Chair for Computer Science 10 (Media Computing and Human-Computer Interaction)

RWTH AACHEN UNIVERSITY

Looking for Holograms? Viewing Behaviour in Viewport-limited Augmented Reality

Master's Thesis
submitted to the
Media Computing Group
Prof. Dr. Jan Borchers
Computer Science Department
RWTH Aachen University

by
Niklas Kaulitz

Thesis advisor:
Prof. Dr. Jan Borchers

Second examiner:
Prof. Dr. Torsten Kuhlen

Registration date: 30.01.2018
Submission date: 30.07.2018

Eidesstattliche Versicherung

______________________________ ______________________________

Name, Vorname Matrikelnummer

Ich versichere hiermit an Eides Statt, dass ich die vorliegende Arbeit/Bachelorarbeit/ Masterarbeit* mit dem Titel

____

____

____

selbständig und ohne unzulässige fremde Hilfe erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt. Für den Fall, dass die Arbeit zusätzlich auf einem Datenträger eingereicht wird, erkläre ich, dass die schriftliche und die elektronische Form vollständig übereinstimmen. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

______________________________ ______________________________

Ort, Datum Unterschrift

*Nichtzutreffendes bitte streichen

Belehrung:

§ 156 StGB: Falsche Versicherung an Eides Statt

Wer vor einer zur Abnahme einer Versicherung an Eides Statt zuständigen Behörde eine solche Versicherung falsch abgibt oder unter Berufung auf eine solche Versicherung falsch aussagt, wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft.

§ 161 StGB: Fahrlässiger Falscheid; fahrlässige falsche Versicherung an Eides Statt

(1) Wenn eine der in den §§ 154 bis 156 bezeichneten Handlungen aus Fahrlässigkeit begangen worden ist, so tritt Freiheitsstrafe bis zu einem Jahr oder Geldstrafe ein.

(2) Straflosigkeit tritt ein, wenn der Täter die falsche Angabe rechtzeitig berichtigt. Die Vorschriften des § 158 Abs. 2 und 3 gelten entsprechend.

Die vorstehende Belehrung habe ich zur Kenntnis genommen:

______________________________ ______________________________

Ort, Datum Unterschrift

Contents

List of Figures

Abstract

Augmented reality allows perception of and interaction with virtual objects in the real world. Especially with head mounted displays, with which a user sees the world through a glasses like augmented reality device, the augmentations become ubiquitous. This creates the opportunity for new ways of interacting with digital content. However, first we need to know how users interact with these systems to design pleasant user experiences.
The view behaviour of users is a core part of their interaction with an any system. Interacting with digital content which is embedded into the real world is bound to change the way users look at the world. We presume that the user's view behaviour is also influenced by the limited field of view of current generation's head mounted displays. In these devices, augmentations can only be seen in a limited area in the centre of a user's visual field.

In this thesis, we present a study in which we tested how an augmented reality headset changes the view behaviour. Participants were asked to search for objects in a designated area. They either wore an eye tracker or a head mounted augmented reality display. Afterwards, we compared where participants looked, how long they needed to complete the objective and different aspects of the viewing behaviour, such as height distribution of areas looked at and how fast they moved on average.

In general, we found that the head movement while wearing an augmented reality headset resembles the eye movement under normal conditions, although the movements were slower over all. Additionally, completing the search task took longer when participants wore an augmented reality headset.
As our study was designed to do fundamental research, these results can not be applied directly on augmented reality application design. However, our results show the existence of effects that can be explored further in future work.

Überblick

Erweiterte Realität ermöglicht es virtuelle Objekte in der echten Welt wahrzunehmen und mit ihnen zu interagieren. Besonders durch am Kopf befestigten Bildschirmen, die wie eine Brille getragen werden und durch die der Nutzer die Welt betrachtet, werden diese Erweiterungen allgegenwärtig. Dadurch entsteht die Möglichkeit vollständig neuer Interaktionstechniken mit digitale Inhalte. Allerdings muss man erst verstehen wie Nutzer mit solchen Geräten interagieren um dann Anwendungen mit einer angenehmen Nutzererfahrung zu entwickeln.
Wie sich Nutzer verhalten ist Kernbestandteil der Interaktion mit jedem System. Mit digitalen Inhalten zu interagieren, die in die echte Welt eingebunden sind, verändert zwangsläufig das Blickverhalten der Nutzer. Zusätzlich erwarten wir, dass die Einschränkung des Sichtfeldes, die heutige am Kopf befestigte Bildschirme für erweiterte Realität aufweisen, das Sichtverhalten beeinflusst. Bei diesen Geräten können die Erweiterungen nur in einem begrenzten Bereich des Sichtfeldes gesehen werden.

In dieser Arbeit präsentieren wir eine Studie, in der wir getestet haben wie ein am Kopf befestigter Bildschirm für erweiterte Realität das Sichtverhalten beeinflusst. Wir haben Teilnehmer gebeten in einem Bereich nach Objekten zu suchen. Dabei haben sie entweder einen Eyetracker oder einen am Kopf befestigten Bildschirm getragen. Dann haben wir verglichen wo die Teilnehmer hingeschaut haben, wie lange sie brauchten um die Objekte zu finden, in welche Höhen die Teilnehmer schauten und wie schnell sie sich dabei bewegten.

Wir haben herausgefunden, dass die Teilnehmer in der Testbedingung erweiterte Realität ihren Kopf so bewegen, wie sie normalerweise ihre Augen bewegen, jedoch langsamer. Zusätzlich haben sie länger gebraucht um die Aufgabe zu beenden.
Unsere Studie war auf Grundlagenforschung ausgelegt und die Ergebnisse können nicht direkt für die Entwicklung von Anwendungen verwendet werden. Jedoch haben wir die Grundlagen für zukünftige Forschung gelegt.

Conventions

Throughout this thesis we use the following conventions:

Box plots used in this thesis display the median, or second quartile, of a data set as a thick black line and first quartile ($Q1$) and third quartile ($Q3$) as upper and lower boundary of the box. The three quartiles ($Q1$, $Q2$ and $Q3$) are the values that split a set of data into four parts with the same number of values. The end of the whiskers, the dotted lines above and below the box, display the corresponding quartile ($Q1$ or $Q3$) plus $1.5xIQR$, with IQR being the interquartile range ($Q3-Q1$). If the the whisker would be outside the range of the data it is capped at the highest or lowest data point. Dots outside the whiskers display outlier values.

The whole thesis is written in British English.

Chapter 1

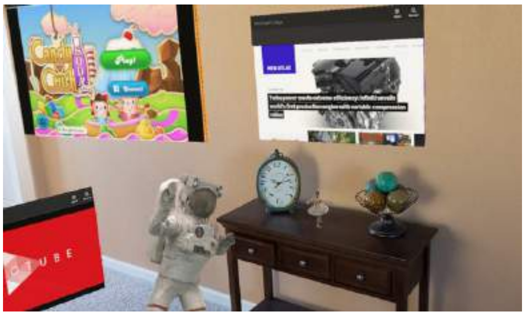
Introduction

Augmented reality, the process of embedding digital content in the real world, changes the way we interact with computers. Especially, when we see the whole world through an augmented reality device our surroundings become part of our interaction with the computer. In the case of glasses like devices, also called head mounted displays, a screen replaces our view of the real world and we can see digital content, also called augmentations, in a hands free manner. The move from isolated two dimensional windows into the digital world to an always present three dimensional representation, that can interact with the real world surrounding us, opens up new possibilities for interaction with computers and experiencing the digital world we live in.

Augmented reality creates new possibilities for interaction with digital content.

We are interested in how head mounted augmented reality displays change the behaviour of the users. Humans are used to interacting with the real world, but artificially placed intangible augmentations are something new. We presume that these new aspects will change the way we act compared to the way we do in an unaugmented world. Insights into these behavioural shifts can help to create more natural interaction techniques. Techniques which are easy to learn and do not feel intrusive.

Our theory is that head mounted displays change the behaviour of users.



(a) a scene with augmentations

(b) same scene with limited field of view

Figure 1.1: These images show what a limited field of view looks like. In a) a sample scene with some augmentations is shown and in b) one can see how this would look like with an augmented reality device with limited field of view. [https://newatlas.com/hololens-fov-field-of-view-illustrated/44903/ (28.07.2018)]

We conducted a study focusing on the view behaviour in augmented reality settings.

The part of the behaviour we are mainly interested in is the viewing behaviour. The way people look at the world (when, how much and where) is one of the most basic parts of interaction. Thus, it has implications on almost all areas of an application's experience.
We conducted a study in which we compared the viewing behaviour of participants in a setting with augmentations to a baseline we recorded in a setting in which no augmentations were shown. The goal of this study was to find out how exactly the viewing behaviour changes.

Head mounted displays have a limited field of view, which might change people's view behaviour.

Current head mounted displays come with technical limitations. One of which is a limited area in which the display can show augmentations. This leads to the users seeing digital contents only in the centre of their visual field, while the periphery consist of the unaugmented real world. This effect is also called a limited viewport or, more commonly, a limited field of view. Figure 1.1 shows a side by side comparison in a mock-up scenario. While some augmentations are cut off, others cannot be seen at all.
Because of the unfamiliarity of augmentations and the limited field of view we expect the viewing behaviour of users of head mounted display to change.

Insights into where people look with their field of view will help to place digital content in a way that they will be seen. Augmentations need to be placed in a way that they are easily found, because the peripheral vision, the area towards the edges of one's field of view, cannot be used to look for them.

A person's gaze direction, the direction in which the person is looking, is a good measure for their spatial attention [Hoang Duc et al., 2008]. Knowing on what point in space a user focuses is a useful context information for applications. If a person wearing a head mounted display with limited field of view moves their head like they would usually move their eyes, the head direction, which is easier to measure, could be used as a simple approximation of a user's spatial attention.

We expect that knowledge of the view behaviour can be used for augmentation placement or attention approximation.

1.1 Structure of the Thesis

The structure of this thesis is as follows: In chapter 2, we present previous research. Among the presented work, we outline work on the effects of a limited field of view in augmented reality headsets [Ens et al., 2016, Kishishita et al., 2014, Ren et al., 2016]. Others have researched gaze and spatial attention under different conditions [Castelhano et al., 2009, Mills et al., 2011, Risko and Kingstone, 2011, Yarbus, 1967] and some have already combined the two and analysed gaze and attention in augmented reality scenarios [Vidal et al., 2014].

Following that, chapter 3 gives a collection of fundamentals. Aside from topic related issues, such as the visual system, attention and augmented reality in general, we detail some aspects of tracking technology, such as motion capture, head tracking and eye tracking.

After that, we present our conducted experiments in chapter 4. First, we discuss our preliminary study, in section 4.1, which consisted of a questionnaire with the goal to collect information about people's offices. We used the results to inform the design an office like room for our main study.
Second, we present our main study in section 4.2. We dis-

cuss each aspect of our study, such as the hypotheses, task, room, technology, variables and implementation, in detail.

Then, we describe the evaluation procedure and the results in chapter 5. For the evaluation, we present our preprocessing and data filtering steps and the evaluation of our data. Following that, we discuss the results and give our interpretation.

Finally, in chapter 6, we summarise our work and give an outlook into our ideas for future experiments.

Chapter 2

Related work

In this chapter we will present the work of other researchers in the field of head mounted augmented reality systems as well as different gaze behaviour and attention shifts. In section 2.1 we discuss work which focuses on the influence of the limited field of view on performance and behaviour in augmented reality settings. Following that section 2.2 presents papers that focused on changes in gaze behaviour induced by different conditions and lastly section 2.3 is comprised of research combining the aspects of attention and augmented reality.

2.1 Limited Field of View Effects

Ens et al. [2016] focused on the aspect of limited field of view in current augmented reality headsets. They used a CAVE environment to simulate augmented reality with varying fields of view. In that context they compared two selection techniques for virtual objects: direct pointing, in which the subjects used their tracked finger to tap at a target on a virtual plane and raycasting where the subjects pointed at the intended target with a hand-held tracked device. They found that direct pointing was the faster method, especially for smaller objects, but the differences diminished with a smaller field of view.

Ens et al. [2016] compared two pointing techniques under varying fields of view.

Kishishita et al. [2014] compared two augmentation highlighting methods under varying fields of view.

Another analysis of a limited field of view's influence on tasks in an augmented reality setting was done by Kishishita et al. [2014]. Their subjects were asked to find highlighted virtual objects while also focusing on a real world puzzle. Additionally to the size of the field of view, they changed the labelling technique of the target to be found. The two techniques were in-view labelling, in which the label for a target was displayed in the field of view with a guiding line to the target, if it was outside of it, and in-situ labelling, in which the label was only displayed if the target was inside the field of view. They found that with increasing field of view the two methods became more and more similar but that there was no definitive increase in search performance in bigger fields of view.
Contrary to our approach they were not interested in how the view behaviour of a user was affected by the field of view. Only task dependent metrics were analysed.

Ren et al. [2016] tested the influence of the different field of view sizes on an information gathering task.

Also Ren et al. [2016] tested the influence of the size of the field of view on different tasks. Similar to Ens et al. [2016] they used a full surround virtual reality environment to simulate augmented reality with limited field of view. They used a scene of the Luxor Temple enriched with charts providing information about the statues. Additionally the statues and their corresponding information were linked with guidelines. Users were asked to retrieve different information for which they needed to follow the lines and collect the information in different charts and the scene in general. They found that a limited field of view increases the task completion time. Additionally they looked at the head movement of the participants and found that while the condition does not affect the average velocity of the subjects, they did look at different regions. In the limited field of view condition, more time was spent looking high up.
Again, in contrast to our study, they were not interested in the regions subjects looked at, but rather used task dependent metrics for analysis.

2.2 Gaze and Attention behaviour changes

A highly researched aspect, in the field of attention behaviour, is the influence of a viewing task on the gaze behaviour of people. One of the earliest and most influential studies was conducted by Yarbus [1967]. He showed images to people and asked them to look at these images with different tasks in mind. Among the various tasks were free examination, reporting the ages and remembering different aspects like clothes and positions of people in the images. He found that the task heavily influenced the viewing behaviour of his participants and reasoned for a top-down attention guiding model.

Different research groups studied the influence of different task formulations on participants' viewing behaviour.

Numerous groups have replicated and expanded on that idea. Such as Castelhano et al. [2009], who compared search and memorisation tasks. Due to modern eye tracking technology they were able to analyse measures, such as number of fixations and gaze durations on specific objects or saccade amplitudes. Their findings were similar to Yarbus' [1967] findings: the parameters regarding specific objects in the scene changed, while global metrics, like average fixation duration and saccade amplitudes, did not.
In a similar study Mills et al. [2011] analysed the development of these measures throughout the trial. They found that the number of eye movements as well as their distance was lower towards the end of the trial.
While we did take their findings into consideration when deciding on a task for our study, we did not compare different tasks and used other methods of analysis.

Risko and Kingstone [2011], on the other hand, analysed the change in a person's gaze behaviour when they felt observed. They conducted a study in which a participant, after conducting a fake trial, was left alone in a room with a swimsuit calender, as a provocative stimulus, among the decorative features. They measured how many participant looked at the calender with the conditions: participant wore an eye tracker, participant wore an eye tracker that was off and participants were observed by a hidden camera. They found that significantly less people looked at the calender while feeling observed.

Risko and Kingstone [2011] researched the influence of knowing that one is being observed on the viewing behaviour.

While our study has different conditions, we too are interested in the change of the view behaviour regarding different areas of interest.

2.3 Gaze and Attention in Augmented Reality

Vidal et al. [2014] tested different use cases of eye tracking data, as approximation for spatial attention, for augmented reality applications.

Vidal et al. [2014] analysed the use of eye tracking data for interaction in the context of a head mounted augmented reality display. They proposed some measures to determine if the user is looking at the real- or the virtual world and tested them with promising results. On top of that they presented some methods how an interface could stay up to date without drawing too much attention to the virtual world by using the effects of change blindness and the inability to see colours with the peripheral vision. In contrast to our work they are more interested in using eye- and head tracking methods for interaction purposes rather than using them to investigate how the augmented reality changes the behaviour of the users, like we do here.

Chapter 3

Background

In this chapter we present background information to the topics and technologies which are content of this thesis. The fundamentals regarding the visual system show how people look around and what influence the peripheral vision has on the viewing behaviour. Spatial attention models are another aspect influencing the viewing behaviour. After that we describe the technologies we used, such as augmented reality devices and various movement tracking methods, which we used to implement our study.

3.1 Vision

While the peripheral vision is worse than the central vision, it is important for deciding what to focus one's attention on.

The human vision is a broad topic and covering the whole process from light hitting the eye to understanding what was perceived is not in the scope of this thesis. But there are some aspects of vision, which are important for our work. One such aspect is, that while the human field of view is big (up to 180° width and 125° height [Kishishita et al., 2014]), not all of it is perceived with the same quality. The central 5° of the field of view are perceived by the fovea. Then up to 9° are handled by the parafovea and up to 17° the perifovea. These three combined are called the macula [Wandell and Thomas, 1997]. Everything outside of that, we call peripheral vision. Due to the distribution of the cells that detect light, called cones and rods, there is a smooth degradation of vision of colours and shapes from the centre of the retina to its edges. Anstis [1974] demonstrated this by measuring how big a letter needs to be that it can be read at different angles in the field of view. A model for visual search by Wolfe [1994] presumes that there are two patterns involved: one that interprets only coarse data gathered mainly in the peripheral vision and one that does high level analysis, like face detection or reading. Being a kind of preprocessing step in a new scene, the first pattern has a high influence for this kind of task and so does the peripheral vision.

3.2 Attention

Attention is focusing on specific pieces of information to be able to process the load of information the senses gather.

The amount of data which is gathered by the human senses is enormous. While the brain definitely is an astounding construct capable of complex processes, it is not able to analyse all the influences at a high level. To deal with this issue, we have the concept of attention. From the current mass of stimuli, a subset is taken and prioritised for processing [Chica et al., 2013]. But despite the amount of research dedicated to this topic [Chica et al., 2013, Duchowski, 2007, Hoang Duc et al., 2008, Yarbus, 1967], it is not yet fully understood how it works. Attention is a high level brain function and we do not have good sensors for recording mental processes yet. Attention is too vast of a field to be fully discussed here, that is why we focus on the subject of spatial attention which is of particular interest for us.

Spatial attention is the process of focusing on certain spatial areas for detailed processing.

Spatial Attention Spatial attention focuses on the way we process and focus on external stimuli. Specifically any form of perception which can be located in the world around us. Mainly this is what we see and where we look. There seem to be at least two mechanisms for spatial attention guiding [Chica et al., 2013].

The endogenous (also known as top-down or voluntary) system allows people to guide their focus of attention consciously towards aspects they are interested in. And the exogenous (also known as bottom-up or involuntary stimulus-driven attention) system which is triggered by external stimuli (e.g. movement, luminance changes, etc.) and pulls the attention of a person towards something. Both of these systems have their purpose. Without the endogenous attention, one would not be able to guide one's eyes towards task relevant objects and thus it would not be possible to gather specific information (e.g. read) or locate objects for interaction (e.g. tools). On the other hand without exogenous attention, one would not be able to react to changes of the environment which do not happen in the currently focused area (e.g. an appearing predator).
While it is well understood that these two systems exist, it is not quite clear how exactly they interact with each other. It seems to be the case that both are different systems in the brain with contrasting ways of information processing that sometimes interact with each other to determine which controls the resulting behaviour [Chica et al., 2013].

Spatial attention can be controlled voluntarily (endogenous) or involuntarily (exogenous).

A person's gaze direction is a good measure of where that person's spatial attention lies [Hoang Duc et al., 2008]. The gaze is the direction in which a person is looking. It can be measured by eye trackers or manual annotation by an examiner.

Spatial attention can be approximated by the person's gaze direction.

3.3 Augmented Reality

Augmented reality is the process of displaying digital content (then called augmentation) in a real world setting. It lies on the spectrum between reality, where everything is real, and virtual reality, where everything is virtual (e.g. video games). This area is called mixed reality continuum and is shown in Figure 3.2. As usually the majority of the scene is real and only some digital contents appear, it is closer to reality than to virtual reality [Poelman and van Krevelen, 2010]. The approach were real content is shown in a mostly virtual reality is called augmented virtuality. Augmented reality is much more common and thus the term mixed reality is often synonymous for it.

Augmented reality is the process of embedding digital content in the real world.

Figure 3.1: This shows an example application of augmented reality. In this scene the luxor temple is enhanced with historical information. [Ren et al., 2016]

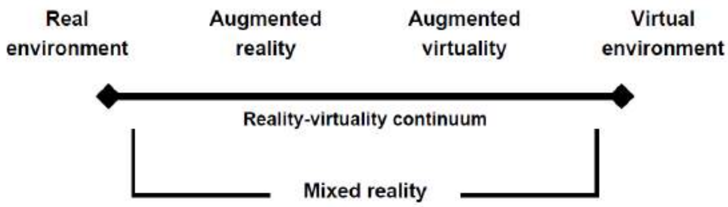


Figure 3.2: The mixed reality continuum. A graph displaying the relationship between different stages of the mixed reality.[Poelman and van Krevelen, 2010]

Display Technology

Video see-through systems show a video stream of the world, with augmentations already embedded. Optical see-through systems have a transparent display, which lays digital content over the real world.

There are two main ways to implement augmented reality: Video- and optical see-through. In the former case, one sees a monitor on which a life feed of the real world is displayed, which is then enriched with digital contents. Many augmented reality applications for mobile phones work that way. In optical see-through systems, on the other hand, one sees the real world as is and the digital contents are added in other ways. Most augmented reality headsets, like the Microsoft Hololens[1], employ this strategy. Usually the user looks through some kind of glasses which also function as a projection screen.

Video see-through systems have the advantage that the digital content can be embedded very well [Takagi et al., 2000]. The optical quality of content and real world can be

[1]https://docs.microsoft.com/de-de/windows/mixed-reality/hololens-hardware-details

matched and there is less relative movement between the two. On the other hand seeing the real world as a video is less pleasant for the users, especially because in most cases there is a certain delay between their movements and the movement of what the system displays. Optical see-through systems have the advantage of a more natural perception of the real world but the augmentations look less real [Zhou et al., 2008]. This is because, transparent displays can only add light, which leads to the augmentation, especially dark ones, being somewhat transparent. Additionally, the augmentations have the same delay all video see-through systems suffer from. The augmentations seem to be not quite attached to the world when the users move their head too fast.

Projective systems use the world as a screen and project augmentations on real objects.

Additionally, augmented reality can be done projective. In this case digital augmentations are projected onto real world surfaces [Poelman and van Krevelen, 2010]. Similar to optical see-through systems, this type offers a natural view of the surrounding. However, it is not possible to create floating hologram like augmentations and the systems are usually limited to indoor use because of the brightness of current projection technology [Poelman and van Krevelen, 2010].

Display Types

Augmented reality devices can also be distinguished by their size and where the display technology is placed. The three main categories here are hand-held, head-worn and spatial displays [Carmigniani et al., 2011, Poelman and van Krevelen, 2010]. As Figure 3.3 shows their main difference is size. All display technologies can be used for each type.

In hand-held a small movable display is used that augments the real world.

Hand-held augmented reality displays are, as the name suggests, displays which the user has in their hands and through which he can see an augmented reality. These are usually video see-through systems. A good example for this are mixed reality smartphone applications. The camera of the smartphone records the real world and then augmentations are displayed on the screen.

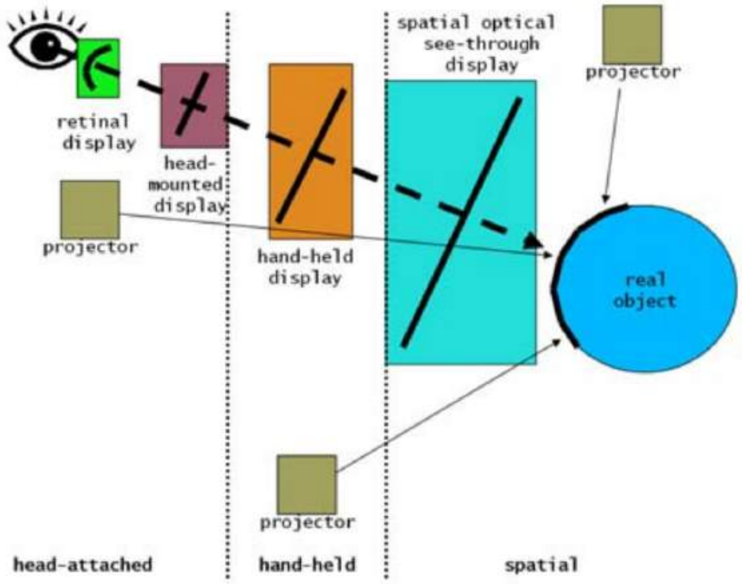


Figure 3.3: This graph displays the relationship between different augmented reality display types. Mainly how size and placement influences the name.[Bimber and Raskar, 2006]

Head mounted displays are glasses like devices in which a display is placed directly in front of the eyes.

Head-worn displays, also called head mounted displays, are devices worn similar to glasses but the lens is replaced with a display. This is the most immersive type of augmented reality, as the augmentations are embedded into the user's field of view. Additionally, the hands are free for interaction purposes.
All three display technologies can be used in this case. In video and optical see-through systems the user looks through the display at the real world, while in the projective systems the projector is placed on the user's head and displays the content on any surface in front of the user.

Spatial displays are locally fixed displays which are used to show some form of augmented reality. For instance like a mirror with digitally added accessories.

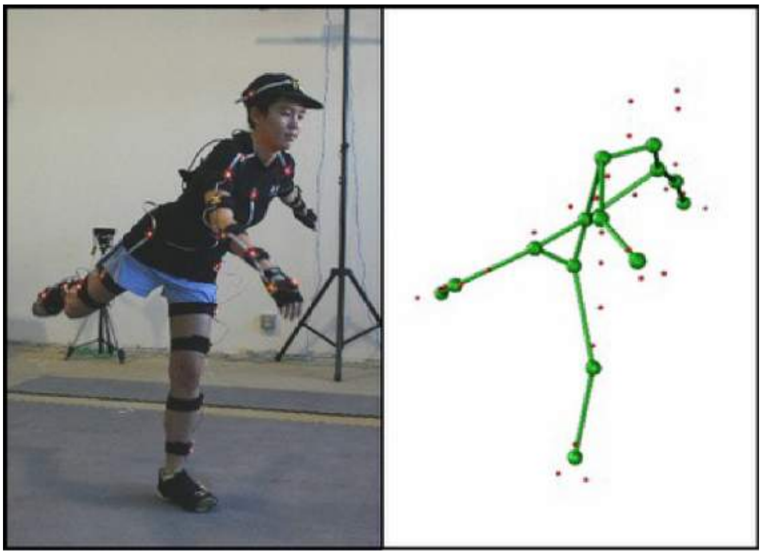


Figure 3.4: This picture shows a side by side comparison of an optical motion tracking set-up and its digital representation. [Öllerer, 2011]

3.4 Motion Capture

The goal of motion capture systems is to determine the position and orientation of objects in three dimensional space. This has multiple applications like transferring the movement of a real actor onto a computer generated replacement in films, observing the movement of subjects under different conditions in research or using the movement of a person to interact with computers. There are two ways to classify motion capture systems [Öllerer, 2011]. They can either be divided by the techniques used to implement the systems or classified by their operational parameters. The classification by operational parameters is based on the assumption that a system has an active part, which sends some kind of signal, and a passive part, which receives the signal and uses it for location purposes:

Motion capture systems record the three dimensional position of objects or people. These recordings can then be used to recreate the movements in a virtual environment.

- **outside-in**: The receivers are in the area surrounding the subject and the transmitters are located on the subject's body.

- **inside-out**: The receivers are on the subject's body and the transmitters surround the subject.
- **inside-in**: Both the receivers and senders are on the subject's body.

Outside-in and inside-out systems have the inherent disadvantage that they need a fixed set-up, called an arena, and thus are not mobile[Öllerer, 2011]. Inside-in systems usually lack the external reference points for absolute position measures and can be much more intrusive.

Aside from this classification they can be ordered in terms of implementation technologies.

Sensors which detect their own movement can measure the movement of objects.

Movement sensor based methods are usually inside-in systems. They use sensors which are attached to the subject's body and measure the movement of certain body parts. Their measurements have a high precision, but they lack an external reference point, which leads to increased errors over time [Öllerer, 2011].

Image processing software can be used to detect the position of objects or persons.

Computer vision based methods use image processing software on normal or three dimensional camera recordings to locate people and their position. These systems can not be classified the previous way, as they do not have a signal sending part. Many different approaches exist in this category. Summarizing these methods is beyond the scope of this work, but has been done in other papers [Cheng et al., 2015, Weinland et al., 2011].

Reflective markers can be used to detect their location by triangulation with multiple camera feeds.

In optical tracker based methods good visible markers are attached to objects or body parts, Cameras surrounding the arena detect these markers and triangulate their position in space. Usually the cameras and markers use infra-red light, as it is not distracting the users [Öllerer, 2011]. These systems have a low latency and a simple flexible set-up. Additionally, they are able to track all objects to which markers can be attached and multiple objects simultaneously. These systems are currently the most widespread.

Different approaches can be combined to cancel out disadvantages.

Additionally, there exist hybrid methods that combine different approaches. This is usually done with methods that counteract each others disadvantages. For instance, a movement based model could be used with a computer vision based model. The former would supply precision while the latter would provide the external reference point.

Head Tracking The term head tracking describes a group of approaches to measure the position and orientation of the head in space. As the head is a body part, this topic is a subsection of motion capture systems. All methods detailed previously can also be used for head tracking. In some cases, for instance computer vision based methods, different algorithms are used for head tracking.
In this scenario optical tracker based methods are not that common, as they are intrusive and lack mobility. Computer vision based methods [Al-Rahayfeh and Faezipour, 2013, Czupryński and Strupczewski, 2014, Murphy-Chutorian and Trivedi, 2009] are used in some applications as they are the least intrusive. In cases where intrusion is no issue, for instance when wearing a head mounted display, movement sensor based methods can be applied.
The external marker based approach is unique to head tracking. In them a camera is placed attached to the head that records the surrounding and tries to locate the head based on certain features or specially designed markers. The Dikablis Eye Tracker[2], for instance, uses this approach.

Head tracking is the process of detecting the position of people's heads in three dimensional space.

With head tracking, the head can be used as a form of input device, an approximation of the gaze direction and, in combination with an eye tracker, to compute the actual gaze direction [Al-Rahayfeh and Faezipour, 2013, Czupryński and Strupczewski, 2014, Murphy-Chutorian and Trivedi, 2009]. The gaze direction can in turn be used as an approximation of a person's focus of attention [Hoang Duc et al., 2008].

3.5 Eye Tracking

Eye tracking has similar applications as head tracking. Most notably, it can be used as an input method (e.g. for users with impaired movement) or as a measure of people's attention [Hoang Duc et al., 2008]. Additionally, many eye tracking methods compute the direction a person is looking relative to the head and need to be combined with head tracking to find the gaze direction in space.

Eye tracking is the detection of where a person looks. It is used for attention approximation or as an input method.

[2]https://www.ergoneers.com/mess-software-und-analyse-software/d-lab/head-tracking/

Figure 3.5: This picture shows an application example for a mobile head mounted eye tracker. This device records the eyes with two cameras placed directly in front of them and computes the gaze direction based on the location of the pupil. [https://www.ergoneers.com/en/newsroom/applications/ (27.07.2018)]

While the first eye tracking systems are fairly old, new methods are still in active research. A vast amount of different approaches has been developed over the years.

In the past, physical measurements (e.g. movement of muscles or a contact lens) were used to locate the eye.

Electro-OculoGraphy is the process of determining the eye direction by measuring the muscle movement around the eyes, which is done by electrodes placed on the skin. This was among the most used methods some time ago. However, these methods lack mobility and are rarely used today [Duchowski, 2007].

The methods with the most accurate results use contact lenses, which are equipped with mechanically or optically locatable features. These methods have a high degree of intrusion and discomfort and lack mobility as well [Duchowski, 2007, Young and Sheena, 1975]

Today's methods use image processing software to locate the eye.

Today most measures are video based. This means the eyes are filmed by one or two cameras and computer vision algorithms detect distinct features which they use to locate the eye and compute the gaze direction. This group of methods is huge, as the methods use different features as well as different algorithms. For further reading see [Al-Rahayfeh and Faezipour, 2013].

These methods can either be implemented as head mounted systems, in which the user wears a glasses like apparatus to which the cameras are attached, or as a table mounted system, in which the cameras are placed on a table in front of the subject. Head mounted systems are more robust and precise [Duchowski, 2007, Young and Sheena, 1975]. On the other hand, they require additional head tracking and are intrusive. With the combination of some features table mounted systems are able to do point of regard measurement, which means that they can compute the eye and its direction in space, without needing to know the head position.

Chapter 4

Experiments

Our research question was: How does the view behaviour of users wearing a head mounted augmented reality display with limited field of view change? The view behaviour is a core part of a user's interaction with the world. It defines where users look and how much time is spent studying certain areas. Due to the field of view limitation of current augmented reality headsets we expected this behaviour to change.

We conducted a study designed to answer this question. In this study participants were asked to search for items in an office like room under two different conditions. In one condition the items were real and the room was presented without any augmentations. In the other the participants wore an augmented reality headset and looked for objects which could be real or hologram like augmentations. We decided on an office like environment as we expect that to be a likely use case for augmented reality systems.

We conducted a study comparing the view behaviour between an augmented reality setting and normal conditions.

To determine how an office like room looks we conducted a preliminary survey. We asked people of different fields of work to answer a questionnaire regarding their office and used the results to design our main study.

In this chapter we will first present our preliminary survey and its results. After that we will present the study we conducted. The results of our main study will then be presented and discussed in chapter 5.

4.1 Preliminary Survey

We wanted to find out how an average office looks like.

We wanted our main study to take place in an office like environment. Tiller and Veitch [1995] detailed the rooms they used for their studies, but as the context of their study was an office stock we could not apply their results. We did not find any data which helped specify how an average office looks like. Thus, we conducted a small preliminary study, consisting of a short questionnaire, to answer this question. In the questionnaire we asked our participants various questions about their current office. The goal of this survey was mostly to give us rough guidelines to design the room for our main study, instead of providing statistically significant findings. Because of this, we do not present the study in detail. The full questionnaire can be seen in Appendix A

Questionnaire

Our questions regarded general working conditions, room layout, furnishing, workstation and decorations.

We used a web based platform to create and evaluate our questionnaire. The questionnaire was split into 5 sections. For ease of analysis most questions were multiple choice. In total the questionnaire took about $10\ min$.

The first section was designed to gather information about the general working conditions. For instance regarding field of work or business size. This data was used as contextual information to interpret some unique answers to other questions.

Following that we had questions regarding the room layout of the office. We gathered information for the following features: size of the room, number of people in that room and size of the window surface.

In the section furnishing we wanted to get information about the bigger contents of the office. For instance kind and number of seating accommodations, size of the desk and kind, size and contents of storage spaces.

The fourth section consisted of questions regarding the workstation in particular. Among them were questions about technical appliances or other office related items (e.g. pens, post-it notes,...).

The last section regarded decorative objects in the office.

For instance paintings, plants or personal trinkets.
Additionally, we had one free text question in each section asking for aspects we missed in the design of our survey.

Participants

We sent out the questionnaire to friends, family and colleagues who are working in offices. We tried to reach different businesses in different fields. Contacting people we knew allowed us to target areas we did not cover before. The participants got a link to the questionnaire and filled it out on their computer.
In total we got 14 responses and did not collect further demographic data.

Our participants were office working family members, friends and colleagues.

Results

Among our answers were 6 different fields of work, with computer science being the most common (6 participants), and business sizes ranging from 12 to 25000 people with a median of 50. Only 2 participants stated that they did not have their own workstation but share them with their colleagues. Times at work ranged from $3\ h$ to $11\ h$ with an average of $8\ h$. While this information is no direct factor for designing our study it helped us to select characteristics we wanted our office to represent.

Most participants were computer scientists.

In the section room layout, we got reported office sizes of $12\ m^2$ to $150\ m^2$ with a median of $24\ m^2$. 1 to 50 people shared a room with a median of 2 persons per office. In average each participant had $8\ m^2$ for herself. Additionally, we asked how much window surface the office had with results ranging from $4\ m^2$ to $50\ m^2$ with a median of $12\ m^2$.

The average room has $24\ m^2$ and 2 workstations.

The third section was about furniture. On average each participant had $2\ m^2$ of desk space and 3 chairs. Additionally 9 participants stated that there was another table in their office. Most of these were meeting tables with some mentioning a coffee table or a secretary. The storage space per office ranged from $0\ m^2$ to $50\ m^2$ with a median of $3\ m^2$. We also asked which type of storage containers factor into that value. With 12 participants stating to have one, a rolling file cabinet was the most widespread form of storage container,

$2\ m^2$ desk space and $3\ m^2$ of storage space are common.

with regular file cabinets (9 answers) and bookshelves (7 answers) on position 2 and 3. We were also interested in the contents of these storage containers. Most people filled their storage space with books, folders, files, technical appliances, food, drinks, personal trinkets and office supplies. Additionally, 10 of our participants had a whiteboard or flip chart in their office and 4 reported having a wall mounted clock.

A usual workstation has a computer, two monitors, a telephone and various other appliances.

In the fourth section, we asked for details regarding the workstations and their equipment. Each workstation had at least 1 computer. Among the 14 participants, 8 stated to have a tower PC, 7 a laptop with external appliances and 3 a laptop without appliances. The number of monitors per PC ranged from 1 to 3 with 2 being the most common. The second most common technical appliance was a telephone with 13 participants stating to have one. Other items which were sometimes present were: Printer (5 answers), clocks (3 answers) and audio equipment (2 answers). Of the analogue items, most people had pens, post-it notes, markers, notepads and personal trinkets, while staplers, hole punches, tape and picture frames were less common.

Most participants had decorative elements like pictures, posters, paintings or plants.

The last section of interest was used to collect data on decorative elements of the office. Most offices had some form of plant, while small plants ($< 50\ cm$) were the most common with counts ranging from 0 to 10 with an average of 2.4 small plants. Regarding larger plants, people seemed to have either none or at least 2 with 4 large plants being the most. We grouped paintings, pictures and posters. Their counts displayed a sizeable variance. The numbers ranged from 0 to above 10 paintings with no number having more than 4 votes. On average participants had 3.3 paintings, pictures or posters in their room. Last but not least participants had on average 1.5 other art pieces in their offices. This number, however, is affected by some outliers. 8 of 14 participants did not report having any other pieces of art.

While the free text questions are not quantifiable they indicate that the mentioned aspects might be common. Mentioned objects were fans, delivery boxes and plush toys. Additionally, a sink and kitchens were mentioned but these seemed to originate in unusual circumstances, namely the medical field and a business with just one huge office.

4.2 Main Study

Based on our preliminary questionnaire, we designed our main study. The concept was to ask participants to perform a task in an office like room, while we recorded in which direction they looked. We did this for two conditions: while wearing an augmented reality headset with limited field of view (augmented view condition) and without any vision altering equipment (unaugmented/normal view condition). We used a between subject design, because we could only get one measure per participant. Knowing the room beforehand would alter their behaviour and we only had one room available for our study.

In our between subject design study, we asked participants to perform a task in an office like room. Conditions were: seeing an augmented reality and normal conditions.

We split the process of deciding on the finer details into six parts. In section "The Hypotheses" 4.2.1 we outline our research question and hypotheses for the study. In section 4.2.2, we detail the task the participants carried out and the design rational behind our choices. Section 4.2.3 presents the technology we used for tracking our subjects, recording their gaze data and displaying the augmented reality. Then in section "The Room" 4.2.4 we describe the room we used for our study. It details which information we took from our questionnaire and how we integrated the tracking setup. Following that section "The Variables" 4.2.5 defines the independent and dependent variables we isolated for our study as well as name the variables we controlled. Finally in "The Implementation" 4.2.6, we describe in detail how we conducted the study and some aspects of it which did not come up in the concept phase.
The evaluation of this study is presented in chapter 5.

4.2.1 The Hypotheses

Our research question: Do subjects wearing an augmented reality headset with limited field of view look around differently? Because of the field of view limitations, we assumed that the lack of peripheral augmentation detection would have the biggest influence. Usually the peripheral vision helps to create a rough overview of an area which

Without seeing augmentations in the peripheral vision, people's attention should be more spread out.

is then used to guide attention to the interesting regions [Wolfe, 1994]. This lead us to the first hypothesis:

1. Hypothesis
People wearing an augmented reality headset with limited field of view look more often at less interesting regions (e.g. walls, floor, ceiling,...) as people under normal conditions.

The participants head should make up for the limitations imposed on the eyes.

If we presume that users of augmented reality headsets need to cover the whole field of view with the central field of rendering, their head would need to compensate for the movement the eyes usually do. Which resulted in our second null hypothesis:

2. Hypothesis
People wearing an augmented reality headset with limited field of view will move their head more as under normal conditions.

We presume vertical movements of the gaze are mainly achieved by moving the eyes.

In line with the last hypothesis, the third one focuses on the movement of the head. In this case the vertical movement. We presumed that people tend to use more eye movements when looking at high or low objects. Due to the limited field of view this is not possible with an augmented reality headset. This could lead to a different height distribution of the view points. View points are point where either the eye gaze or the head direction hits an object.

3. Hypothesis
View points from the actual gaze or approximated gaze from the head direction in the augmented view condition have a larger height distribution than the approximated gaze from the head direction under normal conditions.

Participants in the augmented view condition need to split their attention on real and virtual objects.

The focus of our experiment is comparing the view behaviour regardless of the task they perform. To give participants a reason to look around our office like area (or test area), we implemented a search task where virtual and real objects need to be found. We elaborate on that in section

Figure 4.1: The task of our study consisted of finding these five animals. In the augmented view condition 3 were shown as augmentations (upper/right row), while the other two were 3D printed figurines (lower/left row). In the unaugmented view condition all five animals were shown as 3D printed figurines.

4.2.2. However, when piloting our study we noticed that wearing an augmented reality headset somewhat distracts from the real world and thus real objects got overlooked. This led us to the fourth null hypothesis:

> **4. Hypothesis**
> People wearing an augmented reality headset with limited field of view need the more time to find a combination of real and virtual objects, as people searching for only real items under normal conditions.

4.2.2 The Task

As stated above our main objective was comparing how people look at a room under different conditions. No hy-

(a) augmented view condition **(b)** unaugmented view condition

Figure 4.2: We replaced three search targets with augmentations in the augmented view condition. Additionally, we replaced some objects (e.g. the telephone) with augmentations. This way the augmented search targets did not stand out and were as hard to find as the real ones. Picture a) shows a hiding spot in the augmented view condition and b) shows the same hiding spot in the unaugmented view condition.

pothesis required a specific task.

The chosen task influences the view behaviour. We chose a search task and handle its influence as a controlled variable.

Our goal in deciding on a task was to keep participants engaged and looking at the room for collecting enough data. This is because we had only one room available and wanted to get the most data out of each trial. Previous research has shown that the task design has a high influence on viewing behaviour [Mills et al., 2011, Yarbus, 1967]. This indicates that the only way to record the natural way people look at a room would be to make the task free-view. Doing our comparisons on data recorded under natural conditions would increase the generalisability of our data. However, we presumed that this task would not keep participants engaged long enough. We reasoned that the task is a controlled variable with equal influence on all conditions. Thus the task itself should not have a noteworthy impact on the results of our comparisons. Compared to a memory task, in which participants would need to memorise as much about the room as possible, we expected a search task to facilitate a less organised viewing behaviour.

Small animal figurines were used as targets. Being out of context they did not influence the search locations.

In a search task participants are asked to look for objects or markings in a distinct area. In our case participants were asked to look for 5 objects simultaneously in the room we arranged. The objects were figurines of animals. The targets were roughly 5 cm long or tall.

When searching for items in a given situation the context of said objects has a high influence on our search behaviour [Wolfe, 1994]. For instance in an office like environment one would look on the desk for a computer mouse, instead of on the floor or at the wall. We mitigated this effect by using targets, which have no context in an office like environment and are not expected to appear in certain areas.

The targets were augmentations or 3D printed plastic figurines of same size and colour.

Depending on the condition, these animals were either 3D printed plastic figurines or hologram like augmentations of the same shape, colour and size. In the normal view condition, all objects were plastic figurines, while in the augmented reality condition 2 were real objects while the other 3 were displayed as digital augmentations. All five targets can be seen in Figure 4.1. The augmentation representation of the targets does not look exactly the same as the targets, because augmentations are created by adding light to the scene, not absorbing it. To prevent the augmented targets to stand out, we replaced some of the other objects, which were not part of the search task, with augmentations.

All targets were hidden in plain sight.

Each object was hidden in a way that it could be easily detected without interacting with the environment (e.g. open drawers) or moving in atypical ways (e.g. crawling beneath tables). On the other hand, the objects were hidden well enough for the trials to go on long enough for data collection. The latter we achieved by placing the targets behind somewhat occluding objects (e.g. plants), in similarly coloured areas or in areas with other distracting objects. Figure 4.2 shows a hiding spot in the augmented and the unaugmented view condition.

We used three different arrangements of the targets' hiding spots.

We created three different arrangements of the targets to reduce the influence of target placement on our results. In each trial of one arrangement the same animals were augmented in the augmented view condition, but we switched them between different arrangements. All animals were 2 times virtual and one time real, with the exception of the snail, which was virtual in every arrangement, and the dog, which was never virtual. The dog was black and could not be displayed by our mixed reality headset. Each arrangement was used as often as every other one and they were balanced between the conditions.

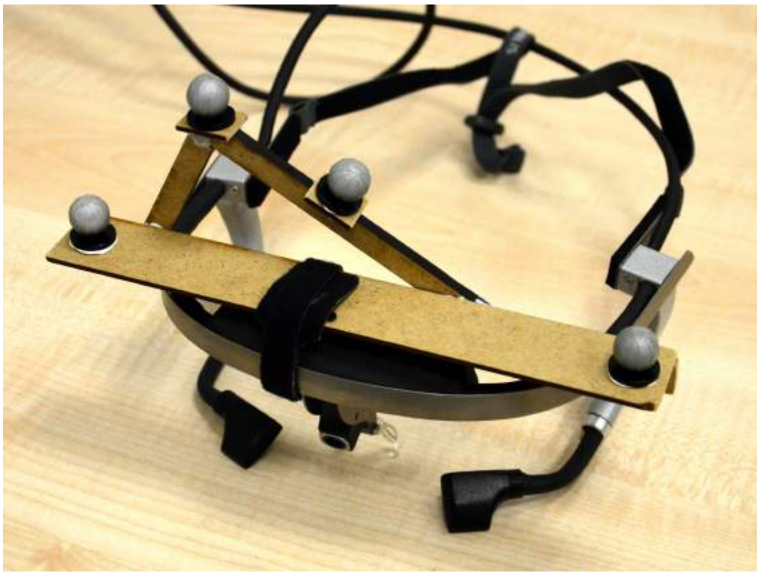


Figure 4.3: We attached a rig, that could be tracked by the motion tracking system, to the eye tracker and the hololens. That way we could record the head direction of the participants. This picture shows the rig attached to the eye tracker.

Participants were looking for all targets at once till all were found.

Before the trial started the participants saw a picture of the objects. We only had 5 objects which were easily distinguishable from all other objects, because they did not fit into the context of an office. We reasoned that forgetting the objects would be no issue. Participants were asked to look for all of the objects at the same time and when they found one just state that they have seen one and go on to look for the next one. The trial was concluded when all objects were found or when a time limit was reached. We introduced the time limit to keep the trials in a comparable interval. Additionally, this was to avoid frustration among our participants. The participants were not told of the time limit, as we did not want to introduce an element of stress.

4.2.3 The Technology

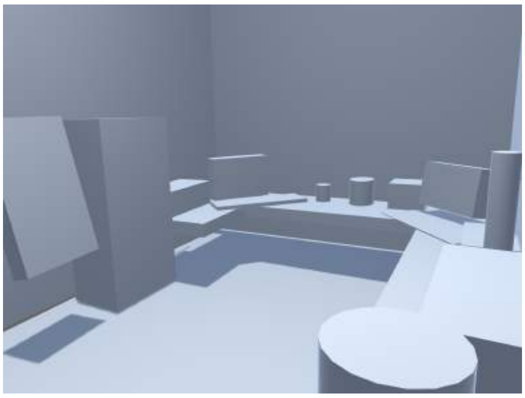
We selected the technology we used based on two criteria: availability and suitability. The first is owed to the price of technical instruments and the latter to the goals we tried to achieve. Our final system consisted of a Vicon motion tracking system, a Microsoft Hololens mixed reality display, a Dikablis eye tracker and various computers running software such as Vicon Nexus, D-Lab and Unity.

Head tracking was done with an optical motion tracking system.

The Vicon motion tracking system is an optical tracker based outside-in system. Cameras are placed around the arena and infra-red reflecting markers are attached to the subject. Our system consisted of 6 Vicon Bonita cameras and Vicon Nexus 2.6. We build a rig, which could be attached to the Hololens as well as to the Dikablis eye tracker. A picture of the rig attached to the eye tracker can be seen in Figure 4.3. As both, the Hololens and the eye tracker, are fixed to the head, this resulted in a comparable head position measures in all conditions. These measures were used for two things: computing the actual gaze in the room in combination with the eye tracking and tracking the head. Aside from being available, the system is a good fit for our study as optical tracker based methods are fast and precise, we already have an immobile set-up and the system is capable of communicating with D-Lab, the software used to receive the eye tracking data. The direct connection between D-Lab and Vicon Nexus has two advantages. First, the position and gaze data are exported in a single file. Second, the data are already synchronised, which removes the need for a later synchronisation step which might introduce errors.

We used a head mounted eye tracker, which allowed participants to move around the room.

For gaze data collection, we used a Dikablis Professional Wireless eye tracker . Being a head mounted and wireless system, this allowed the participants to move around the room while doing the search task. Additionally it was a good fit to the Vicon motion tracking system, because, as previously mentioned, they can work in tandem. However, the Hololens and the Dikablis eye tracker are both worn like glasses which made it impossible to wear them both at the same time. Because of this, eye tracking data is only collected in the unaugmented view condition. The eye tracking system itself consisted of the eye tracker, which is worn similar to glasses. Despite this, it is compatible with corrective eye wear and the limitation to the field of view is relatively small. That system is connected to a portable



(a) simplified

(b) normal

Figure 4.4: This is a side by side comparison of our simplified model (a) and the real room (b). Each box in a) represents one area of interest.

computer, in our case a Microsoft Surface, which streams the data to a stationary laptop via a WiFi connection. That laptop in turn receives the data in D-Lab 3.0 and streams it to Vicon Nexus via LAN. Due to the WiFi connection, the system is limited to a sampling rate of 30 frames per second.

The head mounted display with limited field of view was a stand alone system, that allowed the participants to move.

In the augmented view condition said augmentations were created with the Microsoft Hololens , a wearable holographic computer. This means it is a self contained head mounted mixed reality display capable of running all augmentation software locally instead of resorting to streaming solutions. While, like most other optical see-through systems, the normal vision is hardly limited, holograms are only displayed in a central area of roughly 40° width and 22.5° height [Xiao and Benko, 2016]. This limitation on the visual field as well as its mobility makes the Hololens perfect for our intents and purposes.

We used the three dimensional room scan of the Microsoft Hololens to compute the points participants looked at.

After the trial we analysed the data which we exported from Vicon Nexus. The first step of that process was projecting the gaze vector data in a 3 dimensional model of the room used for the study. That room model was created by exporting the spatial map of the Microsoft Hololens. We used this model to create a simplified version of the room by placing cubes and cylinders of different sizes around each object making up the room. The result of this process can be seen in 4.4 Two chairs, which were also in the test area, were not represented in the model as they could be moved and were excluded as a hiding spot for the targets. The accuracy lost by using a simplified model were negligible, because these errors were smaller than the ones introduced by the recording process.

The model needed to be aligned with the data gathered by the Vicon system. We achieved this alignment by recording a sample trial where we tracked an object which we moved along distinctive features of the room, like the wall, corners of a table/file cabinet and the floor. Then we manually matched this 3 dimensional position data with our model to obtain the transformation parameters. Because of the aforementioned accuracy loss this process is precise enough for our purpose.

The two coordinate systems were matched with the use of prominent features of the furniture.

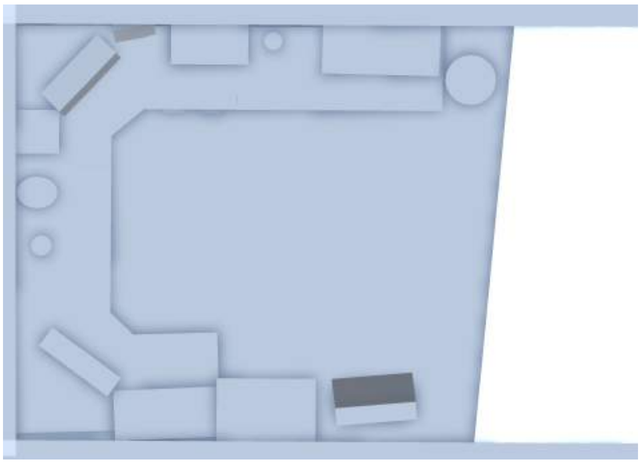
4.2.4 The Room

Based on our design informing survey, described in section 4.1, we chose to use a set-up with two workplaces and roughly $20\ m^2$. The room we used was $4\ m$ x $6\ m$. One of the short sides was a window surface. One of the two longer sides incorporated a large mirror but the remaining two sides were regular white walls. To enhance the stability of the Vicon system and exclude the outside as an influence on our trial, we closed the blinds resulting in the visual appearance of a grey wall.

We conducted the study in a rectangular room of $24\ m^2$.

The study room contained the preparation and the test area. The former needed to be included in the study room because some calibration steps required the Vicon system, which needed to be distributed around the whole room for best coverage. We used a removable curtain to divide the whole room in the two separate parts. Figure 4.6 shows the curtain in the half opened and completely opened state. We introduced the participant to the system in the preparation area while the curtain was closed and the volunteer did not see the test area. Then we asked the participant to stand with their backs to the curtain and removed it. Afterwards, we could do the calibration with the whole Vicon system but without the participant seeing the test area.

The room contained the preparation and test area, which were divided by a curtain.



(a) top down perspective

(b) left side of the room

(c) right side of the room

Figure 4.5: For our analysis we used a simplified three dimensional model of our study room; a) shows rendering in top-down perspective. The white area to the right was the preparation area, that was not included in the model. Below that are two pictures of the test area: one (b) showing the left side, when looking from the preparation area, and one (c) showing the middle and right side.

The preparation area housed equipment used for set-up and calibration.

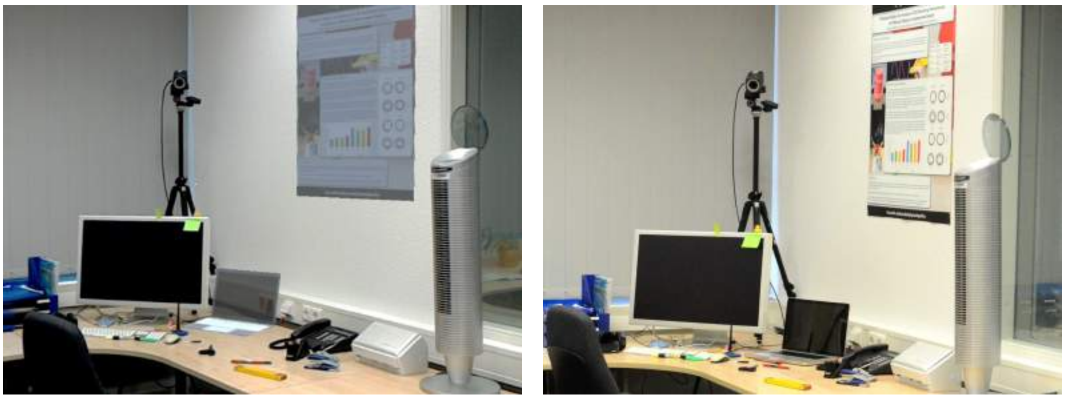
The preparation area was comprised of a table, a chair and other devices needed for the calibration procedure. The table was used to store the technical devices while no participants wore them and situate the laptops used to operate the various systems. The chair was mainly for the participants' comfort while they read the consent form or were fitted with the Microsoft Hololens or eye tracker.

The test area's design was guided by the preliminary survey.

Guided by the results of our previous study, we designed the test area to look like an average office. The furniture consisted of three tables, a file cabinet, a flip chart, a rolling

(a) half closed curtain **(b)** opened curtain

Figure 4.6: Because participants should not see the room before the trial, we divided the room with a curtain. These pictures show the curtain half open (a) and completely open (b).



(a) augmented view condition **(b)** unaugmented view condition

Figure 4.7: We replaced some real objects with augmentation in the augmented view condition. This way the augmented search targets did not stand out and were as hard to find as the real ones. Picture a) shows a rendering of the augmented view condition and b) shows a comparable section in the unaugmented view conditions set-up.

file cabinet and two chairs. One of the tables was used only for office appliances, such as an air filter, a cutting machine and a fax. The other two were each equipped to look like workplaces. Both workplaces consisted of one big monitor and computer related input devices, a telephone, office

supplies (e.g. pens, text markers, post-it notes, folders) and various trinkets. Additionally, one also had a laptop and earphones. On a small table area connecting the two workplaces, we placed two small plants. Other objects around the room were a dustbin, a shredder, a bigger plant and a big poster at the wall. Two chairs were placed in front of the workspaces.
The Vicon system could not track the area directly in front of the desks, where the chairs stood. Because of this we prohibited participants to enter that area.

Four of the six Vicon cameras were placed in the corners of the room and the remaining two roughly in the middle of the longer walls. This way we had the most reliable tracking around the room.

Some task unrelated objects were replaced with augmentations to hide the targets.

In the augmented, we replaced some objects in the room with augmentations. These objects were: a telephone, a laptop and a poster. This way the augmented search targets do not stand out as much and are as hard to find as the unaugmented ones. Figure 4.7 and Figure **??** show the differences.

4.2.5 The Variables

Independent variable levels: eye tracking, head tracking and hololens

The two conditions, unaugmented view and augmented view, resulted in one independent variable with a total of three levels: eye tracking in unaugmented view, head tracking in unaugmented view and head tracking in augmented view. We call these conditions eye tracking, head tracking and hololens.
The lack of eye tracking in augmented view was due to technical difficulties, as detailed in section 4.2.3

Dependent variables: task completion time, area of interest counts, view points, rotational movement and translational movement

We measured three variables: the task completion time, participants' viewing direction and the position of the participants in the room. The task completion time was the time elapsed between the moment we allowed the participants to start the trial and the end of a trial. When the participants needed longer as $2.5\ min$ that was the task completion time, despite the task not actually being completed.

The direction was either the actual gaze measured by the eye tracker or the forward vector of the participant's head. We used the measured data to compute the dependant variables. We divided the room in areas of interest and counted which area got viewed for how many frames. We computed the point in space where the participant looked at, which we called view point. And we computed the movement of the participants by change in their position and view direction.
Thus, we had a total of five dependent variables: task completion time (in s), area of interest counts, view points (3D coordinates in m), rotational movement (in degrees) and translational movement (in m).

Controlled variables: the outside, target arrangement, task order over course of study and room design

Additionally, we had some controlled variables. The blinds in the study room were closed to reduce the effect of the current time and occurrences outside. We also had three different arrangements for the search targets. Each was used the same number of times as the other arrangements and between the conditions. Over the course of the experiments we switched what condition we tested as to remove order effects regarding the dates or the conduction of the study. All trials were conducted in the same room without any changes done to it.

4.2.6 The Implementation

We had 42 participants in total. Most of which were students with an age between 18 and 31.

The participants for our study were recruited on the grounds of the computer science department of the RWTH Aachen. Thus most participants were students of the computer science faculty. Each volunteer was offered sweets as a reward for participation. In total 42 individuals took part in our study; 21 in each condition. Their ages ranged from 18 to 31 with a mean of 23. The majority (35) were male with another 6 being female and the remaining 1 reporting as other. We also asked for their experience with augmented reality devices on a Likert scale from 1 to 5; 1 being no experience at all and 5 being as much experience as one can get. The average reported value was 1.5 and the median was 1. Indicating that most volunteers had almost no experience. Additionally every participant reported perfect

or corrected to perfect vision. Participants were randomly assigned to one of the two conditions.
The rest of this section will give a more detailed summary of each step our trials consisted of.

Welcome and Introduction

Participants were introduced to the trial and asked to sign a consent form.

After recruiting a participant, we gave them a detailed rundown of the procedure, including how they were allowed to move and where the search targets could be hidden. Depending on their choice this was done verbally or by means of a consent form detailing the trial's execution. The full consent form can be seen in Appendix B. Additionally, we collected some demographic information consisting of age, sex, experience with augmented reality in general and acuteness of vision. Last but not least, every participant was shown a photo of the search targets. In the augmented view condition they were also shown their digital representation.

Calibration and Familiarisation

Calibration of the eye tracker consisted of three steps: adjusting the eye tracker, calibrating it in D-Lab and calibrating it in Vicon Nexus.

In the unaugmented view condition, participants were asked to put on the eye tracker and adjust its fit to their liking. Afterwards, we moved the cameras, filming the eyes, into the correct position. Additionally, we measured the location of the left eye in regard to a certain infra-red marker. This was needed for Vicon Nexus to place the eye correctly. For the second step, calibrating the eye tracker in D-Lab, participants were asked to stand at a headrest facing a wall. On that wall, we had marked four points with clearly visible, coloured strips of tape. Then we asked the participants to look at these points one after the other, while we entered the actions into D-Lab. A picture of the headrest with a participant can be seen in Figure 4.8. Further, we needed to calibrate the eye tracker in Vicon Nexus. Therefore, it was necessary to remove the curtain, which hid the test area, because only then all cameras of the Vicon system could see the participant which is needed for a reliable tracking. However, we asked the participants not to move around for this step, because otherwise they would have seen the test area before the trial started. The calibration procedure itself consisted of moving another tracked object in front of

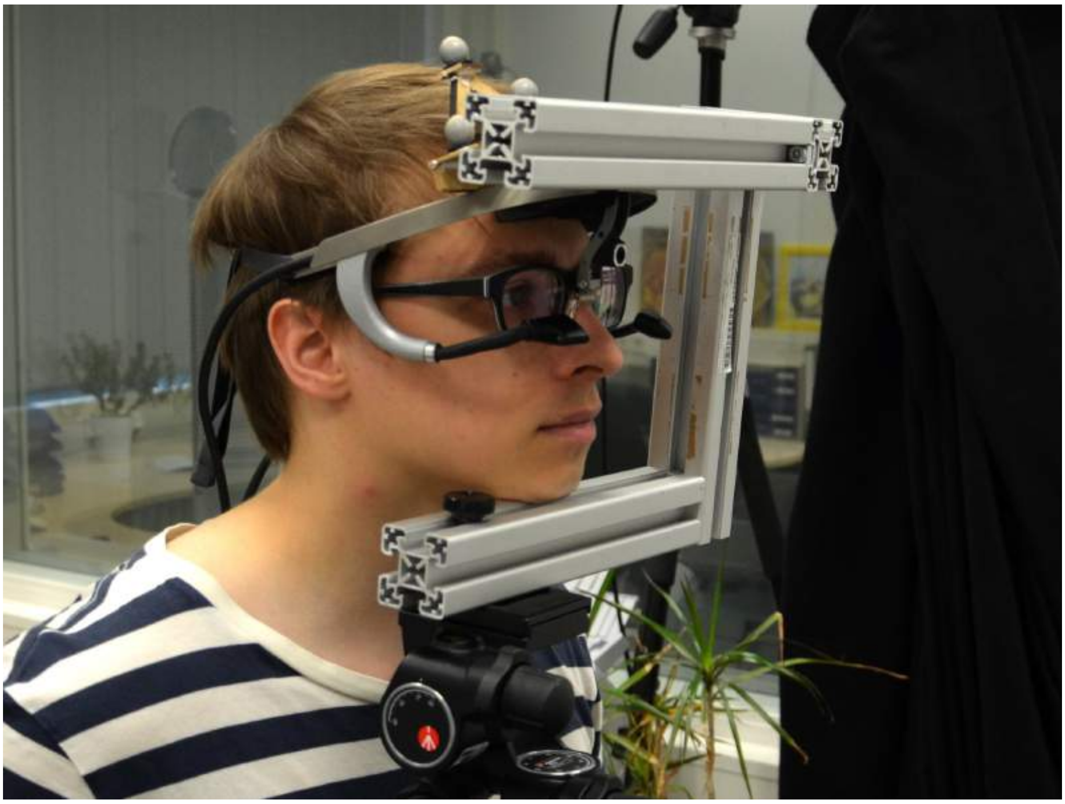


Figure 4.8: A participant, who wears the eye tracker, stands at the headrest while the eye tracker is calibrated. The headrest helps holding the head still, while the participant looks at distinct points on a wall, for calibration purposes.

the participant and asking them to follow a certain marker on it with their eyes. This information was used by Vicon Nexus to rotate the gaze vectors, provided by D-Lab, correctly.

Participants were shown augmentations to familiarise with the head mounted display.

For the augmented view condition, the familiarisation procedure consisted of only two steps. Firstly, participants were asked to put on the Microsoft Hololens and adjust its fit to their liking. Secondly, they were shown some augmentations, located in the preparation area of the study room, to get accustomed to their existence and how they looked. In this step, they were asked to stand with their back towards the actual test area because the augmentations placed in said area would have been visible behind the curtain.

(a) eye tracker

(b) Microsoft Hololens

Figure 4.9: These pictures show a participant with the eye tracker (a) and Microsoft Hololens (b) set-up. In the left picture, the blue box contains a tablet, which records the eye tracking data.

When they confirmed to have familiarised with the system, the curtain was removed behind them as final preparation of the trial.
A participant wearing the devices can be seen in Figure 4.9.

Trial

Participants looked for the search targets until all were found or a time limit was reached.

The trial itself was the same for both conditions. Either way after the calibration, participants stood in the centre of the room with the curtain removed still facing the preparation area. We gave them one final overview of the trial as a reminder. Then we started the recording of the trial. For the final calibration step, which is listed here because it is done while already recording the trial, we held out another tracked object in front of them and asked them to focus a marker at eye level while standing relaxed. This was used to record the direction their head was facing. Then we gave participants the signal to turn around and start looking for the search targets. This marked the actual beginning of the trial. The recorded part before that point was only used for calibration but not for analysis.
When all objects were found or a time of $2.5\ min$ elapsed,

we stopped the trial. Finally, we asked the participants to take off either device, thanked them for their participation and released them from the study room.

Chapter 5

Evaluation

The evaluation was a two step process. In the first step the data was preprocessed and prepared for statistical tests. Afterwards, we applied these tests. For both steps we created our own tools. While we do not describe the code in detail, we explain the procedure in the following sections. We discuss the results of our analysis in section 5.3. We will conclude this chapter with a section answering our initial research question and one describing the limitations of our work.

5.1 The Preprocessing

Our collected data consisted of a file for the gaze data and the answers to the short questionnaire. The gaze data file was in comma-separated-values (*.csv) format and detailed the location and rotation of each detected segment in three dimensional space. A segment is an object which is defined by a distinct arrangement of infra-red reflective markers. In our case, this was the head-unit, which was attached to the device worn by the user, and the calibration object, which was used for the various calibration steps. Additionally, in the unaugmented view condition, the file contained the gaze vector recorded by the eye tracker.

A *.csv file for each trial contained all data exported from Vicon Nexus.

Data Extraction

We computed the view points by intersecting the gaze vectors with a three dimensional model of the room.

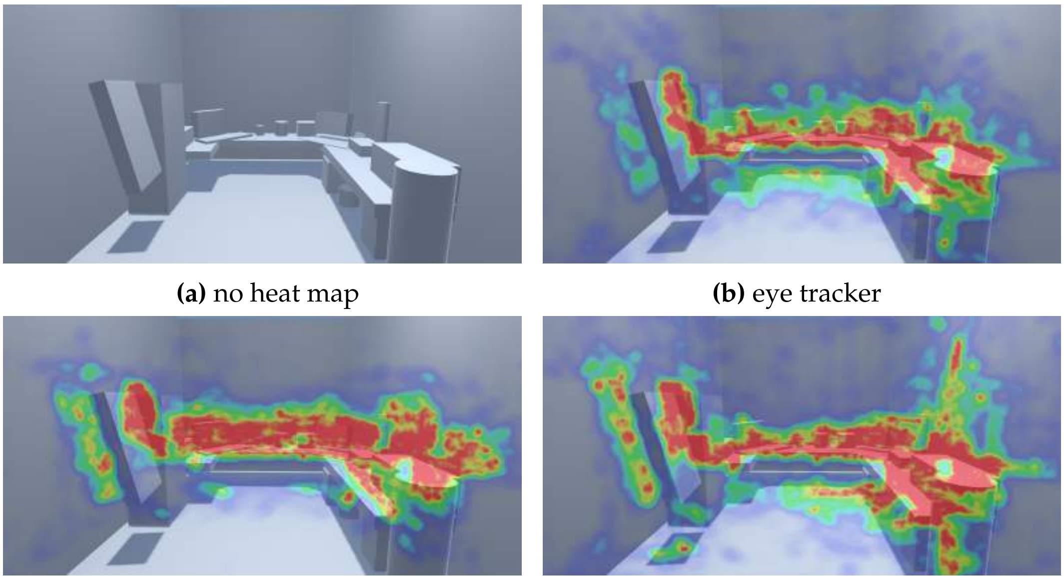
We build a tool with Unity that read the file, placed the head in a simplified three dimensional model of the study room and then used ray casting to compute the points where the gaze and head direction data intersected the room, thus where the participants were looking. The coordinate space of the Vicon data and our model were matched using the method detailed in section 4.2.3. In either condition, we evaluated the head direction data and, in the unaugmented view condition, we additionally incorporated the actual gaze data, which was measured by the eye tracker. As detailed in section 4.2.6, we started each trial in a position where the participant looked straight ahead on another tracked object, which we held at face level. We used this position and defined the head direction as the vector from the head segment to that calibration object.

The simplified model of the room was created by placing bounding boxes around areas of interest.

We used a simplified model of our study room for data analysis. This version was created by importing the model, which we generated with the Microsoft Hololens, into Unity and placing cubes of different sizes around objects in the room to mark areas of interest. This way we created a model of the whole room including walls, floor and ceiling. For filtering purposes, we used one big cube to contain the whole preparation area. This way we could just ignore any gaze trajectories which fell into that direction. We did this for three reasons. Firstly, these areas of interest classified what the user was looking at and thus, we could run our data analysis on semantically enhanced data. Secondly, as previously mentioned, our gaze data lacked accuracy. Using these areas as preprocessing step reduced the influence of that. Thirdly, the model created by the Hololens is not perfect. The walls are not completely flat, smaller objects are not always detected properly and sometimes artefacts appear in mid air. By using a simplified model we lost accuracy in correctly recorded areas but gained overall robustness.

The tool exported *.csv files that could be imported into R.

Our data extraction tool in turn exported three files per trial. In the first file, the total count of hits on each given area of interest was saved. The second file stored the coordinates of the actual hit point as well as the area of interest hit. This way, we did store the area of interest data twice but



(a) no heat map **(b)** eye tracker

(c) head tracker **(d)** hololens

Figure 5.1: Picture a) shows a rendering of our simplified model, while pictures b-c) visualise the view point data as heat maps. While the differences are small the eye tracker data (b) are more focused on interesting areas, while the head tracker (c) and hololens (d) data are more evenly distributed.

this was done to simplify further steps of the analysis process. In the last file, we stored the position of the head unit and a vector for the gaze direction. In the unaugmented view condition, we exported six files per trial in total: one set for the eye tracking and one set for the head tracking data. The second and third type of file did not get the advantages of dividing the room into areas of interest. We did further filtering steps on them to account for accuracy losses.

Data Visualisation

The tool we created for data extraction was also capable of visualising the view point data. We used a heat map approach for visualisation as these show the data in an easily comprehensible form. The heat maps can be seen in Figure 5.1. One can see that the view points of the head tracker data were more evenly distributed, while the ones of the eye tracker data were more focused on interesting regions. The distribution hololens data's view points is between the

We visualised the view point data as heat maps.

other two distributions; to a degree focused but on more areas. Additionally, the hololens data also shows some further exploration of more remote areas, like the right wall and the floor.
These maps are purely data visualisation and were not used for statistical analysis.

Data Filtering

We filtered the data in R and removed unwanted parts of the trials.

The second step of preprocessing was done with the programming language R, which is designed for statistical analysis. As stated in section 4.2.6, the last calibration step, which was measuring the head direction, took place while the recording of the trial was already running. The data from that time frame was still in the data which we exported from Unity. However, as not being part of the trial they got filtered out in R.

Instability of the Vicon system caused loss of data or introduction of errors, which needed to be filtered out.

Aside from accuracy issues, we also had a problem with the stability of detection. The Vicon system was sometimes losing track of single markers quite. Sometimes this led to the whole head unit disappearing and other times it led to a misplacement of the head unit.
In the first case, we are losing data but no errors are introduced. We missed the head tracking data in, on average, 33% of frames per trial in either condition. For the eye tracking data this loss was slightly higher (on average 42% of the frames), because this data could not be recorded when the eye tracker could not find the pupils, which was the case when the participant was blinking. This data is also shown in Figure 5.2.

In the second case however, wrong data is created. We used different filtering approaches to reduce this effect. Because the chosen methods are highly dependent on the data in question we describe them in the upcoming section 5.2.

In some trials, we lacked the data of the beginning of the trial, which is necessary to compute the head direction. These trials needed to be excluded from the analysis. Thus, we excluded one trial in the unaugmented view condition and three in the augmented view condition. In total, we had 38 trials remaining for the analysis.

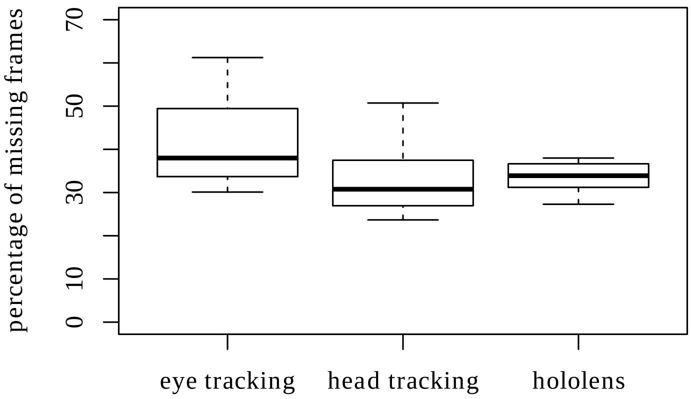


Figure 5.2: This graph shows the percentage of frames, in which no data could be recorded for each condition. The percentages are highest for the eye tracking data, which is to be expected as blinking and undetected pupils are additional missing frames. Head tracking and hololens percentages are similar to one another.

5.2 Analysis

We tested if the different arrangements of the search targets had any influence on the data. We did this by using the arrangement as independent variable and test for differences in the results. However, there were no statistically significant effects found and we regarded this variable as controlled by randomisation.

The arrangement could be regarded as controlled variable.

Areas of Interst
For the first test, we used the counts of the views hitting the areas of interest. Because we had trials with varying lengths, we computed the percentages of views fore each trial and area of interest. We split the areas of interest in two groups depending on how interesting or spatial attention

Visually stimulating areas were regarded as more interesting and compared to less interesting areas.

drawing we regarded them. For this, we took into account how visually stimulating the area was. The less interesting group consisted of the walls, the floor, the ceiling, the windows, the flip chart, the file cabinet and the two dustbins. As the blinds on the windows were closed they looked similar to a grey wall.

The eye tracking view points had the highest amount in interesting regions, followed by the hololens and then head tracking data.

Considering our independent variables with three levels (eye tracking, head tracking and hololens) and the area of interest counts, which was normal distributed, we chose to do a one-way between subjects ANOVA. The test reported a statistically significant difference between the percentages of views on less interesting areas of interest depending on the condition ($F(2) = 4.423, p < 0.05$). Figure 5.3 suggests that, in the unaugmented view condition, the eyes are directed more on interesting areas ($mean = 50.3\%, SD = 14.4$) than the head is ($mean = 37.5\%, SD = 16$). In the condition in which the participants wore a Microsoft Hololens ($mean = 42\%, SD = 9.6$) the head direction was directed at interesting areas more than it was in the unaugmented view condition, but not as much as the eyes were. However, these differences are small with a high distribution among our samples.

Movement of the Participants

Rotational and Translational movement were analysed separately.

The movement data was split into translational and rotational movement. For the former, we only have two conditions, because our eye tracking and head tracking trials come from the same participants. The rotational movement however, was measured within the gaze and the head direction vectors and thus resulting in three conditions again.

We used the median of half second intervals to filter outliers out.

For this, we used the file which stored the position of the head and the direction of the head or gaze per frame. By computing the distance between the positions or the angles between the frames we could compute the magnitude of the participants' movement. We computed the differences between frames which were half a second apart. This was done to filter out small jitters in the movement to get an accurate measure for the bigger changes in position. However, as stated earlier the recorded position and rotation data had issues with their robustness. Vicon sometimes de-

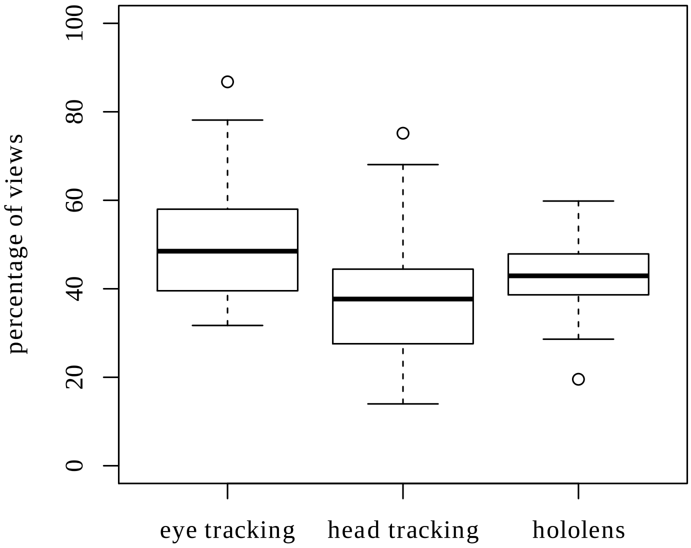
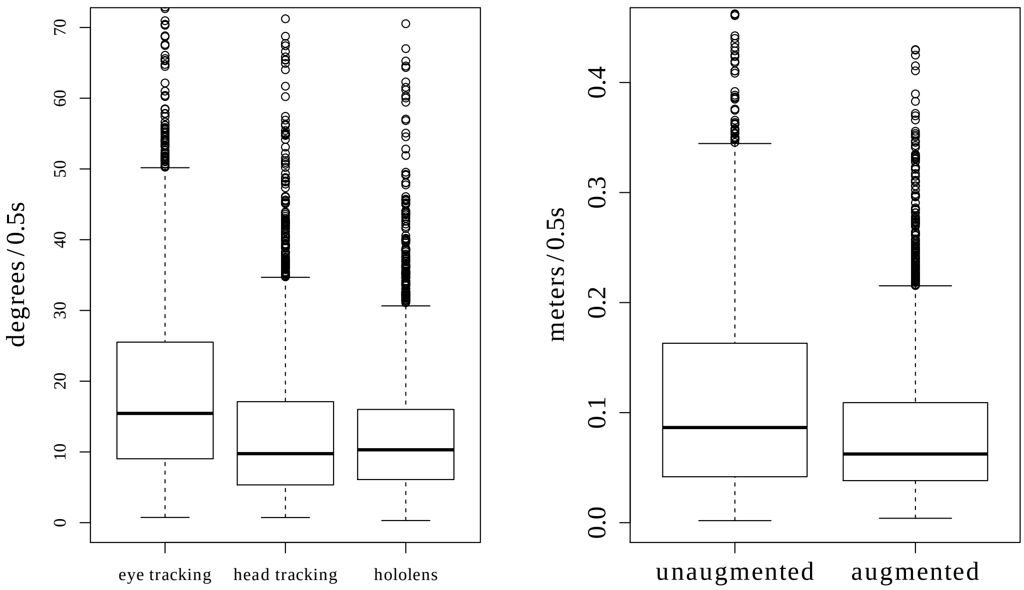


Figure 5.3: This graph shows the percentage of views on areas of interest we regarded as more interesting. This percentage is highest for the eye tracking data and the lowest for the head tracking data. The hololens data, while also being head tracking, is more similar to the eye tracking data.

tected the head unit in a false position or rotation which led to sudden jumps in their detection and thus a high amount of outliers in our movement data. To reduce this effect, we only took the median for every half a second interval. However, in some time frames there were too many outliers compared to precise values, which means that some outliers stayed in the data.
For the rotational movement data we chose to do a Kruskal-Wallis Test, as we had unmatched data and it was not normal distributed. While the results are statistically significant ($\chi^2(2) = 431.04, p < 0.001$) the graph in Figure 5.4 shows comparably small differences. The eyes ($mean = 19.3°/0.5s, SD = 15$) seemed to move more than the head, but the latter did not change between



(a) rotational movement **(b)** translational movement

Figure 5.4: These graphs show the rotational (a) and translational (b) movement per half a second. In a) one can see that the eyes are moved more than the other two conditions, but their differences are negligible. Graph b) shows that participants moved more in the unaugmented view condition but the differences are slight.

the augmented ($mean = 12.8°/0.5s, SD = 11.5$) and unaugmented($mean = 13.3°/0.5s, SD = 12.4$) view condition.

For the translational movement data, we only had two conditions with no normal distribution and thus, we chose to do a Mann-Whitney-Wilcoxon Test. There was a statistically significant difference between the translational movement of participants in the unaugmented view condition ($mean = 0.11m/0.5s, SD = 0.09$) and the augmented view condition ($mean = 0.09m/0.5s, SD = 0.13$) ($U = 4342400, p < 0.001$). The graph, in Figure 5.4, shows small changes of the data compared to their variance.

Height of the View Points

Here, we compared the height profiles of the view point data between the conditions. We did not filter the height data for robustness issues. As we do not compare frames with each other, their effect is reduced and the majority of data points were correctly detected. Due to the fact, that the eye tracker did not have such robustness issues, the whole effect was introduced by the Vicon system and thus was the same for all conditions. This way we could handle it as a variable controlled by randomisation.

Height data was not filtered, because the outliers' influence was small.

Considering our three levelled independent variable and non normal distributed data, we again chose to apply a Kruskal-Wallis Test. It indicated that statistically significant difference exist between the view points' height of the eye tracking ($mean = 0.85m, SD = 0.54$) data and the head tracking data ($mean = 1.07m, SD = 0.42$) and the hololens data ($mean = 1m, SD = 0.73$)($\chi^2(2) = 15099, p < 0.001$). A graph comparing the normalised height profile of the three conditions can be seen in Figure 5.5. The profile of hololens data is quite similar to the one from the eye tracking data . However, the data seems to be a little bit more distributed towards the higher regions of the room. The head tracking data on the other hand has a different distribution. Its peak is much higher than the one of the other two conditions and less people have pointed their head at floor near regions.

Task Completion Times

The task completion time was defined as the time between the moment participants were allowed to turn around and the instance we stopped the recording of the trial. We stopped the trial instantly when the last object was found or, as previously stated, when the trial took longer than $150sek = 2.5min$. Our independent variable had only two levels again, as eye and head tracking in the unaugmented view condition were recorded from the same trials and thus have the same lengths. For this reason, and because the data was not normal distributed, we chose to do a Mann-Whitney-Wilcoxon Test to compare the task completion times between the two conditions. It indicated that participants in the augmented view condition ($mean = 140.2s, SD = 23.9$) took longer to find the targets than participants in the unaugmented view condition

Most participants in the augmented view condition did not complete the task in time.

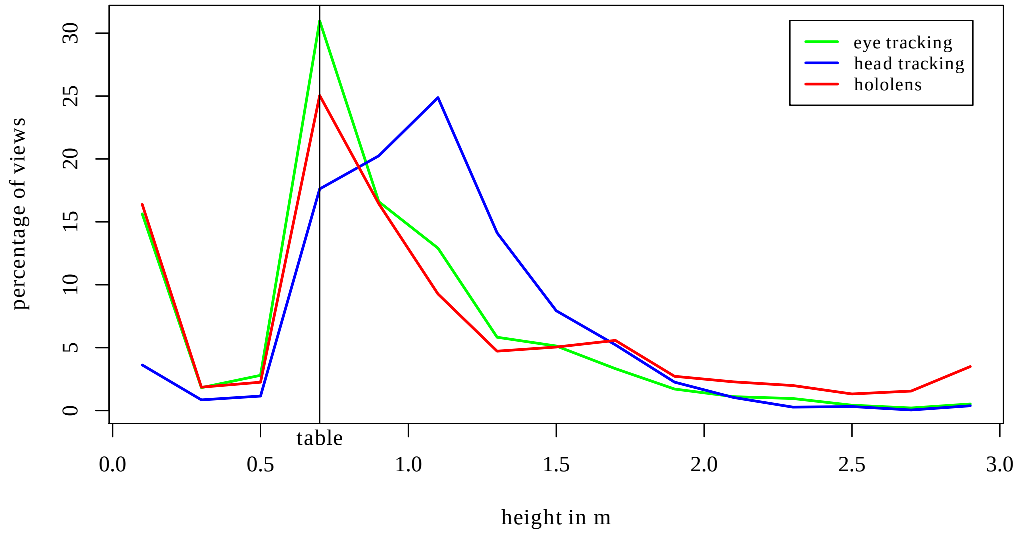


Figure 5.5: The lines display the normalised histogram of the view points' height for the conditions. The vertical line marks the height of the desks. While the distribution of eye tracking and hololens are similar, the hololens data is more distributed (especially at the higher end). The head tracking data on the other hand shows a different distribution with a higher peak and almost no views on the floor.

$(mean = 87.2s, SD = 42.1)(U = 50, p < 0.001)$. These findings were statistically significant. As most trials, in the augmented view condition, would have taken more than $150sek$, the values for the augmented view condition are not accurate. Figure 5.6 shows the distribution of the data.

Because of this, we additionally compared the success rate of the two conditions. We used a Fisher's Exact Test to compare the success rate in the unaugmented view condition (80%) to the success rate of the augmented view condition (25%) and the test reported our results to be statistically significant $(oddsratio = 10.98, p < 0.01)$.

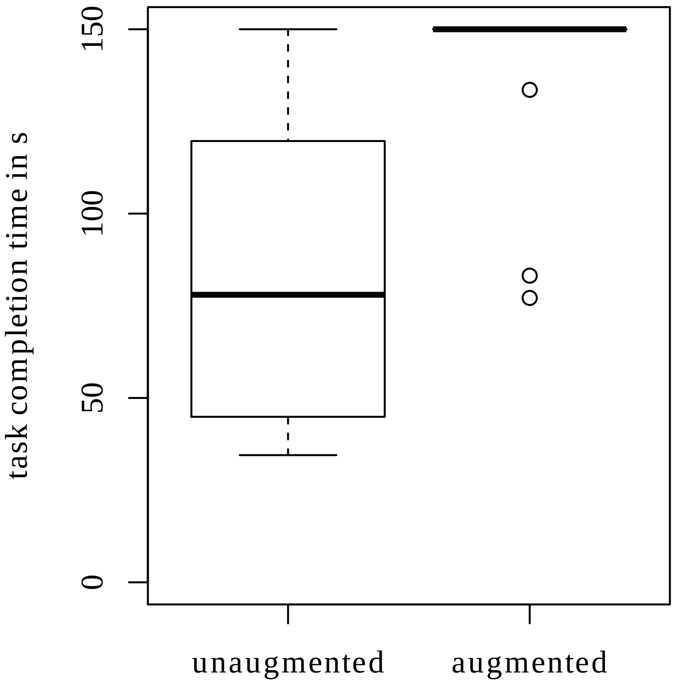


Figure 5.6: This graph shows distribution of the task completion times for the unaugmented view and the augmented view conditions. We stopped each trial that exceeded $150\ s$. As most trials in the augmented view condition did that, their times are not accurate. It can be seen, that participants in the unaugmented view condition were faster in completing the task.

5.3 Discussion

In this section, we return to the hypotheses we detailed in section 4.2.1. For each hypothesis, we review the results we presented in the last section and present our interpretation.

1. Hypothesis
People wearing an augmented reality headset with limited field of view look more often at less interesting regions (e.g. walls, floor, ceiling,...) as people under normal conditions.

In the augmented view condition participants' view points were more evenly distributed.

The first hypothesis we tested showed that people seemed to direct their eyes less on uninteresting regions as they would their head. While the differences are small compared to the variance of our results, a tendency can be seen. A difference between eye and head tracking was to be expected, but it is interesting to see that the hololens data is similar to the eye gaze behaviour of the unaugmented view condition. Especially Figure 5.5, which displays the height the participants looked at, seems to support this change of the viewing behaviour. Under the assumption that the hololens data is more similar to the eye tracking data than the head tracking data , we reject the null hypothesis and assume that users wearing a head mounted augmented reality device with limited field of view spend more time looking at uninteresting areas. The heat maps shown in Figure 5.1 support these findings as well.

The difference between the eye tracking data and the hololens data could result from the inability to use ones peripheral vision to find virtual objects. Thus, more uninteresting areas needed to be scanned in order to find the targets. This could also lead to the differences in task completion time.

2. Hypothesis
People wearing an augmented reality headset with limited field of view will move their head more as under normal conditions.

The average movement speed was not different, but the trials had a higher duration.

The amount of head rotation did not seem to vary between the two conditions. In both cases, it is less than the amount of eye movement of the unaugmented view condition. The latter part is not surprising. The eyes are much easier to move and while the head determines the general direction, the eyes are the ones actual looking for objects in the scene. However, in the augmented view condition more of this burden should have fallen on the head, as the field of view in which holograms can be found is much smaller than the field of view in which real object can bee seen. That this is not the case could have multiple explanations. It could be the case that even in unaugmented view the eyes mostly remain within the central area of the field of view. It is also conceivable that more time is needed to perceive an area, because both real world and digital influences need to be processed. Or maybe it is inconvenient to move the head faster than it already usually does. For us, the latter two interpretations are more likely as they would also explain the much higher task completion times. Of course, it is possible that multiple influences are combined to result in this outcome. Regarding the hypothesis, the similarity of the average head movement per half a second between the two conditions leads us to accept this null hypothesis. However, when taking into account that the trials of the augmented view conditions were much longer the total amount of head movement is higher in that condition.

Additionally, there seems to be a minor difference in the translational movement of participants depending on the condition. Participants wearing the eye tracker moved more around the room than participants in the augmented view condition. We actually expected this effect to be the other way around. This could have been explained by the attempt to counter the viewing limitations by also moving more around the room. The actual behaviour could be justified by a different comfort while wearing the device. A heavier device or one with a less stable attachment method could lead to the participants moving more carefully. If that is the case this effect would also influence the rotational movement.

> **3. Hypothesis**
> View points from the actual gaze or approximated gaze from the head direction in the augmented view condition have a larger height distribution than the approximated gaze from the head direction under normal conditions.

Eye movements are used more when looking downwards.

The height analysis of the view points is a bit complicated as the layout of the room has an influence on the measurements. The spike at roughly $70cm$ for instance is on table height. And being an office like environment most of the interesting areas are on said height and slightly above. The increase on ground level on the other hand is due to most participants also searching below the tables. While the profile of the eye tracking and the hololens data are quite similar, the head tracking data is not. The highest amount of points is almost $50cm$ above the ones from the eye tracker. And also on the floor are less head direction hit points. On the other hand, while there is less data above $2m$, in general the curves of eye and head tracking data become quite similar. This suggests that people use their eyes more when looking downwards.

Eye tracking and hololens height profiles were similar, but hololens heights were more distributed.

The similarity between the eye tracking data and the hololens data suggest that the head movement behaviour of users wearing an augmented reality display with limited field of view is comparable to the eye movement behaviour of people viewing unaugmented scenes. The main difference between these two seems to be that the data of the unaugmented view condition is spread out more evenly. Especially in higher regions more points were recorded from participants wearing the Microsoft Hololens. This can be explained by the limited field of view. Participants are forced to move their head similar to the way they usually move their eyes to be able to find all augmentations. Additionally, they can not exclude areas from their search based on the information gathered via peripheral vision.

A similar behaviour can be seen in Figure 5.1, which shows a visualisation of the view points.
In conclusion, we reject the null hypothesis. The head tracking is quite different to other two cases. While the differences between them are subtle we think that the wider spread of the hololens data is evidence of a different behaviour.

4. Hypothesis
People wearing an augmented reality headset with limited field of view need the more time to find a combination of real and virtual objects, as people searching for only real items under normal conditions.

Last but not least we analysed the task completion time and success rate of the conditions unaugmented view and augmented view. Participants wearing the Microsoft Hololens needed a lot longer to find the search targets. In most cases, they needed longer than we wanted the trials to last, resulting in them not succeeding in the task. Thus, we reject the null hypothesis. Aside from the possible explanations, we already mentioned other aspects that could lead to this result. In our tests for the trials, we observed that participants had trouble focusing equally on the real and virtual objects. Leading to some cases were real object got overlooked because the participants were too focused on the digital augmentation. Another possible explanation is that the augmentations are less visible than the real objects, especially when hidden in an area of similar colours. However, this effect is counteracted by the holograms consisting of additional light and thus seeming to glow. In some circumstances this made the augmentations much easier to see. We observed all these effects while conducting the study.

Searching targets in augmented reality needs longer than under normal conditions.

Limitations

When planing and conducting our study we faced multiple limitations.

We focused on proof for effects instead of measuring their magnitude.

Our experiment mainly was designed to prove the existence of changes in the viewing behaviour. Therefore, we can only speculate about their magnitude. Understanding the magnitude of chagnes is out of the scope of this thesis and subject of future work. For the upcoming limitations this needs to be kept in mind. They limit the accuracy of our results, however their influences are not strong enough to influence our interpretation of them.

One limitation is the lack of robustness of our motion tracking system. The Vicon system is designed to be placed outside the arena, but because of the room constraints we had to place it within. Additionally, it would have been best to place the cameras above head height, which was not possible with our tripods. The resulting occlusion issues also reduced the reliability of our system. This led to the high amount of missing frames reported in Figure 5.2. However, it is unlikely that these missing frames would report a considerable different viewing behaviour of the participants.

Data points were lost, because our Vicon set-up lacked robust tracking.

Additionally, both conditions are influenced equally and thus the results are comparable.

A long pipeline led to inaccurate eye tracking data.

The accuracy of our eye tracking data was less than desirable. The eye tracker we used achieves its best accuracy if the objects which are viewed have the same distance to the participant and the eye tracker is calibrated for that distance. This was not possible in our set-up.
Additionally, a minor error in the calibration could also lead to loss of accuracy. The data was processed by two programs. Both had their own calibration step. This resulted in a long pipeline with multiple steps that could introduce errors.
We adjusted for this by using areas of interest for the analysis and doing other filtering steps. Furthermore, the introduced errors are different for each participant and they balance themselves out. In the final results their influence should be limited enough, that this effect does not alter our interpretation.

Chapter 6

Summary and future work

In this chapter, we summarise this thesis and the results we got form our experiments. Afterwards we outline some ideas for future research, which will conclude this thesis.

6.1 Summary

The view behaviour is influenced by the setting and important for application design.

The view behaviour of people has a huge impact on everyday tasks as well as interaction with digital devices. Understanding it can influence the design process for software and allow to create more natural interaction methods.
Today's optical see-through head mounted displays come with restrictions that limit the field of view in which augmentations can be shown. Because of this, the view behaviour of users is likely to change.
These two effects lead us to being interested in finding out what changes occur, because we believe that this understanding allows for a better interaction design

Similar research has been done, but not as fundamental as ours.

First, we gave insights to our field of research, attention in augmented reality, in chapter 2. We outlined influences of the gaze data on interaction with augmented reality that have been found [Vidal et al., 2014] and what other researchers tested in regards to the limited field of view of head mounted displays [Kishishita et al., 2014, Ens et al., 2016, Ren et al., 2016]. Additionally, we have compiled a selection of relevant background information, which we used for creating our hypotheses and our experiments in chapter 3. Here we presented aside from the technology we used, other aspects, such as the visual system, attention and augmented reality.

Our preliminary study informed the design of the main study.

In chapter 4.1, we described our preliminary study, which we used to design an office like room for our main experiment. We found that most office workers share a $24m^2$ room with one colleague. Additionally, we gathered detailed information on their workplace and other furniture which can be found in their office.

We conducted a study in which participants performed a search task while seeing an augmented reality or not seeing one.

After that we presented our own study in chapter 4.2. We started by defining the hypotheses. For the first we reasoned that the lack of peripheral guidance when looking for augmentations would lead to the participants viewing uninteresting regions more. In line with that argumentation we also expected the participants to move their head more when looking for digital content and used this as our second hypothesis. We also expected this effect to show in the height of the objects the participants looked at. Thus, our third hypothesis was, that our volunteers would look at different heights when wearing the mixed reality display. We observed that volunteers struggled to find real and virtual objects in the preliminary tests of the search task. This lead us to presume that participants searching for real and virtual objects would need longer to find all of them. This was our fourth and only task dependent hypothesis.
With these in place, we designed and conducted a user study. Participants were asked to look for small objects within an office like room while either wearing an eye tracker or an augmented reality display with limited field of view. We measured their viewing behaviour and task completion time.

After that we analysed the data in chapter 5, and found that the condition had a significant influence on the viewing behaviour of the participants. Participants moved their head in the augmented view condition similarly as they would

usually move their eyes. Albeit, slower and with a wider distribution around the whole room.
Additionally, the task completion time was significantly longer in the augmented view condition.

6.2 Conclusion

Based on the discussion of the individual hypotheses we have got an understanding of how the viewing behaviour of users of augmented reality displays with limited field of view differs from their normal viewing behaviour.

People move their head in a limited field of view scenario like they usually move their eyes.

The movements of the head in an augmented view condition become similar to the movements the eyes would usually do. The heat maps, the height distribution of view points and the area of interest percentages support this claim. However, the eye movements in the unaugmented view condition seem to be more focused on interesting regions, while the hololens data show a slightly higher distribution. The average angular distance of the head movement is not different between the two conditions. This aspect of the eye's movement behaviour seems not to be transferred. This could be because of a different comfort of the headwear in the conditions or just because the head has more mass to move, compared to the eyes.
We reason that this similarity exists because participants can only use the centre of their visual field to see holograms. Thus, they need to move the centre of their visual field over an area and cannot just move their eyes there. The wider distribution could stem from the loss of the peripheral vision when it comes to seeing holograms.

Search time in augmented reality is higher.

The time needed to find a mix of augmented and real objects in an augmented reality environment is much higher, than the time needed to complete a comparable task under normal conditions.
This could be, because of the lack of peripheral vision when searching for augmentations or because the difficulty to focus on real and virtual objects at the same time.

We proved the existence of effects but did not measure their magnitude.

Our thesis was designed to do fundamental research. We found proof of the aforementioned effects, but our results are not detailed enough to have a thorough understanding of them. In the upcoming future work section we present our ideas for experiments that can deepen the grasp on these effects.

6.3 Future work

Measuring the eye movement in augmented reality settings could improve the understanding.

If possible it would be interesting to incorporate eye tracking in the augmented view condition. This would allow for a better understanding how exactly the relationship between eye and head movement changes. Additionally, different sizes of field of view limitations could be tested. That would allow application of the results in different scenarios.

It could be possible to use head tracking instead of eye tracking.

We found a similarity between head tracking in the augmented view condition and eye tracking in the unaugmented view condition. This suggests that one could use head tracking instead of eye tracking for attention approximation in the context of mixed reality displays with a limited field of view. One could test this by comparing them in the context of attention sensitive applications.

The difference in task completion time could have different origins.

We did find a rather sizeable difference between the task completion times of the conditions. However, we have multiple explanations for that effect: inability to use the peripheral vision for search, difficulty to concentrate on real and virtual objects simultaneous and needing to move the head instead of the eyes but without an increase of the average head movement speed. Designing a test with these as independent variables could give insight into the interaction between the effects.

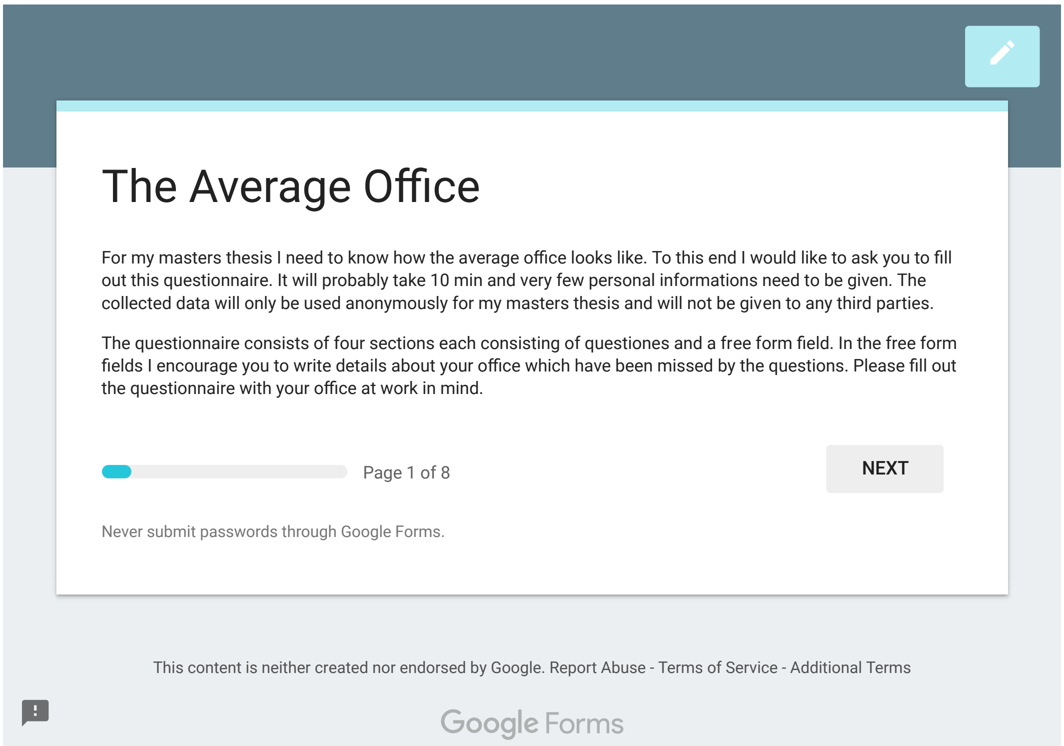
One could analyse finding rates and speeds of search targets.

Last but not least, we found that more uninteresting regions were viewed when the participant's field of view was constrained by the head mounted display. However, the difference was minor. As we did not record which search target was found how many times, we could not analyse how the placement affected the rate of finding. Another experiment focusing on arrangement and presentation of the targets could give more insight.

Appendix A

Room Content Survey

This appendix will show the survey we sent out for our preliminary study. The survey was created with Google Forms and was filled out online.


The Average Office

For my masters thesis I need to know how the average office looks like. To this end I would like to ask you to fill out this questionnaire. It will probably take 10 min and very few personal informations need to be given. The collected data will only be used anonymously for my masters thesis and will not be given to any third parties.

The questionnaire consists of four sections each consisting of questiones and a free form field. In the free form fields I encourage you to write details about your office which have been missed by the questions. Please fill out the questionnaire with your office at work in mind.

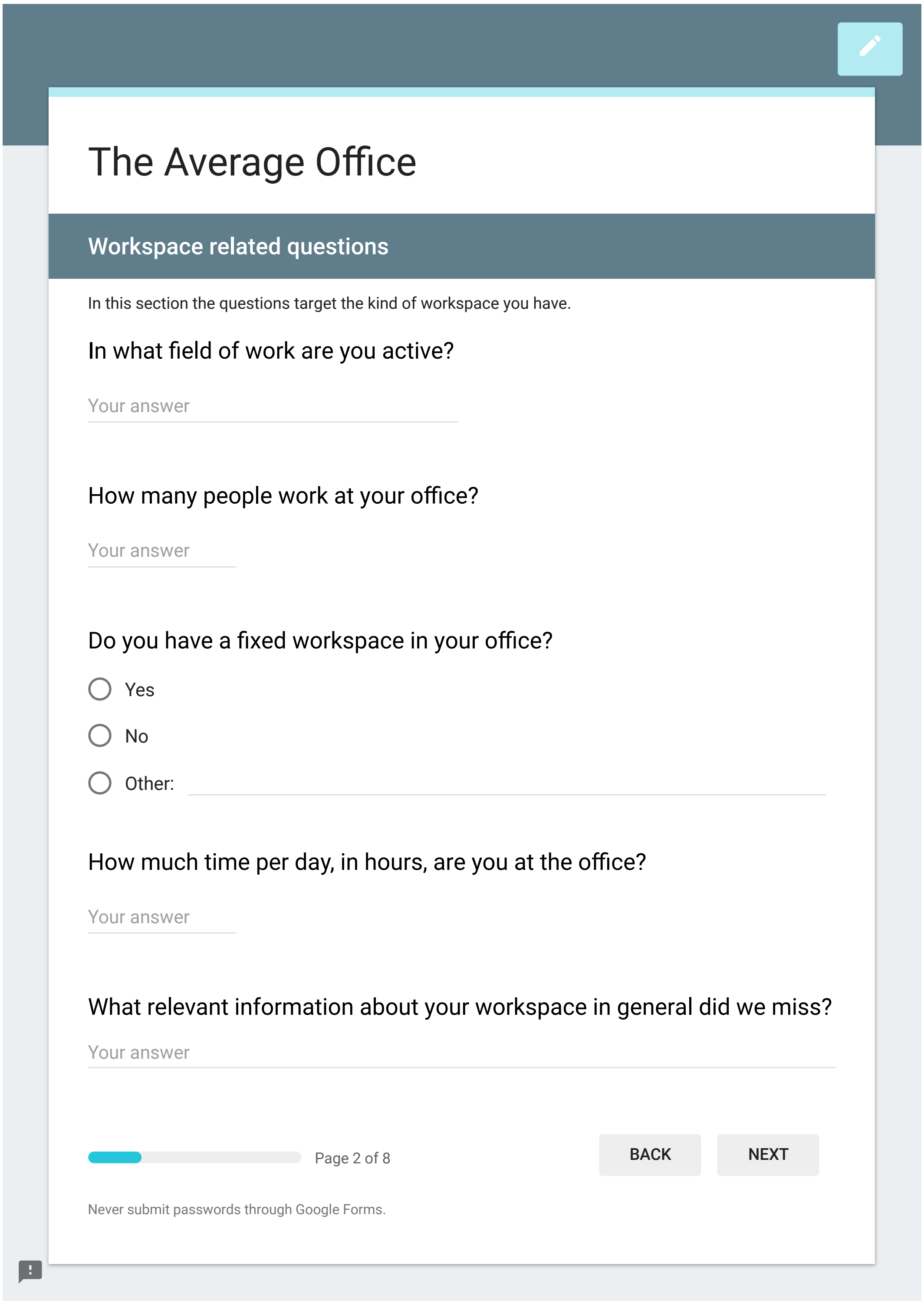
Page 1 of 8

NEXT

Never submit passwords through Google Forms.

This content is neither created nor endorsed by Google. Report Abuse - Terms of Service - Additional Terms

Google Forms



The Average Office

Workspace related questions

In this section the questions target the kind of workspace you have.

In what field of work are you active?

Your answer

How many people work at your office?

Your answer

Do you have a fixed workspace in your office?

- Yes
- No
- Other:

How much time per day, in hours, are you at the office?

Your answer

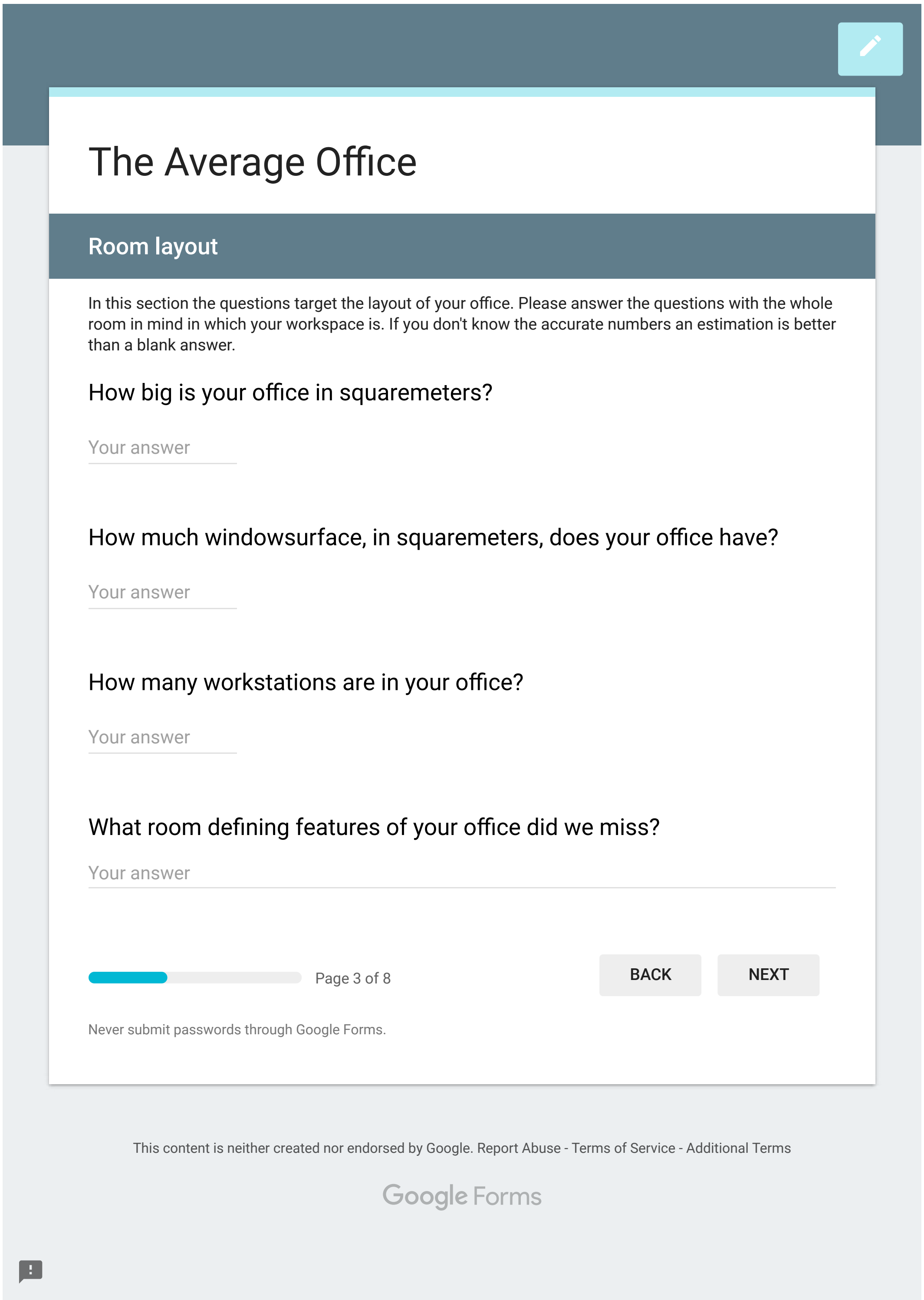
What relevant information about your workspace in general did we miss?

Your answer

Page 2 of 8

BACK NEXT

Never submit passwords through Google Forms.



The Average Office

Room layout

In this section the questions target the layout of your office. Please answer the questions with the whole room in mind in which your workspace is. If you don't know the accurate numbers an estimation is better than a blank answer.

How big is your office in squaremeters?

Your answer

How much windowsurface, in squaremeters, does your office have?

Your answer

How many workstations are in your office?

Your answer

What room defining features of your office did we miss?

Your answer

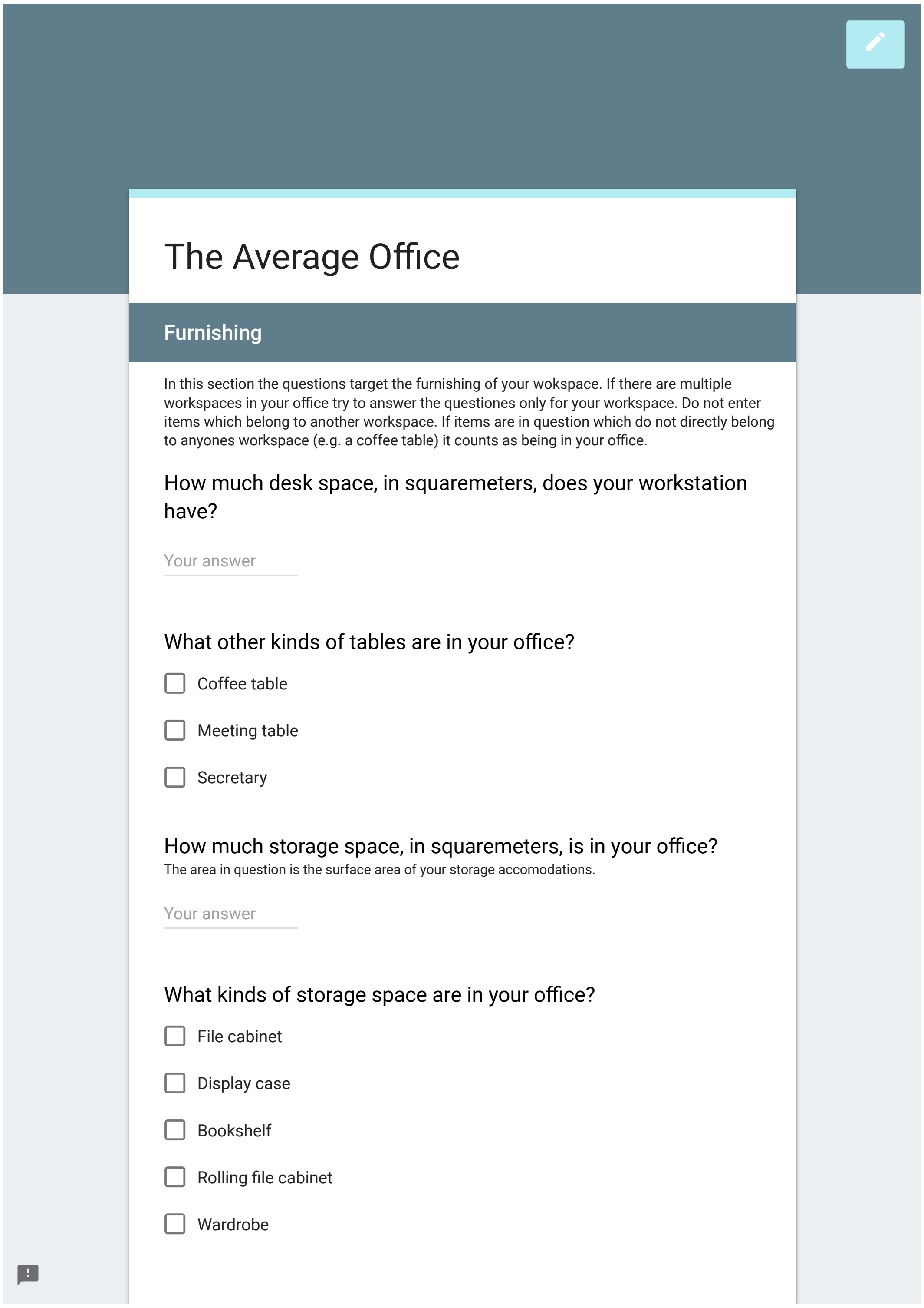
Page 3 of 8

BACK NEXT

Never submit passwords through Google Forms.

This content is neither created nor endorsed by Google. Report Abuse - Terms of Service - Additional Terms

Google Forms



The Average Office

Furnishing

In this section the questions target the furnishing of your wokspace. If there are multiple workspaces in your office try to answer the questiones only for your workspace. Do not enter items which belong to another workspace. If items are in question which do not directly belong to anyones workspace (e.g. a coffee table) it counts as being in your office.

How much desk space, in squaremeters, does your workstation have?

Your answer

What other kinds of tables are in your office?

- [ ] Coffee table
- [ ] Meeting table
- [ ] Secretary

How much storage space, in squaremeters, is in your office?
The area in question is the surface area of your storage accomodations.

Your answer

What kinds of storage space are in your office?

- [ ] File cabinet
- [ ] Display case
- [ ] Bookshelf
- [ ] Rolling file cabinet
- [ ] Wardrobe

What fills your storage space?

Don't waste too much time on precise numbers or the percentages adding up to 100. An estimation is enough.

	0%	1-20%	21-40%	41-60%	61-80%	81-100%
Stacks of paper						
Plants						
Files						
Folders						
Picture frames						
Technical appliances						
Books						
Office supplies						
Magazines						
Empty space						
Personal trinkets						
Food and Drinks						
Spare clothing						

How many seating accomodations are in your office aside from your desk chair?

Your answer

What kinds of seating accomodations are in your office?

- [ ] Stability ball
- [ ] Desk chair
- [ ] Sofa
- [ ] Armchair
- [ ] Balans chair
- [ ] Normal chair

How many lamps are in your office?

Your answer

What kinds of lamps are in your office?

- [ ] Indirect lighting
- [ ] Floor lamp
- [ ] Desk light
- [ ] Ceiling light
- [ ] Hanging light
- [ ] Candles

What other items furnish your office?

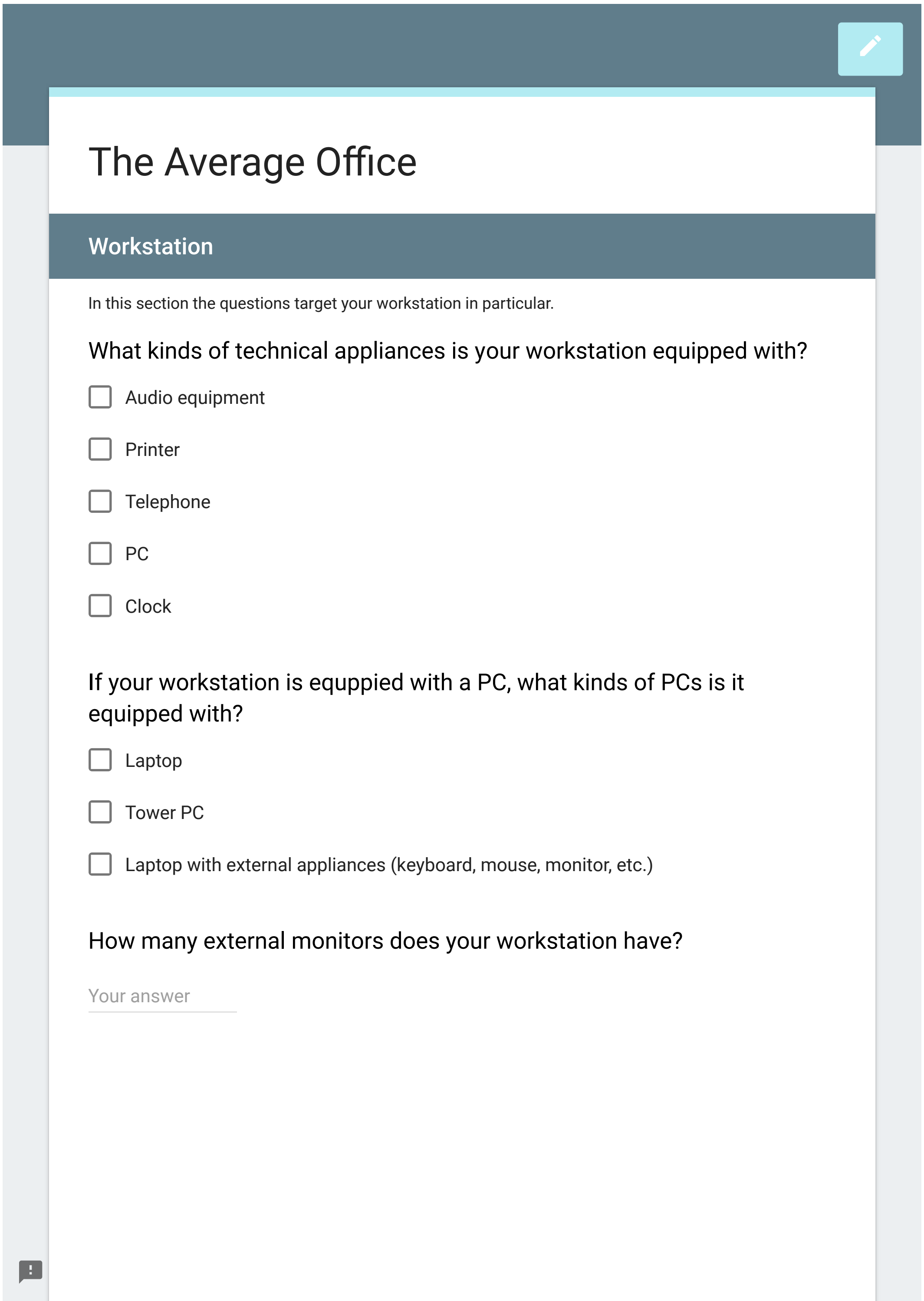
- [ ] Clock
- [ ] Flip chart
- [ ] Whiteboard

What furnishing in your office did we miss?

Your answer

BACK NEXT Page 4 of 8

Never submit passwords through Google Forms.



The Average Office

Workstation

In this section the questions target your workstation in particular.

What kinds of technical appliances is your workstation equipped with?

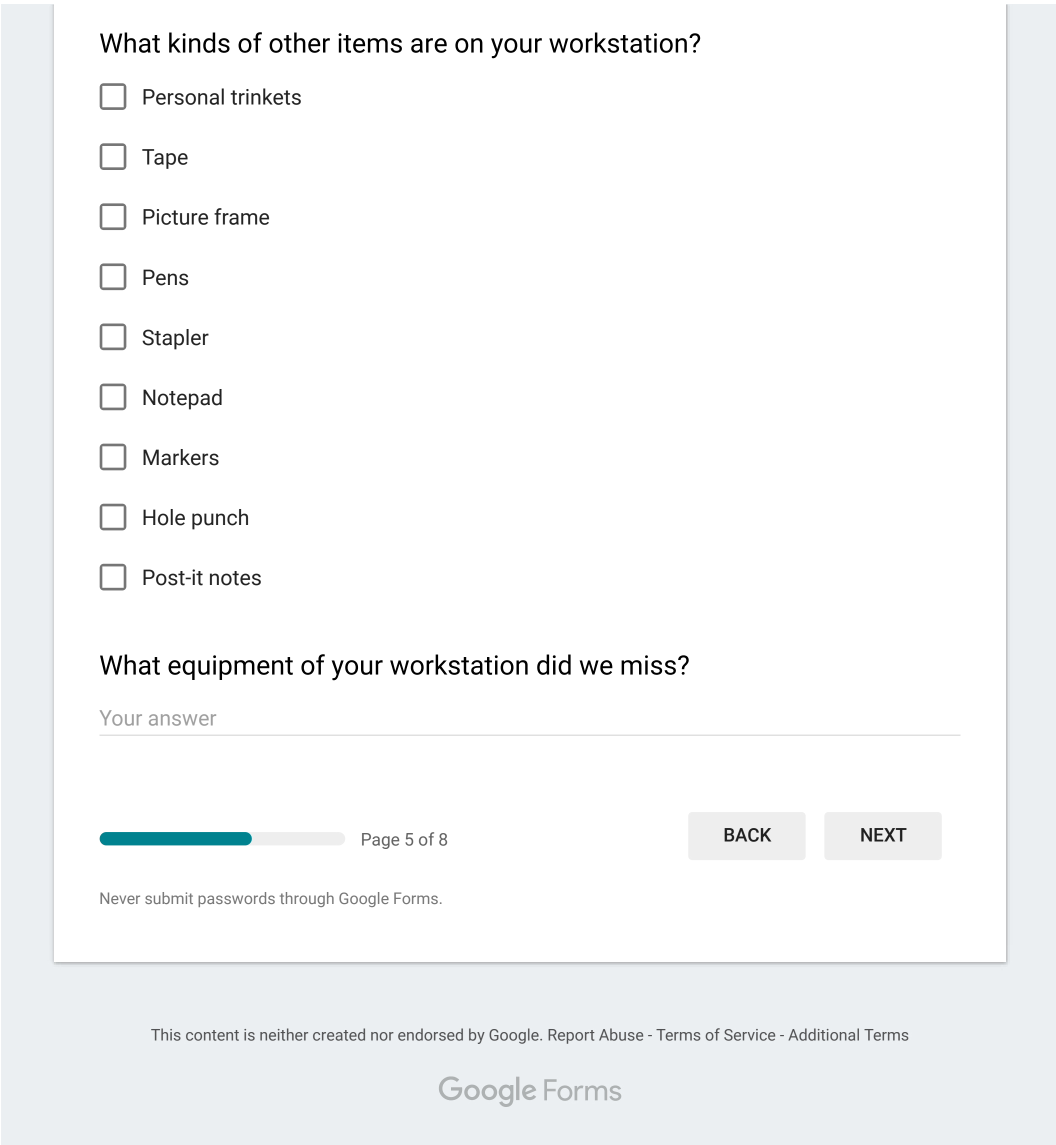
- [ ] Audio equipment
- [ ] Printer
- [ ] Telephone
- [ ] PC
- [ ] Clock

If your workstation is equppied with a PC, what kinds of PCs is it equipped with?

- [ ] Laptop
- [ ] Tower PC
- [ ] Laptop with external appliances (keyboard, mouse, monitor, etc.)

How many external monitors does your workstation have?

Your answer


What kinds of other items are on your workstation?

- [ ] Personal trinkets
- [ ] Tape
- [ ] Picture frame
- [ ] Pens
- [ ] Stapler
- [ ] Notepad
- [ ] Markers
- [ ] Hole punch
- [ ] Post-it notes

What equipment of your workstation did we miss?

Your answer

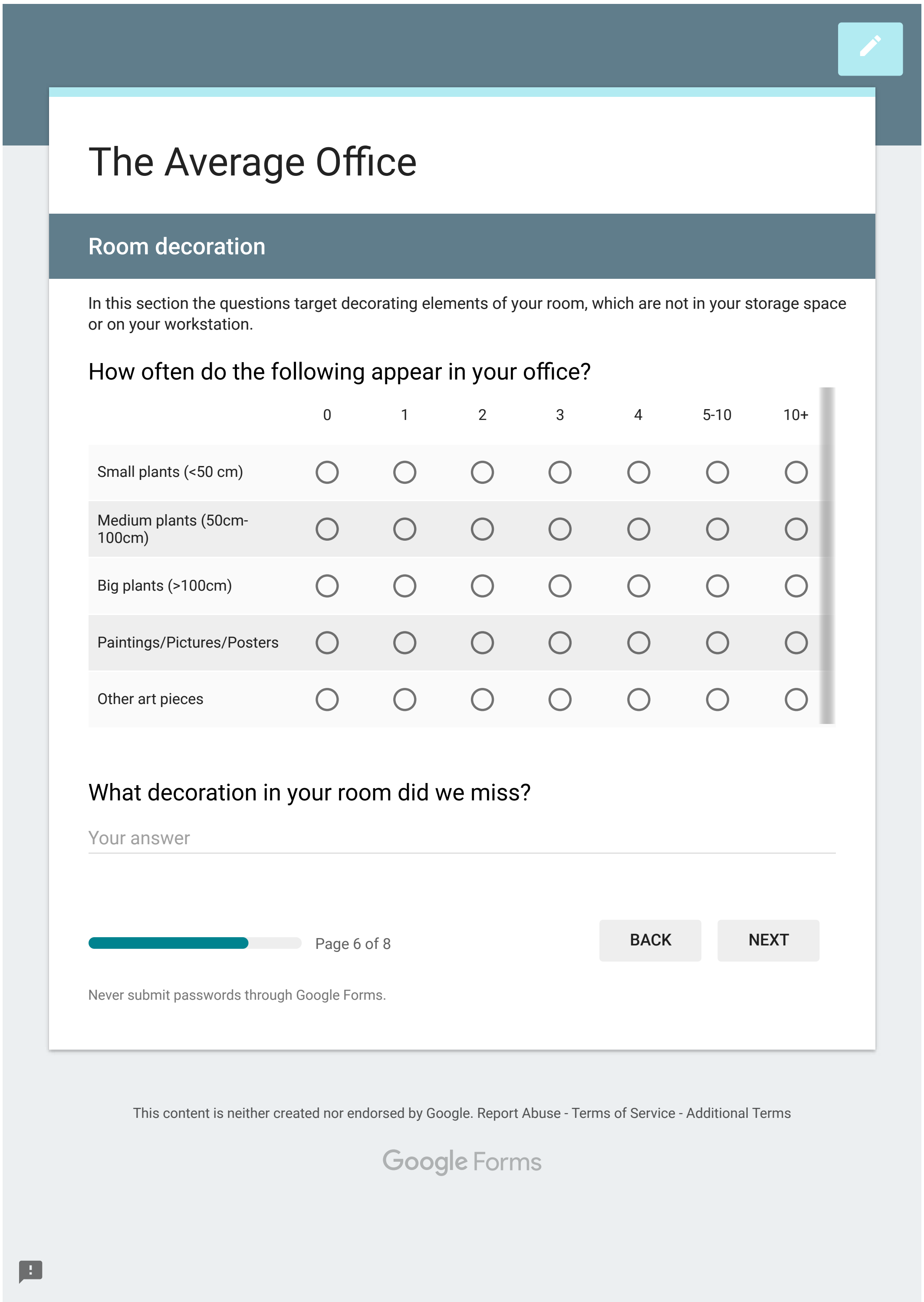
Page 5 of 8

BACK NEXT

Never submit passwords through Google Forms.

This content is neither created nor endorsed by Google. Report Abuse - Terms of Service - Additional Terms

Google Forms



The Average Office

Room decoration

In this section the questions target decorating elements of your room, which are not in your storage space or on your workstation.

How often do the following appear in your office?

	0	1	2	3	4	5-10	10+
Small plants (<50 cm)							
Medium plants (50cm-100cm)							
Big plants (>100cm)							
Paintings/Pictures/Posters							
Other art pieces							

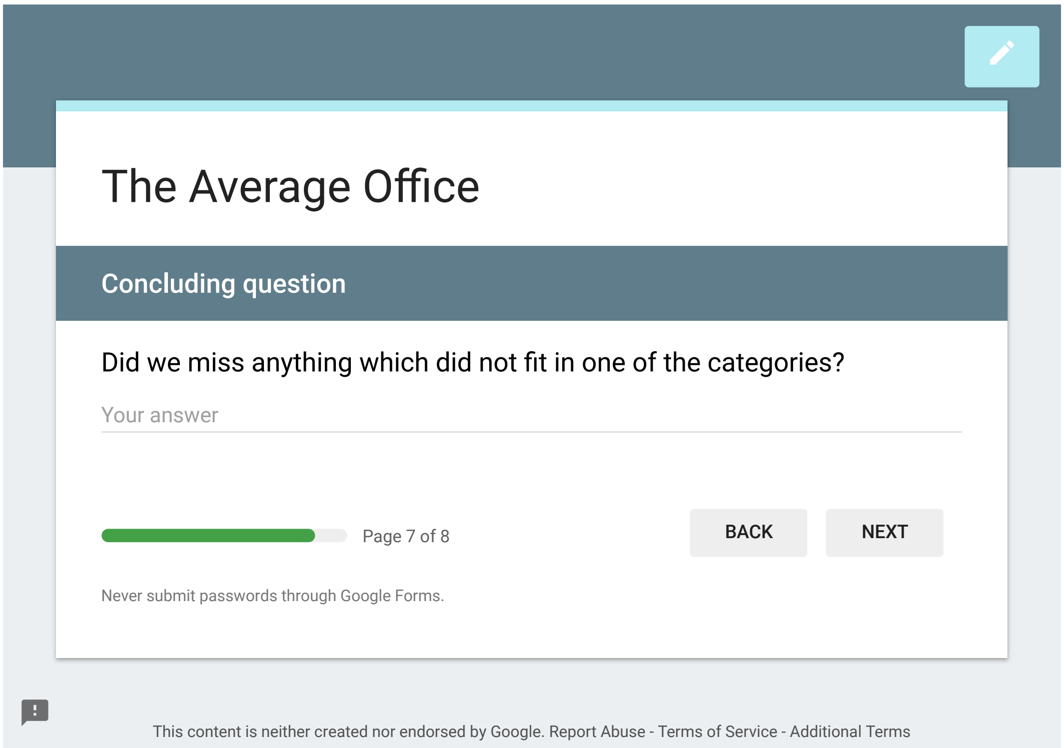
What decoration in your room did we miss?

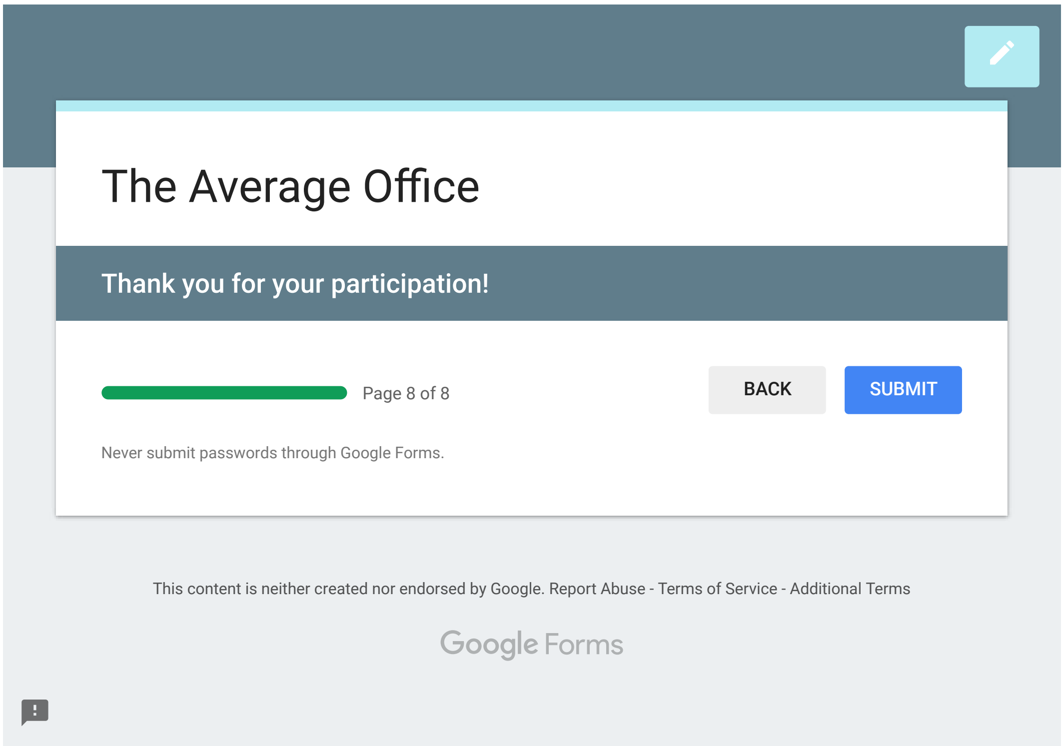
Your answer

Page 6 of 8

BACK NEXT

Never submit passwords through Google Forms.


The Average Office
Concluding question
Did we miss anything which did not fit in one of the categories?
Your answer
Page 7 of 8
BACK
NEXT
Never submit passwords through Google Forms.
This content is neither created nor endorsed by Google. Report Abuse - Terms of Service - Additional Terms


The Average Office
Thank you for your participation!
Page 8 of 8
BACK
SUBMIT
Never submit passwords through Google Forms.
This content is neither created nor endorsed by Google. Report Abuse - Terms of Service - Additional Terms
Google Forms

Appendix B

Consent Form

Here we show the consent form we asked the participants of our main study to sign.

Informed Consent Form

Analysing view behaviour in the presence of holograms

PRINCIPAL INVESTIGATOR Niklas Kaulitz
Media Computing Group
RWTH Aachen University
Email: niklas.kaulitz@rwth-aachen.de

Purpose of the study: The goal of this study is to understand how holograms affect how users view an unfamiliar room. Participants will be asked to find objects in a defined area while wearing an eye tracker. Head and eye motion will be used for analysis.

Procedure: Participation in this study will involves two phases. In the first phase you will be asked to put on the eye tracker or Hololens and get acquainted with the system. In the second phase, you will be asked to look at a office like room and point out certain, previously shown, objects. This study should take about half an hour to complete.

After the study, we will ask you to fill out a questionnaire. In this questionnaire, we will gather some personal data.

Risks/Discomfort: You may become fatigued during the course of your participation in the study. There are no other risks associated with participation in the study. Should completion of either the task or the questionnaire become distressing to you, it will be terminated immediately.

Benefits: The results of this study will be useful for understanding the way users change their viewing behaviour if holograms are present. Which in turn may help design interactive systems with holograms.

Alternatives to Participation: Participation in this study is voluntary. You are free to withdraw or discontinue the participation.

Cost and Compensation: Participation in this study will involve no cost to you. There will be snacks for you during and after the participation.

Confidentiality: All information collected during the study period will be kept strictly confidential. You will be identified through identification numbers. No publications or reports from this project will include identifying information on any participant. If you agree to join this study, please sign your name below.

_____ I have read and understood the information on this form.
_____ I have had the information on this form explained to me.

Participant's Name	Participant's Signature	Date
	Principal Investigator	Date

If you have any questions regarding this study, please contact Niklas Kaulitz email: niklas.kaulitz@rwth-aachen.de

Bibliography

A. Al-Rahayfeh and M. Faezipour. Eye Tracking and Head Movement Detection: A State-of-Art Survey. *IEEE Journal of Translational Engineering in Health and Medicine*, 1: 2100212–2100212, 2013. ISSN 2168-2372. doi: 10.1109/JTEHM.2013.2289879.

S. M. Anstis. A chart demonstrating variations in acuity with retinal position. *Vision Research*, 14(7):589–592, 1974. ISSN 0042-6989. doi: 10.1016/0042-6989(74)90049-2.

O. Bimber and R. Raskar. Modern approaches to augmented reality. In *ACM SIGGRAPH 2006 Courses*, SIGGRAPH '06, 2006. ISBN 1-59593-364-6. doi: 10.1145/1185657.1185796.

J. Carmigniani, B. Furht, M. Anisetti, P. Ceravolo, E. Damiani, and M. Ivkovic. Augmented reality technologies, systems and applications. *Multimedia Tools and Applications*, 51(1):341–377, 2011. ISSN 1380-7501, 1573-7721. doi: 10.1007/s11042-010-0660-6.

M. S. Castelhano, M. L. Mack, and J. M. Henderson. Viewing task influences eye movement control during active scene perception. *Journal of Vision*, 9(3):6–6, 2009. ISSN 1534-7362. doi: 10.1167/9.3.6.

G. Cheng, Y. Wan, A. N. Saudagar, K. Namuduri, and B. P. Buckles. Advances in Human Action Recognition: A Survey. *arXiv:1501.05964 [cs]*, 2015.

A. B. Chica, P. Bartolomeo, and J. Lupiáñez. Two cognitive and neural systems for endogenous and exogenous spatial attention. *Behavioural Brain Research*, 237:107–123, 2013. ISSN 0166-4328. doi: 10.1016/j.bbr.2012.09.027.

B. Czupryński and A. Strupczewski. High Accuracy Head Pose Tracking Survey. In *Active Media Technology*, Lecture Notes in Computer Science, pages 407–420. Springer, Cham, 2014. ISBN 978-3-319-09911-8 978-3-319-09912-5. doi: 10.1007/978-3-319-09912-5_34.

A. Duchowski. *Eye Tracking Methodology - Theory and Practice | Andrew Duchowski | Springer*. Springer, 2 edition, 2007. ISBN 978-1-84628-609-4.

B. Ens, D. Ahlström, and P. Irani. Moving Ahead with Peephole Pointing: Modelling Object Selection with Head-Worn Display Field of View Limitations. In *Proceedings of the 2016 Symposium on Spatial User Interaction*, SUI '16, pages 107–110, 2016. ISBN 978-1-4503-4068-7. doi: 10.1145/2983310.2985756.

A. Hoang Duc, P. Bays, and M. Husain. Eye movements as a probe of attention. In *Progress in Brain Research*, volume 171, pages 403–411. Elsevier, 2008. ISBN 978-0-444-53163-6. doi: 10.1016/S0079-6123(08)00659-6.

N. Kishishita, K. Kiyokawa, J. Orlosky, T. Mashita, H. Takemura, and E. Kruijff. Analysing the effects of a wide field of view augmented reality display on search performance in divided attention tasks. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 177–186, 2014. doi: 10.1109/ISMAR.2014.6948425.

K. Öllerer. How to make a PIXAR movie - Motion Capturing, 2011.

M. Mills, A. Hollingworth, S. van der Stigchel, L. Hoffman, and M. D. Dodd. Examining the influence of task set on eye movements and fixations. *Journal of Vision*, 11(8):17–17, 2011. ISSN 1534-7362. doi: 10.1167/11.8.17.

E. Murphy-Chutorian and M. M. Trivedi. Head Pose Estimation in Computer Vision: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(4):607–626, 2009. ISSN 0162-8828. doi: 10.1109/TPAMI.2008.106.

R. Poelman and D. W. F. van Krevelen. A Survey of Augmented Reality Technologies, Applications and Limitations. page 22, 2010.

D. Ren, T. Goldschwendt, Y. Chang, and T. Höllerer. Evaluating wide-field-of-view augmented reality with mixed reality simulation. In *2016 IEEE Virtual Reality (VR)*, pages 93–102, 2016. doi: 10.1109/VR.2016.7504692.

E. F. Risko and A. Kingstone. Eyes wide shut: implied social presence, eye tracking and attention. *Attention, Perception, & Psychophysics*, 73(2):291–296, 2011. ISSN 1943-3921, 1943-393X. doi: 10.3758/s13414-010-0042-1.

A. Takagi, S. Yamazaki, Y. Saito, and N. Taniguchi. Development of a stereo video see-through HMD for AR systems. In *Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)*, pages 68–77, 2000. doi: 10.1109/ISAR.2000.880925.

D.K. Tiller and J.A. Veitch. Perceived room brightness: Pilot study on the effect of luminance distribution. *International Journal of Lighting Research and Technology*, 27(2):93–101, 1995. ISSN 1365-7828. doi: 10.1177/14771535950270020401.

M. Vidal, D. H. Nguyen, and K. Lyons. Looking at or Through?: Using Eye Tracking to Infer Attention Location for Wearable Transparent Displays. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers*, ISWC '14, pages 87–90, 2014. ISBN 978-1-4503-2969-9. doi: 10.1145/2634317.2634344.

B. A. Wandell and S. J. Thomas. *Foundations of Vision*. 1997. doi: 10.1002/col.5080210213.

D. Weinland, R. Ronfard, and E. Boyer. A survey of vision-based methods for action representation, segmentation and recognition. *Computer Vision and Image Understanding*, 115(2):224–241, 2011. ISSN 1077-3142. doi: 10.1016/j.cviu.2010.10.002.

J. M. Wolfe. Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, 1(2):202–238, 1994. ISSN 1069-9384, 1531-5320. doi: 10.3758/BF03200774.

R. Xiao and H. Benko. Augmenting the Field-of-View of Head-Mounted Displays with Sparse Peripheral Displays. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 1221–1232,

2016. ISBN 978-1-4503-3362-7. doi: 10.1145/2858036.2858212.

A. L. Yarbus. *Eye Movements and Vision*. Springer US, 1967. ISBN 978-1-4899-5379-7.

L. R. Young and D. Sheena. Survey of eye movement recording methods. *Behavior Research Methods & Instrumentation*, 7(5):397–429, 1975. ISSN 1554-351X, 1554-3528. doi: 10.3758/BF03201553.

F. Zhou, H. B. Duh, and M. Billinghurst. Trends in Augmented Reality Tracking, Interaction and Display: A Review of Ten Years of ISMAR. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, ISMAR '08, pages 193–202, 2008. ISBN 978-1-4244-2840-3. doi: 10.1109/ISMAR.2008.4637362.

Index

Typeset July 28, 2018