 39–65. Springer, 2014.

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

[108] E. Marcello, F. Gardoni, and M. Di Luca. Alzheimer's disease and modern lifestyle: what is the role of stress? *Journal of neurochemistry*, 134(5):795–798, 2015.

[109] S. M. Marcora, W. Staiano, and V. Manning. Mental fatigue impairs physical performance in humans. *Journal of applied physiology*, 106(3):857–864, 2009.

[110] B. Martinez and M. F. Valstar. Advances, challenges, and opportunities in automatic facial expression recognition. In *Advances in face detection and facial image analysis*, pages 63–100. Springer, 2016.

[111] S. Mayer, G. Laput, and C. Harrison. Enhancing mobile voice assistants with worldgaze. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–10, New York, NY, USA, 2020. Association for Computing Machinery.

[112] Y. K. Meena, H. Cecotti, K. Wong-Lin, and G. Prasad. A multimodal interface to resolve the midas-touch problem in gaze controlled wheelchair. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 905–908. IEEE, July 2017.

[113] T. Megaw and T. Sen. Visual fatigue and saccadic eye movement parameters. In *Proceedings of the Human Factors Society Annual Meeting*, volume 27, pages 728–732. Sage Publications Sage CA: Los Angeles, CA, 1983.

[114] M. MELCHIOR, A. CASPI, B. J. MILNE, A. DANESE, R. POULTON, and T. E. MOFFITT. Work stress precipitates depression and anxiety in young, working women and men. *Psychological Medicine*, 37(8):1119–1129, 2007.

[115] Y. Mineshita, H.-K. Kim, H. Chijiki, T. Nanba, T. Shinto, S. Furuhashi, S. Oneda, M. Kuwahara, A. Suwama, and S. Shibata. Screen time duration and timing: effects on obesity, physical activity, dry eyes, and learning ability in elementary school children. *BMC public health*, 21(1):1–11, 2021.

[116] D. Miniotas, O. Špakov, I. Tugoy, and I. S. MacKenzie. Speech-augmented eye gaze interaction with small closely spaced targets. In *Proceedings of the 2006 symposium on Eye tracking research & applications*, pages 67–72, 2006.

[117] L. Mowatt, C. Gordon, A. B. R. Santosh, and T. Jones. Computer vision syndrome and ergonomic practices among undergraduate university students. *International journal of clinical practice*, 72(1):e13035, 2018.

[118] T. W. C. MSW and E. M. Higgins. Workplace stress. *Journal of Workplace Behavioral Health*, 21(2):89–97, 2006.

[119] S. Munshi, A. Varghese, and S. Dhar-Munshi. Computer vision syndrome—a common cause of unexplained visual symptoms in the modern era. *International Journal of Clinical Practice*, 71(7):e12962, 2017.

[120] C. Myers, L. Rabiner, and A. Rosenberg. Performance tradeoffs in dynamic time warping algorithms for isolated word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(6):623–635, December 1980.

[121] T. E. of Encyclopaedia Britannica. Morse code, March 2018. [Online; accessed September 14, 2018].

[122] T. E. of Encyclopaedia Britannica. Hours of labour. Encyclopædia Britannica, inc., May 2019. `https://www.britannica.com/topic/hours-of-labour`.

[123] T. E. of Encyclopaedia Britannica. Industrial revolution. Encyclopædia Britannica, inc., May 2019. `https://www.britannica.com/event/Industrial-Revolution`.

[124] F. D. of Idioms. nine-to-five job. (n.d.), May 2019. `https://idioms.thefreedictionary.com/nine-to-five+job`.

[125] T. A. I. of Stress. Workplace stress, May 2019. `https://www.stress.org/workplace-stress`.

[126] Oracle. Audioformat (java platform se 7), June 2020. [Online; accessed March 30, 2021].

[127] W. H. Organization. Occupational health, May 2019. `https://www.who.int/occupational_health/topics/stressatwp/en/`.

[128] A. Papoutsaki, P. Sangkloy, J. Laskey, N. Daskalova, J. Huang, and J. Hays. Webgazer: Scalable webcam eye tracking using user interactions. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 3839–3845. AAAI, 2016.

[129] M. Parisay, C. Poullis, and M. Kersten-Oertel. Felix: Fixation-based eye fatigue load index a multi-factor measure for gaze-based interactions. In *2020 13th International Conference on Human System Interaction (HSI)*, pages 74–81, 2020. doi:`10.1109/HSI49210.2020.9142677`.

[130] M. Parisay, C. Poullis, and M. Kersten-Oertel. Felix: Fixation-based eye fatigue load index a multi-factor measure for gaze-based interactions. In *2020 13th International Conference on Human System Interaction (HSI)*, pages 74–81, 2020.

[131] M. Parisay, C. Poullis, and M. Kersten-Oertel. Eyetap: Introducing a multimodal gaze-based technique using voice inputs with a comparative analysis

of selection techniques. *International Journal of Human-Computer Studies*, 154:102676, 2021.

[132] M. Parisay, C. Poullis, and M. Kersten-Oertel. Idea: Index of difficulty for eye tracking applications - an analysis model for target selection tasks. In *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 1: HUCAPP,*, pages 135–144. INSTICC, SciTePress, 2021.

[133] S. N. Patel and G. D. Abowd. Blui: Low-cost localized blowable user interfaces. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology*, UIST '07, pages 217–220, New York, NY, USA, 2007. ACM.

[134] M. R. S. Paul A. Freiberger and Others. Computer. Encyclopædia Britannica, inc., May 2019. `https://www.britannica.com/technology/computer`.

[135] M. Perry. The life & work of charles dickens, May 2019. `https://www.charlesdickensinfo.com/`.

[136] M. Perry. Oliver twist, May 2019. `https://www.charlesdickensinfo.com/novels/oliver-twist/`.

[137] N. Peter. Practitioner, 1979.

[138] K. Pfeuffer, J. Alexander, and H. Gellersen. Partially-indirect bimanual input with gaze, pen, and touch for pan, zoom, and ink interaction. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 2845–2856, New York, NY, USA, 2016. ACM.

[139] K. Pfeuffer and H. Gellersen. Gaze and touch interaction on tablets. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, pages 301–311, New York, NY, USA, 2016. ACM.

[140] K. Pfeuffer, B. Mayer, D. Mardanbegi, and H. Gellersen. Gaze+ pinch interaction in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, pages 99–108, 2017.

[141] K. Pfeuffer, L. Mecke, S. Delgado Rodriguez, M. Hassib, H. Maier, and F. Alt. Empirical evaluation of gaze-enhanced menus in virtual reality. In *26th ACM Symposium on Virtual Reality Software and Technology*, pages 1–11, 2020.

[142] J. Pi and B. E. Shi. Probabilistic adjustment of dwell time for eye typing. In *2017 10th International Conference on Human System Interactions (HSI)*, pages 251–257. IEEE, July 2017.

[143] pierre rouanet. pierre-rouanet/dtw, 2020. Accessed on 16.Oct.2020.

[144] V. Rajanna and T. Hammond. A gaze-assisted multimodal approach to rich and accessible human-computer interaction. *CoRR*, abs/1803.04713, 2018.

[145] N. Ramanauskas. Calibration of video-oculographical eye-tracking system. *Elektronika ir Elektrotechnika*, 72(8):65–68, 2006.

[146] P. Renner and T. Pfeiffer. Attention guiding techniques using peripheral vision and eye tracking for feedback in augmented-reality-based assistance systems. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 186–194. IEEE, 2017.

[147] A. S. Richardson, J. E. Arsenault, S. C. Cates, and M. K. Muth. Perceived stress, unhealthy eating behaviors, and severe obesity in low-income women. *Nutrition journal*, 14(1):122, 2015.

[148] M. Rosenfield. Computer vision syndrome (aka digital eye strain). *Optometry in Practice*, 17(1):1–10, 2016.

[149] M. Roser. Economic growth. *Our World in Data*, 2019. `https://ourworldindata.org/economic-growth`.

[150] M. Roser. Working hours. *Our World in Data*, 2019. `https://ourworldindata.org/working-hours`.

[151] D. Rozado, J. Niu, and M. Lochner. Fast human-computer interaction by combining gaze pointing and face gestures. *ACM Transactions on Accessible Computing (TACCESS)*, 10(3):10, 2017.

[152] J. F. Ruiz-Rabelo, E. Navarro-Rodriguez, L. L. Di-Stasi, N. Diaz-Jimenez, J. Cabrera-Bermon, C. Diaz-Iglesias, M. Gomez-Alvarez, and J. Briceño-Delgado. Validation of the nasa-tlx score in ongoing assessment of mental workload during a laparoscopic learning curve in bariatric surgery. *Obesity surgery*, 25(12):2451–2456, 2015.

[153] H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1):43–49, February 1978.

[154] N. Sammaknejad, H. Pouretemad, C. Eslahchi, A. Salahirad, and A. Alinejad. Gender classification based on eye movements: A processing effect during passive face viewing. *Advances in cognitive psychology*, 13(3):232, 2017.

[155] H. Saul. Staring at computer screens all day ’changes your eyes’, scientists say. `https://www.independent.co.uk/news/science/staring-at-computer-screens-all-day-changes-your-eyes-scientists-say-9543939.html`, 2014.

[156] S. Schenk, M. Dreiser, G. Rigoll, and M. Dorr. Gazeeverywhere: Enabling gaze-only user interaction on an unmodified desktop pc in everyday scenarios. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 3034–3044, New York, NY, USA, 2017. ACM.

[157] S. Schenk, M. Dreiser, G. Rigoll, and M. Dorr. Gazeeverywhere: enabling gaze-only user interaction on an unmodified desktop pc in everyday scenarios. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 3034–3044, 2017.

[158] Schenk, Simon and Tiefenbacher, Philipp and Rigoll, Gerhard and Dorr, Michael. Spock: A smooth pursuit oculomotor control kit. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2681–2687. ACM, 2016.

[159] K. Sengupta, M. Ke, R. Menges, C. Kumar, and S. Staab. Hands-free web browsing: enriching the user experience with gaze and voice modality. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, pages 1–3, 2018.

[160] L. E. Sibert and R. J. K. Jacob. Evaluation of eye gaze interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '00, page 281–288, New York, NY, USA, 2000. Association for Computing Machinery.

[161] N. Sidorakis, G. A. Koulieris, and K. Mania. Binocular eye-tracking for the control of a 3d immersive multimedia user interface. In *2015 IEEE 1st Workshop on Everyday Virtual Reality (WEVR)*, pages 15–18. IEEE, March 2015.

[162] J. Siegrist. Chronic psychosocial stress at work and risk of depression: evidence from prospective studies. *European Archives of Psychiatry and Clinical Neuroscience*, 258(5):115, Nov 2008.

[163] G. Soleil. Workplace stress: The health epidemic of the 21st century, December 2017. HuffPost News.

[164] O. Spakov, H. Istance, K.-J. Räihä, T. Viitanen, and H. Siirtola. Eye gaze and head gaze in collaborative games. In *11th ACM Symposium on Eye Tracking Research and Applications, ETRA 2019*. ACM, 2019.

[165] O. Špakov and D. Miniotas. On-line adjustment of dwell time for target selection by gaze. In *Proceedings of the third Nordic conference on Human-computer interaction*, pages 203–206. ACM, 2004.

[166] T. M. Spruill. Chronic psychosocial stress and hypertension. *Current Hypertension Reports*, 12(1):10–16, Feb 2010.

[167] sr research. Visual angle calculator, January 2020. `https://www.sr-research.com/visual-angle-calculator/`.

[168] D. M. Stampe. Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. *Behavior Research Methods, Instruments, & Computers*, 25(2):137–142, 1993.

[169] S. Stellmach and R. Dachselt. Look & touch: gaze-supported target acquisition. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2981–2990, 2012.

[170] S. Stellmach and R. Dachselt. Still looking: Investigating seamless gaze-supported selection, positioning, and manipulation of distant targets. In *Proceedings of the sigchi conference on human factors in computing systems*, pages 285–294, 2013.

[171] W. Stewart and J. Barling. Daily work stress, mood and interpersonal job performance: A mediational model. *Work & Stress*, 10(4):336–351, 1996.

[172] V. Sundstedt. Gazing at games: An introduction to eye tracking control. *Synthesis Lectures on Computer Graphics and Animation*, 5(1):1–113, 2012.

[173] J. Sweller. Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12(2):257–285, 1988.

[174] Thomas, JG. The dynamics of small saccadic eye movements. *The Journal of physiology*, 200(1):109–127, 1969.

[175] K. Tsubota. Tear dynamics and dry eye. *Progress in Retinal and Eye Research*, 17(4):565–596, 1998.

[176] K. Ukai and P. A. Howarth. Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations. *Displays*, 29(2):106–116, 2008.

[177] Van Beers, Robert J. The sources of variability in saccadic eye movements. *Journal of Neuroscience*, 27(33):8757–8770, 2007.

[178] M. Vasiljevas, T. Gedminas, A. Ševčenko, M. Jančiukas, T. Blažauskas, and R. Damaševičius. Modelling eye fatigue in gaze spelling task. In *2016 IEEE 12th International Conference on Intelligent Computer Communication and Processing (ICCP)*, pages 95–102. IEEE, 2016.

[179] B. B. Velichkovsky, M. A. Rumyantsev, and M. A. Morozov. New solution to the midas touch problem: Identification of visual commands via extraction of focal fixations. *Procedia Computer Science*, 39:75–82, 2014.

[180] E. Velloso, M. Wirth, C. Weichel, A. Esteves, and H. Gellersen. Ambigaze: Direct control of ambient devices by gaze. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, DIS '16, pages 812–817, New York, NY, USA, 2016. ACM.

[181] R. Vertegaal. A fitts law comparison of eye tracking and manual input in the selection of visual targets. In *Proceedings of the 10th International Conference on Multimodal Interfaces*, ICMI '08, page 241–248, New York, NY, USA, 2008. Association for Computing Machinery.

[182] M. Vidal, A. Bulling, and H. Gellersen. Pursuits: Spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '13, pages 439–448, New York, NY, USA, 2013. ACM.

[183] O. Špakov and D. Miniotas. On-line adjustment of dwell time for target selection by gaze. In *Proceedings of the Third Nordic Conference on Human-Computer Interaction*, NordiCHI '04, page 203–206, New York, NY, USA, 2004. Association for Computing Machinery.

[184] M. Ward. A brief history of the 8-hour workday, which changed how americans work, May 2017. `https://www.cnbc.com/2017/05/03/how-the-8-hour-workday-changed-how-americans-work.html`.

[185] J. Wertz. Open-plan work spaces lower productivity and employee morale. `https://www.forbes.com/sites/jiawertz/2019/06/30/open-plan-work-spaces-lower-productivity-employee-morale/#62aa2a7361cd`, 2019.

[186] J. O. Wobbrock, K. Shinohara, and A. Jansen. The effects of task dimensionality, endpoint deviation, throughput calculation, and experiment design on pointing measures and models. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1639–1648. ACM, 2011.

[187] C. M. Worthman. Evolutionary biology of human stress. In *Basics in Human Evolution*, pages 441–453. Elsevier, 2015.

[188] M. Yu, Y. Lin, D. Schmidt, X. Wang, and Y. Wang. Human-robot interaction based on gaze gestures for the drone teleoperation. *Journal of Eye Movement Research*, 7(4):1–14, 2014.

[189] J. Zagermann, U. Pfeil, and H. Reiterer. Measuring cognitive load using eye tracking technology in visual computing. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, BELIV '16, pages 78–85, New York, NY, USA, 2016. ACM.

[190] D. G. Zhao, N. D. Karikov, E. V. Melnichuk, B. M. Velichkovsky, and S. L. Shishkin. Voice as a mouse click: Usability and effectiveness of simplified hands-free gaze-voice selection. *Applied Sciences*, 10(24):8791, 2020.

[191] B. Zheng, X. Jiang, G. Tien, A. Meneghetti, O. N. M. Panton, and M. S. Atkins. Workload assessment of surgeons: correlation between nasa tlx and blinks. *Surgical endoscopy*, 26(10):2746–2750, 2012.

[192] Zimmermann, Eckart and Lappe, Markus. Visual space constructed by saccade motor maps. *Frontiers in human neuroscience*, 10:225, 2016.

Appendix A

Source Codes

All developed applications for the user studies reviewed in this thesis are available on GitHub.

1. Circle Button: the source code for the Dart-based test application:
   `https://github.com/MohPar2020/CircleButton`

2. SpeechRecognitionSample: the source code to demonstrate voice recognition using Windows 10 Speech Recognition engine:
   `https://github.com/MohPar2020/SpeechRecognitionSample`

3. EyeTapStudy: the source code for the Matrix-based test application:
   `https://github.com/MohPar2020/EyeTapStudy`

4. AudioRecording: the source code to demonstrate real-time audio monitoring to detect noise in the input applied in the EyeTAP prototype:
   `https://github.com/MohPar2020/AudioRecording`