

MAPPING THE VISUAL FIELD – AN EMPIRICAL
STUDY ON THE USER EXPERIENCE BENEFITS OF
GAZE-BASED INTERACTION IN VISUAL FIELD
TESTING

Pauliina Kiviranta

University of Tampere
Faculty of Communication
Sciences
MDP in Human-Technology
Interaction
M.Sc. thesis
Supervisor: Poika Isokoski
September 2017

Visual field testing is a medical procedure used for detecting impairments in a patient's field of view. While plenty of research and different applications concerning visual field testing exist, gaze-based interaction in visual field testing is quite an understudied subject. This study's goal was to investigate, whether it is possible to build a functional visual field test application with a relatively simple pc + monitor + eye tracker setup that utilizes gaze input from the user, and if so, whether the gaze input could have potential user experience -related or other advantages over more traditional mouse click input.

To acquire a basic understanding on the human visual field and the principles of visual field testing, a literature review was conducted as the basis of this study. After that, a test software with available gaze input and mouse click input was designed and built. User tests with overall 8 participants were arranged in the university's gaze research laboratory. In the tests, the participants completed two test runs with each input method, and were asked to evaluate and compare the following user-experience related attributes, *the mental and physical load, effort, discomfort and learnability*, during and after the test. Finally, a lightweight interview where the participants were allowed to elaborate their evaluation was conducted.

While the participants did not consider either of the tested interaction techniques too difficult to use, the gaze interaction generally gained more favourable evaluations from the participants. Based on the results, gaze input shows promise, although a lot more work, for which some ideas are proposed in the discussion section of this study, is still required to make any definitive conclusions.

Keywords and terms: Visual field test, Eye tracking, User experience, Gaze input, Empirical study

Contents

1	INTRODUCTION	1
2	THE HUMAN VISUAL FIELD	4
	2.1 Human visual field	4
	2.2 Effect of age.....	5
	2.3 Blind spots	6
	2.4 Reasons to test the visual field	7
3	VISUAL FIELD TESTING.....	10
	3.1 How the visual field can be measured	10
	3.2 Fixation.....	14
	3.3 Defining stimuli as seen or not seen	15
	3.4 Qualitative and quantitative visual field testing	16
	3.5 Stimuli in visual field tests	17
	3.6 Computerized visual field testing	18
	3.7 Home applications for visual field testing	20
	3.8 Gaze-tracking and visual field testing	22
	3.9 General user experience challenges in visual field testing	24
4	RESEARCH METHOD	25
	4.1 The test setup	25
	4.2 The test software.....	27
	4.3 Mouse input test	29
	4.4 Gaze input test	30
	4.5 Test output	31
	4.6 Participants	32
	4.7 The test procedure	34
5	TEST RESULTS	37
	5.1 Test accuracy	37
	5.2 Feedback collected during the test.....	42
	5.3 Feedback collected after the test.....	50
6	DISCUSSION.....	55
	6.1 Implications and interpretation of the test results.....	55
	6.2 Suggestions for future research	58
	6.3 Suggestions for future software development	59
7	CONCLUSIONS	63
	REFERENCES	65

APPENDIX

1 INTRODUCTION

Visual field testing is an important medical procedure used for detecting ailments that have an effect to the extent or to the intactness of the area that a person is able to see. While visual field testing is a heavily studied subject, a relatively small amount of research exists on the user experience and the potential advantages that gaze-based interaction can provide in applications built for visual field testing. In this study, these matters are investigated .

The human visual field refers to the whole area that can be seen above, below and to the both sides while the eye remains fixed [Werner, 1991]. There are several reasons to test the visual field - most importantly, a non-intact visual field is a certain sign of some kind of an illness or trauma that either affects the retina or the visual pathways. Also, an EU-directive defines treshold values above which a person's visual field must be in order for them to be eligible to have a driver's licence - in unclear cases a visual field test might be necessary before a driver's licence can be granted. The analysis of visual field test results is often a useful tool for diagnosing illness, and in addition to pointing out the existence of a visual field deficit, it can also help point out the location of deficit. For example, glaucoma is a relatively common illness that directly effects the visual field in a negative manner. [Walsh, 2011, Werner, 1991]

There are numerous techniques for performing *perimetry*, which is the common term used for visual field testing. At it's simplest, a visual field test can be a simple confrontational test in which the test's administrator moves their fingers or other physical stimuli on the sides of the patient's head [Johnson et al, 2011]. These days, automated perimetry is used for more detailed analysis. Most common visual field tests are built so that the person taking the test receives visual output in the form of some kind of stimuli from the test and provides input about whether they detected the stimulus or not either verbally or with a mouse or button click. While gaze would intuitively sound like a natural interaction modality for an eye-related application, and while there are also numerous accurate and affordable eye tracking devices available these days, gaze-based interaction seems to be very underused in visual field testing.

The main goal of this study was to investigate whether gaze input would be usable or even advantageous compared to more traditional mouse interaction in a visual field test

application. The goal of this study was crystallized into the following two research questions:

- Can gaze tracking be successfully used as an input method in a visual field test application built in a simple pc + monitor + eye tracker setup?
- In terms of user experience or otherwise, does the gaze tracking -based visual field test application have advantages over a more traditional mouse-based test application?

The work for this thesis study began by conducting a literature review on the basics of the human visual field and visual field testing. After that, a test software with two different test settings, one in which test participants would interact with gaze, and one with mouse interaction, was designed and built. A series of laboratory-based user tests was conducted to evaluate and compare the user experience provided by the two interaction techniques. Eight participants were recruited to the tests. The participants subjectively evaluated user-experience related attributes including the *physical and mental load, effort, discomfort* and *learnability* during and after the tests. Also, at the end of the test sessions, a lightweight interview was conducted so that the participants could explain and elaborate their evaluation. The software's functionality was validated by checking whether the participant's physiological blind spot was detected successfully or not on each test run.

The setup that the test software was designed for consisted of a personal computer, a monitor and a commercially available gaze tracking device. The selected setup was considered relevant as the required hardware is inexpensive in comparison to medical perimetric devices, while it still makes objective measuring possible. A visual field test setup such as the one presented in this study could suit home use, such as disease monitoring, and on the other hand it could be used by professionals in situations and locations where the investment on an expensive perimetric device is not desired or even possible, e.g. in countries where the budget for healthcare is small. Several examples of softwares designed for a similar setup were reviewed for this study (see section 3.8). The designed and tested software aims for the same benefits that those softwares had, but in addition to that, the possibility of the gaze input improving the accuracy and user experience is also explored.

This thesis consist of overall seven chapters. Chapters 2 and 3 are focused on the theory of visual field testing. In chapter 2, the human visual field is explained in detail, and in chapter 3 the principles and current practices of visual field testing. Chapter 4 presents the used research method, including the test software that was designed and built for this study and the settings and the procedure of the user tests that were conducted. Chapter 5 presents the results of this study. The implications of the results, the test setting's success, and potential future research and development are discussed in chapter 6, and finally, chapter 7 sums up the conclusions.

2 THE HUMAN VISUAL FIELD

In order to design, develop and test a visual field mapping application, it is important to be aware of how the human visual field functions. This section provides a theoretical background to how the human visual field works. It also describes the key variables and their typical values in visual field testing. The basic definitions and dimensions of the visual field are presented in section 2.1. Following that, the effects of age to the visual field are discussed in section 2.2. The physiological blind spot and its dimensions are presented in section 2.3 and section 2.4 discusses visual field related issues and reasons why testing for them is important.

2.1 Human visual field

The term *visual field*, also known as *the island of vision* [Spector, 1990], refers to the entire area that can be seen above, below and to the both sides while the eye remains fixed. The visual field can be measured in degrees of the viewing angle. As seen in figure 1, the viewing angle can be illustrated as a right triangle, in which the viewed object's height or width (A) is the right angle's opposite side and the viewing distance (B) is the right angle's adjacent side. Respectively, the viewing angle (C) is the angle between the adjacent side and the hypotenuse. [Wandell, 1995; Werner, 1991]

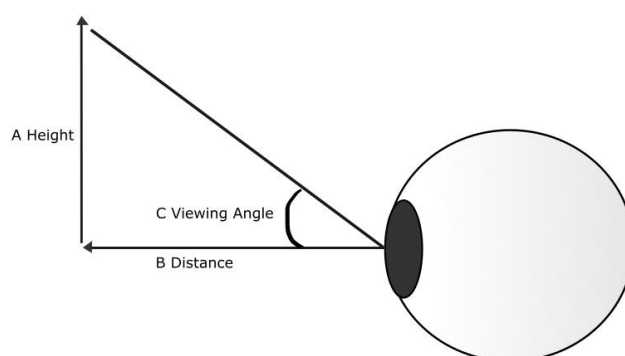


Figure 1. The viewing angle, illustrated based on [Wandell, 1995].

The *normal* human visual field cannot be defined by a set of exact values, as there is a considerable variability in the so-called normal visual field, and the term *normal* rather describes a range of values associated with the majority of the nondiseased population.

There is variance within the population, and the visual field also steadily declines with age. In short, normally, the approximated full visual field of a single eye or *the monocular field of view*, measured from the central fixation, is 60 degrees superiorly/upwards, 75 degrees inferiorly/downwards (see figure 2), 100 degrees to the temporal side and 60 degrees towards the nose (see figure 2). Thus, the visual field approximately covers 160 degrees in width and 135 degrees in height. Meanwhile, *the binocular field of view*, which refers to the area that can be seen while both eyes are open, is approximately 200 degrees wide with 120 degrees of binocular overlap. These measurements are not definitive, and for example according to Spector [1990], a monocular visual field that is ~10 degrees narrower in each direction than the aforementioned, is still considered normal. [Wandell, 1995; Werner, 1991]

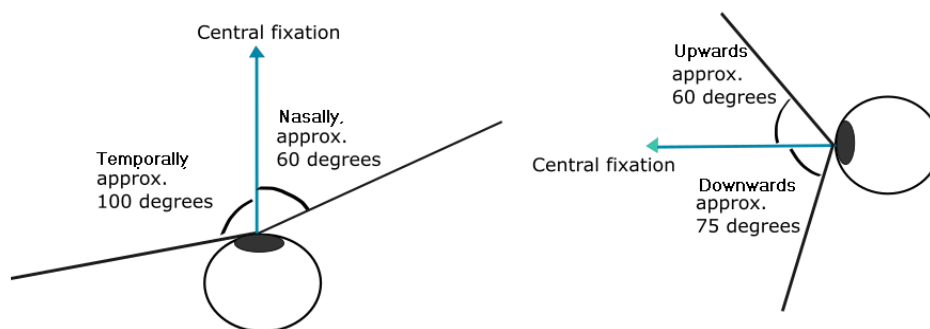


Figure 2. The monocular field of vision, illustrated based on [Werner, 1991].

While the human visual field covers a large area, it is not uniform, and the *visual acuity* varies within the island of vision. The most *peripheral vision* is only capable of discriminating movement, while the most *central vision* is clear and sharp. The visual acuity in different areas of the visual field depends on the amount and distribution of the two different types of *photoreceptor cells*, the rods or the cones, around the *retina*. The small central region of the highest visual acuity is called the *fovea*, and it has the highest count of cone receptors on the entire retina. [Spector, 1990; Wandell, 1995]

2.2 Effect of age

As mentioned, age has an effect to the normal visual visual field. As the interpretation of findings in a visual field test heavily depends on understanding the normal visual field, it is important to take the effect age into account. In general, the normal visual

field values presented previous apply best to the population of adults between 20 and 30 in age. The decline of the visual field with age has been studied in a large extent. The sensitivity of the different areas of the visual field gradually declines with age and the temporal peripheral field gets narrower after the sixth decade of life. [Patel et al., 2015; Walsh, 2011; Werner, 1991]

Patel et al. [2015] have studied the visual field in children aged 5-15 years. The writers argue that most of the normative visual field data available is based on adult subjects, and more guidance on interpretation of visual field data of children is needed. Their study shows that the visual field, tested with two different perimetric approaches, indeed does increase up to 12 years of age, after which adult values are reached and maintained. The increasing visual field patterns showed to be specific to the perimetric approach that was used, so it is important to also notice that the *normal* also depends on the measuring and analysis method used. [Patel et al., 2015]

2.3 Blind spots

Every human eye has a natural blind spot. As seen in figure 3, the blind spot is normally located 12 to 17 degrees temporally and 1,5 degrees inferiorly from fixation. Normally, the blind spot is roughly 5,5 degrees in width and 7,5 degrees in height. Any additional blind spots or depressions in the visual field, as well as enlargening of the physiological blind spot, are abnormal. [Spector, 1990; Werner, 1991]

The physiological blind spot has several mentionworthy characteristics. The blind spot is caused by the end of the optic nerve, also known as the optic disc, which is an area on the retina that contains no photoreceptors [Wandell, 1995]. The physiological blind spot does not disturb the vision when both eyes are open because the other eye's visual field covers the blind area and the visual cues received from the other eye help the brain fill in the missing visual information. Thus, the blind spot of one eye can only be detected with the other eye occluded. With the other eye occluded, an object of a suitable size can fully disappear into the blind spot. As an object disappears into the blind spot, instead of actually appearing as a spot without sight, the visual system fills the blind spot in with the colour and the texture of the background to create a full and coherent image. For example, if a black dot on a white background was used as a stimulus, the area covering the black dot would also appear as the white background when the dot disappears into the blind spot. [Ramachandran, 1992]

In visual field testing, the physiological blind spot has several uses. In clinical tests, the physiological blind spot is often tested to make the patient aware of possible blind spots in their vision, as well as to help them accommodate to the test procedure. The blind spot can be used to validate the fixation by putting the target into the blind spot every now and then. In clinical testing, it is also important to test the physiological blind spot since some visual defects can also affect its size and the area around it [Walsh, 2011; Werner, 1991]

Scotomas are abnormal areas of lowered or missing acuity in the visual field, surrounded by normal sight. Scotomas can be caused by a variety of diseases or lesions affecting the visual pathways. Areas of total visual loss are called *absolute scotomas*, while areas of lowered visual acuity are called *relative scotomas*. [Martin, 2015]

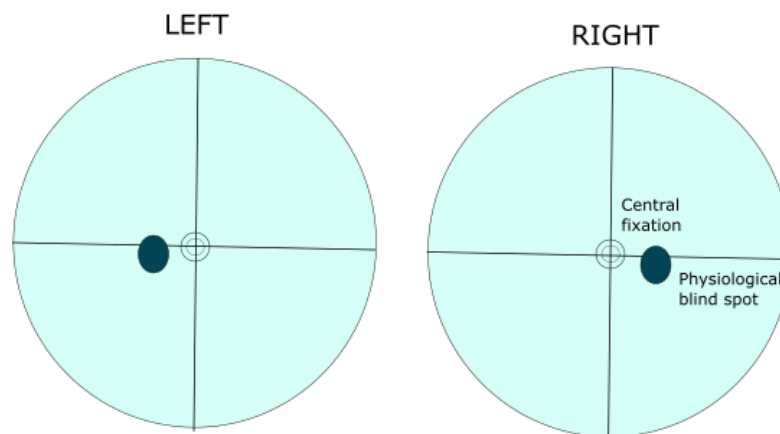


Figure 3. The location of the physiological blind spot in relation to the central fixation. Any other blind spots or depressions in the visual field are abnormal. Illustrated based on [Broadway, 2012]

2.4 Reasons to test the visual field

There are several reasons to test a person's visual field. Firstly, visual field testing can help detect a great variety of illnesses. If a person's visual field is not intact, it is certain that there is something wrong with their retina or the *visual pathway*, which in short is the bundle of nerves that extends from the eye to the visual center of the brain. By testing the visual field, it is possible to detect lesions in the retina or the choroid, defects in the optic nerve fiber bundle, diseases affecting the optic nerve or the optic chiasm, and defects in the retrochiasm visual pathways, which includes the optic tract, the

lateral geniculate body and the visual cortex. Illnesses affecting the visual field can cause distinguishable local scotomas, general depression in the sensitivity of the visual field or a combination of the two. The enlarging of the physiological blind spot and the overall constriction of the visual field are also possible. The analysis of visual field test results, either automated or performed by a professional, can often be a useful diagnostic tool. In addition to providing information on the existence of a lesion, analysing the visual field test results can also provide information on the location of the lesion and its nature. Naturally, in addition to detecting defects, visual field testing can also be used to monitor the progress of several affecting diseases. [Walsh, 2011; Werner, 1991]

Glaucoma is a good example of the various illnesses that effect the visual field. Glaucomas are a group of progressive eye diseases that are characterized by the death of retinal ganglion cells, also known as optic neuropathy, and intraocular pressure. Glaucoma affects more than 70 million people all over the world and it is the world's leading cause of incurable blindness. Glaucoma can be a primary disease or it can be secondary caused by various other defects, including trauma, certain medications or tumors. Detecting glaucoma in its early stages would be important since slowing the disease's progression is the main goal in treating the disease. [Weinreb et al., 2014] According to Walsh [2011], detecting glaucoma is more complex than just identifying the increased intraocular pressure and it is also important to measure the functional abnormalities that are caused by the disease. In glaucoma, both, a general depression of the visual field and local scotomas, as well as often the enlargement of the physiological blind spot are present in the visual field tests. [Walsh, 2011; Werner, 1991]

In addition to diagnosing illness, a visual field test might be necessary to make sure that a person is fit to drive a car in the traffic. An adequate visual field is a requirement for safe driving, and even intuitively, it can be concluded that a person with a serious tunnel vision or scotomas affecting the visual field is not fit to drive. There are certain legal minimum requirements for a person's visual field that they need to meet in order for them to achieve or renew their driving licence. The requirements vary somewhat from country to country. In the EU area, as is stated in the *Directive 2006/126/EC of the European Parliament and of the Council of 20 December 2006 on driving licences*, apart from exceptional cases that are justified by a favourable medical opinion and a positive practical test, a person must have a normal binocular vision and a horizontal field of vision of at least 120 degrees. In case the person has a progressive eye disease, regular examinations are required to validate their fitness to drive. Meanwhile, Keeffe

and Charlton argue that while legal guidelines for acceptable visual field values do exist, it is not as simple as that to define what is the actual acceptable visual field minimum for safe driving. As restricting a person from having a driving licence can have a large impact on their quality of life, the writers argue that much more future work to define the actual effect of visual impairment and visual field restrictions to driving safety should be performed to find evidence-based guidelines that clinicians can base their recommendations on. Overall, visual field testing is important in defining whether a person is fit to drive or not. [Keeffe and Charlton, 2007]

3 VISUAL FIELD TESTING

The test settings and the selected procedure for visual field testing vary heavily depending of the method in use. While some measurement techniques have been around for a very long time, new tests and improvement features to the old ones are constantly developed. As beign aware of existing solutions and their applications is important in order to design a novel visual field testing system, the different visual field testing methods are reviewed in detail in this chapter. Section 3.1. serves as a brief introduction to what visual field testing is about. Section 3.2 discusses the fixation, which has an important role in many visual field testing methods. Section 3.3. is about the user input – the section presents several examples on the interaction techniques used in different tests that allow the test participant to signal whether a stimulus was seen or not. Section 3.4. briefly discusses the differences between qualitative and quantitative visual field testing. Different stimuli and their applications are presented in section 3.5. Section 3.6. presents examples of different computerized visual field tests, from the modern “gold standard” testing device called the Humphrey Field Analyser to some more experimental applications. In section 3.7, several new experimental application for home-administered visual field testing are reviewed. The current use of gaze-tracking in visual field testing is reviewed in section 3.8, and finally, some of the most common user experience challenges related to visual field testing are discussed in section 3.9.

3.1 How the visual field can be measured

Visual field testing has a long history - first accounts of the mapping of peripheral vision took place over 2000 years ago, and quantitative measurements of the sensitivity of the visual field are almost 200 years old. The most commonly used method of visual field testing – detecting small targets against a uniform background in varying location – has stayed the same for over 150 years. While many modern computerized solutions for visual field testing exist, older manual testing techniques still hold their place as well. With many patient groups, the human element of the test situation is still very important and can affect the reliability of the results greatly. For example, complex technology can often be overwhelming for neuro-ophthalmic patients, who often have many other deficits too than just those that concern their visual field. At it’s simplest, a visual field test can be performed for example with a *confrontational visual field test*,

which can be performed by the patient counting how many fingers the tester currently has visible or by assessing the detection and vividness of coloured items that the tester shows to them. [Johnson et al, 2011; Walsh, 2011]



Figure 4. Visual field testing at it's simplest – a confrontational visual field test. © Heiko Philippin. [Broadway, 2012]

The visual field can basically be mapped with two different approaches – the *campimetry* or the *perimetry*. While there are semantic differences, the two terms are often used interchangeably as a synonym to a visual field test. *Campimetry* refers to the testing of the central parts of the visual field on a flat surface. The tool used for campimetry is called a *campimeter*. Many different applications of a campimeter are used for visual field mapping, and a simple *tangent screen* is a good example of such an application. The modern tangent screen was introduced by Jannik Bjerrum in 1889. The original Bjerrum Tangent Screen was 2 meters high and wide and covered 30 degrees of the visual field. The wall-mounted screen included a constant dark background. By displaying white pins, varying in size and attached the end of a stick, to the test subject, Bjerrum was able to map their visual fields in good quantitative detail. [Johnson et al, 2011; Werner, 1991; Farlex Partner Medical Dictionary 1]



Figure 5. The Bjerrum tangent screen for manual visual field testing. © Sussex Vision International. [Broadway, 2012]

Meanwhile, *perimetry* is not only limited to the area of the central vision but refers to the measurement of the limits of the visual field. Unlike campimetry, perimetry is usually performed on a curved surface to cover as large area as is needed. The tool used for perimetry is called a *perimeter* and as for campimetry, numerous different applications exist. *The Goldmann perimeter*, introduced in the 1940s by the Swiss scientist Hans Goldmann, is one of the most used perimetry devices through all times, with which quantitative perimetric tests can be performed. The Goldmann perimeter is a manually operated perimetric device in which the patient gazes into a hemispherical bowl that covers their entire field of vision. By using the operation system on the backside of the device, the examiner displays stimuli of varying brightness onto the bowl's surface. The Goldmann perimeter allows the examiner to measure the patient's visual field in great detail. Meanwhile, it's greatest fallback is that everything from fixation monitoring, stimulus detection signalling and stimulus placement needs to be done manually and thus the test can be prone to human error depending on how skilled the examiner is in running the test. Also, with large data sets, the test can be time-consuming and fatiguing. The Goldmann perimetry also requires large, specific equipment, so it cannot be used for portable testing. [Johnson et al, 2011; Werner, 1991; Farlex Partner Medical Dictionary 2]



Figure 6. The Goldmann perimeter in use. The patient gazes into the hemispherical bowl while the examiner displays stimuli on it's surface. [Fortenbaugh et al., 2012]

There are two major types of visual field testing (applies to campimetry and perimetry alike). In *static visual field testing*, the patient is required to detect stationary targets, while in *kinetic visual field testing*, the targets move. Both test types have their strengths and weaknesses, as moving targets are perceived differently by the human visual system than the static targets. Kinetic tests are generally faster and allow the plotting of a larger area in less time than static tests. Meanwhile, static tests are more accurate and superior in testing for scotomas and visual field defects caused by glaucoma. Both types of tests can, for example, be performed with the previously presented Goldmann perimeter, although the movement of the kinetic targets must in that case be produced manually.

[Broadway, 2012; Werner, 1991]

In most visual field tests, each of the test subject's monocular fields of view are tested separately from each of their eyes. A few tests, e.g. *the Esterman's test* which is available for *the Humphrey field analyser* (described in section 3.6.), are designed for testing the whole binocular field of view [Esterman, 1982].

3.2 Fixation

In most of the commonly used visual field test's, e.g. in the previously mentioned manual tangent screen test, the Goldmann perimetry, or in tests administered with the Humphrey field analyser (described further in section 3.6) the test subject is required to keep their gaze fixated to a predefined *fixation point* during the test. To assure accuracy, holding the fixation is very important and it needs to be monitored by some means. The EyeWiki of the American Academy of Ophthalmology lists three methods that are mainly used to monitor correct fixation. Firstly, the physiological blind spot can be utilized to keep check of the fixation. When using this method, the physiological blind spot is mapped in a preliminary test, after which the results are used in the actual test by placing the stimulus occasionally into the blind spot area to verify the fixation. Secondly, an eye-tracking device can be used to monitor the fixation. Thirdly, the fixation can be monitored manually by the person running the test. [EyeWiki, 2017] As holding the fixation can be laborious (see section 3.7 for further discussion about the common issues related to visual field testing), some *moving eye campimetry tests* have also been developed. *The Damato campimeter* is a portable device which can be used to examine the central 30 degrees of vision. The device has a hole in it's center, inside which a selected stimulus is displayed, and the area around the central whole has 60 numbered and marked points. During the Damato campimetry test, the test participant is asked to sequentially look from number to number and report, in case the stimulus displayed in the center disappears. The physiological blind spot is used as a control to verify the correct performance. For example, Rowe et al (2010) have tested the Damato device against the commonly used Humphrey Field Analyser to find out whether the Damato test can reliably be used as a screening method and whether it is easy to understand for the test participants. The tests suggested that the device could be a useful tool in test situations that require quick testing and portability, as only 9,5% of the test participants were not able to understand or take the test, and the test results of the ones that could take the test were adequately accurate in comparison to the Humphrey test. [Rowe et al., 2010]

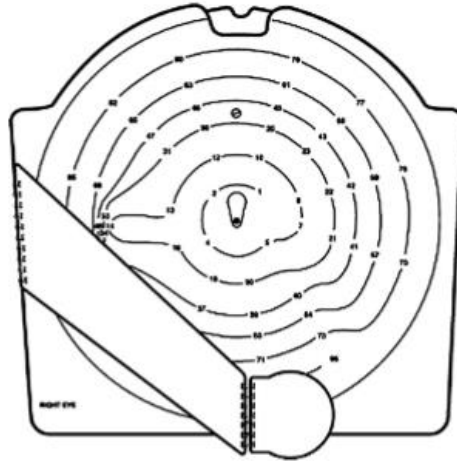


Figure 7. An illustration of the Damato 60-point campimeter. The stimuli are presented in the hole in the device's center while the patient moves their gaze from one numbered point to another. [Rowe et al., 2010]

3.3 Defining stimuli as seen or not seen

As mentioned in previous sections, to chart a visual field, it is necessary to somehow define whether a stimulus was seen in a particular area of the visual field or not. In the simplest traditional test settings, the participant might give vocal feedback to the person running the test. In modern computerized test settings, the feedback is often given by pressing a button or by clicking a mouse when a stimulus is detected. While commonly used, these traditional measurement techniques rely on subjective feedback from the participant, which naturally provides some challenges and can compromise the test's reliability. To tackle the problem, various objective measurement tools have been developed. For example in *the pupil campimetry*, stimuli are displayed to the patient's eye and the reactions of their pupil are recorded with an infrared camera. In this measurement technique, a reduced or completely missing pupil response is considered as a sign of a visual field defect. While new objective methods are constantly developed, and they certainly have their benefits, there is still a solid need for visual field measurement techniques that are based on the patient's subjective feedback. While many types of visual field deficits can be reliably measured with the pupil campimetry technique, some types of visual field losses do not result in a reduced pupil response, and thus do not show on the pupil campimetry test. [Schmid et al., 2005]

3.4 Qualitative and quantitative visual field testing

The visual field can be tested with varying stimuli and test settings, depending on the desired data and results and the available resources. Qualitative visual field tests do not provide specific threshold values or other qualitative measurements. Instead they provide crude binary information on whether a stimulus was seen in a specific location or not. Qualitative visual field tests can be performed for example with a Goldmann style visual field test in which the intensity of the stimulus is set higher than any expected detection threshold.

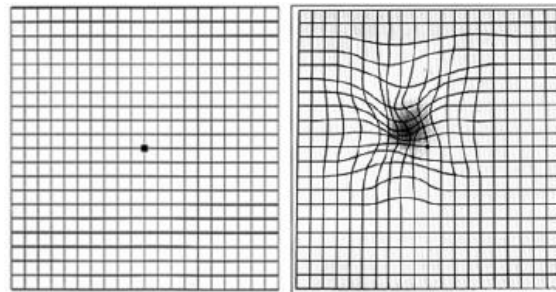


Figure 8. The Amsler grid as seen by a person with normal vision (left) and by a person with visual field defects (right). https://www.researchgate.net/figure/262380348_fig5_Fig-5-Amsler-Grid-normal-vision-left-AMD-patient-sees-right

A qualitative visual field test can also be performed by using an Amsler grid – a grid consisting of vertical and horizontal lines on an uniform background. If the person being tested reports that some parts of the grid are missing or seem distorted, it can be defined that there is something wrong with their visual field. [Johnson et al., 2011]

While qualitative measurements can give hints of underlying visual field issues, quantitative methods are needed for more detailed analysis. Quantitative data can be acquired, for example, by controlling the brightness of the stimuli presented. For example, in the classical Goldmann perimetry, the visual field is mapped by starting from stimuli that are quite dim and then moving on to brighter and brighter ones until the test subject signals that they saw the stimulus. The tester seeks for the brightness' threshold values on different areas of the visual field. The values acquired from the test can then be used to draw a map of *isopters* that present the test subject's visual acuity in different areas of their visual field. Generally, a healthy visual field has the highest

visual acuity in the most central isopter, and the visual acuity gradually lowers towards the periphery. [Werner, 1991]

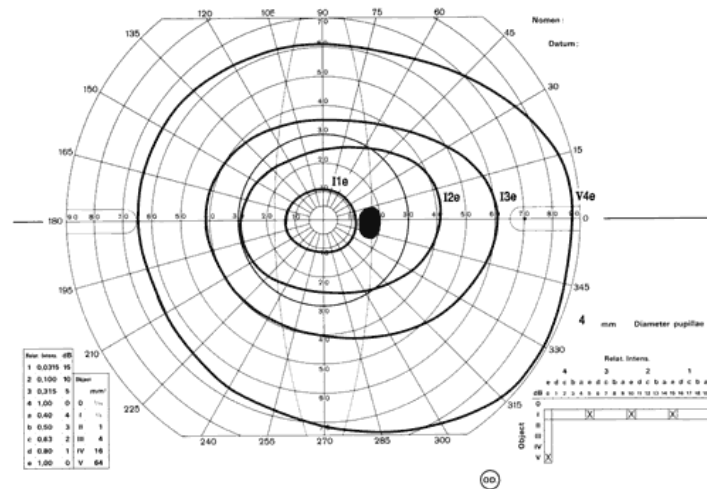


Figure 9. A visual field visualized as isopters. The pupil is depicted in the center, surrounded by isopters marked with black rings. The black spot depicts the physiological blind spot.
<http://www.oculist.net/downaton502/prof/ebook/duanes/pages/v3/ch049/006f.html>

3.5 Stimuli in visual field tests

Currently, white-on-white perimetry, in which a white, luminated target is displayed on a white background in order to find the detection threshold value, is the most common type of perimetry that is used. In the most commonly used modern devices, five different sizes of stimuli, ranging from 0,25 mm² to 64 mm² are used, depending on the patient's visual acuity. [Eyewiki, 2017]

The type of perimetry performed sets some requirements to the target. When performing kinetic perimetry, it is important that the target does not move too fast, or the test results could turn out distorted. In static perimetry, it is important that the target duration is not too short or long to avoid distortion in the results. For example according to Werner [1991], a target duration between 0,1 and 0,25 seconds is suitable. [Werner, 1991]

Modern automated perimetry allows the presentation of even more varying and complex visual stimuli. For example McKendrick [2005] argues, that the traditional white-on-white is really not the ideal stimulus for all visual function specific measurements, and presents various optical visual field testing techniques with different stimuli. For example, in *Short Wavelength Automated Perimetry* (SWAP), blue targets are displayed

on a bright yellow background, which, for example, allows the detection of glaucoma-related visual field loss earlier than the white-on-white perimetry. In *Frequency Doubling Technology Perimetry* (FDT) and in *Flicker Perimetry*, the targets displayed on surface flicker on an increasing rate, and the patient is asked to report when they still perceive the target as flickering in order to find proper threshold values. In *High Pass Resolution Perimetry*, or otherwise known as ring perimetry, ring-shaped targets (see figure 10) are displayed in different locations of the visual field in order to find out threshold values for the smallest rings the patient is able to detect in the location.



Figure 10. ring-shaped stimuli used in High-pass Resolution Perimetry.
<http://www.oft.gu.se/webdiagnos/Perimetry/perimetry.html>

As the targets in traditional white-on-white perimetry are relatively large, *Rarebit Perimetry* can be helpful in test cases that require the detection of minor visual field deficits. In rarebit perimetry, test targets that present the minimum amount of information for the visual system are used. For a very brief duration of 200 milliseconds, the test presents two tiny, bright stimuli on a black surface and the patient is required to indicate, whether they saw two, one or zero stimuli at each round. Rarebit perimetry cannot provide threshold values, but only a hit-and-miss rate for each presented pair of stimuli. [McKendrick, 2005]

3.6 Computerized visual field testing

Since the early 1970's, numerous computerized solution for visual field testing have been developed. These days, probably the most commonly used clinical visual field test device is the Humphrey field analyser (see figure 11). Other similar devices in common clinical use include the Henson, Tubingen and Octopus perimeters. The Humphrey field

analyser consist of a similar hemispherical bowl that is used in the Goldmann perimeter. The stimuli are projected on the bowl's surface and the patient signals detection by pressing a button. A great variety of screening, threshold testing and analysis software is available for the device. For example, with the Humphrey field analyser, the visual field can be screened with a programme consisting of 120 targets with a luminance higher than the expected detection threshold. The targets are displayed and missed targets present areas that require further analysis. In this programme, a positive test is defined by 17 missed target altogether, or by 8 missed targets in any particular quadrant of the visual field. [Sundaram, 2009; Werner, 1991]



Figure 11. The Humphrey field analyser.

<https://www.zeiss.com/meditec/int/products/ophthalmology-optometry/glaucoma/diagnostics/perimetry/humphrey-field-analyzer-hfa3.html>

The results acquired with the humphrey field analyser are all in all quite reliable, as the device and the test softwares come with built-in mechanisms that help tackle the most common issues. Fixation is monitored by placing the stimulus in the blind spot area from time to time, false positives can be detected from the result visualization and false negatives are confirmed with retesting. While the Humphrey field analyser is a robust tool, it also has it's challenges. First of all, all of the automated tests take some time from the patient's side to learn. According to Sundaram [2009], three or four test repeats might be needed until actually reliable results are obtained. The device is also very expensive, requires a lot of space and cannot be used in situations where a portable visual field testing device would come in handy. [Sundaram, 2009]

3.7 Home applications for visual field testing

All of the visual field testing techniques reviewed thus far were solutions for clinical settings. Most of these tests need to be administered by a professional and require large and expensive special equipment, which mostly restricts visual field testing to special settings like optometrist/ophthalmologist clinics. While the most sophisticated visual field testing needs to be left for the experts, desktop computers and mobile devices could easily be used as cheap alternative platforms for preliminary screening test softwares for general practitioners, nurses, and even for self-administered home tests. Already, many easily accessible and easily administrable desktop and mobile visual field test applications have been developed. Several interesting examples of such tests are reviewed in this section.

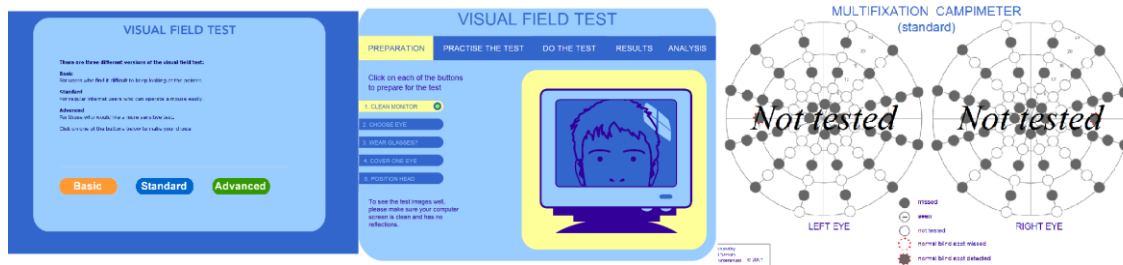


Figure 12. Screenshots of Online Campimetry. www.testvision.org .

Online Campimetry (www.testvision.org) is an online visual field test based on Damato campimetry (see section 3.2 for further details) for measuring the central vision. The only requirements for the test are an internet connection and a computer with a monitor and updated Adobe Flash. Three preset test modes designed for regular internet users, people with difficulties at looking at the mouse pointer, and for people who wish to take a more detailed test are available. Before the test, the user is asked to clear their monitor, to cover one eye, and to position themselves properly. The test can also be taken with glasses on. During the standard test, a black dot stimulus briefly flashes somewhere on the screen. To signal detection, the user should move their mouse and their gaze over to the area where the black dot appeared. When the mouse cursor is in the right area, a green smiley face appears and the user is required to click on it, after which a new stimulus will appear. Blind spot detection is used to validate the results,

which are displayed in a visualization after the test is done (see figure 12).

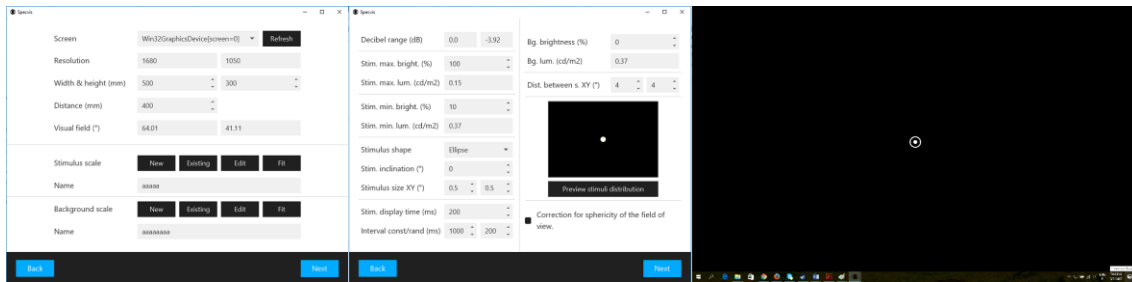


Figure 13. Screenshots of Specvis. <http://www.specvis.pl/>.

Specvis (<http://www.specvis.pl>), originating from Poland, is another free, open-source perimetry software available online. The software is still under development and currently it is available for desktop computers with Java runtime environment. In the upcoming versions, the system is planned to be made available for virtual reality goggles, and it will also accept gaze-tracking input later in the future. Like Online Campimetry, *Specvis* allows the mapping of the central visual field. While the Online Campimetry had three preset tests available, *Specvis* allows the user to configure their own test in great detail (see figure 13), from stimuli luminance to stimuli displaying speed. The test itself is quite straightforward – as in many previously presented visual field tests, the user is required to fixate to a fixation mark in the center of the screen and stimuli are displayed one after another as the user signals detection. As the application works on desktop, personalized tests and personal results can be saved and viewed later.

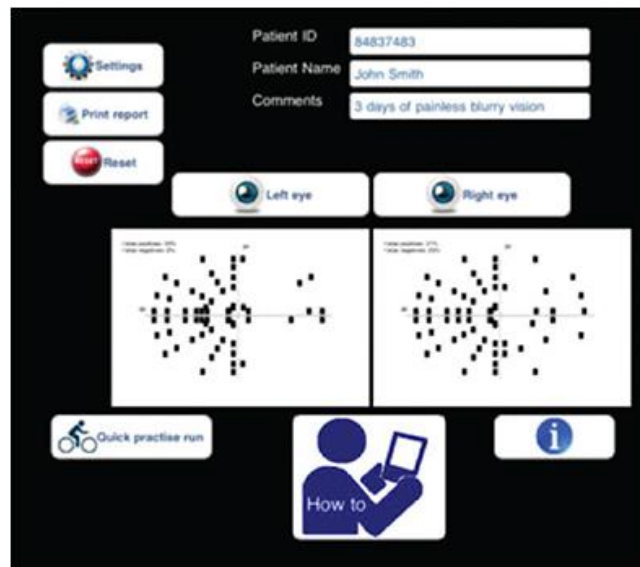


Figure 13. VisualFields easy. <http://www.retinalphysician.com/issues/2015/may-2015/smartphones-for-visual-function-testing> .

There have also been attempts to create visual field testing applications for mobile devices and *VisualFields Easy* (<https://itunes.apple.com/us/app/visualfields-easy/id495389227?ls=1&mt=8>), a perimetry application for iPad, is an interesting example that. In *VisualFields Easy*, the user is required to hold the tablet 33 centimeters in front of them. In the test itself, the visual field is mapped quadrant by quadrant – a red dot that works as the fixation point is placed in one of the corners of the screen, and the rest of the space is used for displaying stimuli. The user signals stimulus detection by tapping the screen or with a Bluetooth keyboard if that is what they prefer. This is repeated until all four quadrants have been mapped. The results are displayed in a visualization after the test (see figure 13).

3.8 Gaze-tracking and visual field testing

Intuitively, gaze-tracking features could be useful in visual field testing for fixation validation and also for providing input to keep track of stimulus detection. As mentioned in section 3.2, many modern commercial visual field testing devices are integrated with gaze-tracking features for fixation validation, including the modern versions of the Humphrey Field Analyzer presented in section 3.6.

While gaze-based input is quite underused in visual field testing, several examples of existing solutions were found for the review in this study. For example, Murray et al. (2008) have created and tested an experimental visual field assessment technique for performing automated static perimetry that utilizes gaze tracking. The developed system consisted of a personal computer, an external display, and a Tobii eye-tracking device. The technique was especially designed for testing the visual field in children. As children might not be able to concentrate and comply to the test process in a more traditional visual field test, in this test setting, the participants were not explicitly asked to follow the displayed stimuli, but their naturally occurring saccadic eye-movements were used as input instead. The eye-tracking data was used to track reflexive saccadic eye movement in order to evaluate stimulus detection, and it was also used to calculate a suitable location for the next stimulus to be displayed. At least to some extent, the results of the tests were promising: the technique seemed to be quite close in accuracy to a Humphrey screening test when tested with healthy adults, and even with children, the results seemed highly agreeable with an expected normal field of vision. [Murray et al., 2008]

Meanwhile, Wroblewski et al. (2014) have developed and clinically tested a head-mounted perimetry device called VirtualEye, which also uses gaze tracking. The system is a custom made device, consisting of a computer software, microdisplays and a binocular eye-tracker, all of which are integrated to a head-mounted visor. As there are separate displays for each eye, no eye patch is needed in order to test one eye at a time. A full, qualitative threshold test can be performed with the system. The system was tested with two different inputs – manual mouse input and visual input. Similarly, as in the solution by Murray et al. (2008), the test participants were not asked to give the visual input explicitly, but the system relied on the natural saccadic response caused by the visual stimuli. The preliminary clinical testing indicated that with both inputs, the system could measure the visual field in a manner comparable to the standard automated perimetry performed with the Humphrey Field Analyzer. Even though, in the tests, the visual input did result into poor fixation and/or a large amount of false negatives or negatives, and the writers argue that the unfamiliarity of the input method might play a role in successful testing. [Wroblewski et al., 2014]

3.9 General user experience challenges in visual field testing

While specific articles or similar material that would have explicitly evaluated visual field testing from user experience's point of view were not found to be reviewed in the information search for this study, multiple user-experience related challenges were mentioned in the reviewed material. Many of the currently existing automated testing methods require that the patient holds a steady fixation for the test's duration to acquire reliable test results. To achieve a steady fixation for the whole test's duration, the patient needs to use their eye in an unnatural manner, and for some, it might be a very difficult task. Prolonged test situations and unnatural (lack of) eye movements can cause fatigue in patients. Especially prolonged test situation can cause other types of discomfort as well, and while patient's comfort should be considered as a value itself, it should also be noted that fatigued and uncomfortable patients might perform worse in the test than would be expected normally. In short, the patient's comfort should be considered when designing a visual field testing system. [Werner, 1991]

Additionally, one thing that creates challenges to visual field testing is that in addition to working test and testing equipment, the test's reliability also heavily depends on the test subject's personal performance. The learning effect is a real thing in visual field testing - some patients with visual field deficits have been tested since early age so that they've become quite adept at taking a visual field test, and thus their results are generally more reliable than of those who are new to getting their visual fields tested. There are many different user groups for visual field testing. For example, many patients, like those with acute neural damage don't often have experience in visual field testing, and might provide more unreliable results than other patients. Varying user groups with varying capabilities and experience pose a challenge to the visual field testing equipment and software, emphasizing the importance of good user experience and usability. [Walsh, 2011; Werner 1991]

In the next chapter, a novel visual field testing software designed to address these aforementioned challenges will be presented. The software was tested in a series of user tests, which will be described in chapter as well. The test results will be presented in chapter 5, and finally, the implications of these results, including how the software answered to the challenges presented in this section, will be discussed in chapter 6.

4 RESEARCH METHOD

This chapter presents what was done in the empirical part of this study. A test software with a simple qualitative visual field test and two alternative interaction techniques was designed and developed to evaluate the differences of the user experience provided by the two. In this chapter, the design, implementation and technical setup of the test software are described in detail. After that, the test settings, participants and procedure are presented.

4.1 The test setup

The hardware used for the test consisted of a laptop computer, a computer mouse, an external display and a gaze tracking device. The laptop computer that was used for developing and running the test software was an Acer Aspire V3-575G with Windows 10 as the operating system. A Samsung display screen was used to display the software. A regular 3-buttoned Logitech USB mouse was used for click input. The gaze tracking device used was a Tobii EyeX gaze tracking device that utilizes near infra-red light-based illumination and sensors in order to track the user's eye movement. While it is acknowledged that the Tobii EyeX is not an official, fault tolerant medical measuring device, there are several reasons why it was chosen for this study. The device was selected due to its easy commercial availability, low price, portability, ability to track only one eye at the time, and its easily approachable developer kit. If a working test software similar as the one described in this study existed, the selected gaze tracker could potentially be the device of choice for the users of the software. [Tobii Developer Zone, 2017]

In the test situation, each participant was seated in front of the computer display and positioned so that their eyes were at the distance of 55 centimeters from the display. On each test run, the participant was asked to cover one of their eyes with an eye patch, and as needed, their position was adjusted so that their uncovered eye was vertically and horizontally in line with the fixation point in the middle of the display. The eye-tracking device was attached to the bottom edge of the display and the mouse used for input was offered for the participant to use with their right or left hand as they preferred. An illustration of the test setting is presented in figure 13.

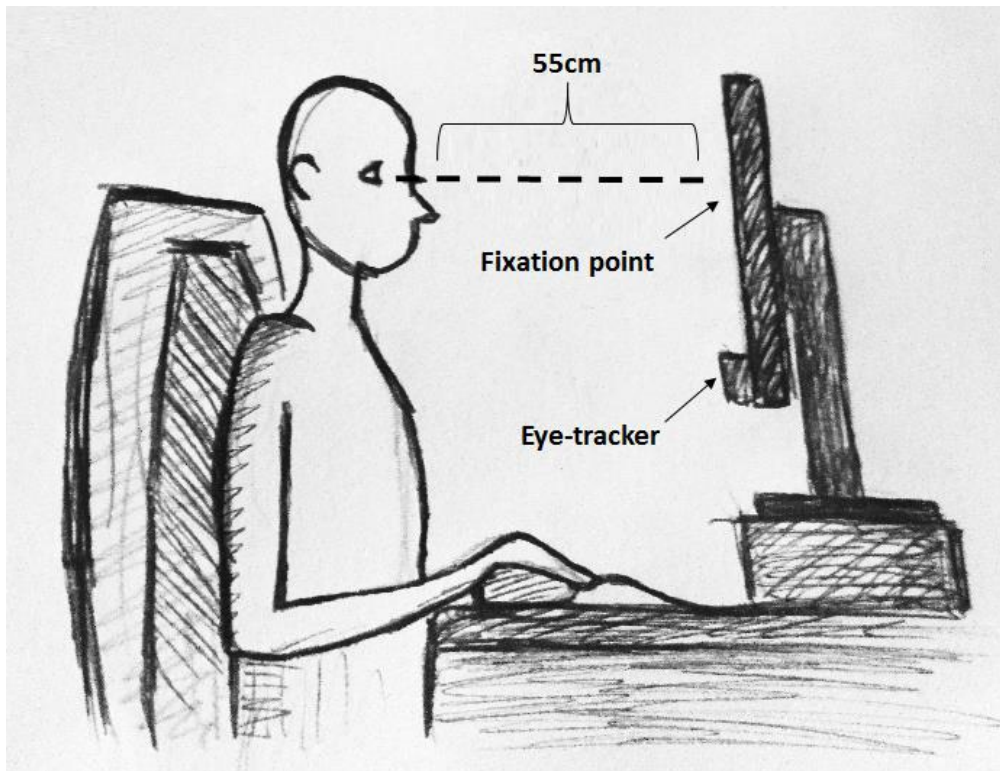


Figure 13. An illustration of the laboratory test setting.

The tests for this study were performed in the university's gaze-research laboratory, which provided a distraction-free and steady setting for all of the tests. The environmental conditions were not manipulated much during the tests, even though for all participants, the laboratory's lights were dimmed during the test performance to ensure that external light did not disturb the test performance. In these tests, laboratory equipment was not used in other ways, and the test performances were not recorded with any media. In the first test, the use of a chin rest from the laboratory was attempted, but it turned out to be a distraction, and the gaze-tracking device could not be calibrated while it was used. While a chin rest would have been a useful tool for regulating the participant's position, the chin rest was abandoned during the actual tests due to those issues. Several other resources were also used in the test situation – a black, light-blocking, rubber stringed medical eye-patch with a sterile swab tucked underneath it for hygiene reasons, was used to cover the participant's eye that was not currently tested. Consent forms, background information forms and evaluation forms were provided in print format, alongside with writing equipment. All in all, the goal was to keep the test settings as simple and as replicatable as possible.

4.2 The test software

The test software was implemented with Java on the Netbeans software development platform, and the gaze-tracking data was provided by a separate C++ program. With the current solution, the software allowed the mapping of 25 degrees of the visual field to each side from the fixation point, and 15 degrees of the visual field upwards and downwards from the fixation point. The extent of the mapped area was selected based on the size of displays available, and also by the sitting distance required by the gaze tracker. In the future, the mappable area could be extended by using a larger screen, by using a gaze-tracker that allows a shorter distance from the display, or by using the available screen space in a more versatile manner, for example as it is used in the VisualFields Easy -solution presented in section 3.7.

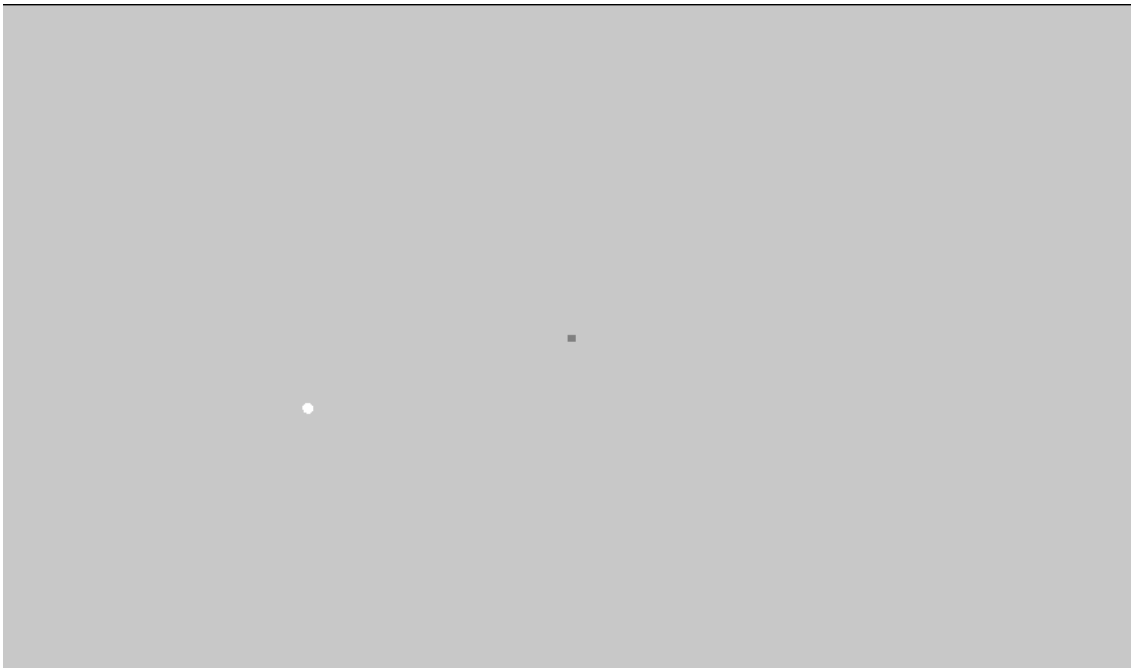


Figure 14. The graphic user interface of the test. The dark rectangle in the middle is the fixation point and the white dot is a stimulus.

The graphic user interface of the test software was built from simple components. The stimuli appeared on a grey background that filled the entire screen and has the fixation point marked in the middle (see figure 14). White dots with a diameter of 0,47 centimeters (15 pixels) were used as stimuli and the fixation point was marked with a dark grey rectangle with a width and a height of 0,31 centimeters (10 pixels). The fixation point's location was calculated so that it was placed exactly into the middle of the screen. The background area consisted of an active area (covering 50 x 30 degrees

of visual field as mentioned above) on which stimuli were displayed, and some padding was added around the edges to also fit the area into the middle of the screen. The colour values of the user interface components were selected by approximating suitable values for a qualitative test in which the participant should not miss stimuli because of a lack of contrast. The grey hues were used also in order to avoid too sharp contrast that could cause fatigue in the long run.

For this study, a set of 42 preset stimuli locations were selected so that they would cover the mapped area of the visual field thoroughly. The stimuli locations were the same on each test run, but the displaying order was scrambled to avoid a learning effect. Eight of the stimuli were located on the calculated blind spot area, and the rest were placed evenly around the mappable area. The blind spot area was calculated to cover an area of 7 x 7 degrees of the visual field either to the left or to the right from the fixation, depending on the eye being tested. As the software supported testing both, the left and the right eye, an identical set of stimuli was used for both eyes with the exception of the location of the blind spot stimuli's coordinates. A full graphical presentation of the stimuli's and the blind spot's locations can be found in Appendix 1.

To start a test, the software inquired with dialogue selections whether the left eye or the right eye would be tested, and whether the test would be performed with mouse input or gaze input. Also, in this implementation, the gaze-tracking feature needed to be turned on separately before actually starting the test, due to the selected technical approach. The gaze-tracking part of the software was implemented by modifying an existing available C++ solution for receiving the stream of gaze data, and the gaze data was transferred to the Java solution through TCP. After the dialogue selections, in both, the mouse input test and the gaze input test, the participants were allowed to start the test with a mouse click as they felt ready and comfortable. The mouse and gaze input tests are described in more detail in the next sections.

The feedback provided by the test was implemented quite roughly and even though graphic feedback would have been nice, it was left out of this project's scope. In this implementation, after all the stimuli are displayed and the test is done, the results were displayed in a dialogue in a text format, listing which stimuli were missed and which one's were successfully detected by the participant. Timestamped results were also

saved into a text file for later viewing. The full description of the feedback provided by the test software is later described in section 4.5.

4.3 Mouse input test

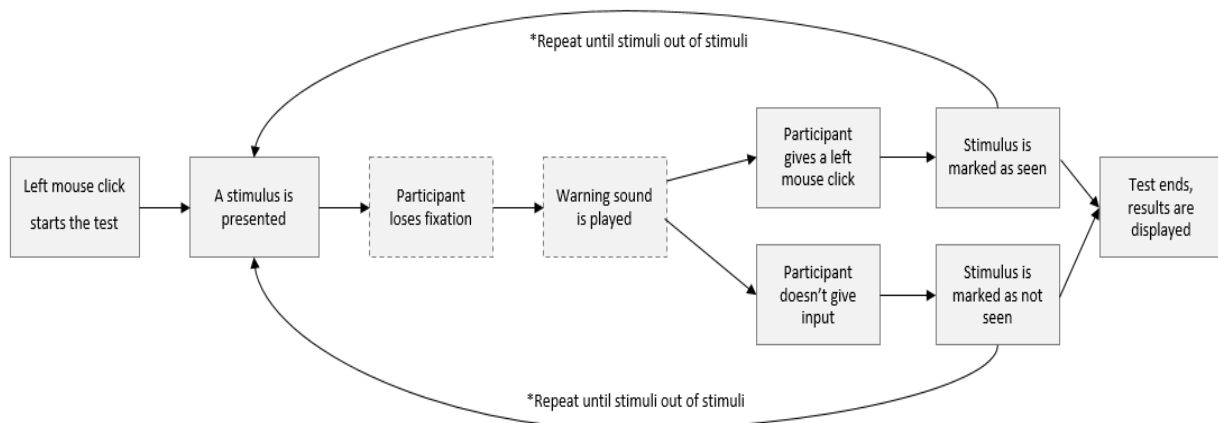


Figure 15. The test procedure of the mouse input test.

In the mouse input test, an explicitly given left mouse button click was used as an input method to signal stimulus detection. After the participant started the test with a mouse click, the stimuli (described in the previous section) appeared and disappeared one after another in a randomized order until all the stimuli had been presented. The participant had 3 seconds to react with a mouse click to each stimulus in order to mark it as *seen* – first, the stimulus was visible for 1,5 seconds after which it would disappear. At this point, the participant had additional 1,5 seconds to provide input until the stimulus was marked as *unseen* and the next stimulus was presented. This procedure was repeated until there were no unrepresented stimuli left. It is acknowledged that in future implementations it would be beneficial to add some randomized variance to the length of the reaction period in order to avoid false positive clicks that can be the result of the participant recognizing the even rhythm of the test. Anyhow, for the evaluation of the user experience of different interaction techniques, it was considered adequate to have a reaction period of a standardized length for each of the stimuli.

In the mouse input test, gaze-tracking was also used non-explicitly to monitor the participant's fixation, since the participant was required to hold their fixation in the marked fixation point at all times while the test was running. If the participant's gaze wandered away from the fixation point, the system would react by playing a warning sound. Also, if the participant's gaze was detected wandering, the stimulus that was currently being displayed was be marked with a special mark (see section 4.5 for the full description of the test's output).

4.4 Gaze input test

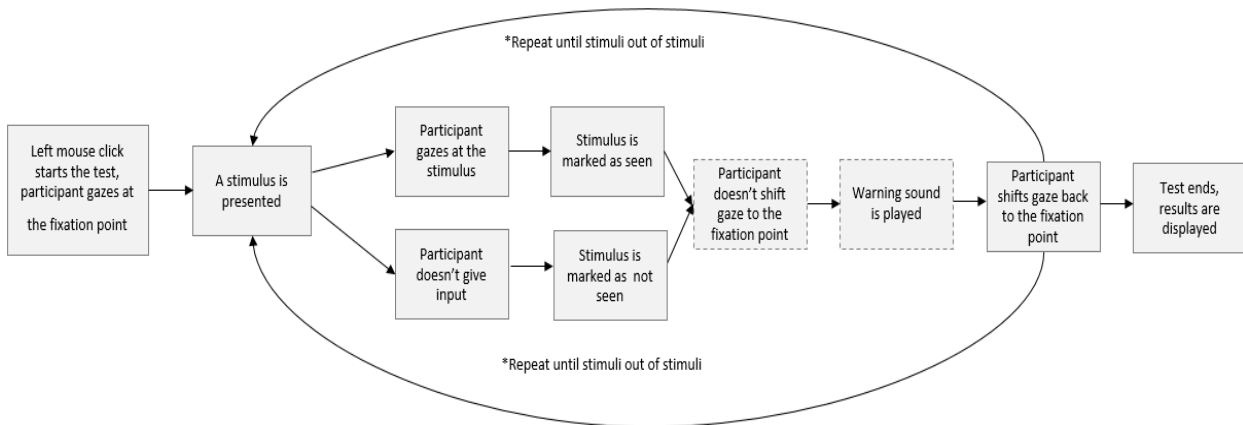


Figure 16. The test procedure of the gaze input test.

In the gaze input test, the participants used their gaze to signal stimulus detection. As in the mouse input test, the participant was again asked to start the test with a mouse click, after which the stimuli would appear one after another. At the start of the test, the participant was asked to fix their gaze into the fixation point, and as they would see a stimulus appear, they would signal detection by looking at it. Similarly as in the mouse input test, each stimulus appeared for 1,5 seconds, after which the stimulus disappeared and the participant had additional 1,5 seconds to react to it. In case of missing input, the stimulus was marked as *unseen*. If at any point of the 3 second long reaction period the coordinates from the gaze data stream would match the coordinates of the stimulus plus/minus a 100 pixel wide buffer zone around it, the stimulus was marked as *seen*.

After looking at the stimulus, the participant was required to turn their gaze back into the fixation point. The next stimulus would only be displayed, after the participant's gaze was detected to be back in the correct location, and in case the gaze was detected wandering, a warning sound was played as a reminder to look back into the fixation point. This procedure was repeated until there was no more stimuli left to show.

4.5 Test output

As was stated in section 4.2, graphic feedback features were not implemented for the test software, but the feedback was provided in a text format. The feedback that described the participant's performance consisted of nine data variables, which are also presented in figure 17. The variables written in grey text ended up not being used in final analysis, although they could provide useful in future iterations of the software.

Data	Format
Amount of stimuli presented	Number
Seen stimuli	List of stimuli codes + pixel locations
Missed stimuli	List of stimuli codes + pixel locations
Seen/missed ratio	Percentage value
Stimuli seen in the blind spot area	List of stimuli codes + pixel locations
Stimuli missed in the blind spot area	List of stimuli codes + pixel locations
Seen/missed ratio in the blind spot area	Percentage value
Missing fixation (in the mouse input test)	Listed stimulus marked with ' ! '
Reaction time	Time in milliseconds

Figure 17. The data provided by the test.

Naturally, the number of stimuli was recorded into the feedback. In addition to that, the successful stimuli detection was recorded by listing which stimuli were seen and which one's were missed. As the blind spot detection was used as a validator of the test working in this study (see section 5.1), the seen and missed stimuli in the blind spot area were also recorded separately from all the stimuli that were presented. To ease the

monitoring of the seen and missed stimuli, the mapped area was split into columns marked with numbers and rows marked with alphabetical characters, providing each stimulus an alphanumeric code. For example, the stimulus coded as A5 was placed 15 degrees above and 5 degrees to the left of the fixation point. As an exception, the stimuli in the blind spot area were coded as BSL/BSR1-8. Additionally, the pixel coordinates of the seen/missed stimuli were recorded. In the initial feedback, percentage ratios of the seen and missed stimuli were calculated, and in the later analysis this was expanded further (see section 5.1.). The full illustrated chart of the stimuli and their codes can be found in Appendix 1.

In the mouse input test, the stimuli that were detected while the fixation was not on point were marked with an exclamation mark. This variable was not used in the analysis as it could not meaningfully be compared with anything in the gaze input test. Even though, with some iteration this information could provide useful in future versions of the software. Also, reaction time was recorder with each successfully detected stimulus. As the variable signaling fixation loss, this variable could provide useful input for more sophisticated features in the future, although it was left out of the analysis made for this study. In addition to these variables, the feedback also recorder the timestamp of the test's completion time, and which eye and which input the test was performed with.

4.6 Participants

The software described in the previous sections was tested in a series of laboratory tests to shed some light on the research questions of this study Overall eight people participated to the tests. Since the test's purpose was to evaluate the user experience of two different input methods, and not to evaluate the diagnostic accuracy, participants with a normal visual field were recruited for this test. The participants were required to be able to see the computer screen at least reasonably well from the distance of 55 centimeters without eye glasses or with contact lenses on. Due to the small pool of participants, there were no specific demographic requirements. For convenience reasons, the participants were recruited from among the university's students and staff.

Five men and three women, between the ages of 19 to 55, participated to the test. None of the participants had experience in taking part to a computerized visual field test

before this test, although one of the participants recalled that he had taken a simple confrontational visual field test (see figure 4) at some point of his life. While it would have been acceptable to wear contact lenses during the test, none of the participants did. In other remarks, one of the participants had had his near-sightedness fixed in a laser surgery some time before and two of the participants reported to have a lowered visual acuity in one of their eyes. None of these things had a significant effect to the test performance in this particular test setting. A summary of the participants is presented in figure 18.

#	Age	Gender	Has participated in a visual field test before	Contact lenses	Other notes
1	39	Female	-	-	-
2	29	Male	-	-	-
3	35	Male	-	-	-
4	19	Female	-	-	Lowered eyesight in right eye
5	39	Male	-	-	Eyesight fixed with laser surgery.
6	55	Male	A confrontational test (see figure 4)	-	-
7	36	Female	-	-	Lowered eyesight in left eye
8	28	Male	-	-	-

Figure 18. Summary of the test participants.

4.7 The test procedure

Each participant performed the test individually, and thus overall eight test sessions were held. The tests were held at the university's gaze laboratory and each test session took between 30 to 45 minutes. Before the eight test runs, two separate informal pilot tests were held in order to validate that the test software worked and that the test goals were set correctly.

Each of the eight official test runs was performed in the same manner. First as a test session started, the participant was greeted and guided to their seat in the laboratory. First of all, they were handed two forms – the consent form (see appendix 2) and the background information form (see appendix 3). As the participant was filling the forms, the upcoming test procedure was also explained to them orally, and they were asked to state any questions that they might have about the procedure.

After the participant was ready with filling the forms, they were provided with an eye patch, with which they were asked to cover one of their eyes. In this test, the order between left and right eye was not controlled and thus the participant was allowed to select, which eye they wanted to do the test with first. Five of the participants chose to start with their left eye and three with their right eye. Before the actual test, the calibration functionality of the Tobii EyeX software was performed in order to accommodate the gaze-tracking system to the participant's eye. After a successful calibration, the participant was set to the correct distance from the screen and the first test setting was ran (see chapter 4 for a further description).

The order of testing the different inputs was controlled in this test – each odd-numbered (1,3,5,7) participant did the mouse input first and each even-numbered (2,4,6,8) participant did the gaze input first. In each test session, both inputs were first tested with one eye, and then with the other. Thus, each participant performed four test runs. The test orders for each participant are presented in figure 19.

#	1	2	3	4	5	6	7	8
Test order	1L-2L- 1R-2R	2L-1L- 2R-1R	1R-2R- 1L-2L	2L-1L- 2R-2L	1L-2L- 1R-2R	2L-1L- 2R-1L	1R-2R- 1L-2L	2R-1R- 2L-1L

Figure 19. Test order for each participant, where 1 = mouse input, 2 = gaze input, L = left eye and R = right eye.

After each test run, a copy of evaluation form 1 (see appendix 4) was handed to the participant to fill. After the first two runs, the participant was asked to switch the eyepatch over the other eye and the gaze-tracking system was also recalibrated. Also, the participants correct distance from the screen was monitored through the test and adjusted as necessary. After all four test runs, as the participant had filled the evaluation form 1, they were handed the evaluation form 2 (see appendix 5) in order to compare their experience of the mouse and gaze-tracking input. After the participant had filled the form, the answers were discussed together in a lightweight interview to find out whether they had something else to add, and to collect some deeper qualitative feedback on the software. After the discussion, the participants were thanked and escorted out of the laboratory. To summarize how the tests were ran, a nutshell representation of the test procedure's workflow is illustrated in figure 20.

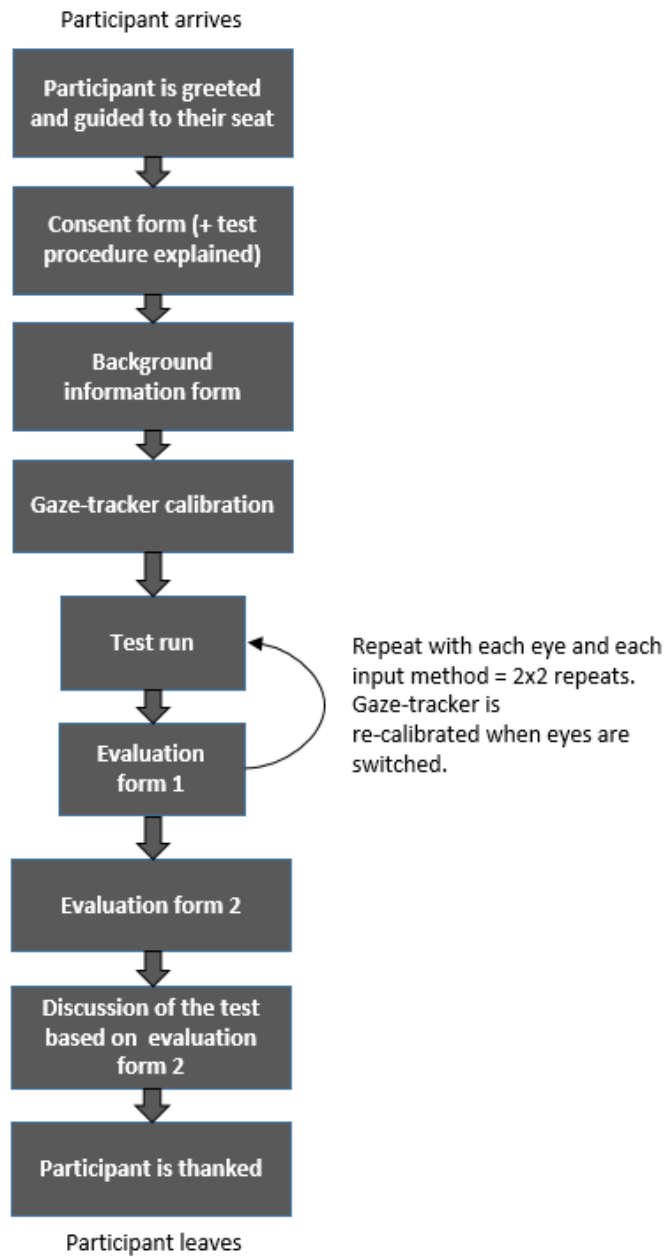


Figure 20. The test procedure.

5 TEST RESULTS

The results from testing the software are discussed in this chapter. In order to find out whether it is possible to build a visual field testing system with gaze-tracking input, the test's accuracy was evaluated based on the results provided by the test software. The tests provided three different types of results that are discussed in their own sections. Accuracy-related observations were made based on the participants' performances during the test runs. The user experience of the test software was evaluated with form surveys and a brief interview with each participant. Feedback was collected during the test with Evaluation form 1 (see appendix 4) after each test run, and comparative feedback of the two tested interaction techniques was collected with Evaluation form 2 and open interviews after the tests.

5.1 Test accuracy

While the the laboratory tests' emphasis was on measuring the user experience, several light-weight measures were collected to validate that the test software worked as intended. Based on the output provided by the software (see figure 17), successful blind spot detection was calculated from each test run. In this analysis, blind spot detection was consider successful when at least one stimulus from the blind spot area was missed during the test. This level of accuracy was considered adequate for this particular test setting, but it is acknowledged that more exact meters are needed in order to develop a clinically valid software (see chapter 6 for further discussion). Additionally, the amount of so-called false negatives, or in other words stimuli missed outside the blind spot area, was calculated.

As anticipated, a learning curve was visible in the test results. Regardless of the type of test that was performed, for multiple participants, the first test runs yielded worse results than the latter ones (see figure 21). Overall, on the first test run, blind spot was detected from 5 participants. On the second test run, the blind spot was successfully detected from 6 participants. The blind spot was successfully detected from 7 participants on the third and fourth test run. It is worth noting that the unsuccessful detections concerned two different participants on runs three and four (see figure 23 for more details).

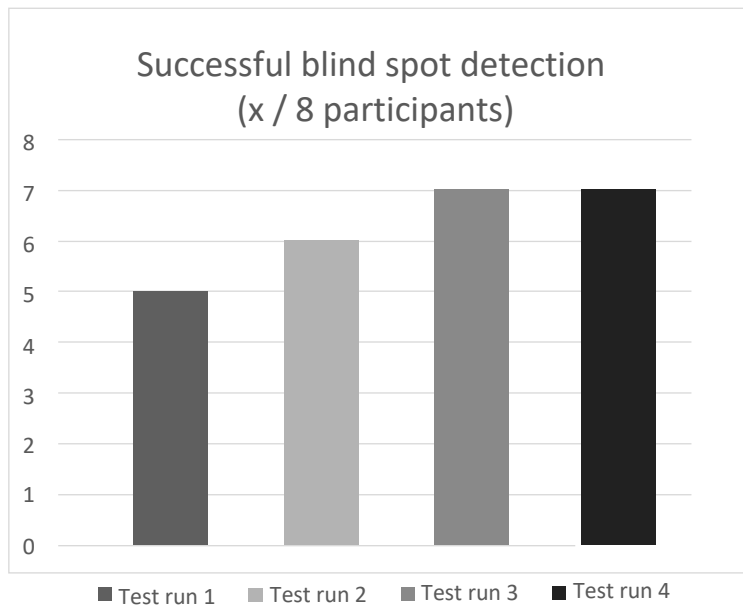


Figure 21. Successful blind spot detections by overall test runs.

Test	Blind spot detected on x / 8 participants	Avg. FN
Mouse input, 1. round	6/8	0.88
Mouse input, 2. round	6/8	0.38
Average	6/8	0.63
Gaze input, 1. round	5/8	1.13
Gaze input, 2. round	8/8	0.25
Average	6.5/8	0.69

Figure 22. Successful blind spot detections by different test settings.

Figure 22 compares the success of blind spot detection between the two different test settings. With the mouse input, the blind spot was successfully detected from 6

participants on both, the first and the second test run. As can be seen in figure 23, participant #7's blind spot could not be detected on either of the test runs, participant #3's blind spot could not be detected on the first test run, and participant #4's blind spot couldn't be detected on the second test run with the mouse input. With gaze input, the blind spot could not be detected successfully from three participants, but on the second run, it was successfully detected from every participant. Within the different test settings, learning was also notable - with both input types, the amount of false negative detections was higher on the first test run with the particular input. With the mouse input, the amount of overall false negatives on the first test run was slightly lower than with the gaze input. On the other hand, on the second test run, the amount of false negatives was slightly lower with the gaze input compared to the mouse input. As a whole, the gaze input provided few more false negative detections, but all in all the differences were quite small. A more detailed representation of the distribution of blind spot detections and false negatives by participant and by test run can be seen in figure 23.

In addition to the obvious learning effect, there are several other recognized reasons for the unsuccessful blind spot detections and false negatives during the test. First of all, as explained in section 4.1, the participant's distance from the computer screen and the gaze tracking device was regulated only by measuring and adjusting the participant's posture as needed. Thus, it is possible that the participant's posture shifted during the test, causing the expected viewing angle to be incorrect. In several cases, while the blind spot could not be detected, the participant missed stimuli (counted as false negatives here) that appeared very close to the calculated blind spot area, suggesting that their position in front of the screen was off. As the gaze-tracking device also provides the z-coordinate (the user's distance from the device), with the gaze-tracking input, position-related issues could easily be diminished by making the screen content responsive to the z-coordinate. Another solution would have been a proper chin rest, as for these tests there was no suitable one available.

Another easily distinguishable source of false negative detections was the stimuli presented near the edges of the screen. In several cases, participants missed stimuli near the edges, most probably not due to vision-related issues but due to the visual disturbance caused by the edges of the screen. Even though a "buffer area" of ~1-2 centimeters between the test area and the screen's edges was designed to avoid this, in

some cases this was not enough. This is definitely a problem that needs to be tackled somehow, as there shouldn't be any software-related accuracy issues in a software like this. There are several solutions that could provide help.

- The buffer area could slightly be extended. The problem with this solution is that the more the content on the screen is extended, the larger the physical screen is also required. The computer screen used for the testing was already quite large compared to most common commercial devices, and it would be preferable that the it was possible to operate the software in as compact settings as possible.
- Some experimenting could be done with the physical test settings and the contrast of the objects on the screen to see if it helped the issue. As described in section 4.1., the participants performed the tests in a dimly lit space and as described in section 4.2., the objects on the screen (the background and the stimuli) had a particular contrast. It is possible that objects with a higher contrast between them might not be affected by artifacts caused by the screen edges. This solution also has it's problems – the test software should not be limited to qualitative testing only (which it will be if only a certain contrast combination allowed), but it should allow different contrast combinations for qualitative treshold testing as well.
- The screen area could be used more efficiently. Similarly as in the VisualField Easy concept (presented in section 3.8), different areas of the visual field could be mapped separately by placing the fixation point in other areas of the screen than in the middle of it. This approach would allow somewhat larger buffer areas on the screen edges, and it would also make it possible to map larger areas of the visual field than with the fixation point in the center.


In addition to the aforementioned things, further and more detailed software improvement suggestions are discussed in chapter 6.

While only a small data set with 8 participants and 2 x 2 test runs was collected for this study, these results could hint that gaze tracking could provide an input method that does not lack in accuracy compared to a more traditional click input. Of course, more specific and extensive testing with a larger pool of participants, more repeats and more accuracy-oriented test tasks is required to make any actual conclusions.

Participant	#1	#2	#3	#4	#5	#6	#7	#8	Avg.
Test run 1	1L	2L	1R	2L	1L	2L	1R	2R	
BS	2/8	3/8	0/8	0/8	4/8	4/8	0/8	5/8	2.25
FN	0	0	5	0	1	0	1	3	1.25
Test run 2	2L	1L	2R	1L	2L	1L	2R	1R	
BS	0/8	4/8	2/8	3/8	5/8	5/8	0/8	5/8	3.00
FN	3	0	0	0	1	0	2	1	1.00
Test run 3	1R	2R	1L	2R	1R	2R	1L	2L	
BS	6/8	4/8	3/8	2/8	4/8	4/8	0/8	4/8	3.38
FN	0	0	1	0	0	0	0	1	0.25
Test run 4	2R	1R	2L	1R	2R	1R	2L	1L	
BS	4/8	4/8	3/8	0/8	2/8	4/8	2/8	4/8	2.88
FN	0	0	1	1	0	0	0	1	0.38
Avg. BS	3.0	3.75	2.0	1.25	3.75	4.25	0.5	4.50	2.89
Avg. FN	0.75	0.0	1.75	0.25	0.50	0.0	0.75	1.5	0.69

BS = Number of stimuli missed in the blind spot area.

FN = False negatives (expecting that the participants' visual fields were normal), stimuli missed in other than the blind spot area.

 = False negatives


 = No stimuli missed in the blind spot area / blind spot not detected

Figure 23. a detailed presentation of the overall blind spot detections and false negatives by participants and by test runs.

5.2 Feedback collected during the test

As explained in section 4.7, feedback was collected from the participants after each test run with an evaluation form (see appendix 4). The evaluation form consisted of five closed-ended questions that were derived from NASA's Task Load Index to suit the task at hand (<https://humansystems.arc.nasa.gov/groups/tlx/index.php>). The questions were designed to measure five subjectively evaluatable attributes that count towards the user experience. The measured attributes were *physical* and *mental load*, *required effort*, resulting *fatigue/discomfort/stress/frustration* and the test's *learnability*. On a scale from zero to nine, the participants were guided to mark the value they felt described their current feelings the best after each test run. As explained in section 4.7, each participant did four test runs altogether, and also filled the evaluation form four times. The order in which each participant performed their test runs can be reviewed in figure 19 in section 4.7. Only numeric evaluation was provided with this mid-test form, and the participants were allowed to elaborate their feelings in the form and the interview that were administered after the test (see section 5.3).

For the analysis, the score provided by the participants were evaluated and compared in two ways. First of all, arithmetic mean values of the score given to each attribute were calculated and compared based strictly on the order of the test runs. In other words, the score from the participants overall first, second, third, and fourth test run were compared to see, whether repeating the test regardless of the input method had an effect on the evaluation. This type of comparison can be seen, for example, in figure 25, which represent the average physical load score by the overall order of test runs. Secondly, the arithmetic mean values were calculated and compared based on the input and the order of test runs that considered the same input to see whether the input method had an impact to the evaluation. As can be seen, for example, in figure 26, the average score for this comparison was calculated by combining and dividing the score from each participant's first and second test round with each input.

To evaluate the implications of the results, a brief statistical analysis was performed to see whether the differences in the dependent variable could be interpreted to have been caused by chance, or by the changes in the independent variable. Repeated measures ANOVA (analysis of variance) was selected as the suitable tool for the analysis. In each analysis, the dependent variable was the mean of scores provided by

the test participants to evaluate each attribute. The independent variable was the input type paired with the test order (see figures 26, 28, 30, 32 and 34). A summary of the analyzed sets of test results and the results of the analysis are presented in the following table (figure 24).

Dependent variable	Independent variable	Sig. (p-value)	Statistically significant
Physical load avg.	Input type + test order	0.095	-
Mental load avg.	Input type + test order	0.038	yes
Effort avg.	Input type + test order	0.132	-
Discomfort avg.	Input type + test order	0.081	-
Learnability avg.	Input type + test order	0.051	-

Figure 24. Statistic significance of the results based on repeated measures ANOVA.

In this analysis, the differences in the dependent variable were considered significant, when $p < 0.05$. As can be seen in figure 24, only the differences in mental load ($p = 0.038$) could be interpreted as statistically significant based on the repeated measures ANOVA. Based on the analysis, the other results were not statistically significant. Next, all five attributes and the acquired score results are presented in more detail.

First of all, the test participants were asked to evaluate the physical load caused by each test run. In this study's context, the physical load referred to the evaluated amount of physical effort required to perform successfully, and to the possible physical discomfort caused by the test situation. On a scale from 0 to 9, 0 represented minimum physical load, and 9 represented a high physical load. The presumption was that the test with mouse input could cause a heavier physical load, especially due to having to hold the eye still unnaturally for a prolonged period.

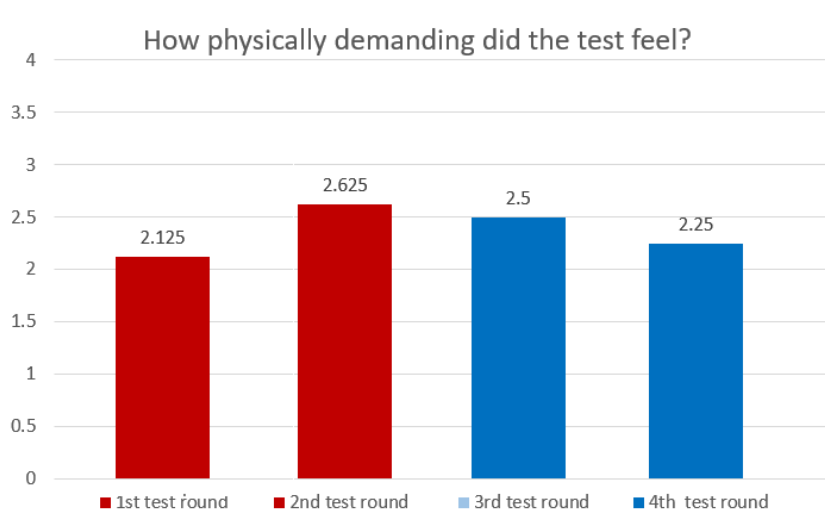


Figure 25. Mean score of the evaluated physical load of each test setting based on the overall order of the tests (on a scale from 0 to 9).

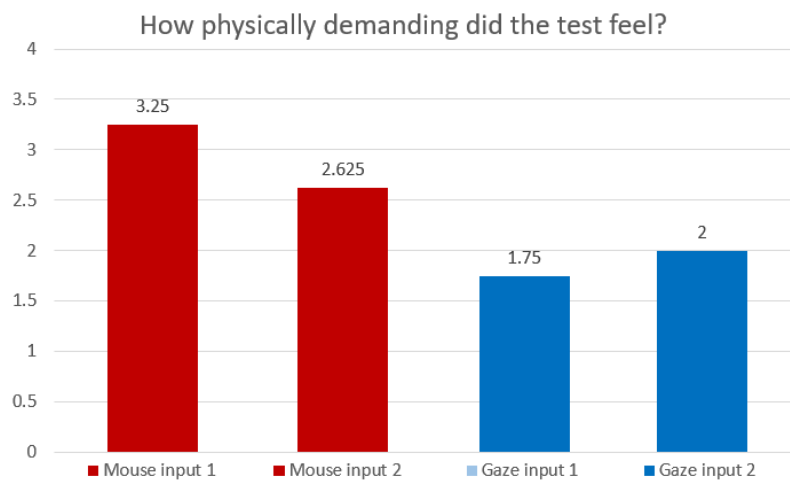


Figure 26. Mean score of the evaluated physical load of each test setting based on the input method (on a scale from 0 to 9).

As can be seen in figure 25, there were mild differences between the overall test rounds, and the first round was evaluated with a slightly lower score than the rest of the rounds. As can be seen in the chart presented in figure 26, the participants evaluated the mouse input rounds as more physically demanding than the gaze test rounds. These differences did not turn out to be statistically significant (see figure 24). It should be noted that all in all, the participants mostly gave quite low scores regarding the physical load – even the highest mean in the chart is only 3.25.

Secondly, the participants were asked to evaluate the mental load caused by each test run. In this context, the mental load referred to the mental effort, for example the amount of concentration, required to perform the test, and to the amount of mental discomfort caused by the test. Similarly as for the physical load, on a scale from 0 to 9, 0 represented minimum mental load, and 9 presented a very high mental load. Before the test, it was assumed that the mouse input test might require more concentration than the gaze input test, which result to a higher mental load.

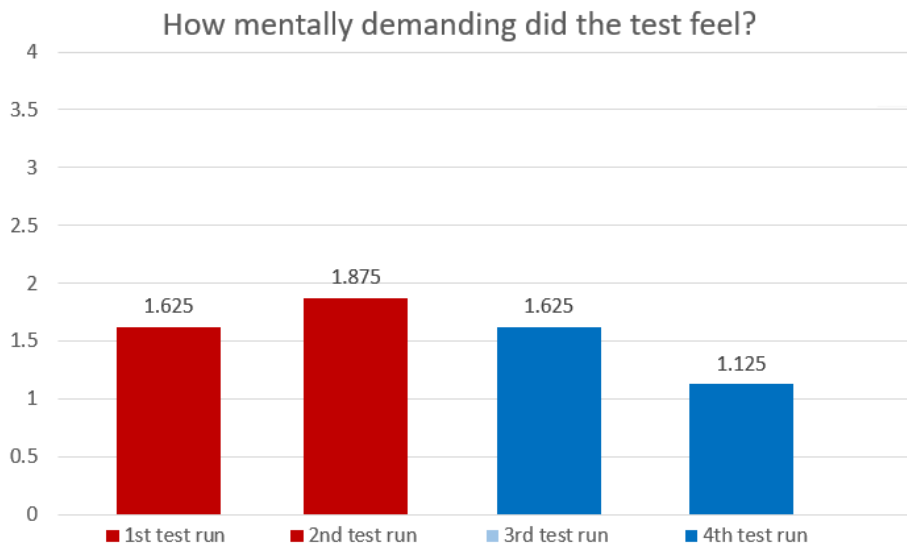


Figure 27. Mean score of the evaluated mental load of each test setting based on the overall order of the tests (on a scale from 0 to 9).

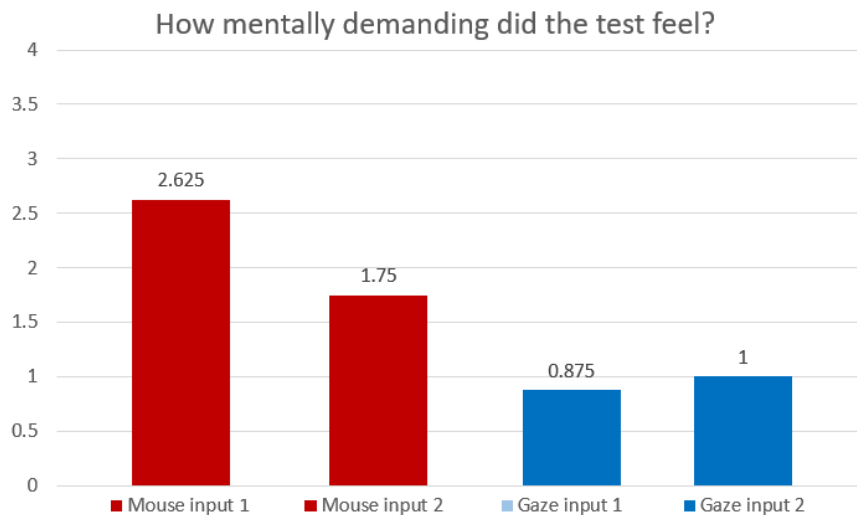


Figure 28. Mean score of the evaluated mental load of each test setting based on the input method (on a scale from 0 to 9).

As can be seen in figure 27, the evaluated mental load did not vary heavily between the overall test rounds, although the final round was considered somewhat mentally "easier" than the previous rounds. As can be seen in figure 28, the participants evaluated that both of the mouse input test rounds caused a higher mental load than either of the gaze input test rounds. Unlike with the physical load, the differences in the average mental load score between the different inputs (figure 28) turned out to be statistically significant ($p = 0.038$). As this significance was further analyzed with a paired samples t-test, statistical significance was shown for the mouse 1 - gaze 1 pair ($p = 0.021$) and for the mouse 1 - gaze 2 pair ($p = 0.035$). While it is not meaningful to compare the mouse 1 - gaze 2 pair due to the fact that the participants already had experience with the test when doing gaze 2, the difference between mouse 1 and gaze 1 is mentionworthy. As with the physical load, it should be noted that participants also gave generally low scores to the mental load despite the available scale.

The third attribute that the participants were asked to evaluate was the overall estimated effort that was required to perform the test successfully. The participants were allowed to evaluate the effort they used as a whole, including physical, mental and cognitive effort. Similarly as with the physical and the mental load, the participants were asked to evaluate the effort on a scale from 0 to 9, where 0 represented minimum required effort and 9 represented very high required effort. An assumption was made that as long as the working principle of the gaze input is understood by the user, the mouse input test should require more effort than the gaze input test, since the gaze input test aims to a more natural way of interacting with the test with a single modality instead of two. It was also presumed that lower physical and mental load should result to lower required effort.

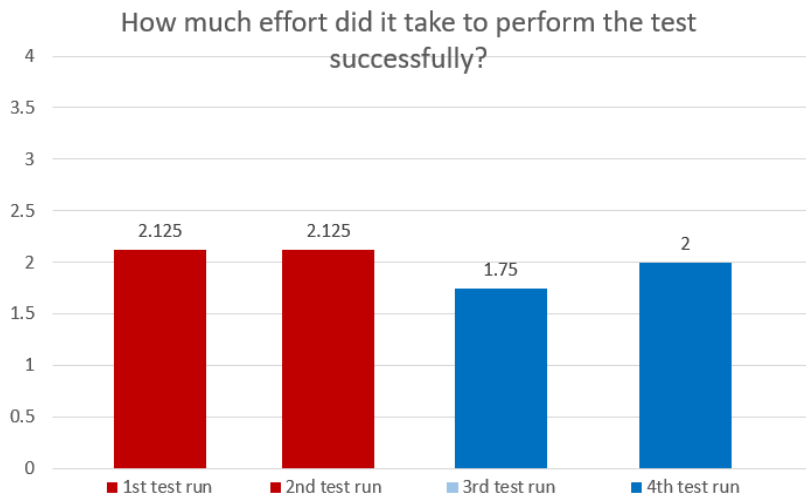


Figure 29. Mean score of the estimated effort required by each test setting based on the overall order of the tests (on a scale from 0 to 9).

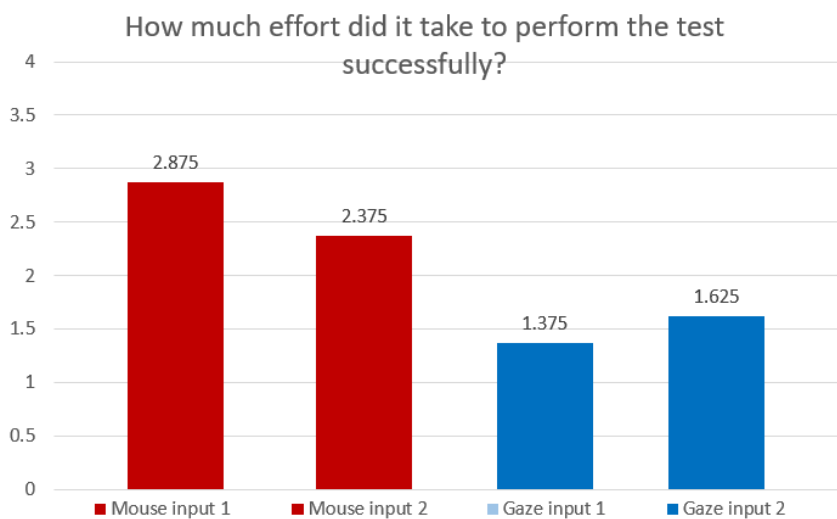


Figure 30. Mean score of the estimated effort required by each test setting based on the input method (on a scale from 0 to 9).

As can be seen in figure 29, there were minimal differences between the estimated effort in each of the overall test rounds. When it comes to differences between inputs (figure 30), the same trend that was visible in the physical and mental load continues here as well. In terms of effort, both mouse input test rounds received a higher mean scores than the gaze input test rounds. These differences did not turn out to be statistically significant and as with the physical and mental load, the participants evaluated effort with quite low scores in general.

The fourth attribute that the participant were asked to evaluate was the overall discomfort caused by the test. The participants were asked to evaluate the feelings of physical and mental discomfort as a whole, including the feelings of fatigue, stress and frustration. Similarly, as with the previous attributes, the evaluations were made on a scale from 0 to 9, where 0 stood for minimum discomfort and 9 for very high levels of discomfort. It was assumed that the mouse input test might cause somewhat more discomfort due to having to hold the eye still in the fixation point during the test

Did the test make you feel fatigued/uncomfortable/stressed/frustrated?

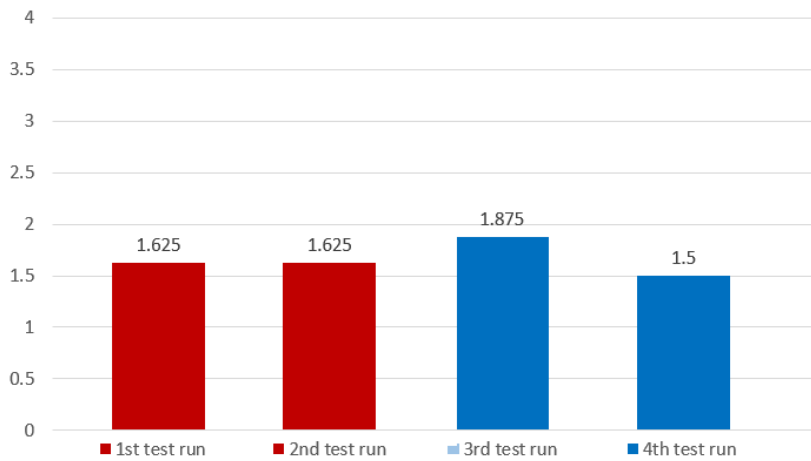


Figure 31. Mean score of evaluated discomfort caused by each test setting based on the overall order of the tests (on a scale from 0 to 9).

Did the test make you feel fatigued/uncomfortable/stressed/frustrated?

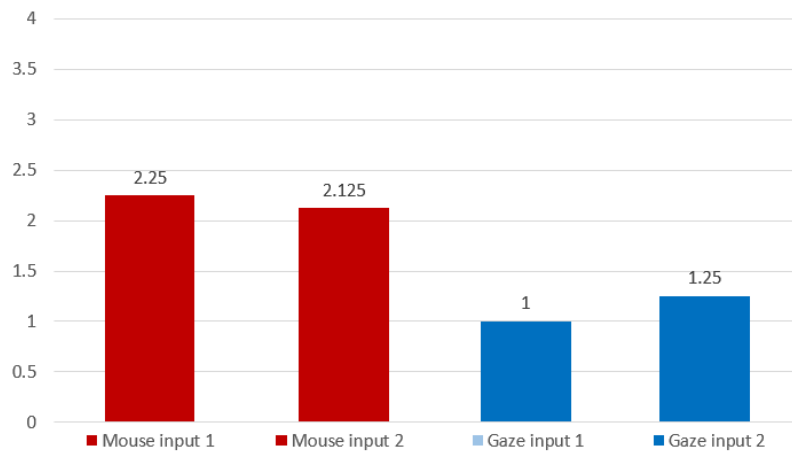


Figure 32. Mean score of evaluated discomfort caused by each test setting based on the input method (on a scale from 0 to 9).

As can be seen in figure 31, there were no large differences in the evaluated discomfort between the overall test runs. Similarly, as with the previously presented attributes, both of the mouse input test rounds were evaluated to have caused more discomfort than the gaze input test rounds (figure 32). The differences in the data did not turn out to be statistically significant and as with the other evaluated attributes, the participants did not give very high scores for discomfort in general, which could be interpreted as that they did not consider either of the inputs drastically discomforting.

The final attribute that was evaluated in between the tests, was the learnability. In this study's context, it referred to how easy the current test situation felt to learn and to understand in order to perform successfully. The participants were asked to evaluate the attribute on a scale from 0 to 9, where 0 stood for very low learnability and understandability and 9 stood for very high understandability and learnability. Before the tests an assumption was made that the gaze input test might be easier to learn since there is only one modality (the gaze input + visual output) to control instead of multiple modalities (mouse input + visual output). On the other hand, it was also acknowledged that the gaze input test could have its difficulties, especially at the start, as it is probably not as familiar modality as the mouse input is.

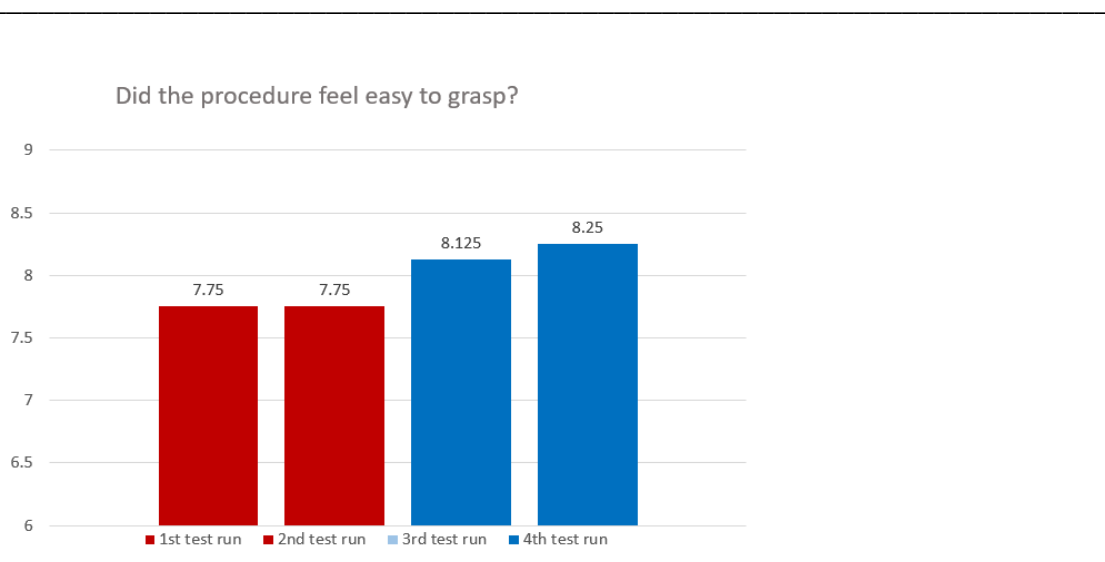


Figure 33. Mean score of the estimated learning effort required by each test setting based on the overall order of the tests (on a scale from 0 to 9).

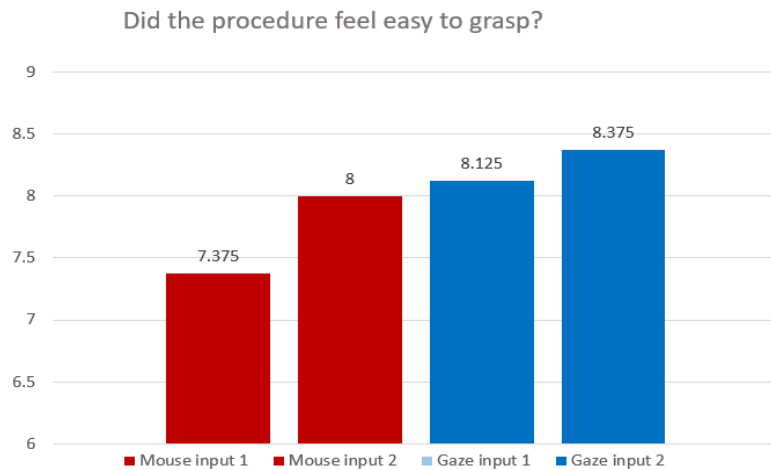


Figure 34. Mean score of the estimated learning effort required by each test setting based on the input method (on a scale from 0 to 9).

As can be seen in figure 33, the overall learnability improved with time as was expected. As can be seen in figure 34, both mouse input rounds were considered slightly less easy to learn and understand than the gaze input test rounds. These differences didn't turn out to be statistically significant either.

5.3 Feedback collected after the test

As was explained in section 4.7, the participants were asked to compare the user experience provided by the two different input methods after all four test runs were completed. The evaluation was performed so that the participants filled a comparative evaluation form (see appendix 5) and afterwards their answers were discussed together in a brief open interview. The evaluation form consisted of seven questions in which physical strain, mental strain, required performance effort, resulting discomfort, understandability, the feeling of naturalness, and overall preference of the two input methods were compared based on the participant's subjective retrospection of the test experience. To compare the different attributes, the participants were asked to select a suitable answer from three checkbox options ("Mouse input test", "Gaze input test" and "No preference") under each question statement, and if necessary they were allowed to elaborate their selection with a short open answer. Next, the collected results to all seven questions are presented in a chart (figure 35) and the open answers collected from the open questions and the interview are explained in more detail.

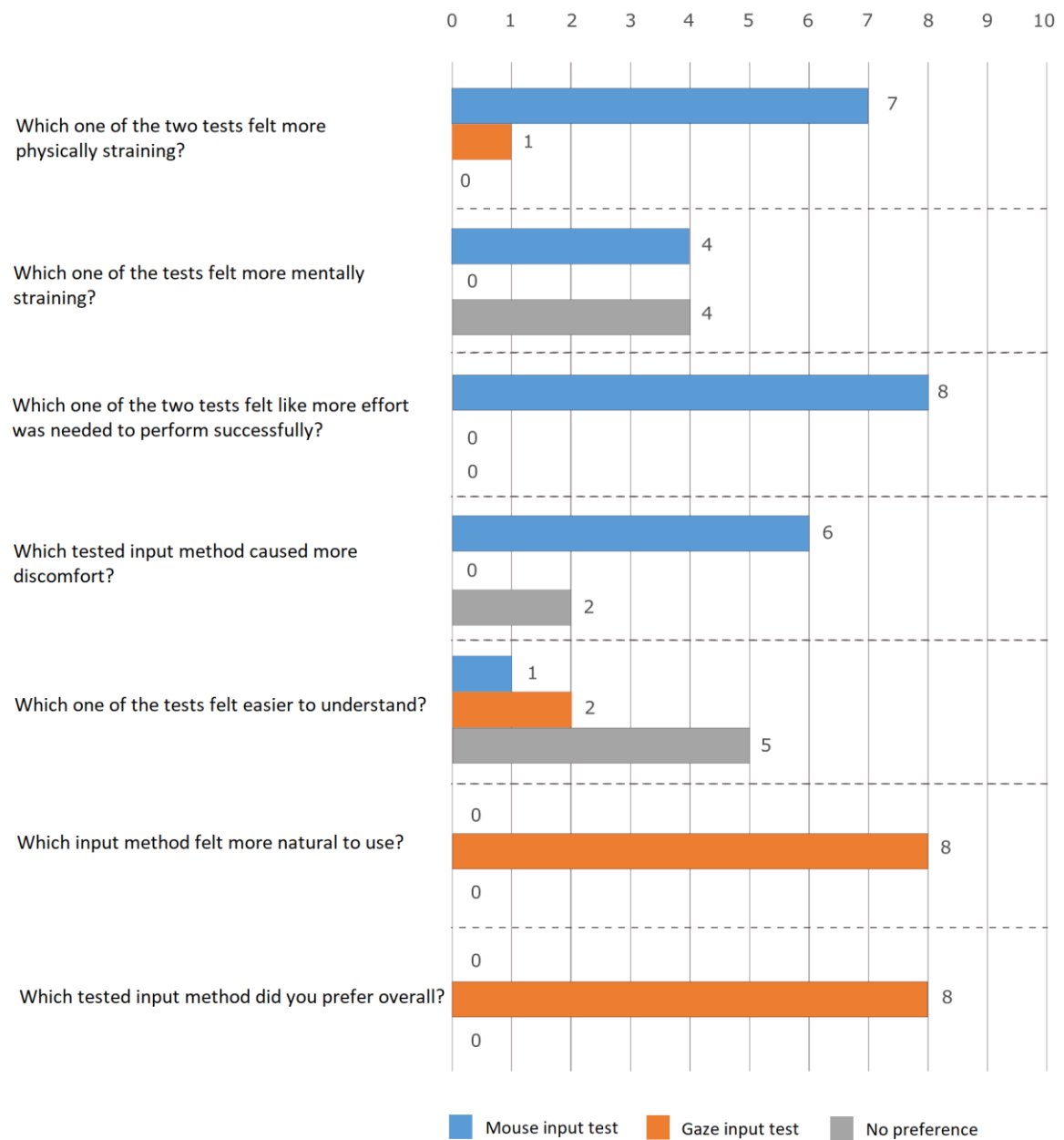


Figure 35. The distribution of participant’s answers to the questions that were presented in the evaluation form administered after the test.

As can be seen in figure 35, seven out of the eight participants considered the mouse input test more physically straining than the gaze input test, while one participant considered it the other way round. Holding the gaze still seemed to be the major source of physical strain in the mouse input test. Many of the participants reported that they had to struggle to some extent to hold their gaze still as stimuli appeared, as it required

them to battle the natural reflex of turning the gaze to the direction of the stimulus. Some participants also reported that in a long run, staring at the same spot caused disturbing visual after effects. The participant who considered that the gaze input test felt more physically straining of the two tests reported that in a long run repeating the same movement over and over again felt fatiguing to the eye. Overall, the participants did not consider either of the tests too physically straining and a few mentioned that they could have virtually selected either of the options to be the more physically straining one.

The participants' opinions about the mental strain caused by the test were equally distributed between the mouse input test and "no preference". Those with no preference considered that both tests were quite easy to accomplish without any mental strain. Similarly as in the first question, the participants who felt that the mouse input test was more mentally straining felt that having to hold the eye still for an extended period was the cause of the strain as more concentrating was needed to hold the eye still. Several participants reported that the warning sound played after fixation loss in the mouse input test felt stressful and disturbing, suggesting that the feature needs iteration. In the current implementation, the warning sound feature had some technical flaws - for example it would not always turn off immediately as the fixation was restored.

The participants were unanimous that the mouse input test was the more effort-requiring of the two tests. The participants explained their selections with similar arguments as the ones that they explained the two previous questions with: extra concentration and effort was required to successfully hold the fixation still for the whole test's duration. A few of them mentioned that apart from the strain caused by the forced fixation, using the mouse as an input device did feel effortless per se, and that neither of the tests felt excessively difficult to perform. One of the participant felt that as their eye began to feel fatigued with time, especially the stimuli displayed further away from the fixation felt more difficult to detect.

Six of the participants considered that the mouse input test generally caused more discomfort than the gaze input test, while two participants did not have a preference. Again, holding the fixation was considered the major source of discomfort. Some of the participants also reported that while performing the mouse input test, they felt unsure whether their clicks had actually been recorded, while the gaze input test did not cause

such uncertainty. This was a curious remark, as the software did not actually give visual or auditory feedback in either of the tests when a stimulus was successfully detected, and the stimuli overall behaved identically in both tests. The participants elaborated that they usually expect an instant reaction and tangible feedback when controlling a user interface with a mouse, and not having such feedback made them feel unsure. Meanwhile, the gaze input did not cause a similar effect. Further studying would be required to explain what happened, as the participants were not single-handedly able to explain whether they had different expectations towards the two interaction techniques, whether the gaze input was just simply slower to apply than the mouse input so that the disappearing of the stimulus felt immediate, or whether there was something else that caused the difference.

The participants had varying opinions about which one of the tests felt easier to understand. Mostly, the participants thought that both tests were easy to understand with the provided instructions – five out of the eight participants did not have a preference on which one of the test was more understandable. The two participants who considered the gaze input more understandable complimented it's naturalness. One of them also mentioned that in the mouse input test, it would have felt more intuitive to move the mouse to the direction of the displayed stimuli, or even click on them, which confused them at first. One participant considered that the mouse input test felt easier to understand due to the mouse's familiarity as an input device.

The test participants were unanimous on which one of tested inputs felt more natural to use. All eight reported that the gaze input felt more natural. The participants mostly backed this selection up with similar arguments that had already come up before: it felt more natural to follow the natural reflex and move the gaze towards the direction of a detected stimulus, than to forcibly hold it still. One of the participants thought that a combination of the two inputs might have felt most natural to them, since the physical sound of the mouse click would have provided them better feedback on successful interaction. Some of the participants did mention that while the gaze was more comfortable to use in this particular task, the mouse is overall more familiar as an input device, while the gaze input is not, which could possibly result into challenges with less tech-savvy user groups.

All of the participants considered that the gaze input felt more preferable out of the two tested methods, mostly due to the overall naturalness. Some other arguments that had already been mentioned in the previous answers also came up in connection with this final question. One of the participants stated that the gaze input felt overall better because it used a single main modality for the output and input instead of two. Some of the participants thought that the gaze input also had some novelty value, and it was "more fun" to use compared to the traditional mouse input.

To summarize the answers, it seemed that the gaze input test was overall less straining to do, and that the participants preferred it over the mouse input test. Several factors should be taken into account again, when reading into the results. First of all, most of the participant's answers were not that strict, and in many situations they explained that they did not feel that the differences between the two interaction techniques were that big. The test runs were relatively short, so the participants did not feel that either of the tests would have felt intensely straining. Had the test durations been somewhat longer, the differences might have appeared more steep. Secondly, it should be noted that the participants of this study, as described in section 4.6, were all healthy, working-age, and relatively familiar with technology and laboratory testing, which probably had an impact to their answers. None of the participants struggled mentionably with the test, and they were not stressed by the overall test situation. If a more representative group of participants had taken the tests, the answers could have appeared different. Thus, while the gaze input seems to have potential, further testing is required and these results should be regarded as suggestive rather than definitive. The implications of the test results are discussed further in the next chapter.

6 DISCUSSION

The user test results presented in the previous chapters suggest that gaze-based interaction could have certain user experience benefits. While the results promisingly seemed to be in favour of the gaze input, any definitive interpretations should not be made at this point, as this study was the first one on this subject. Much more future research and development work, hopefully inspired by this study, will still be required to get deeper into this subject. In this chapter, the implications of this study's results and the used research methods are discussed and reflected. The additional goal of this chapter is to provide suggestions and possible next steps for future study and software development.

6.1 Implications and interpretation of the test results

While the test participants did not consider either of the tested interaction techniques too difficult or uncomfortable to use, they generally gave slightly more favourable evaluations to the gaze-based interaction. The participants considered that compared to the mouse interaction, the gaze interaction felt more natural to use and that the mouse input test required more concentration and effort due to the unnatural requirement of having to hold the fixation in one spot for a prolonged time. Between test settings, the participants were asked to give score to five different attributes during the tests, and after analysis, the average score for mental load turned out to have some differences between the two inputs that showed statistical significance, in favour of the gaze input, as was presented in chapter 5. The statistical differences were especially mentionworthy between the first mouse test round and the first gaze test round. Also, as the participants were asked to compare and evaluate the two interaction techniques freely after the tests, the gaze input test turned out to be a cautious crowd favourite. When reflecting the test application and the test results against the recognized user experience challenges that were presented in section 3.9, the gaze-based interaction seems to have potential in tackling some of the issues. The most apparent benefit is that the gaze-based interaction does not require the holding of the fixation, which can be difficult for some people. The participants also evaluated that the gaze input test caused slightly less fatigue than the mouse input test. While this sounds promising, it should be noticed that the participants only took two relatively short test runs with each input. Based on these results, it's not

possible to deduce which one of the inputs methods will be more fatiguing or uncomfortable in a longer run. This should be studied, as longer test situations might be necessary for a thorough mapping of the visual field. The results of this study do not provide answers to whether the gaze interaction serves varying user groups better than the more traditional mouse input, as the group of test participants was quite homogenous in terms of their testing experience and health. Based on the participant's subjective evaluations, the tests gave some promising results about the gaze-based test's learnability, although any definitive conclusions cannot be made without further research. As gaze-based interaction is yet unknown for the vast majority of people, to better understand which input method is easier to learn, it should be tested with different, and especially with more challenging user groups than the one that was tested for this study, and if possible, rather objectively than subjectively. While it is difficult to find objective measures for some of the tested attributes, e.g. for the mental load which is heavily experience-based, it should be possible to measure the learning curve by more objective means. Of course, the subjective measures are important in terms of user experience, but for a medical test application, such as this one, it is important to acquire objective measures whenever it is possible. As traditional visual field tests also require practice from the patient's side to be accurate, comparing the learning curves of traditional devices and the interaction technique presented in this study could be worth the try.

When evaluating the results of this study, it is important to examine the test settings that were used to understand what implications the results have. As was mentioned previously, the group of participants that was recruited for this study was, while not unrealistic, definitely not inclusive. The test participants were picked based on convenience reason, and thus the sample might not be representative of the real user groups of this kind of an application. The answers present the subjective experience of a group that consisted of relatively healthy-eyed and computer-savvy people who were familiar with eye-tracking on at least on a level of concept. While these participant might represent a small sub-group of potential users of the visual field test application, they definitely do not represent the entire pool of potential users. Had the participants represented different types of user groups, the results might have turned out different.

This study did not go very deeply into charting the realistic user groups and their special requirements. In order to build a successful test application, it would be

important to chart the intended patient user groups and their specific needs with the help of literature and subject matter experts, and then iterate the software accordingly. When evaluating the effect of the selection of participants, it should also be noticed that the test situation and the motivation of the participants of this study were very different than in an actual test situation in a clinical setup. While laboratory tests probably simulated a clinical environment reasonably, the participants were not the subjects of the test similarly as they would be in a real life situation, and their mood and motivations were likely to be different than those of an actual patient. In this study, the participants entered the test situation relatively relaxed and with an open mind, and their most obvious motivation was to be helpful with evaluating an experimental software. In a real life situation, the test situation might be a lot more anxiety-inducing, as the patients are the ones under scrutiny, and the test result could lead to a diagnosis of an illness, the loss of a driving licence or other undesired consequences. This whole nature of the test situation could have also had an effect to the participant's performance and evaluation. While this study mostly focused on testing a test software, it is important to keep in mind that the user experience is a sum of all factors affecting the test situation, including external non-software factors and the participant's inner mental state. Also, due to the small sample size ($n = 8$) of this study, the overall results and the results of the statistical analysis, while promising, should be interpreted cautiously. With a larger sample size the results of the statistical analysis might have turned out differently.

Based on the tests, the 10-stepped scale used in evaluation form 1 worked adequately for this study, although further experimenting with suitable user experience measurement scales will definitely be beneficial in future research. In this study, the original NASA's 21-stepped task load index scale was toned down to 10 steps to ease the filling of the form and to simplify the analysis. It turned out to be a good thing that the scale was not simplified even further into a 5-stepped Likert scale, as the participants' answers mostly revolved around the lower end of the scale - a smaller scale would not have necessarily caught differences between test as well as the 10-stepped scale did. In retrospect, it would have been a good idea to flip the question on learnability the other way round to resemble the previous questions, so that the "positive" options would have been at the lower end of the scale and vice versa. Some of the test participant got slightly confused as they first expected that the lower number options represented the positive options as they did in the previous questions.

6.2 Suggestions for future research

As no proper previous studies on the subject of using gaze input for visual field testing were available, this study should be viewed as the first experimental steps into a subject that could and should be studied further. In short, this study only scraped the surface of a huge amount of work that is required to build a clinically valid test software. In the future, the results of this study should be replicated with a more versatile user groups. Also, statistical data of proper volume should be collected and analyzed in order to make any definitive conclusions on the user experience value of gaze control of the visual field testing software. In the future, the interaction techniques should also be evaluated with more complex test tasks. For example, the results of this study do not tell anything about the potential of the test application in actually discovering visual field deficits, as all of the participants had a healthy field of vision. While at some point it will be necessary to test the system with participants that have actual visual field issues, it is also possible to simulate visual field deficits with specifically designed goggles, which can make the test setting easier to design and arrange. Another option is to simulate a visual field flaw with the software itself by not displaying stimuli in certain locations of the screen.

To better understand the potential of an application such as the one that was designed and tested in this study, the professional users' needs should also be researched further in the future. The application was not even intended to compete with some of the current "gold standard" testing devices like the Humphrey field analyzer (presented in section 3.5) that are used for highly detailed diagnostics. Instead, being affordable and portable, a setup and an application similar to what was designed for this study could be used as a cheap and lightweight preliminary tool for detecting the existence of potential visual field defects. The system could be used in facilities where an investment to larger equipment is not desired or possible – it would be affordable and space-saving in more general settings than in a specialized optometrist's office as well. It should be noted that these suggested uses are only educated speculations based on the literature review made for this study, and further studying is required to better understand the context of use of such system. Overall, co-operation with subject matter experts and the charting of their needs and operative environment were out of the scope of this study. In future studies, professional optometrists should be included in the development to actually chart their

testing equipment-related user needs, and to find out whether a system like the one presented in this study would bring additional value to their daily work in their own operative environment. Any necessary analytic features can be added in the future, and in case the potential professional user group turns out to be nurses and general practitioner instead of specialized optometrists, it is also possible to design and create an iteration of the software that is highly guiding and easy to administer. The first thing to do is to just chart the user needs first.

In this study, aspects related to the application's accuracy, reliability and validity were not studied in depth. While user experience is also important to medical applications, the aforementioned attributes are mandatory. In future studies it would be important to study the accuracy and reliability related to the gaze-based interaction in more detail by recreating test data sets and a test environment that is as close to a realistic setting as possible. The method that was used in this study to validate that the test was performed successfully was quite a crude implementation of blind spot detection. For more accurate validation in future studies, more accurate validation methods should be revised from literature and implemented to the next iterations of the test software.

6.3 Suggestions for future software development

In addition to future research suggestions, the results of this study helped generate some ideas on what the future development steps could be. Based on the lessons learned from this study, to improve the experience of the person who is taking the test, the following improvements and new features could be included in the next version on the application:

- As it is obvious that a learning curve exists with visual field test applications, it might be beneficial to include *a short tutorial dataset* to be displayed before the actual test is performed – while oral instructions worked reasonably, hands-on experience would work even better. Optionally, the participants could be allowed to perform one or more non-recorded test runs before the actual tests.
- In the mouse input test, *randomizing the duration between stimuli* is an easy-to-implement future feature that could help improve the test's accuracy. While none of the participants in this study admitted to recognizing a rhythm between the appearance of the stimuli, it is possible that for some individuals it could

cause false positive results. It is quite unlikely that this risk exists with the gaze input test though, since the participant also needs to be aware of the stimulus' location in order to signal a positive detection. With the gaze input it is also possible to monitor the route the participant's eye travels to the target, which helps to define whether they actually saw the target or whether they sought for the target.

- In the gaze input test, some kind of *visual feedback for successful detection* would improve the user experience. While the participants did get the (false) feeling that the stimuli disappeared immediately as they looked at it, they would have wanted actual confirmation on successful interaction. This could be implemented with an animation or e.g. by changing the colour of the stimulus.
- To keep the algorithm for displaying stimuli simple, in this test application, they were always displayed in relation to the fixation point in the middle of the screening. By developing a slightly more complex *algorithm that calculates the location of the next stimulus based on the point the user is currently looking at*, the need for returning to the fixation point could be removed. Even though, based on the results of this study, this might not be problem free: it turned out that the participants saw stimuli that were close to the screen's edges less accurately – the more free of distracting elements the field of vision is, the easier it is to perform the test accurately. Also, not having a steady point to fixate to from time to time could cause more wandering of the eye which could cause more inaccurate answers. Perhaps this could be a feature that could be tested in a future study.
- In this study, the participant was required to hold their head still in one place in order to secure the correct viewing angle and the distance to the screen. As the eye tracker also provides the z coordinate, it would be easy to implement a feature that calculates *the angle and the location of the stimuli also based on the user's distance* from the screen. It should be notified that this does not allow the participants to move totally freely, as there are limits to the distance from which the eye tracker functions, and there are also limits to monitor sizes.

In addition to the aforementioned improvements, in order to improve the test's accuracy, testing capabilities and the experience of the person administering the test, the following new features and improvements could be implemented in the next iteration of the test application:

- First of all, for convenience reasons, the test data set consisting of the stimuli's locations and appearance was very simplified in this study. In the future, *more complex test settings* could be created with relatively simple additions to the program. For example, in this study, all of the stimuli had the same size and brightness, which only allows simple qualitative visual field testing (see section 3.4.). It would be quite simple create a test program that also allows qualitative testing. When designing such test program, some work needs to be done to calculate which colour contrasts match the differences in brightness that are utilized in traditional white on white perimetry. As stimuli cannot be displayed by manipulating the computer monitors' brightness in certain locations of the viewed area, the manipulatable features will be the colour values of the background and the stimuli, instead of the brightness of the stimulus, which is used for example in the Humphrey visual field analyzer. [Werner, 1991]
- It is possible that the computer monitor could be used *to chart even larger areas of the visual field* if the fixation point was placed closer to the monitor's edges instead of only placing it in the middle. Research is needed to find out how the eye tracker reacts to such positioning of the user, and to find out whether the edges of the screen cause unwanted artifacts to the field of vision. If the screen edges do cause issues, the physical test setting could be manipulated to see if a properly darkened room helps.
- Also to enhance the application's reliability, a *retest* feature would be a good addition in the future. At it's simplest, the retest feature could work so that after the participant does not successfully signal stimulus detection, the same stimulus would be displayed again (either right away or even better, after a while). In this case, only if the detection was missing on the second round as well, the stimulus would be marked as unseen in the results.
- During the user tests, it was acknowledged that technically the participants could have taken advantage of the stimulus' display time to scan at least part of the screen to see if there was a stimulus somewhere. In future iterations, *the eye tracking data could also be used to check how much the participant's gaze wanders* during the test by recording the distance it travels before stimulus detection. Some design work will be required to define things like the maximum allowed wandering distance and the consequences that happen when it is exceeded.

- *Analysis algorithms* were left out of the scope of this study. All in all, many such well-researched algorithms exist, so in the future it would be a good idea to review them and possibly apply one of them to the application as well. The results in the test application created for this study were only displayed in text form, but *graphical feedback* would also be a good addition to ease the interpretation for the test's administrator and for the participant as well.

In this chapter, the implications of this study's results were discussed with a heavy emphasis on potential future research and development. Even though the gaze-based interaction showed some promise based on this study, there is still a plenty of work to do before it is full clinical potential can be defined and a truly usable test software developed and possibly released for market. There are many viewpoints from which new research and development on this topic could be continued based on this thesis study, and it is likely that the ideas presented in this chapter cannot even capture the majority of them. Hopefully this study will provide grounds and inspiration for future research, including new interesting thesis topics for other students.

7 CONCLUSIONS

In this study, a visual field testing application with two available input methods, the novel gaze-based interaction and the more traditional mouse-based interaction, was designed, built and tested in laboratory settings. The goal of this study was to investigate whether the gaze input would be usable or even advantageous compared to the more traditional interaction in a visual field test. To get familiar with the principles and current practices of visual field testing, first, a literature review was conducted. Based on what was learned during the review, a test software was designed and built. One major design goal was to create a prototype of a test software that could perhaps someday work as a cheap and portable tool for visual field testing as a preliminary screening device and in places where the traditional expensive and heavyweight test equipment are not affordable or otherwise worth the investment. A series of user tests were organized to evaluate the software and overall eight people with healthy vision participated to the tests. During the tests, the participant took overall four test rounds with the software, two with each interaction technique and two with each eye. The participants were asked to evaluate the user experience after each test run, and after the tests they also compared the interaction techniques against each other.

This study was motivated by and built around two main research questions:

1. Can gaze tracking be successfully used as an input method in a visual field test application built in a simple pc + monitor + eye tracker setup?
2. In terms of user experience or otherwise, does the gaze tracking -based visual field test application have advantages over a more traditional mouse-based test application?

Based on this study, the answer to the first research question is affirmative. Successful blind spot detection was used to determine whether the test worked or not. In these tests, on the first round of using the gaze input, the blind spot was successfully detected from 5/8 participants, and on the second round from 8/8 participants. This compared very well against the results with the mouse input, as the blind spot was detected successfully on both rounds of mouse input from 6/8 participants.

As an answer to the second question, based on this study's results, gaze-based interaction could have some advantages, although additional research is required to find out whether the results apply to other user groups. As the sample size for this study was very small, the results should also be validated with a much larger pool of participants so that proper statistical analysis and conclusions can be achieved. The participants of this study gave the gaze interaction generally better evaluations than they did to the mouse interaction, although they did consider that both of the interaction techniques felt relatively easy to use and learn. The participants especially complimented that the gaze interaction felt more natural to use, and that it required less concentration than the mouse interaction. The requirement of holding the fixation at one place for a prolonged duration made the mouse input feel somewhat more uncomfortable than the gaze interaction. The difference between the two interaction techniques' evaluated mental load, in the gaze interaction's favour, also showed statistical significance when tested with Repeated Measures ANOVA.

As this study was the first one on this subject, and as the test settings were quite lightweight for convenience reasons, any definitive conclusions cannot be made based on the test results. The main conclusion of this study is that a simple gaze-interaction based visual field testing system is possible to build, and user experience-wise it could have some advantages over more traditional interaction. Yet, a lot of additional research is required to validate the user-experience related findings and to study features that were left out of the scope of this study, including specific user needs and accuracy. In addition to the immediate test results, ideas and suggestions for future research and development, which will hopefully happen, were presented in the discussion section of this study.

REFERENCES

[Broadway, 2012] David C. Broadway, Visual field testing for glaucoma – a practical guide. *Comm Eye Health*. **25**, 79 (2012), 66-70.

[Esterman, 1982] B. Esterman, Functional Scoring of the Binocular Field. *Ophthalmology*. **89**, 11 (1982), 1226-1234.

[EyeWiki, 2017] EyeWiki. 2017. Eyewiki.aao.org/Standard_Automated_Perimetry (Accessed on 13.3.2017).

[Farlex Partner Medical Dictionary 1] Farlex Partner Medical Dictionary. S.v. "campimeter." Retrieved 6/3/2017 from <http://medical-dictionary.thefreedictionary.com/campimeter>.

[Farlex Partner Medical Dictionary 2] Farlex Partner Medical Dictionary. S.v. "perimetry." Retrieved 6/3/2017 from <http://medical-dictionary.thefreedictionary.com/perimetry>.

[Fortenbaugh et al., 2012] Francesca C. Fortenbaugh, Shradha Sanghvi, Michael A. Silver and Lynn C. Robertson, Exploring the edges of visual space: The influence of visual boundaries on peripheral localization. *Journal of Vision*. **12**, 19 (2012).

[Johnson et al., 2011] Chris A. Johnson, Michael Wall and H. Stanley Thompson, A History of Perimetry and Visual Field Testing. *Optometry and Vision Science*. **88**, 1 (2011), E8-E15.

[Keeffe and Charlton, 2007] J.E. Keeffe and J.L. Charlton, Visual Fields and Driving. *Clinical & Experimental Ophthalmology*. **35** (2007), 594-595.

[Martin, 2015] Elizabeth Martin, *Concise Medical Dictionary*. Oxford University Press. 2015.

[McKendrick, 2005] Allison M. McKendrick, Recent Developments in Perimetry: Test Stimuli and Procedures. *Clin Exp Optom*. **88**,2 (2005), 73-80.

[Murray et al., 2008] I. C. Murray, B. W. Brash, M. E. MacRae and R. A. Minns, An Automated Visual Field Assessment Technique for Children Utilising Eye Tracking. *Investigative Ophthalmology & Visual Science*. **49** (2008), 2543.

[Patel et al., 2015] Dipesh E. Patel, Philippa M. Cumberland, Bronwen C. Walters, Isabella Russell-Eggitt, Mario Cortina-Borja and Jugnoo S. Rahi, Study of Optioma Perimetric Testing in Children (OPTIC). *Ophthalmology*. **122**, 8 (2015), 1711-1717.

[Ramachandran, 1992] Vialyanur S. Ramachandran, Blind spots. *Scientific American*. **266**, 5 (1992), 86-91.

[Rowe et al., 2010] Fiona J. Rowe, Henri Sueke and Sandra D. Gawley, Comparison of Damato Campimetry and Humphrey Automated Perimetry Results in a Clinical Population. *Br J Ophthalmol*. **94**, 6 (2010), 757-762.

[Schmid et al., 2005] R. Schmid, H. Luedtke, B.J.Wilhelm and H. Wilhelm, Pupil Campimetry in Patients with Visual Field Loss. *European Journal of Neurology*. **12** (2005), 602-608.

[Spector, 1990] R.H. Spector, Visual Fields. In: Walker H.K., Hall W.D. and Hurst J.W., editors. *Clinical Methods: The History, Physical, and Laboratory Examinations*. 3rd edition. Boston: Butterworths. 1990. Chapter 116. Accessed through: <https://www.ncbi.nlm.nih.gov/books/NBK220>

[Sundaram, 2009] Venki Sundaram, *Oxford Speciality Training : Training in Ophthalmology*. OUP Oxford. 2009.

[Tafaj et al., 2011] Enkelajda Tafaj, Thumas C. Kubler, Jörg Peter, Wolfgang Rosenstiel, Martin Bogdan and Ulrich Schiefer, Vishnoo – An Open-Source Software for Vision Research. *24th International Symposium on Computer-Based Medical Systems (CBMS)*. 2011.

[Tobii Developer Zone, 2017] Tobii Developer Zone. 2017. <http://developer.tobii.com/documentation/> Accessed on 20.5.2017.

[Walsh, 2011] Thomas J. Walsh, *Visual Fields: Examination and Interpretation*. Oxford University Press. 2011.

[Wandell, 1995] Brian A. Wandell, *Foundations of Vision*. Sinauer Associates Inc. 1995.

[Weinreb et al., 2014] Robert N. Weinreb, Tin Aung and Felipe A. Medeiros, The Pathophysiology and Treatment of Glaucoma: A Review. *JAMA*, **311**, 8 (2014), 1901-1911.

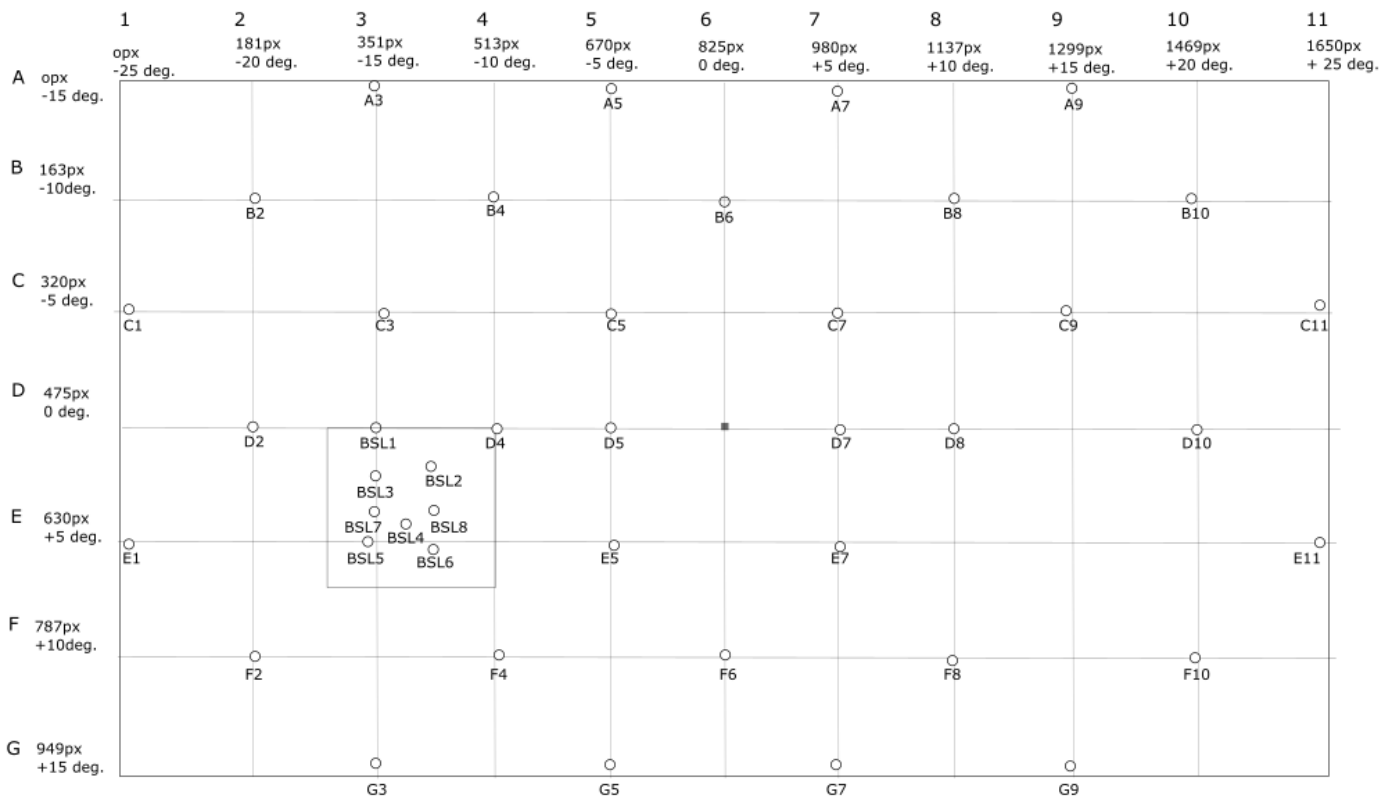
[Werner, 1991] Elliot B. Werner, *Manual of Visual Fields*. Churchill Livingstone Inc. 1991.

[Wroblewski et al., 2014] Dariusz Wroblewski, Brian A. Francis, Alfredo Sadun, Ghazal Vakili and Vikas Chopra, Testing of a Visual Field with Virtual Reality Goggles in Manual and Visual Grasp Modes. *BioMed Research International*, (2014).

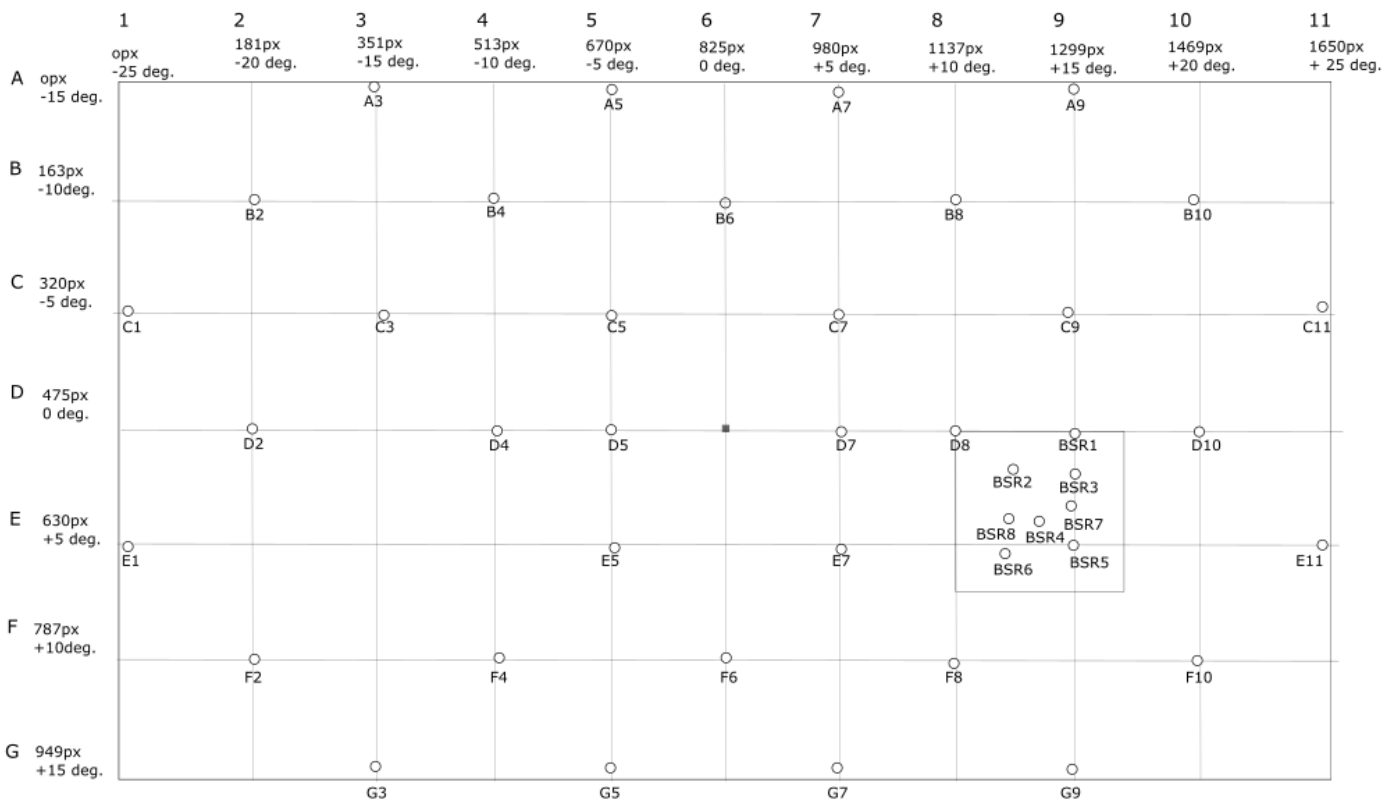
Directive 2006/126/EC of the European Parliament and of the Council of 20 December 2006 on driving licences Accessed 19/11/2016: <http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32006L0126>

APPENDIX 1: STIMULI CHARTS

Left:



Right



APPENDIX 2: CONSENT FORM

CONSENT FORM

Participant: _____

What will happen in the test?

The participant will perform altogether four test runs with a visual field mapping software in order to evaluate two different interaction techniques. The test session will take approximately 30 minutes. While the goal in the long run is help in the development of diagnostic tools, at this point this particular system is not reliable. Any issues identified by the system might be caused by problems in the system rather than in the participant's visual field. The purpose is to study the interaction techniques in this context, not to provide a medical diagnosis.

Risks & participant rights

While there are no particular risks is participating to this study, some temporary discomfort or fatigue might be caused due to having to fixate one's gaze to one point for an extended period. If the participant wishes to stop at any point during the test, they are free to do so without any consequences.

I have read and understood this consent form. I understand that participation is voluntary and I can refuse to participate at any time without any consequences. By signing I agree to participate in this study.

Date & Place: _____

Signature: _____

Thank you – your participation is appreciated!

APPENDIX 3: BACKGROUND INFORMATION FORM

BACKGROUND INFORMATION

Participant: _____

The purpose of this questionnaire is to collect demographic background information of the participants of the test. The questionnaire will be handled so that the identity of the participants will be known only by the researcher. Individual participants cannot be recognized from the reports that may be published.

Check the suitable answers

Gender

Female

Male

Age : _____

Have you ever participated in a visual field test before?

Yes

No

.. If you answered "yes", describe in what kind of a visual field test?

Are you wearing contact lenses?

Yes

No

APPENDIX 4: EVALUATION FORM 1

Participant: _____

Test: 1 2

Eye: Left Right

[Circle your answer]

How physical demanding did the test feel?

Not at all Very demanding
0 1 2 3 4 5 6 7 8 9

How mentally demanding did the test feel?

Not at all Very demanding
0 1 2 3 4 5 6 7 8 9

How much effort did it take to perform the test successfully?

None at all A lot
0 1 2 3 4 5 6 7 8 9

Did the test make you feel fatigued/uncomfortable/stressed/frustrated?

Not at all A lot
0 1 2 3 4 5 6 7 8 9

Did the test procedure feel easy to grasp?

Not at all Very easy
0 1 2 3 4 5 6 7 8 9

APPENDIX 5: EVALUATION FORM 2

Participant: _____

[Check the suitable options]

Which one of the two tests felt more physically straining?

Mouse input Gaze input No preference

Why?

Which one of the two tests felt more mentally straining?

Mouse input Gaze input No preference

Why?

Which one of the two tests felt like more effort was needed to perform successfully?

Mouse input Gaze input No preference

Why?

Which tested input method caused more discomfort?

Mouse input Gaze input No preference

Why?

Which one of the two tests felt easier to understand?

Mouse input Gaze input No preference

Why?

Which input method felt more natural to use?

Mouse input Gaze input No preference

Why?

Which tested input method did you prefer overall?

Mouse input Gaze input No preference

Why?
