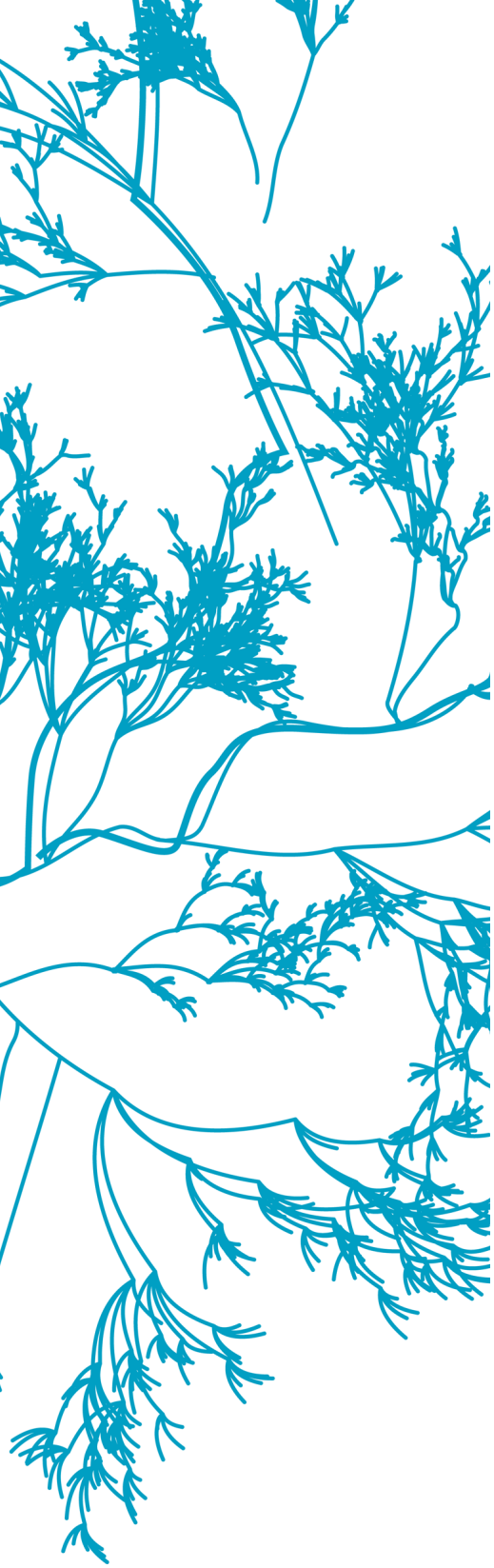


MSc Interaction Technology
Master Thesis

Creating an Augmented Reality Digital Mirror: Testing its Feasibility and Exploring Applications for Self-Perception Research

Rens van der Werff

Supervisors:
M.A. Gómez Maureira, PhD
dr. M.A. Friehs
dr.ir. D. Reidsma

May, 2025

Department of Human Media Interaction
Faculty of Electrical Engineering,
Mathematics and Computer Science,
University of Twente

UNIVERSITY OF TWENTE.

Contents

Abstract

This thesis presents the ideation, design, and evaluation of an augmented reality digital mirror for self-perception related research. It is intended as a more accessible and physically grounded alternative to head-mounted display (HMD) Virtual Reality (VR), which, despite its widespread use, can lead to reduced embodiment, increased cognitive load, and other limitations. The proposed system, called MIRA (Mirror for Interactive Reality Augmentations), was designed to reproduce physically accurate reflections while enabling real-time visual augmentations. A user study ($n = 35$) explored the technical and perceptual feasibility of the system and its capacity to trigger self-referential behaviour. Results showed that 63% of participants exhibited behavioural reactions to a virtual stain placed on their clothing in the mirror image, suggesting effective attention capture and self-association. However, the system suffered from high latency ($\sim$500 ms), which disrupted the natural feel of the mirror and limited participants' belief in the augmentations. Over half of the participants misinterpreted the mirror's behaviour, possibly due to this limitation, or a fundamental misunderstanding of mirror workings. These issues can be overcome, indicating strong potential for the system as a tool for self-perception research. The thesis concludes by formulating practical design guidelines for the creation of similar digital mirror systems and highlights avenues for future work.

Acknowledgements

I would like to express my sincere gratitude to my daily supervisors Maro Gómez Maureira and Max Friehs for sticking with me throughout my graduation process. They remained positive and eager to provide great feedback and encouragement, even if the work was sometimes progressing (very) slowly. I really appreciated this, you were amazing.

Admittedly, the work presented here went beyond the original scope, as I didn't want to rush something out just to finish. That's why Chapter 3 turned out so long. This thesis has been a valuable learning experience, and I can genuinely say I'm proud of what I've accomplished.

I would also like to thank the Interaction Lab for allowing me to use the lab facilities as and when I needed. It allowed me to try many different approaches to my research, which was invaluable. Special thanks to Daniel Davison, who consistently came up with great ideas and suggestions throughout the project, and kept me employed. The lab also provided a fantastic place to study and work (and occasionally do neither).

Finally, I would like to thank my friends, family, and girlfriend, who kept asking me when I'd finally be done, but never stopped supporting me. My gratitude is also extended to everyone that participated in my user study.

Chapter 1

Introduction

1.1 Problem statement

Extended Reality (XR) technologies, especially head-mounted-display (HMD) Virtual Reality (VR), have gained significant attention in various domains over the last few decades. These technologies have solidified themselves as versatile and powerful tools, and offer new possibilities in many areas of research, an important one being psychology and neuroscience [1]. This thesis focuses on one particular subdomain of psychology research, namely self-perception, in which HMD VR, often referred to as just VR, is used extensively.

Self-perception is a psychological phenomenon that involves individuals observing and interpreting their own actions, emotions, and cues in their environment to develop a self-image [2]. Current self-perception research heavily relies on HMD-based VR, which, despite its flexibility, often disconnects users from the real world and introduces issues like reduced embodiment and increased cognitive load [3]. This raises the question of whether alternative XR technologies, particularly those more integrated into the real world, can provide more natural, accessible, and equally effective platforms for studying self-perception.

1.2 Research objective

This thesis will describe the ideation, design, and evaluation of a novel alternative to head-mounted-display VR, in the domain of self-perception research. This takes the shape of an augmented reality digital mirror, named MIRA (Mirror for Interactive Reality Augmentations). Due to its novelty, design guidelines will be set up for the (re)creation of such a system.

In order to achieve the above mentioned goals, the following research questions need to be answered, starting with the main research questions:

mRQ1 To what extent can an augmented reality mirror influence individuals' self-perception and body image?

mRQ2 How well can a digital mirror reproduce physically accurate reflections and meet user expectations?

Some sub-research questions are also established to gain further insight.

sRQ1 What types of self-perception research are enabled by a digital mirror?

sRQ2 How do individuals perceive the digital mirror's realism and behaviour compared to a real mirror?

sRQ3 What design guidelines can be set up for creating digital mirrors?

The background chapter will elaborate on the process and reasoning that led to the eventual concept of the augmented reality mirror, with physically accurate reflections. It will describe how existing literature, related technologies, and research gaps motivated the creation of MIRA as an alternative approach to self-perception research.

Chapter 2

Background

To get a better understanding about the design space, the psychological phenomenon of self-perception must first be examined, forming an understanding of its importance and what type of interactive technologies within the XR domain it is used for. Different technologies are then dissected based on their suitability for self-perception research, in order to reveal novel, alternative, design and research opportunities.

2.1 Self-perception

Self-perception is a psychological phenomenon that plays a role in many fields of psychology, especially in social and cognitive psychology. Social psychology research focuses on how a person influences, and is influenced by, others and their environment. Self-perception follows a similar principle, but instead of observing others, individuals construct a self-image by interpreting their own actions, emotions, and environmental cues [2], in turn affecting their behaviour in a social setting. It is not exactly the same as the perception of others, as peoples behaviour changes when they are alone or with others. The focus of attention is also different: when perceiving others, moral contents take priority, while competence is most important in self-perception [4]. The process of self-perception is always active, although the extent to which it happens can differ significantly between individuals. Self-perception can also be invoked manually by reviewing one's actions actively, such as reflecting on past decisions.

The theory of self-perception was originally proposed by Bem [5], as an alternative interpretation of the cognitive dissonance phenomena. Cognitive dissonance occurs when an individual holds two psychologically inconsistent thoughts, which causes psychological discomfort [6]. Cognitive dissonance theory assumes that individuals try to avoid this discomfort by changing one or both of their thoughts to align them. What this theory failed to account for, and what self-perception theory aims to cover, is that people involved in cognitive dissonance experiments are essentially judging themselves and their attitudes. These judgements can lead to changes in the attitude of the participants, and it can therefore not be assumed that people will solely strive for consistency [5]. This effect only exists within the confines of cognitive dissonance experiments, as looking more broadly, high self-perception results in more consistent and predictable behaviour, and solidifies opinions based on past experiences [7].

The broad applicability and relevance of self-perception in many aspects of daily life is evident, as people are confronted with themselves daily, in pictures on their phone, in the mirror or by the consequences of their actions. It is therefore important to explore further into the subdomains of self-perception, namely self-association and self-representation, to get a clearer picture of the different factors influencing the self-image.

2.1.1 Self-association

Self-perception is not physically limited to the observing person, but can extend to external, self-associated stimuli: something a person identifies with, or knows well. These stimuli are prioritised when processing information and the processing happens faster, compared to non-self-associated stimuli [8]. This is called the 'self-prioritisation effect' (SPE). It occurs consistently and can be used to direct the attention of a person [9]. Mundane stimuli that the user is familiar with, will often take priority purely based on the strength of familiarity [10]: people prefer a familiar tool over an objectively better, unfamiliar tool. This is different from more personal, self-associated stimuli, whose physiological response runs deeper and trigger SPE. And while self-associated stimuli are inherently personal, it can also be experimentally induced to achieve the same effect [11].

A prime example of this experimentally induced self-association is the famous rubber hand illusion [12]. In this experiment, Botvinick and Cohen placed a rubber hand in front of participants, while hiding their real hand from sight, then stimulated both hands with the same stimuli (e.g. a feather stroking). This tricked the participants into thinking they were looking at their real hand, which then caused participants to report feeling stimuli that only occurred on the rubber hand. More recent research repeated this experiment using virtual body augmentation and found that it could invoke similar psychological responses [13, 14]. What the rubber hand (and consecutive research) shows, is that a psychological connection can be created with an external self-associated stimulus, as if it were an extension of the body [15].

2.1.2 Self-representation

Games, applications, and research often feature entities that represent the user. Letting the user directly control this representation can be classified as induced self-association [11]: the user creates a connection with the representation of themselves. Even simple geometric shapes can trigger the SPE if the user is informed it represents them [16]. Combining this with the act of 'puppeteering', or controlling a character, which also creates a connection [17], means simple non-human characters can be used for self-representation in games and other applications.

FIGURE 2.1: The default Mii characters as shown on the Nintendo Wii.

According to the definition, any object can be used as an avatar: a virtual representation of a person and something they can see as themselves, though it is often in humanoid shape, as it is the most recognisable shape, with the highest self-association. These avatars create a connection between the user and the virtual environment, even if the avatar shares no resemblance to the user [18]. This is possible in high-immersion environments, as looking more broadly, higher resemblance directly correlates with greater self-association [19], which in turn may yield increased commitment to given tasks [20]. This greater self-association is also found when users were allowed to create and customize their own character [20, 21, 22, 23], giving them an increased sense of ownership.

In all cases, the face and upper body were found most important, as self-associations here caused a stronger effect [20]. Even small self-associations, such as a caricature of someone's head with some key self-relevant facial features (eye colour, nose shape, etc.), like a Mii character shown in Figure 2.1, triggers a higher self-representation [24]. This area of the body is often subjectively considered as the location of the 'self' [25], and it has been shown that when neutral stimuli are presented near this location, they were integrated into the self and given priority over those farther away [26]. As might be expected, they also triggered the SPE [20].

For this research, the key takeaway is that the location of the 'self', where higher self-association occurs, should provide the best resemblance, and thus shows most potential for visual augmentation.

2.2 State of the art

In order to find a novel technological application for self-perspective research, the status quo is explored. Interactive technologies used in self-perception research often fall within the XR domain, as it facilitates augmentation of the user or their environment, as well as testing for corresponding reactions. This makes it suitable for f.e. body-dysmorphia research, the size and shape of the user's body is manipulatable. XR also allows for virtual additions to scenes, f.e. a virtual spider that crawls over the hand of the participant, which can be used in phobia treatment. These two examples form a solid metric that can be used to practically test the suitability of each technology.

XR, or Extended Reality, is an umbrella term for all technologies that augment reality. These technologies can be placed on a linear scale based on their virtuality, starting at physical reality and ending at completely virtual. It is likely due to the relative infancy of the field that research papers and companies utilise inconsistent terms and definitions, as well as categorisations, for these technologies. However, a vast body of research [27, 28, 29, 30] agrees on the two main technologies: augmented reality (AR) and virtual reality (VR). This is portrait in Figure 2.2, which combines the work mentioned before, as well as that of Mujber et al. [31] on VR types and Estrada et al. [32] on AR types, in one figure.

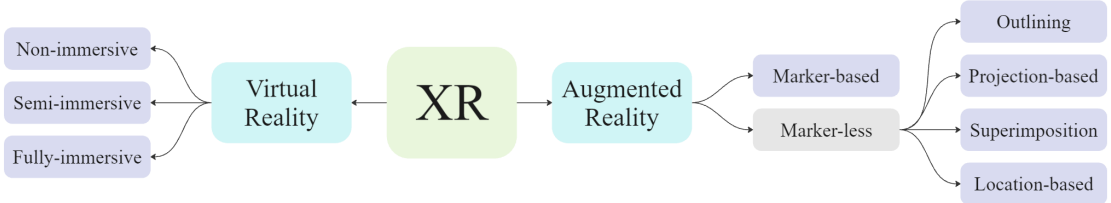


FIGURE 2.2: Combined overview of technology in the XR domain.

2.2.1 Virtual Reality

Due to head-mounted-display VR's rapid rise in popularity over the last decade, it is often used interchangeably with virtual reality, it is however, only a category of VR. Virtual Reality is defined as "an advanced human-computer interface that simulates a realistic environment" [30, p. 20]. This means that a video game on a computer screen also falls within this definition, which would be classified as non-immersive VR as suggested by Mujber et al. [31], who divide VR into three different categories: non-, semi- and fully-immersive (as seen in Figure 2.2).

Related work

All three types of VR are useful tools within many branches of research, as they are versatile and powerful. Each type offers unique possibilities, both technically and within the field of self-perception research. HMD VR (fully-immersive) is the most powerful of the technologies, as it offers the highest virtuality and thus the most customisability and immersion. Non-immersive VR also offers interesting possibilities, despite its relatively lower immersion. Using digital avatars in online meetings has been shown to increase perceived social presence and provide a general increase in user experience, as users feel physically present with their colleagues [33]. Having such an avatar that resembles the user [24] (and by extension others) or one that the user simply controls [17], even on a basic screen setup, can form a sub-conscience connection and trigger self-association responses.

In fully-immersive VR, every part of the real world is blocked out, with the possible exception of the person themselves. Besides the HMDs mentioned before, this can take the shape of a system like CAVE (Cave Automatic Virtual Environment), as originally conceived by Cruz-Neira et al. [34], and shown in Figure 2.3, where all the walls in a room are transformed into screens that change their projection depending on the user's position, giving the illusion of a virtual world. This effect is not seamless however, and can fall apart quickly by moving too close to the walls. This setup, and others like it, are large and expensive installations and a similar effect can be (arguably better) achieved using the more conventional HMD approach.

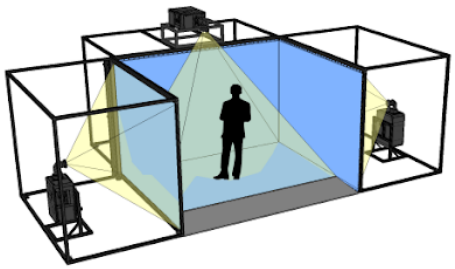


FIGURE 2.3: CAVE Automatic Virtual Environment [34].

HMDs track the user's head and body movement, and display a virtual world accordingly on a display behind two lenses. Because of the lenses, the screen is perceived sharply and view-filling. Recent advances in computer hardware have also made it possible for some HMDs to work 'stand-alone', meaning no external cables or devices are required, freeing up the movement of the user. Mirroring the physical world, actions in HMD VR are performed using head and arm movement. This makes the interactions intuitive and,

combined with the displays that cover (most of) the peripheral vision, offers vastly superior levels of immersion and engagement over standard screens [35, 36, 37].

HMD VR allows virtually altering a person's appearance. This is mostly used in video games to communicate the player's occupation and state. A popular use in research is to find out what effect a change to appearance has on a user: a vast body of self-perception research uses HMD VR for this reason. Interestingly, the psychological phenomena of the rubber hand illusion [12] also applies in a completely virtual environment, as a virtual arm can be perceived as a part of one's self [13]. This also extends to an unfamiliar virtual body [14, 38], which is supported by van Gisbergen et al. [18], finding that it is "not a requirement to develop high avatar-owner resemblance in highly immersive virtual environments". This somewhat contradicts the claims from Ratan and Dawson [24], who stated that higher user resemblance invoke more psychological responses. It appears that this is not required in the case of VR, possibly due to the feedback loop that occurs when the avatar mimics the user's movement: this being the most extreme example of puppeteering [17]. These findings show to facilitate high self-association, either the resemblance or immersion should be high, and one can compensate for the other. While research shows that slowly altering the body into something non-human, such as stone, can be perceived as real by participants when given the proper stimuli [39], having an avatar that resembles the users should still invoke enhanced information processing and self-association over an 'alien' avatar [19, 20, 21, 22, 23, 24].

This research shows that avatar augmentation in VR can retain users' feeling of ownership of their body, suggesting it is a suitable technology in use for the aforementioned body-dysmorphia research example. It is also suitable for the other example: phobia treatment, as projecting virtual spiders crawling on the users' body, using VR therapy was shown effective [40]. Besides this, VR shows significant promise in the domain of exposure therapy in general [41], confirmed by a meta-analysis [42] that suggests that VR is slightly more effective than conventional treatment, with the difference being statistically significant.

HMD pros and cons

All these examples show that fully-immersive VR is a powerful tool and a logical choice for self-perception research, and highlights the reasons for its popularity among XR technologies. While Mujber et al. [31] state in their comparison that fully-immersive VR is far more expensive than the alternatives, this no longer holds true. HMD hardware has developed into a mainstream technology since 2004 and is now affordable for general consumer audiences. The wide availability of tools also make it easy to develop with.

Fully-immersive VR's biggest strength, allowing total manipulation of the world the user is in, is coincidentally its largest downside. As it is at the end of the virtuality scale, there is no more 'real' left, visually. Even with the puppeteering effect [17], it can occur that users distance themselves cognitively from the application, potentially negatively impacting its intended effects. This is especially relevant for self-perception research where the 'self' is typically at the centre, not the 'virtual self'. It is also worth noting that some people suffer from cybersickness in VR, making VR a non-solution for this group.

HMD VR is thus a powerful, but flawed technology when examining it through the self-perception lens. This highlights the potential for another, alternative, technology within the XR domain to suite this area of research better. The other VR types, non- and semi-immersive, did not show qualifiable capabilities. Therefore the AR subdomain should be examined.

2.2.2 Augmented Reality

On the scale of virtuality, augmented reality (AR) would appear in the middle. This is because AR takes reality as its baseline and adds virtual elements. In other words: AR aims to augment reality by providing interesting or useful information. Azuma [43] proposed that AR has three requirements:

1. It combines real and virtual
2. It is interactive in real time
3. It is registered in three dimensions

These requirements were set up in 1997, but still hold true today. Billinghurst et al. [27] elaborate on this by stating that an AR system should be able to generate interactive graphics correctly, overlaid on real images in the right place, in real time. It is a powerful and versatile technology that most popularly takes the shape of 'selfie-filters' in applications like Snapchat, TikTok, or Instagram, or is present in games such as Pokémon GO.

These definitions highlight that the up and downsides of AR for self-perception research are opposite compared to VR: there is 'real', in fact its the main component: AR does not 'extract' the user from the real world [3].

Perhaps more clearly than VR, these definitions and examples show there are many approaches to achieving an AR effect technically. Theses are laid out by Estrada et al. [32] in Figure 2.2, with two main categories, marker-based and marker-less. While both can achieve similar results, marker-based is simpler in its function and use, but requires a physical reference point, the marker, to align the AR additions. Marker-less AR relies on advanced computer vision or other position/location based technologies for its calculations.

Related work

When it comes to self-perception related research, outlining, projection-based, and location-based AR do not show much potential. These technologies are used sparingly, and for specific use cases. Some research using them could be found, but not related to self-perception. As a quick overview: outlining AR, the most vaguely defined type of AR, was found to be used for a children's game [44]; projection-based AR, also called projection mapping, was used for projecting a keyboard on a tabletop [45] and projection internal organs on a mannequin [46] for education; location-based AR, being the least versatile AR type, was used for an app that showed what different places in Christchurch, NZ looked like before the earthquakes of 2010 & 2011 [47].

While these AR types have some interesting research done, they cannot be considered as alternative to VR, due to too many shortcomings and a lack of versatility. Comparatively, the remaining two types, marker-based and (marker-less) superimposition AR, show more potential. In essence, these two AR types are almost identical in its results: they superimpose virtual objects onto a camera feed, producing a seemingly physically present object. The marker-based approach is more simple, as the marker it uses acts as an anchor point, while superimposition uses machine vision to achieve this. Superimposition AR can therefore be seen as an evolution of marker-based AR, as computational power continues to improve.

Marker-based AR uses lightweight image recognition algorithms, specifically tuned on triangulate the marker. This means it does not know anything about the physical space surrounding the marker. While it is fast and very accurate, the requirement of a physical

marker can be problematic, since it needs direct line-of-sight. This can also result in a decrease in sense of immersion, as the presence of the marker becomes the source of distraction or impediment. This was found by [48], where the system they created allowed doctors to look inside a human body, provided trackers were well-placed (Figure 2.4a). To combat this, markers can be used that blend in more to their environment, or even made invisible to humans [49].

In a controlled lab setting, these downsides are less prevalent and thus the technology can be used more effectively, for example in cockroach phobia therapy. In their study, Botella et al. [50] placed a marker on a table and showed participants cockroaches crawling op top of it and over their hands, see Figure 2.4b. The marker used here was small and not distracting, as the users' attention was directed towards the cockroaches, and the application was confined to the tabletop. This study proved an effective tool in cockroach exposure phobia. Another study using cockroaches [3] suggested that embedding the virtual fear element in the real environment, instead of using VR, allowed a more direct "own-body" perception, increasing the realism of the scenario. Similarly, Suso-Ribera et al. [51] conducted AR therapy with spiders, finding equivalent efficacy between VR and marker-based AR.

Exposure phobia treatment appears to be the main use for marker-based AR literature related to self-perception. The presence of physical markers creates an inherent obtrusion problem for any possible self-manipulation, which negatively impacts immersion. Superimposition AR does not have this problem.

(A) Surgeon using AR [48]. (B) Cockroach phobia therapy [50]. (C) Superimposition example [52].

FIGURE 2.4: Different examples of AR.

The most powerful, and computationally intensive, type of AR is marker-less superimposition. It is powerful as it does not require external aid, such as markers. Instead, it (mostly) uses computer vision to gather environmental context, which creates a 3D representation the environment, making it possible to add additional objects where it would be physically possible [52], for example balls on a tabletop, as can be seen in Figure 2.4c. Izadi et al. [52] also found it was possible to segment individual objects in real time to then manipulate them. One way this is popularly used is for face manipulations, specifically selfie/face filters. Apps like Instagram, Snapchat, and TikTok allow for extensive augmentation of the facial features, by warping the image or superimposing elements. Apart from the technological achievement, face filters also shows greater impact on people's self-perception than retroactive photo editing, as it works in real-time [53]. Still, the more a person's face is augmented and farther deviates from the real, the less the person relates to the image, resulting in a lower self-relevance response [54].

One study specifically explored an AR mirror setup, projecting a self-similar avatar onto virtual mirror [55]. The results indicated that high resemblance significantly improved

users' sense of embodiment, self-identification, and body weight estimation accuracy. However, they also noted that the effect of motor control was less pronounced compared to VR systems, likely due to user's seeing their real body alongside the virtual one.

Due to the relatively recent rise of widely available face-filter applications, and growing concerns about their potentially harmful impact on self-perception and body image, most of the existing literature focuses on this specific use of superimposition AR. These filters have been shown to psychologically widen the gap between the actual self and the ideal self [56], contributing to a distorted sense of reality in which users expect their real-life appearance to match their perfectly filtered digital selves [57].

Best AR alternative

Suso-Ribera et al. [51] found that marker-based AR achieved similar outcomes to VR in phobia treatment, supporting its use as a viable alternative. Additionally, AR can safely expose individuals to feared stimuli by integrating virtual elements into real-world environments, without the downsides of a full virtual settings, such as costs [58]. In both these papers, marker-based AR was discussed, which is quickly phased out in favour of superimposition AR. Computational advancements have made superimposition exceed the possibilities of its marker-based predecessor, without any of marker related limitations. Choosing between the two thus comes down to the following:

Marker-based	**Superimposition**
+ Lightweight	− Resource intensive
+ Easy implementation	− Complex
− Limited	+ Powerful
− Requires marker	+ Stand-alone

Resources and complexity are not a limiting factor for this research, negating the downsides of superimposition AR. More power results in more possibilities, and combined with system working without external dependencies, means superimposition AR offers increased flexibility in prototype development. Requiring a marker can cause the system to be perceived as obtrusive, negatively affecting the immersion of AR systems [59], which is an important metric in the feeling of self-representation [18]. Immersion can also reach greater highs since the system has environmental context and is thus more physically accurate.

2.3 Designing for self-perception

Examining the dominant use of HMD VR in the domain of self-perception and its most promising AR alternative, provides the groundwork for further concept exploration. This section explains the ideation process of the augmented reality mirror concept.

2.3.1 Requirements

In order to streamline the concept exploration process, requirements are set up to inform decision making. These are derived from the literature gathered in this chapter, have a practical rational, or are preferences of the stakeholders.

- **Cost & Complexity** The system should not be too expensive or complex. It must be reproducible, and feasible to be built by one person within a few months.
- **Responsiveness** The system should run smoothly and be intuitive to use. User experience is quickly degraded if the interaction is unresponsive or illogical.
- **Resemblance** Literature suggests that a higher self-association provides better results in self-perception psychology. Technology that facilitates this is therefore preferred.
- **Non-invasiveness** The system should avoid physical or visual obstructions, as these negatively impact immersion. It should be as seamless and plug-and-play as possible.

While some requirements might be subjective or context dependent, responsiveness is quantifiable, and can be measured using latency. In general, lower is better: Samaraweera et al. [60] found that a lower latency between the movements of a self-avatar in a VR mirror and the user resulted in a higher sense of body ownership and a stronger feeling that the movements in the mirror avatar were caused by their own movement, over those with (induced) higher latency.

Besides this, high latency can cause complications. In puppeteering, the control of an avatar, high latency causes users to adopt a start-and-stop strategy to ensure that the 'puppet' performs their actions properly [61, 62]. This can be explained by higher latency causing more overshoot in user's actions, where the start-and-stop method tries to compensate for this.

These effects show that low latency has major benefits, but do not specify an target amount. Guidelines established in 1968 [63] state that any man-computer interaction with a total system latency below 100 ms feels instant to the user, and these guidelines are still referenced today. For example in first-person-perspective games [64], less than 100 ms latency was found to be preferable. While an AR mirror does not clearly fall into this category, both systems directly influence the user's head movements. Moreover, predictable, consistent delay is preferable to variable lag [65]. Lastly, delays greater than 40 ms between sensory inputs (in any modality) are perceived as separate events, meaning a frame rate of above 20 ensures a seamless experience [66].

Taken together, these findings highlight the importance of consistent latency, ideally below 100 ms, to facilitate a natural and intuitive interaction.

2.3.2 'Self-reflection'

Humans interface with their self-image daily and have done so for hundreds of years. The invention of the mirror (and before that, reflections in f.e. water) has made it possible

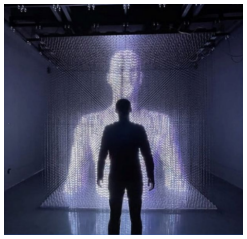
to look directly at one's self. Mirrors are the primary way people interface with their self-image and is the most recognizable object in the realm of self perception, only more recently being challenged by phones, which act similarly, like a pocket-mirror. Compared to HMD's, an external screen or mirror would be unobtrusive, which could lead to it being perceived as more 'real': it is part of the real world, not a virtual one. Augmenting the age-old concept of the mirror to facilitate self-perception research is therefore promising, and although the idea of a 'smart' mirror sounds clearly defined, there are many design directions.

2.3.3 Novel display technologies

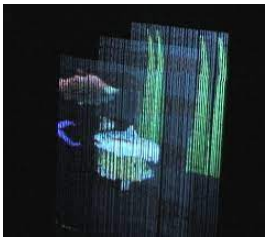
Multiple XR technologies show promise for self-perception applications, although the approaches differ, there is one communality: a type of 'screen' is used to interface with the user. This usually takes the shape of a smartphone screen, projector or TV, but more options are available. The definition of a mirror can be broadly interpreted as something that 'mirrors' the user's image. This can be achieved using different types of, and alternatives to, traditional screens. These can be divided into two categories: high and low fidelity (or resolution).

High-fidelity solutions include screens or projectors capable of displaying images at such high resolution that they become near indistinguishable from reality. The details in these images can get a self-relevant response from the user [24]. Examples of other high-fidelity solutions are: a projection mapped mannequin [46]; augmenting the shadow against a wall of a person; glass reflective 'holograms' (Figure 2.5a); or LED fans that create the illusion of a floating screen.

(A) 3D Holographic glass display[1].



(B) 3D LED matrix[2].



(C) Multi-layered water display [67].

(D) Moving pin display[3].

FIGURE 2.5: Alternative screens examples.

Low-fidelity solutions do not use a screen or projector, but use other methods to display their 'pixels'. These pixels are often physically much larger compared, resulting in a lower resolution, but allowing for more advanced 3D effects. A downside of this lower resolution is that crucial self-relevant details can get lost, especially in the face, which is the most important area for self-perception [20, 26]. However, a self-relevance response can still be invoked by mirroring movements [68]. Examples of low fidelity solutions include: a large 3D LED matrix (Figure 2.5b); multiple layers of falling water projection (Figure 2.5c); a robot that mimics movement; a flip dot display; a moving dot display (such as Figure 2.5d and [69]); and LED drone formations.

[1] www.m.indiamart.com/proddetail/3d-holographic-glass-display-20250005955.html
[2] www.singularityhub.com/2021/04/05/this-huge-hologram-like-3d-display-is-made-of-thousands-of-tiny-led-lights/
[3] www.gadgetmaster.eu/product-eng-13-Pin-art-3D-sculptures.html

The concept of a mirror, can be achieved by most technologies mentioned. However, additions to the mirror image could prove difficult to implement, or not reach the desired impact, on lower fidelity solutions. A real mirror practically has an 'infinite' resolution, which would score perfectly on the resemblance requirement, bar the image being mirrored. This supports using a high fidelity option.

High fidelity environments bring the expectation that the mirror image is of comparable quality. Small imperfections become more noticeable, compared to the simpler low fidelity alternative. This means that the system is much less forgiving, as self-relevance dwindles quicker. Comparatively, a high fidelity solution has more difficulty achieving a proper self-relevance response, but also has a higher overall potential, while a lower fidelity solution should prove easier, with a higher probability of yielding no significant results. Additionally, the higher fidelity solution is easier to built physically, but more difficult in the software department, the opposite being true for the low fidelity solution. Taking these metrics into consideration and comparing them to the requirements set up in the previous section, the high fidelity option shows more potential. It most likely costs less, is a more feasible solution to build and distribute and, suggests providing a greater self-relevance response due to higher resemblance.

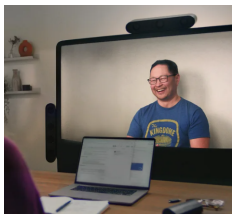
The high fidelity augmented mirror concept falls within the requirements of Section 2.3.1. The cost is mostly in man-hours, as screens and cameras are relatively cheap and augmentation software is (mostly) free. Tuning the complexity properly means the mirror should be feasible to produce, and because mirrors have 'infinite' resolution, resemblance will be high, meaning the self-perception users experience is expected to be high as well.

2.3.4 Existing digital mirrors

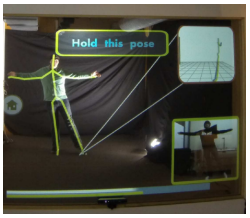
An AR mirror in concept is not novel, though there are many ways to approach it. The most basic approach is to simply apply the face filter applications (as mentioned in Section 2.2.2) to a larger screen and call it a mirror. This is most commonly used in the fashion industry to fit clothes [70], or to try out makeup looks [71] (Figure 2.6a). A downside of such systems is the camera is static, making it unusable from off-angles, only from straight in front. Other research shares this problem, f.e. a screen correctly projecting anatomical data on a person, such as CT data [72]. Latoschik et al. [73] also used a screen, but made the displayed picture entirely virtual, by showing an avatar instead of the user. This avatar takes the appearance of the user and because it is completely virtual, it can provide perspectively correct images. This is similar to another study, where Nimcharoen et al. [74] used Kinect's point cloud data to separate the user's body from the background, and shows it back to the user through AR goggles (Figure 2.6b). Because the body is separated from the background, it can be augmented in many ways, in this research, making the body slimmer or wider. In a study that seemed to combine elements of this research and [73], Fiedler et al. [55] projected a mirror in the environment using an AR headset, showing the mirror 3D avatar mirroring the user. These ideas built on a much older prototype of head tracking to correctly show 3D images [75], here a Wii remote was used, alongside IR lights attached to the head. This works extremely well, but can be considered obtrusive, as the user has to wear something on their head. The natural progression here is to use some unobtrusive type of head tracking, which was not viable at the time of this research. More recent advancements in computer hardware made Project Starline[11] by Google possible, which, similar to [73], is completely virtual, but it appears 'real' (Figure 2.6c). It creates a hyperrealistic avatar of the user and combined with added depth gives the illusion that two people are sitting right across from each other.

(A) Framed MagicFace AR mirror [71].

(B) Point cloud mirror [74].



(C) Google Project Starline[1].



(D) Augmented real mirror [76].

FIGURE 2.6: Augmented mirrors.

Most of these examples use a standard screen as their primary output device, alternatively, it is possible to add AR to an actual mirror. By replacing the back panel of a mirror with a screen, it is possible to project images through the mirror while it stays reflective. This is often done as a do-it-yourself (DIY) project to add widgets to a mirror, but it is also used in research. Using spatial information of the user, Lee et al. [77] were able to control virtual objects on the mirror by placing their hands over it, with completely correct perspective. The same technology was used to track the body of a user and help them exercise with proper form by showing a preview of the movement over the user's body [76] (Figure 2.6d), building upon the framework proposed by Microsoft research[2].

If the digital mirror requires interactions with the user, various interaction methods have been explored. Some systems integrate gesture recognition [72], touchscreens [71], or hover-based inputs [76, 77] to enable users to interact with virtual content on the screen. Others incorporate speech commands or proximity-based inputs [78] to allow for interaction without direct touch.

2.3.5 Opportunities

As made clear in the previous section, a variety of approaches to creating a digital mirror are explored. However, a gap presents itself: a stand-alone screen-based mirror that offers physically accurate perspectives and can augment the image. Despite the seemingly great fit, not much self-perception related research was found using a mirror or something similar. Nimcharoen et al. [74] had a simple version of body morphing and the MagicFace mirror [71] also showed potential for self-perception research. Because of gaps and opportunities like this, there is value in creating a novel installation and testing its potential as a meaningful tool and alternative to HMD VR for self-perception (and potentially other) related research.

Using the requirements, the literature, and the stakeholders as motivation, the most important design decisions have been made. Following decisions will be more specific regarding the building of a functioning augmented reality mirror prototype. Thereafter, the prototype mirror will be tested for its effectiveness, and (future) research opportunities, in a self-perception related user test. The design process and subsequent user tests will then provide a basis on which to answer the research questions set up in Section 1.2.

[1] `www.blog.google/technology/research/project-starline/`
[2] `https://www.microsoft.com/en-us/research/video/holoflector/`

Chapter 3

Method

This chapter will describe the process of designing and building the AR digital mirror prototype. Each design decision is discussed and supported by literature from the previous chapter. The process is described in detail, as to inform future work recommendations. The successes and pitfalls of this research are used to generate technical design guidelines later on. First, some design decisions are discussed that significantly impacts the direction of this development.

3.1 Digital mirror system overview

Creating a digital mirror requires understanding the fundamental physical properties of real mirrors, as well as the technologies available that can achieve these effects digitally. This section explores the key aspects involved in designing such a system, to inform the design later on.

3.1.1 Mirror(less) screen

The related work in Section 2.3.4, in combination with the requirements of Section 2.3.1 suggests two design approaches to the digital mirror. The work either uses a real mirror [76, 77] or does not [70, 71, 72, 73, 75]. Both approaches require the use of a screen, with one using it exclusively, showing a camera feed or virtual representation of one's self, and the other using it to overlay images on the reflection of a mirror, by placing the screen behind a mirror and shining images through it. These descriptions highlight the biggest possibilities and shortcomings of both approaches. Using only a screen allows for the complete control of the output image and thus supports real, virtual and combined content. The downside is that inherent physical obstacles, such as resolution and latency, make it difficult to resemble a real mirror, if that is the goal. Alternatively, placing a screen behind a mirror keeps the instant and physically accurate reflections, but overlaying virtual images to the reflection is technically challenging, and the overlaid image will lag behind the reflection when the user is moving, due to the latency caused by the screen.

When it comes to the reflection, a screen setup will typically [70, 71, 72] only produce a believable reflection while standing straight in front of it, since the camera is static. This could create the illusion of a functioning mirror, but quickly dissolves if a different angle or position is assumed. This contrast a real mirror, which provides an accurate reflection from all angles.

Section 2.3.5 highlights an apparent gap that reveals itself: a stand-alone screen-based mirror that offers correct perspectives when viewed from any angle. This would combine the strengths and possibilities of a screen based solution, combined with the feeling and proper reflections of a real mirror. Looking at related work, this has only been done using a completely virtual representation of the user [73], by tracking the user's eyes in space. Using this same technique applied to camera footage, it is hypothesised that this can create the illusion of a real mirror. This requires the tackling of multiple technical challenges and screen-based limitations, such as latency.

3.1.2 Breaking down a mirror

Mirrors are so common that all humans are accustomed to seeing themselves in them. In order to create a digital counterpart, the physical properties of a mirror should be examined. The most (in)famous aspect of a mirror is that it appears to horizontally flip the image. Most people are aware of this, as their movement in the mirror is reversed, but also because they often prefer their mirror image over their real image in f.e. a photograph [79], possibly due to the mere exposure effect [80], as the mirror image is more frequently viewed than its reverse counterpart. Claiming that the image is horizontally flipped is technically incorrect; rather, it is flipped along the z-axis (front-back). However, this depth inversion creates the illusion of a horizontal flip due to how we mentally interpret reflections [81].

The main appeal of a mirror is that its entire surface is perfectly reflective. The reflection is consistent, always showing the path of light travelling from the user's eyes to the edges of the mirror, bouncing off, and projecting back into the world, capturing everything in between. Consequently, the reflection is influenced only by two factors: the size and shape of the mirror, and the position of the user's eyes, relative to the mirror. This is illustrated by the green lines in Figure 3.1 below.

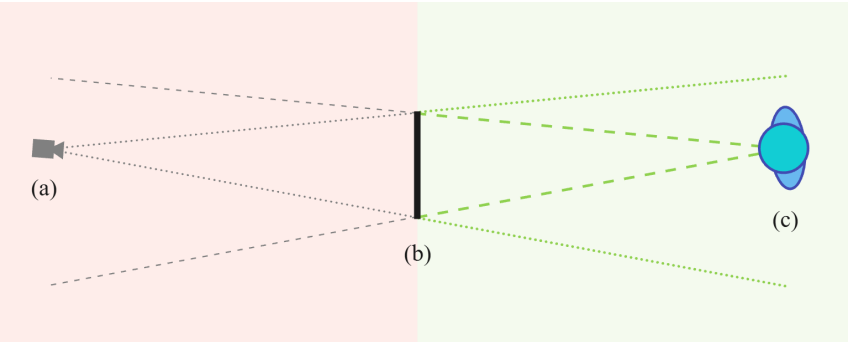


FIGURE 3.1: 'Inner workings' of a mirror. (a) shows the virtual camera, (b) the mirror, and (c) the user. The right (green) part shows the real world, left (red) the virtual world.

The image seen in a mirror is called a virtual image, as it appears inside or behind the mirror. This is illustrated by the dashed lines in the figure above, as they end up behind the mirror. This virtual image is identical to what a person would see if they were standing, or if a camera was placed, directly opposite of the user on the other side of the mirror, looking through it. This is shown in the figure as a virtual camera and its dotted lines of sight that align perfectly with the bounced green lines.

In summary, the user position can be used to compute the virtual camera position. This in turn can be used to calculate what the user should be seeing, without requiring an actual mirror, using only a screen, a camera, and some method of retrieving user position.

3.1.3 User position

There are many technologies that can be used to retrieve user position, and they can be grouped in three categories: marker-based, marker-less, and inside-out. Marker-based technologies utilise markers on the user to create a 3D representation of them. These are complex system, but return accurate results. Marker-less systems are relatively low cost and unobtrusive, that rely on only (specialised) cameras to track the user. They are often much less accurate and come with increased latency compared to the other technologies. Lastly, inside-out systems are attached to the user and determine the position relative to the environment using optical and inertial sensors. Table 3.1 provides an overview of the technologies, including some examples.

Type	**Examples**	**Cost**	**Complexity**	**Intrusiveness**	**Accuracy**
Marker-based	OptiTrack[1], Vive Trackers[2]	High	High	Medium	High
Marker-less	MediaPipe Pose[3], Kinect[4]	Low	Low	None	Medium
Inside-out	Xsens MVN[5], Vive Ultimate[6]	Very High	Medium	High	Very High

Table 3.1: Comparison of different position tracking technologies.

Using the requirements (Section 2.3.1) as a metric, one technology shows most promise: the marker-less approach. It has a relatively low cost and complexity, as a camera is present regardless. It has no intrusiveness, as the user is not obstructed in any way, which also makes the system more plug-and-play. Accuracy is average, however, in the context of a digital mirror, high precision is not required. These technologies also boast a slightly higher latency than the others, but the decrease in complexity and intrusiveness outweigh these negatives.

When this approach is used, the system needs some real world measurements as a reference point. Some systems have this built in, but a camera only approach would require some type is distance sensing.

[1] `https://www.optitrack.com/`
[2] `https://www.vive.com/eu/accessory/tracker3/`
[3] `https://ai.google.dev/edge/mediapipe/solutions/vision/pose_landmarker`
[4] `https://azure.microsoft.com/en-us/products/kinect-dk`
[5] `https://www.movella.com/products/motion-capture/xsens-mvn-link`
[6] `https://www.vive.com/eu/accessory/vive-ultimate-tracker/`

3.2 Building the prototype

Understanding the physical properties and workings of a real mirror will streamline the process of creating a digital mirror. This process can be outlined in five distinct steps, each containing a variety of technical challenges and design decisions that need tackling, and each building upon the previous step. They will therefore be discussed in chronological order and in detail, as it informs the design guidelines set up later. The steps are as follows:

- **Hardware Overview:** What hardware is needed to reach the desired goals while adhering to the requirements?
- **Software Overview:** What needs to be achieved, and what software can be used to achieve this?
- **Video to User Position:** How can the user position be determined accurately and consistently?
- **Perspective Calculations:** How can the digital mirror show the correct perspective from all angles?
- **Augmented Reality:** How can augmentations be applied to the image, and what will be used to test its feasibility?

A high-level overview of the total system can be found in Figure 3.2. Each component shown in this figure will be described in detail within one of the steps mentioned above.

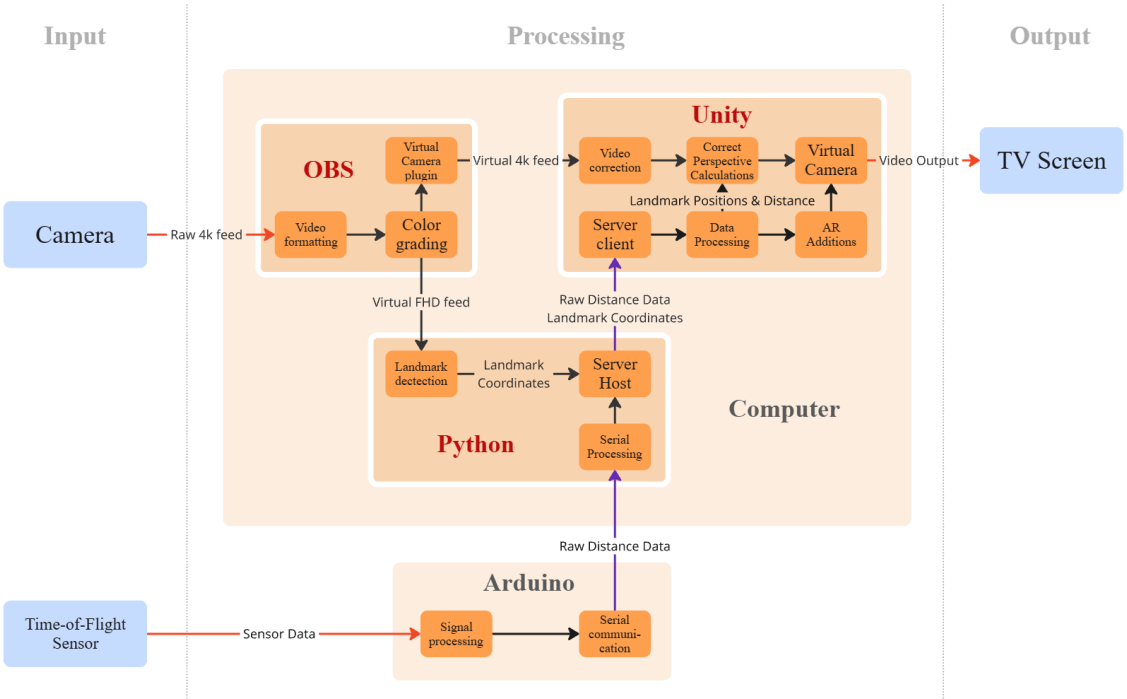


FIGURE 3.2: Complete overview of the final system, broken down into software and hardware components.

3.3 Hardware overview

Required components for this prototype include: a screen the size of a mirror to display the eventual results, a camera that can capture the user and the surrounding environment, a computer that runs all the software as quickly as possible, and a distance sensor that provides some physical context. These items should contain a balanced set of characteristics, to support the flexibility in future design decisions.

3.3.1 Screen

The main attraction of the physical setup is the screen. Users will look at it intently, so it should support high resolution, good colour accuracy, and a size big enough that anyone, regardless of height, can see themselves in it completely. Calculating the vertical height needed for complete body visibility is simple and, interestingly enough, unaffected by the distance of the user to the mirror/screen. Light that bounces off the mirror surface always travels twice the distance between user and mirror, which means that only half the users' height in mirror size is required to fully display the body. The only other requirement is that the top of the mirror is slightly above eye height. Since this has to support users of varying heights, taller users are the bottleneck, but if the mirror is placed too high, shorter users might be unable to view their feet. A mirror of 1 meter vertical with the top at 2 meters high would thus work for most users, supporting a user height range of 1.5–2 meter. Since mirrors are typically taller than they are wide, a screen that is at least 1 meter horizontally (∼45" diagonally) and can be rotated into a portrait position, is required.

The chosen screen is a 55" conference screen from Philips[1], which means it is 121 cm wide and pre-mounted to a roll-able stand, allowing easy displacement. Since this is a decently new screen, the specifications satisfy the requirements mentioned before. The screen exceeds the necessary size, but remains appropriate for mirror standards. The screen is remounted into a portrait orientation.

3.3.2 Camera

The camera greatly impacts the quality of the output image, and by extend the tracking quality of the pose estimation software. The camera should boast a balanced set of attributes, to best satisfy the requirements.

There are three main camera properties, that cannot be addressed in software: resolution, field-of-view (FoV), and latency. Latency and FoV scale granularly, resolution does not. 1080p to 4K is a four times increase in resolution. Higher resolution systems are preferred as they provide a clearer image, which also facilitates zooming in.

In the context of a digital mirror, a minimal viable FoV can be determined. The FoV should support two things: decent horizontal viewing angles and full body visibility at close distance. The latter is more important, as its impact on the user experience is more pronounced. To fully display a person of 2 meters (edge case) at 1-meter distance, the vertical FoV should be ∼83°, translating to a 90° diagonal FoV when using a vertically oriented camera. While this is the minimum, more is better, and a higher FoV of 120° would allow both vertical and horizontal camera orientations.

Another important aspect is the firmware attached to cameras. If possible, accessing the raw camera feed is preferred, as firmware can add more dependencies, increase latency, and (un)distort the image.

[1] `https://www.philips.nl/p-p/55BDL3650Q_00/signage-solutions-q-line-scherm`

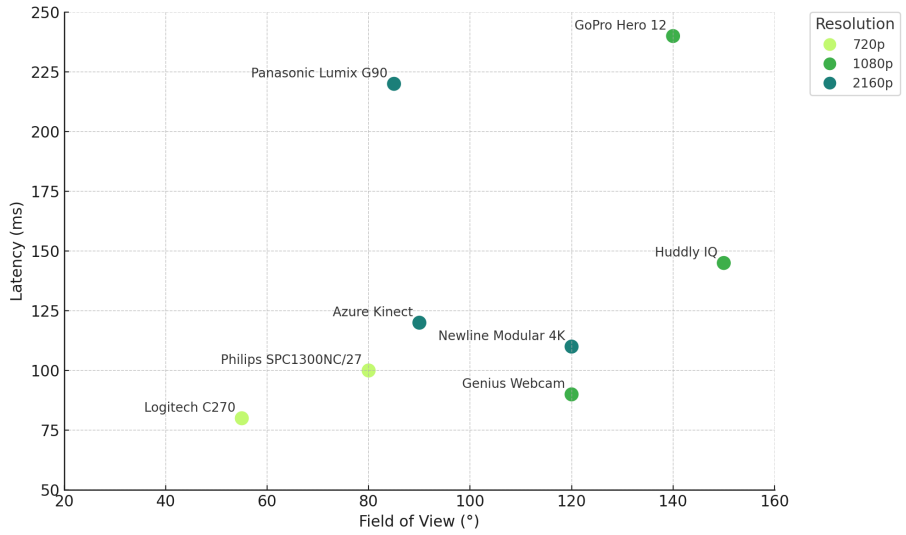


FIGURE 3.3: Overview of different cameras systems. X-Axis shows the field-of-view: higher is better. Y-Axis shows the latency: lower is better. The colours indicate the resolution: higher is better. The Azure Kinect has its own pose estimation built-in, the latency reported is an estimation, which excludes this pose estimation processing time.

Figure 3.3 shows an overview of different camera system tested, arranged by their specifications. These systems have been selected to represent a range of approaches, and each system was tested on its end-to-end latency using a standard technique[1]. This technique also includes the monitor's latency and the computer's processing time, therefore all camera tests were performed using the same hardware. To compensate for any other variables, multiple samples were recorded of each system, the average value is shown.

Webcams

Most systems tested can be categorised as webcams or conference cameras. In total three webcams were tested, the Logitech C270[2], the Genius WideCam F100[3] and the Philips SP1300NC[4], as well as two conference cameras, the Newline Modular 4K Camera[5] and the Huddly IQ[6], with their firmware disabled when possible, to prevent additional processing delay.

Kinect

The Kinect Azure DK[7] is a multisensory camera system specifically designed for pose tracking. This system mainly uses its Time-of-Flight depth sensor for pose estimation. This results in a 3D pose estimation, making it more spatially accurate compared to a computer vision system, which has no context of the physical space. The Kinect comes with two important pitfalls: Vogel [82] found that introducing a vertical angle between

[1] https://hamtv.com/latencytest.html
[2] https://www.logitech.com/en-us/products/webcams/c270-hd-webcam.html
[3] https://www.geniusnet.com/en/products/view/widecam_f100_v2
[4] https://www.usa.philips.com/c-p/SPC1300NC_27/webcam
[5] https://newline-interactive.com/usa/modular-4k-camera/
[6] https://www.huddly.com/conference-cameras/iq/
[7] https://azure.microsoft.com/en-us/products/kinect-dk

the Kinect and the user of 30 degrees resulted in considerably worse skeletal proportion estimations. Placing the camera on the top or bottom of the mirror could thus negatively impact results. The Kinect will also not function properly when turned vertically.

As shown in Figure 3.3, the Kinect shows a competitive latency, however, the value is an estimation of the RGB camera latency. The actual total latency includes the processing time of the built-in pose estimation. The output image is delayed to match this processing time, resulting in a total video latency of roughly 550ms.

Other

A GoPro Hero 12 Black[1] was tested, with more than sufficient resolution and FoV in normal use, unfortunately, when used as a webcam using the official GoPro Webcam Utility[2], the resolution is significantly reduced and the latency is relatively high compared to the other systems. This makes the GoPro only desirable if FoV needs to be prioritised.

Alternatively, a capture card can be used to convert any HDMI output to a virtual webcam feed. This allows for the use of f.e. a DSLR. To test this, a Panasonic Lumix G90[3] DSLR was used in combination with a USB capture card. While the DSLR supports 4K video resolution, the capture card used brings the resolution down to 1080p.

Combining the preferences mentioned before and comparing the different options in Figure 3.3, reveals one camera outperforming the rest: the Newline Modular 4K Camera. This camera boasts a 4K resolution, a good diagonal FoV of 120°, and a decent ∼110ms latency, making it the best all around choice for this project.

Camera placement

Ideally, the camera would be positioned behind the screen, mirroring the user. This is not physically possible, thus the camera should be placed on the edge of the mirror. There are two options, both limit the camera offset to one axis: the top middle, or the side on eye level. Placing the camera on top results in the user looking down on themselves, but does not discriminate on height. Placing the camera on (average) eye level eliminates this effect for people around this height, resulting in the user looking at themselves slightly from the side.

To determine which placement is preferred, the literature was first evaluated. All related work from the previous chapter placed the camera on the top (or bottom) middle, without stating their reasoning. This provided insufficient argumentation, therefore, a quick test was performed. Participants (n=8) were shown a live video feed with the camera in either position, in a similar setup to the final one, and were asked which one they preferred. All participants indicated a preference for the top camera placement. These result combined with the aforementioned up- and downsides, suggest a top mounted approach is preferable.

The camera's FoV supports both landscape and portrait orientations. To facilitates improved full-body visibility, the latter is chosen. The reduced viewing angles were accepted, as they are deemed less important in typical mirror usage.

[1] `https://gopro.com/en/il/shop/cameras/hero12-black/CHDHX-121-master.html`
[2] `https://community.gopro.com/s/article/GoPro-Webcam?language=en_US`
[3] `https://store.eu.panasonic.com/nl-nl/lumix-dc-g90meg-k-systeemcamera-20mp-4k-zwart`

3.3.3 Distance sensor

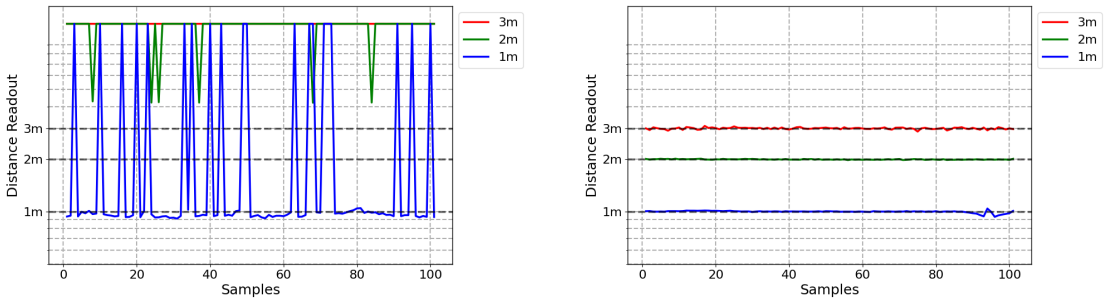
An RGB camera alone can only provide a relative position estimation. Since an absolute position is required, some physical properties are necessary. The simplest approach would be placing a marker on the floor at a specific distance, and using it to anchor user position. However, looking at the requirements (2.3.1), this would interfere with the system's plug-and-play ability, requiring precise measuring if the system is displaced. The best unobtrusive alternative is using a distance sensor. This sensor can return a distance value when the user stands in front of the mirror. It should be reliable, relatively cheap, and working in a 0.5-3 meter range, covering most mirror using scenarios. There are different types of sensor to consider:

- **Acoustic-based:** Distance can be estimated by using sound waves. Using f.e. an ultrasonic sensor (US), which sends out frequencies above human hearing and estimates distance based on the return time of the sound waves. These sensors are inexpensive, but affected by environmental noise and soft surfaces.
- **Infrared-based:** There are two approaches using infrared (IR) light, one is similar to ultrasonic: a time-of-flight (ToF) sensor. This sensor sends out pulses of IR light and estimates the distance based on the travel time of the light. These sensor offer high accuracy, but are affected by reflective surfaces. Another way is looking at the intensity of the sent out refracted IR light, where a higher intensity means a closer object. This approach is inexpensive, but shared many pitfalls of the two sensors mentioned before.
- **Radio wave-based:** This approach measures the return time of radio waves to measure distance. This is very accurate, but also expensive and complex. It is often used in cars, supporting many weather conditions and long-range sensing.

The distance sensor requirements are simple, meaning each of the above options should be valid, an ultrasonic sensor (HC-SR04[1]) was thus tested on its capabilities. It was found not suitable for a multiple reasons, firstly, the value it produced when the user was standing still fluctuated too much, even with added smoothing. Secondly, the 'cone of vision' was found too large, resulting in faulty target separation in ranges over one meter, despite its rated operational range of up to 400cm. Lastly, the signal easily interferes when objects are close to the sensor. The reported weakness of performing worse on soft surfaces (f.e. clothes), is the likely cause of these issues.

To support these observations empirically, a user was positioned in front of the sensor, at 1, 2 and 3 meters distance. One hundred samples were then recorded at each distance and results are plotted in Figure 3.4a. At 1 meter, the sensor can detect the user fairly consistently, with reasonable accuracy. Many spikes occur, where the sensor did not pick up any signal, returning an arbitrarily high number ($\sim$11 meters). These can be filtered out. At 2 meters distance, the sensor only sometimes picks up the correct signal. No signals are picked up at 3 meters distance. It is therefore not recommended to use this sensor on users beyond 1 meter distance.

[1] `https://www.digikey.com/en/products/detail/osepp-electronics-ltd/HC-SR04/11198533`



(A) HC-SR04 US sensor.

(B) VL53L1X ToF sensor.

FIGURE 3.4: US and ToF distance readout test.

An electronics expert at the University of Twente was consulted to determine the best alternative solution, for this research. They recommended using a time-of-flight (ToF) sensor instead, as its cone of vision is considerable smaller, it supports longer ranges, and it works better on soft surfaces. It is also inexpensive compared to the other alternatives. A ToF sensor (VL53L1X[1]) was ordered and testing was repeated. As can be seen in Figure 3.4b, the sensor faired incomparably better, providing consistent and stable distance values at all three distances. The ToF sensor meets all requirements and was therefore selected for use. Additionally, the sensor has a considerably smaller footprint than the US sensor, allowing easier integration into the prototype.

3.3.4 Additional hardware

Next, a computer is needed to connect to the screen and camera, and run the pose estimation, perspective calculations, and image augmentations. Performing these tasks concurrently is computationally intensive, a high-end computer is thus preferred. Faster computers reduce processing time per frame, lowering latency and increasing frame rate, resulting in a more responsive and smoother experience. A powerful computer (32GB RAM, RTX4080) was used.

The ToF sensor that is used also needs to be powered and controlled, and its data sent to the computer. A direct connection is not possible, thus an Arduino UNO is used, forming a bridge between the sensor and the computer.

3.4 Software overview

The system required multiple instances of software, as shown in Figure 3.2, each with a specific task. These different pieces of software, and by extend hardware, need to communicate properly and efficiently in an organized manner. This section describes what software is used for what purpose, and how they communicate.

3.4.1 Unity

Combining processes in the same, centralised software reduces complexity, as well as chances of inter-software failures, future proofing the system. Unity3D[2] will be used as

[1] `https://www.st.com/en/imaging-and-photonics-solutions/vl53l1x.html`
[2] `https://www.unity.com`

the central software, which is a powerful and free 3D engine. It is used by most game developers, but also by researchers, as it provides virtually limitless possibilities for creating applications, by allowing users to write their own code in C#. Unity is free and popular, which means many software platforms have (un)official integration through plugins. The same is true for hardware, such as most camera and display technologies. This prototype, as described in Section 3.1.2, inherently deals with a 3D problem and virtual cameras, which Unity thrives at. These strengths, together with the researcher's prior experience with the software, make Unity ideal as the central software component.

3.4.2 Pose estimation

The pose estimation software should support two features. Firstly, estimating user position, specifically the eyes, to calculate the proper reflection the screen should display. Secondly, it should be able to separate the user from the background, so that user and environment augmentations can be performed. It is efficient to do this processing once and use it for both cases.

OpenPose

OpenPose was the first real-time, multi-person, open-source pose recognition system that became available, created by Cao et al. [83]. It takes a video feed as input, and returns the positions of all bones and joints (commonly called landmarks), as well as facial features, using Machine Learning (ML) algorithms. OpenPose runs in C++, but also has a native Unity solution[1], allowing fast implementation. However, the ML models this software uses are outdated, while other ML models and hardware have improved significantly over the last few years. OpenPose running in Unity on an Nvidia GTX 1660 Graphics Card (GPU), averages around 11 frames per second (FPS), or 90ms per frame. This is subpar, as quick movements can create a disconnect between the users and their 'mirror image' because of the high latency, creating a sense of unresponsiveness. Performance is expected to scale linearly to the processing power of the GPU and with the prototype computer having an Nvidia RTX 4080 GPU, powerful[2], the expected performance would be around three times higher: ∼30 FPS. This is acceptable performance, but the addition of other graphically demanding processes running in parallel, could result in overall worse performance. Therefore, alternative should be explored.

MediaPipe

A popular alternative is Google's MediaPipe Pose Landmark Detection[3], which is an open-source pose estimation software using up-to-date BlazePose ML models. These models outperform OpenPose's models significantly, produces a 15ms frame time compared to OpenPose's 90ms, on the same machine. This is equivalent to 67 FPS, with the prototype machine reaching approximately 200 FPS. This far exceeds the requirements as more than the display's 60Hz should suffice. More FPS remains better, as it means the latency and the experienced delay is lower.

[1] `https://github.com/CMU-Perceptual-Computing-Lab/openpose_unity_plugin`
[2] `https://www.videocardbenchmark.net/compare/4622vs4062/GeForce-RTX-4080-vs-GeForce-GTX-1660`
[3] `https://mediapipe-studio.webapps.google.com/studio/demo/pose_landmarker`

As Mediapipe is open-source, a Unity plugin has been created[1]. It is not official however, and has a far more involved installation process, compared to OpenPose. It does have apt documentation and support, and is therefore implemented and tested. Unfortunately, the plugin did not meet expectation: performing considerably worse than the browser preview. The plugin produces jumpy, inconsistent, landmark locations as well as higher latency. This suggests the implementation is not optimised, it was therefore decided to use the official Python implementation and parse its data to Unity.

3.4.3 Additional software

Besides the two major applications, Unity and Python, some additional software is required. The Arduino UNO code is written in the Arduino IDE and uploaded to it. This will handle distance value processing communication, and runs independently and indefinitely.

Secondly, an application needs to address the following problem: a camera or webcam on Windows can only be used by one application at a time. This prevents conflicts and increases stability, however, Unity and Python require access to the webcam feed at the same time. Windows does not natively provide a solution for this, an application that splits the feed is thus needed. There are multiple applications that can achieve this, the most notable one being OBS[2]. OBS is a free and open-source solution for offline video recording and live-streaming. Most importantly in this context, it can take a webcam feed, convert it to a virtual one, allowing multiple applications to use it.

3.4.4 Communication

Adding additional pieces of software, besides increasing complexity, also require proper internal communication. Windows makes this more difficult, as it lacks some developer options that f.e. Linux has. The hardware communicates with the software through standard Windows protocol, but two lines of communication require a specialised solution.

Arduino → Python (Serial)

The Arduino UNO measures the distance 10 times per second, this data needs to be parsed to Unity. The simplest way to communicate between an Arduino and a computer is via serial communication over the USB bus. The communication is one-way, since the Arduino functions as a sensor in this case. The data in the serial port can be read out by any application that has supports it. It is possible to directly access the serial port in Unity. Some complications arose during implementation, whereafter it was decided to read out the data in Python and integrate it into its existing data stream to Unity, as described below.

Python → Unity (TCP server)

Communication between two applications on Windows is not simple, but doable. The best approach is using a socket server. Through Python, a TCP server is created on the local network. A client is then created in Unity which connects to the server. Once connection is established, Python can start sending data over the socket to Unity. This commination is also one-way, but unlike Serial, this require an active connection between the two programs.

[1] `https://github.com/homuler/MediaPipeUnityPlugin`
[2] `https://obsproject.com/`

The pose estimation data in Python and imported distance data from the Arduino are converted into bytes and a prefix is added, so Unity understands the incoming data type. For both, data gathering, packaging, and sending is run on separate threads, which ensures smooth data streaming and increases system flexibility. The distance data is sent 10 times per second, the pose data 30 times. Unity reads the buffer in a given frame, determines the data type, splits it up if necessary, and processes it accordingly.

3.5 User position

Figure 3.1 showed how proper perspective can be achieved. Only the position of the user, relative to the mirror, and the physical dimensions of the mirror are required for this. 3D user coordinates can be calculated, using two parameters: 1. the angle between the camera and the head of the user, and 2. the distance between the camera and user. The first parameter can be calculated from the pose estimation, using texture coordinates combined with the physical properties of the camera. Pose estimation is not able to provide the second parameter reliably. Each person differs in height and size, and without any physical reference point, the computer cannot predict how far away something is. This means the distance requires another approach, using a distance sensor.

3.5.1 Distance calibration

Pose estimation software, including MediaPipe, often gives the option between a 2D and 3D mode for landmark estimation. As mentioned before, the computer cannot accurately predict the distance to a person, but it can reasonably well estimate a distance relative to a starting position. Coincidentally, a single ToF distance sensor would not be able to measure user distance at all angles, due to its limited cone of vision. Both approaches separately have clear pitfalls, but the combination can solve these problems: the distance sensor can provide the reference point the pose estimation needs. The distance data only needs to be used once, after which the pose estimation takes over. The system first needs to detect a user, so a calibration-like process is required. This prevents accidental detections and communicates clearly to the user that the mirror is functioning properly.

When the system is turned on, a boot up screen appears for a few seconds. This screen is added as artificial delay, to give the ToF sensor time to gather information about the environment. The average distance readout value is taken when boot up is completed, and this value can then be assumed to be a wall or obstacle. Afterwards, if a smaller distance readout is recorded, it can be assumed to be a user. After boot up, the mirror will display instructions, telling the user to step in front of it, thereby entering the cone of vision of the ToF sensor. Based on the distance readout, the user is guided into a predetermined position. When they reach the correct position, the distance value will be recorded for a few seconds. If the person moves too much in this time period, they are instructed to stand still and the recording resets. Once complete, the distance is saved and the pose estimation can take over.

The design of the calibration process is simple and clearly communicate its instructions. In the field of User Experience (UX) design it considered best practice to keep the User Interface (UI) consistent. Since this is the only UI of the the mirror, it has no style restrictions. Besides consistency, it is also known that using multiple UI modalities results in better understanding of the system, as long as it is kept simple, preventing sensory

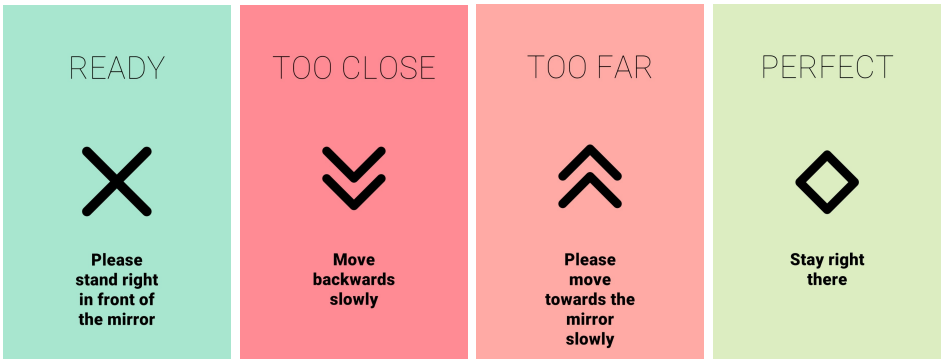


FIGURE 3.5: Calibration screens.

overload [84]. Google's UI design guidelines[1] state that *"to emphasize which information is important, multiple visual and textual cues like colour, shape, text, and motion add clarity"*, which also creates a more inclusive system, as more people can understand it. Each of these modalities is implemented in the design of the calibration screens and shortly described below.

- **Colour** is used to quickly indicate the effectiveness of the current user action. This positive-negative colour association is hard-wired into people's brains [85, 86], and since the mirror is physically large, the colour change will immediately noticeable. The colours are part of a coherent palette[2], two or which can be considered positive, and two negative. No colour is reused, distinguishing each phase. A positive colour is used at the start screen to be invite the user. The negative colours are used when corrective action of the user is required. Finally, a positive colour is used when the person should stand still, signalling a successful action.

- **Text** is most important, as it is the clearest instruction modality. Firstly, text is added at the top, where attention is pulled according to Google's Guidelines. Text is kept as short as possible, with large font size, to signify importance and increase readability. A more detailed explanation is provided at the bottom, a natural place users look for more information. The text is phrased in a neutral to friendly manner.

- **Shapes** and **Motion** are used to assist the instructions of other modalities. Googles Guidelines suggest transitions improve user experience in continues systems and thus, to make the system feel modern, four icons were added, that each consist of the same two elements: two chevrons. By animating the rearranging of these two elements, all the shapes in Figure 3.5 can be created, each thematically fitting to their respective screen. To amplify the effectiveness of these shapes, subtle animations were added. The cross and diamond shape pulse in size, while the 'arrows' move up and down. Finally, to indicate the required standing still time, the diamond slowly fills itself, until the calibration is completed.

- **Sound** is not used, as it is often used to direct attention to something, but since users are focussed on the screen already, sounds are expected to increase chances of sensory overload. Besides that, sound is a medium not associated with real mirrors.

[1] `https://m2.material.io/design/usability/accessibility`
[2] `https://colormagic.app/palette/6719c99b71d46152958e0184`

3.5.2 Pose estimation in Python

Running MediaPipe's Pose estimation API in Python is simple, only requiring limited code. The API runs on the CPU by default, for which a 4K webcam is computationally too demanding. An overview of processing times can be found in Table 3.2, gathered using the built-in preview tool on MediaPipe's website, which uses your computer's components. Here, it is visible that using a GPU can significantly improve performance, freeing up system resources for other applications such as Unity. There are, however, other processes that add to the overall latency, such as OpenCV capturing the webcam image, adding roughly 15ms of delay. The total latency tested for the Lite model on CPU is thus more around 40ms, which would equate to 25 FPS. This is without Unity running. A lower latency is preferred, to ensure a smooth user experience.

From testing, it is decided that running the Lite is preferred over the Heavy model, as latency is more important than accuracy for this research. With the Lite model, switching MediaPipe to GPU reduce the latency by half, so this approach is preferred.

	CPU		**GPU**		
Model	Latency	FPS	Latency	FPS	Difference
Lite	17-22ms	∼52	9-12ms	∼92	+86%
Heavy	88-92ms	∼11	14-17ms	∼60	+485%

Table 3.2: MediaPipe's reported processing time.

Accessing the GPU for computations on Windows can be challenging, especially using ML frameworks like TensorFlow. TensorFlow used to support GPU acceleration natively on Windows, but this feature was deprecated in favour of WSL. MediaPipe, which integrates TensorFlow, benefits from the latest versions, making GPU-accelerated execution on native Windows infeasible. Since Linux still supports GPU acceleration, running a separate Linux computer would be an option, majorly increasing both the cost and complexity of the system. Instead, Ubuntu (a Linux distribution) was installed within Windows as a virtual machine (VM), via WSL. This enabled GPU acceleration, and running the same Python code, the GPU was recognized and performance improved. However, communication methods that were previously established, especially Serial communication, did not function properly within WSL. While some workarounds exist, Serial communication could not be established. Because of this and to simplify the system, an alternative approach is needed.

3.5.3 Virtual camera split

Since the 'optimal' solution proved too complex to implement within the constraints of this research, a workaround method is applied. At this moment in development, Python feeds a 4K resolution video feed to the pose estimation models. Higher pixel counts equals more information to process, resulting in the system slowing down significantly. In this prototype, a high resolution video feed in Unity is important, as it increases the believability of the system. Lowering the resolution system wide is thus not preferred. However, lowering the resolution of the video feed to only MediaPipe would be a solution. MediaPipe return its data as relative texture coordinates, regardless of resolution. Lowering resolution would yield the similar results, but less accurate, allowing for easy implementation.

The first approach to achieving the aforementioned, was utilising down-scaling within Python. This avoids increasing system complexity, as OpenCV, already used in the project, provides a solution for this. While this lowered the processing time of MediaPipe considerably, the additional overhead caused by down-scaling increased the processing time almost equally. Timing all individual processes in the Python code, revealed that a large contribution to the total time was caused by OpenCV simply 'reading' the webcam feed. These findings highlighted the importance of lowering the resolution before reaching Python.

OBS allows any camera feed as input, can change it, and output a virtual version. The current feed is already converted into a rotated virtual 4K stream. All tested virtual camera applications are unable to split the incoming webcam stream into more than one virtual feed with different resolutions. They also do not allow other virtual cameras as input, meaning 'daily-chaining' them is not an option. OBS has the capability to take a virtual camera created by another instance of OBS as its input. However, two instances of OBS cannot have their virtual camera active at the same time. Another solution must thus be found.

Since OBS is open-source, there is a plethora of plugins available. Some of these claimed to support multiple camera outputs, but only allowed it while connecting directly to a broadcasting platform, such as Twitch. Before the official implementation of a virtual camera, there was a plugin that allowed for (multiple) virtual cameras: OBS Virtualcam[1]. The resolution was unfortunately static, and since the official implementation in 2022, this plugin was marked as deprecated, and is no longer supported on newer versions of OBS. However, one older version of OBS supported both the official solution and the plugin. Additionally, since both virtual cameras differ in approach, they can be active concurrently. This results in two virtual cameras outputs using one instance of OBS: a higher resolution version that is fed to Unity; as well as a lower resolution version that is fed to MediaPipe in Python, significantly lowering frame time.

3.5.4 Position in Unity

Unity receives two types of data: a distance measurement in mm, as a float; and a list of all pose estimation landmarks, as Vector3's. The distance is processed simply converted to fit Unity's coordinate system. The landmarks are harder to process, since the vector's x and y are normalised texture values. This makes the system modular, but also means conversion is needed. The z-value of the vector is also arbitrary, as the pose estimation does not have a reference point. This value is not used to calculate user position.

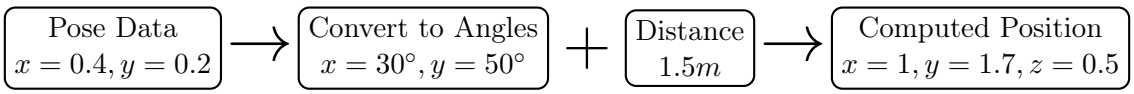


FIGURE 3.6: User Position computation process with example values.

Figure 3.6 above shows the computing process of user position in 3D space, using example values. The normalized pose texture coordinates are converted into a usable data type: angles. The pose estimation returns a set of 33 landmarks, see Figure 3.7, of which two are important at this stage: the eyes. Taking the average value of landmarks 2 and 5, returns the user's point of observation. By combining this with the FoV of the camera and its image distortion (which was not reported by the manufacturer, and was thus tested), the angle from the camera to the user's observation point can be calculated. Other

[1] `https://obsproject.com/forum/resources/obs-virtualcam.949`

physical properties of the camera, such as its positioned height and angle, also influence the algorithm.

Distance, the third step in Figure 3.6, is computed in two ways. First, when the calibration process is completed, distance data gathered from the ToF sensor is used. After this the sensor data is not used, only utilising the calibration value as a reference point. During calibration, the distance between multiple points in the face is also recorded, the size of the face is then also used as a reference. After calibration, the distance is calculated based on the size of the face compared to the reference size. For example: the user's face is half the horizontal size compared to the measurement at 2 meters, so now the users is assumed to be roughly 4 meters away.

Finally, the user's observation point position can be computed. Resulting in an accurate 3D representation of the user, used to calculate the proper reflection. This system is modular and other landmarks can also be tracked if needed.

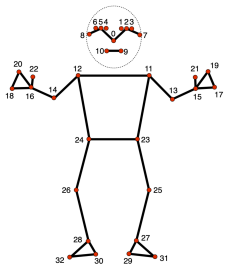


FIGURE 3.7: Overview of the landmarks generated by MediaPipe Pose[1]

3.6 Perspective calculations

Using user position, the perspective calculations are possible. This section described two approaches: the 'faithful' approach, which tackles the ideas described in Section 3.1.2 and illustrated in Figure 3.1, and the 'rules-based' approach, which breaks down a mirror in fundamental properties and recreates them digitally. Two approaches are used, as the first approach was found problematic surrounding its feasibility, which is described below. As this is the straightforward approach, it is important to highlight its flaws and possible improvement for future work. The second approach was found more successful and therefore used for the final prototype.

An important limitation for either approaches is that Figure 3.1 assumes a free moving camera with adjustable optical zoom. This is not possible in the context of this research, as a stationary camera is used. This limitation prevents the system from being truly accurate, as it relies on a 2D representation of a 3D environment. However, this representation can produce a reasonable estimate by projecting the camera's feed (horizontally flipped) on a plane corresponding in size with the camera's FOV, bending it towards the mirror, as shown in Figure 3.8. This results in the camera's 2D feed being accurately represented in 3D space. This implementation is consistent between the two approaches.

3.6.1 Faithful solution

For the faithful solution, three steps are needed, in order. First, the virtual camera is moved to the correct position according to the user position; second, the camera is rotated towards the virtual mirror; and finally, the zoom is adjusted to look through the virtual mirror.

The first two parts are relatively simple, since the user position and mirror dimensions are known in 3D space, as described in Section 3.5.4. The virtual camera is takes the user position and flips the z-axis, the other two axis stay the same. After this, the camera is simply rotated to face the of the middle of the virtual mirror.

The final part is more complex, allowing for some liberties being taken. The virtual camera has the correct rotation, only the zoom needs to be determined. Optically zooming in Unity effectively changes the FOV of the camera: a low FOV means a smaller cone of vision, zooming in the image. Using trigonometry, the correct FOV can be calculated based on the distance between the camera and the mirror's corners. This is a perfectly accurate approach when standing straight in front of the mirror, but using a fixed aspect ratio produces inaccurate images at an angle. The frame of the mirror does not form a perfect rectangle when viewed from an angle, with the camera forcing it regardless. While technically not correct, warping the image at larger angles is not expected to not be noticeable users.

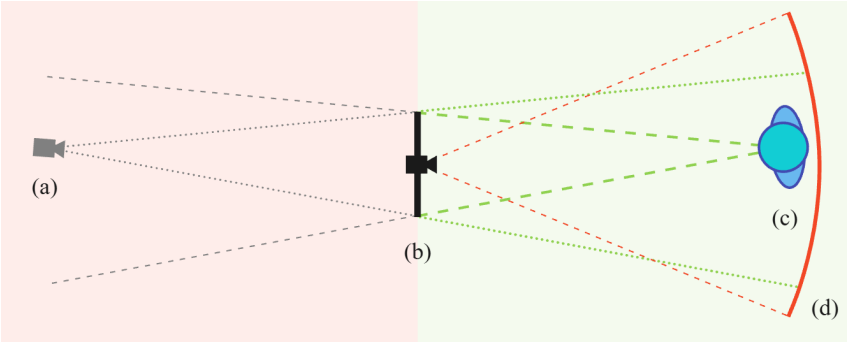


FIGURE 3.8: Setup of the faithful solution in Unity. (a) shows the virtual camera, (b) the mirror and the position of the real camera, (c) the user, and (d) the plane projecting the image of the camera. The right (green) part shows the real world, left (red) the virtual world. Red dashed lines show the FOV of the camera.

While this approach appears to work on paper, and while testing to a certain degree, too many shortcomings arise. If the real camera could physically be moved is a similar manner to the virtual one (mirroring the user), this approach would produce a result extremely similar to a real mirror. It was assumed that by physically accurately projecting the static camera feed, combined with moving the virtual camera properly, a reasonable reflection estimation could be constructed. However, this result did not meet the desired quality. A major consideration is determining the camera plane distance. Close placement resulted in a user moving towards or away causing strange results, and far placement nullified horizontal and vertical head movement. Both options are flawed. Another issue is that users would appear taller or shorter depending on their distance to the mirror, especially for user far from the average height. This is caused by the system's assumption that the real camera is placed in the middle of the mirror, not on top, nor at an angle.

This mismatch of physical and virtual camera produced inaccurate refections. Although difficult to quantify the exact problem, multiple user trying the mirror mentioned it "felt off". On smaller screens these problems were much less pronounced, adding validity to the solution in this context.

3.6.2 Rules-based solution

Since the setup fundamentally differs from a real mirror, described in Section 3.1.2, an alternative solution is required. This is a novel problem, requiring a bespoke solution. After analysing the inner workings of a mirror again and careful consideration, the following approach is established: breaking up the mirror in fundamental rules/laws, and replicating these virtually, resulting in a prototype that mimics the behaviour of a real mirror. The rules are as follows:

- Camera movement should consist of only rotation
- FoV should scale between predefined values
- Eye height should always match on screen

The previous approach moved the virtual camera position based on the movement of the user. This proved to be ineffective with a projected camera feed, resulting in odd perspectives. Since the camera's projection plane is facing the mirror, using this as the camera positions allows for all possible perspective to be achieved using just rotation in the x and y-axis. By only allowing rotation, the system decreases in complexity and provides better insight into possible perspective oddities. While the y-axis rotation requires considerable change in this approach, the x-axis can be easily adjusted. The position component is removed and tweaking the rotation to only mirror the user's x-rotation.

Unlike the other approach, the FoV scaling can be 'hardcoded', it is predictable based on distance. Consequently, changing either the camera or the screen would disrupt this system. To predict the FoV value, three measurements, at close (1.5m), medium (3m) and long (4.5m) distance, were taken. Each measurement calculated the physically correct FoV using trigonometry, and computed the virtual camera's FoV to match the same output, see Figure 3.10. With three measurements, a quadratic equation can be fitted, resulting in predictable FoV at all distances with reasonable accuracy. User height slightly impact FOV, therefore a person of average height was used. Since the system is not a perfect representation of a real mirror, an additional small percentage offset is seemingly unnoticeable to a user. The same applies for horizontal FoV differences at greater angles.

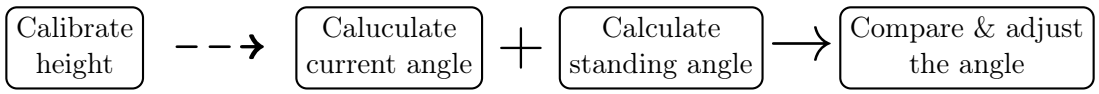


FIGURE 3.9: Eye to mirror height calculation pipeline.

The last, most important, and complex rule will brings the systems together. The downward angle of the camera caused issues in the previous solution: this needs changing. When looking in a mirror, one important observation (assuming an upright mirror) is that the user's eyes will always match their height on the mirror, unaffected by distance. This is only true for the eyes, since they are the observation point. Eye height on the mirror only changes when the user moves it physically, by f.e. ducking or jumping.

Figure 3.9 shows the process overview, where the first part happens during the calibration process described in Section 3.5.1. Upon calibration, besides calibrating the distance to the user, the system also calculates their height. This is done using trigonometry, illustrated in Figure 3.11, and is possible with the physical properties of the mirror and camera. The same technique is then repeated every frame, to calculate the current height.

This was, however, found to be unreliable, with values being off by a noticeable margin. This is presumably due to slightly incorrect distance data and/or distortion in the camera lens. To compensate for this, the same trigonometry was used backwards, to calculate the supposed angle the person needed while was standing up straight.

There are three values that are of importance: the user's height, the current angle, and the standing up straight angle. First, the calibrated height of the user is used to apply the third rule mentioned above. This is achieved by projecting the user's eye height on the mirror on a horizontal plane, as demonstrated by the purple line in Figure 3.11. This is formatted as a value between 0 and 1 indicating vertical placement on the mirror, f.e. 1.70m eye height translates to 0.75 of mirror height. The FoV is calculated separately, leaving only the rotation of the virtual camera until the angle to the eyes is 0.75, or 75% of the virtual image. Lastly, the angle is adjusted up or down based on the current eye height of the user.

This results in a system that, though not fully accurate and less modular than its alternative, estimates the reflection the user should see reasonably well. Additionally, the system is considerably more reliable at longer distances and better handles users of different heights.

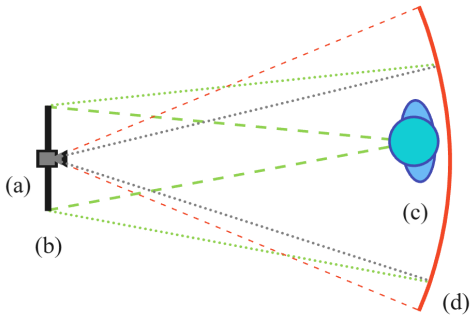


FIGURE 3.10: Setup of the rules-based solution in Unity. See Figure 3.8 for description. The dotted grey lines show the cropped FOV to capture the desired part of the image.

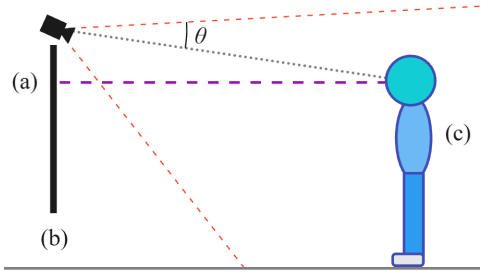


FIGURE 3.11: Eye height calculations. (a) shows the eye height point on (b) the mirror, from (c) the user, projected by the horizontal striped purple line.

With either approach used, standing too close to the mirror (<1m) emphasises the camera's angled image. Additionally, the system struggles to accurately determining user position at this range. To mitigate both these issues, a fade to black is implemented when the user distance becomes too close, signalling the user should move backwards.

3.7 Augmented reality

The second part of implementation regards the augmentation of the output image. Since the image is digital and physically accurate in Unity, everything can be modified or added. This proof-of-concept research aims to verify that a physically accurate reflecting digital mirror is feasible and that image augmentations can be functional and believable, future research is encouraged expand on the latter part. Complexity and added latency are preferably kept low, a solution that directly makes use of the existing systems is thus preferred. MediaPipe provides landmarks, but also the possibility for a segmentation mask, which separates the user from the background. This provides two augmentation opportunities: adding something to the user, using the landmark system; or adding something to the environment, occluding behind the user. As self-perception is the cornerstone of this research, adding something to the person promises more interesting results.

3.7.1 Position

The first task of implementing an addition to the image, is determining the flat position of the object on the user's body. MediaPipe outputs the position of 33 landmarks, see Figure 3.2, of which three or more are required for accurate calculations. From these points, a plane can be formed, allowing for the rotation and scaling of the addition. In this implementation, four landmarks are used: the corners of the torso (landmarks 11, 12, 23, and 24). Any point relative to these landmarks can be used (f.e. the middle), returning the position as a texture coordinate. This coordinate can be translated into an angle using the same method shown in Figure 3.6. Placing an object where this angle intersects with the image projecting creates the illusion of attachment to the user. Smoothing is applied to compensate for pose estimation jitters/inaccuracies.

Currently, the position is exactly in-between the landmarks, which produces a believable result from straight ahead, but when the body is turned the object appears inside the body, breaking the illusion. This means the object requires an offset outwards. To achieve this, the rotation of the user must be determined.

3.7.2 Rotation

Using the same four landmarks, a rotation should be computed. Looking at Figure 3.12, which shows how landmarks are represented in 2D, reveals the lack of an easy solution. A human, with the knowledge that this is a square, can undoubtedly recognise it is rotated to the right, as the vertical line on the right is larger than its left counterpart. This is a simple example, but more complex cases can arise, due to weird user positions, camera angles and distortions. Writing code to accurately detect this is thus difficult, and only applying the above rule did not yield satisfactory results. Other techniques were performed using the texture coordinated of MediaPipe, to no success. As a purely mathematical approach became too complex, an alternative must be found.

Besides the x, y coordinates, MediaPipe also provides a z-axis for its landmarks. For the distance calculations, this was found inaccurate and unstable, and thus unfitting for the problem. However, using the z-value to compare against other landmarks could be beneficial. First, the direction of rotation is required. This can be calculated using the z-value of the shoulders, since the torso is relatively stiff and shoulder should remain clearly visible to the camera at all times. The z-values each frame are difficult to compare, but by storing the z-values upon calibration, when it can be assumed the person stands straight

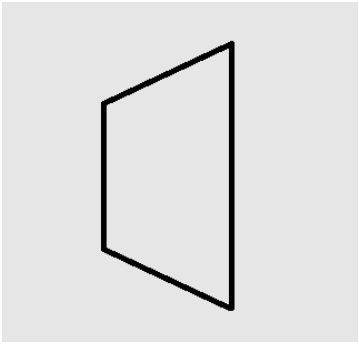


FIGURE 3.12: The torso projected as a 2D quad.

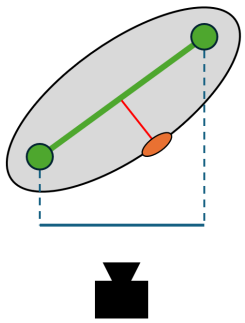


FIGURE 3.13: Offset calculations, top down view.

up facing the mirror, and comparing the values, consistent rotation detection is achieved. This only works for horizontal rotation, vertical rotation produces inaccurate results, due to distortion in the camera. In the context of a mirror, vertical rotation is highly unusual. Therefore, this rotation was abandoned, preventing the interference of other results.

Next, the amount of rotation can be calculated. By looking from the front, as in Figure 3.12, this is difficult to distinguish. Figure 3.13 provides a top-down perspective for better insight. For the camera, the green line appears as the blue line. If the length of the green line is known, the angle between the two can be computed, using trigonometry. Unfortunately, the length of the green line is not known exactly, but a reasonable estimated can be produced, using the calibration phase as a reference point to scale with distance. The user faces the camera during calibration, resulting in a roughly equal green and blue line length. Implementing this provides an angle, which, combined with the rotation direction computed earlier and an offset away from the body, results in the augmented object rotating with the user. Inaccurate distance calculations caused inconsistent angles, multiple band-aid fixes have been applied to compensate for this.

3.7.3 Scaling

Finally, the object's scaling should match the user. This is achieved by linking the scale modifier to a reliable scale indicator on the user's body. The scale factor of the torso is not suited for this, due to shape changing described in the previous section, making it hard to determine whether the person is turning or moving away. The system already has a consistent scaling solution, using the face. It can be assumed that this will always facing the camera and is therefore used for distance, see Section 3.5.4. Requiring few tweaks, this works acceptably well, completing the functionality of the augmentation.

3.7.4 Coupled vs uncoupled video

The video stream is split in OBS, with both data streaming ending in Unity, as seen in Figure 3.2. The data stream through Python takes more time, which results in a de-sync between the two data streams in Unity. The perspective calculations are unaffected by this, but it causes issues for the augmentation. Since the augmented object required data coming in later than the image it's projected on (plus a bit of latency added by smoothing), the object noticeably lags behind the user's movements. This can be reduced by delaying the camera output in Unity to match the Python data stream, increasing latency. It is therefore necessary to determine what takes priority: responsiveness or believability:

perspective calculations or image augmentations. In technical terms, the video and overlay must either run coupled, or uncoupled. This is visualised in Figure 3.14.

- When the two are **coupled**, the video feed holds the processed frame until the overlay finishes, and then displays both at the same time. In this case, the two overlap perfectly, the overall latency now bottlenecked by the overlay processing time. This latency will be noticeable in quicker movements, as the position of the user and the displayed position do not match (temporarily).
- In **uncoupled** mode, the video feed is displayed as soon as it is processed, resulting in a smooth and responsive experience. The overlay processes independently and superimposes its image once finished. The overlay image thus appears a few frames after its corresponding footage, lagging or trailing behind the video feed.

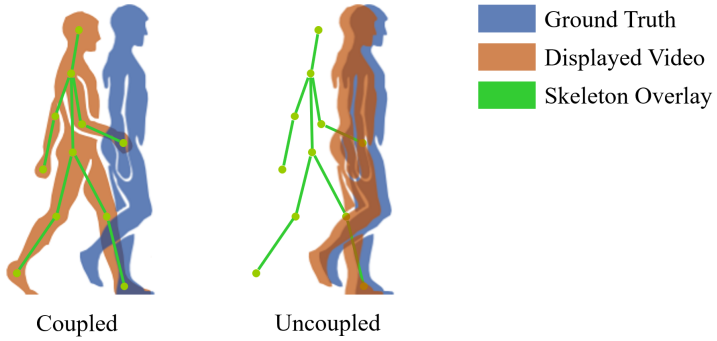


FIGURE 3.14: Coupled versus Uncoupled video comparison

Augmentations not overlaying properly or disappearing, potentially break the 'magic' for many users [87], creating an instant disconnect between the user and the overlaid information. Therefore, the possibly meaningful effect of the augmentation might be negatively effected. Since the augmentation of the image is a key aspect of this prototype and links directly back to self-perception, it is preferable to avoid any disconnect unrelated to the augmentation effect, thus using the coupled version promises to yield better results. Additionally, the latency of the system is already higher than the original goal of $<$100ms. Adding more latency is not expected to negatively influence the results as much as running the augmentation in uncoupled mode would. For these reasons, the system will run coupled.

There is a simple method to achieve this goal: frame buffering. This creates a buffer which stores the last x number of frames, releases them after a specified time. This was first attempted within Unity, but was found not efficient, since the camera is high resolution. Alternatively, the OBS Virtual camera plug-in has a frame buffering feature built-in, which only affects the video feed to Unity. The amount of buffered frames can be set, which at the camera's 30 fps, results in 33 ms of delay per frame. Setting the frame buffer to 4, so 133 ms, was found to yield the best coupled result.

To confirm the coupled feed, a quick test was performed. Two versions, the uncoupled with 0 frames buffered, and coupled with 4 frames buffered, were shown to participants (n=8), asking for their preference, specifically on the tracking quality of the augmentation. 7 out of 8 participants preferred the coupled version, one person stating no preference. Interestingly, none of the participants mentioned, or appeared to notice, the increase in latency in the coupled version.

Chapter 4

Evaluation

4.1 Research goals

To answer the research questions, the digital mirror prototype will be evaluated by real users, in a controlled environment. This evaluation aims to provide insight into two key aspects: the experience of using the digital mirror, comparing it to a real mirror; and the behavioural impact of the augmented reality features. To gather useful data, a controlled example augmentation is required for testing.

4.1.1 Virtual stain

The digital mirror allows for any augmentation of the final image. The augmentation should directly aim to trigger self-perception in the user. The digital mirror can be considered successful when users believe the 'reflections', even when augmented. This would then paves the way for more experimental and impactful future research, as other augmentations are expected to achieve a similar effect, making it valuable for f.e. phobias and body disorder research.

As this experiment aims to provide an evidential basis for future research (and to stay within the scope of the project), the augmentation applied here should be simple, believable and provocative: most participants should physically react to it. It should appear on the person's reflection rather than in the 'scene', creating a self-relevance to the user. Furthermore, the augmentation should make contextual sense. A primary reason to use a (large) mirror is to check if outfits, looks good and, importantly, are clean. Adding a stain to the user's clothing fits these requirements and is expected to provoke a reaction, as this is generally considered unwanted. The reaction this causes, whether verbal or in an action, should be captured and stored, and should provide valuable insight into the possibilities of image augmentation.

The initial implementation of the stain, shown in Figure 4.1 on the left, is roughly 5 cm in diameter, is shaped as a drink or food stain, which is something that most people should be familiar with. The stain is also coloured slightly yellow, to make sure it provides contrast on as many colours of clothing as possible. The stain is placed around the middle of the torso, clearly visible, and where it could realistically occur.

FIGURE 4.1: Picture of the virtual stain projected on a person. Left shows the initial implementation, right the adjusted version after feedback from the pilot test.

4.1.2 Metrics

The research questions, require specific metrics to be collected. The two key aspects of this research differ in this: the experience of using the digital mirror is a usability question, whereas the impact of the augmented reality stain is a behavioural one. The data collected is therefore also different.

To get a grasp on the user experience of the digital mirror and its augmentation, it is important to collect the thoughts of the participants during and after its use. What they think about the prototype, how it compares to a real mirror and whether they believe the stain to be real, are questions that all provide important insight. This data is mostly qualitative, but quantitative data is also be gathered through more normalised means. Data is also partially behavioural: how participants interact with the prototype, how much they move around and how participants react to the stain.

4.2 Study Design

The study consists of a single-condition user test focused primarily on gathering qualitative data, in addition to some quantitative measurements. Each participant interacts with the digital mirror prototype for a short amount of time (a few minutes), which is expected to create a reasonable understanding of its capabilities and limitations, besides noticing the virtual stain. More time is not necessary as the system is inherently simple, and mirrors are a familiar concept to users. After the interaction, observations, questionnaires, and a short interview are used to gather insights on user experience and perception.

This study serves as a proof-of-concept rather than a comparative evaluation, as the system is novel. The direct comparison against a real mirror is not productive, since the augmentation cannot be repeated here. The digital mirror, by nature of being a digital system, also inherently differs from a real mirror, as it introduces a delay, impacting its responsiveness. A direct comparison would almost certainly highlight the limitations of the system, rather than evaluating the potential of the system itself. Instead, the design is centred on exploring whether the digital mirror, in isolation, is perceived as a mirror, and incites meaningful self-perception responses.

The relatively short interaction duration was chosen for theoretical and practical reasons. From a user experience perspective, mirrors are typically used for short tasks, and participants that do not notice the virtual spot within this time frame, are unlikely to benefit from an extended session. Practically, a shorter procedure facilitates easier participant scheduling and recruitment.

4.2.1 Data collection

An overview of the different methods of data collection used during the experiment is listed below. Each entry will be elaborated upon afterwards.

- **Observations:** Behavioural cues, mirror usage
- **Logs:** User position, timestamps
- **Questionnaire:** User experience, perceived realism
- **Interview:** Prototype evaluation, additional context

The first type of data gathered is observations by the researcher, during the participant's interacts with the digital mirror. Close attention will be paid to the behaviour of the participant, especially their head and hand movement when related to the stain. More specifically, occurrence of as looking down or trying to wipe off the stain are noted and their timestamp recorded. These action can possibly provide more context into the subconscience of the participant. They are later labelled and categorised, creating quantitative data. Participants will be asked to perform tasks in front of the mirror: putting on a buttoned shirt, and tying a tie around their neck, tasks that are often performed with the help of a mirror. These tasks are used to keep the participant in front of the mirror longer, letting them use it for something tangible, which should result in more and better impressions of the system. Another benefit is that participants can choose whether or not to use the mirror to aid in their tasks, which will be a point of attention of the observations. During the process, participants are also encouraged to 'think out loud' to aid the researcher in observation note-taking. The observations will generate data for both aspects of this evaluation. This phase is also video recorded for redundancy and possible future cross-referencing with the other data.

During the interaction with the digital mirror, some data can also be logged by the system itself. A logging system was created that saves all data after each test into a file. This data includes: the estimated height of the user; timestamps for important events, such as completing the calibration; and three-dimensional positional data, quantifying movement and allowing the creation of heatmaps, to check for differences between users or groups.

Most quantitative data will be gathered using a questionnaire. It focuses on the user experience of the digital mirror and less on the augmentation, since results here are expected more subtle and personal, making it less suited for a questionnaire. Additionally, other methods of gathering quantitative are present, as mentioned in the previous paragraph, which allows for a shorter questionnaire. A shorter questionnaire has multiple benefits over longer ones: they generate more responses, as less motivation is required [88], and they induce less boredom, creating more consistent attention and question answers [89]. Less data can be gathered, but since there is no testing between conditions, the data primarily serves as an indicator rather than a validation tool.

Standardized questionnaires exist in the user experience domain, such as the System Usability Scale (SUS) [90], Usability Metric for User Experience (UMUX) [91] and User Experience Questionnaire (UEQ) [92], that can generate a score from questionnaire responses. These questionnaires and their scores are great for comparing systems to others that used the same questionnaire, or for testing for significant differences between conditions. Neither use case is applicable here, as there is only one condition, and the digital mirror would undoubtedly score badly against a real mirror. Therefore, the existing questionnaires were

used as inspiration for the final questionnaire, which can be found in Appendix A. It consist of nine questions, eight of which are statements that use a 7-point Likert scale, from 'strongly disagree' to 'strongly agree'. The final question aims to gaze where participants would place the digital mirror, between a static video feed and a real mirror on a 9-point scale. Besides these questions, age and gender of the participants are also recorded.

Ending with an interview should allow the participant to form and summarise their thoughts, and allow any gaps to be filled. It is meant to add additional context to the other data gathered. The interview is semi-structured covering all aspects of the prototype, keeping the flexibility of asking follow-up questions. It consists of six questions, that are kept relatively open, so that the participant can indicate what was important to them. The interview questions, and their possible follow-up questions, can be found in Appendix B. Audio is recorded during the interview, everything is transcribed and labelled.

4.2.2 Participants

From the qualitative data viewpoint, testing with 9 to 17 participants has been found to reach 'saturation' in results [93]. More participants beyond this point is unlikely to yield further insights [94]. However, since quantitative data is also gathered, this number is preferably higher. The number of participants should enable testing for correlations between variables and possible emerging groups. The goal is not to provide definitive statistical proof, but to analyse descriptive statistics.

Every person has interfaced with a mirror, resulting in no specific recruitment exclusion criteria. Therefore, convenience sampling is used, streamlining participant recruitment. Acquirement mostly occurred through promotional online messages in groups of students, as well as asking people in or near the lab facilities. A few participants were recruited through word of mouth.

In total 37 participants were recruited and completed the test, 2 of which were used as pilot tests. Out of the main participant pool, 31 were students at the University of Twente between the ages of 18 and 28. 20 participants identified as male and 15 as female. Furthermore, 15 participants had a comparable background/study program to the researcher, the rest study/work in different fields, most with some technical affiliation.

4.2.3 Physical setup

For this study to be conducted in a controlled manner, the setup is moved to an isolated room, fitting the mirror and enough space to move around. Additionally, a table and chairs are present where the participant and researcher can sit. An information brochure and a consent form are present, as well as a tablet used for filling in the questionnaire, and the materials needed for the tasks. The setup was kept to a minimum to avoid distraction and reduce potential variables. A black sheet covering the mirror was placed before each experiment to conceal its appearance and prevent participants from forming excessive preconceptions. Figure 4.2a below shows a picture of the setup.

(A) The full setup for testing. (B) Participant inspecting the virtual stain.

FIGURE 4.2: Photos of the digital mirror setup and user interaction.

4.2.4 Procedure

The procedures is summarised below in chronological order. As mentioned before, all parts are kept short: the full procedure taking about 15 minutes.

1. **Introduction**
   The participant is welcomed and brought to the testing room. The information brochure is presented for reading and the consent form for signing. They are also verbally reminded about the video and audio recording during the experiment. Any questions are answered by the researcher.

2. **Briefing**
   The procedure, as described below, is explained to the participant. They are encouraged to 'think out loud' and reminded that talking to the researcher is allowed, but a response cannot always be expected.

3. **Experiment**
   The sheet covering the mirror will be removed, and the participant is instructed to follow the instructions on screen. After calibration, the mirror activates, as does the addition of the stain. After about a minute of use, the participant is instructed to perform the tasks: putting on the buttoned shirt and tying a tie. Once completed, they are instructed to take a seat, once they feel like it. The participants are closely observed by the researcher during this phase and all findings are noted.

4. **Questionnaire**
   The tablet containing the questionnaire is handed to the participant. They are reminded that they can ask questions if something is unclear.

5. **Interview**
   A short semi-structured interview is conducted, and the participants are reminded it is audio recorded and live transcribed by a locally run ML model.

6. **Debrief**
   The participant is thanked for their cooperation and is rewarded with a treat. The goals of the study and the reasoning behind the procedure are explained, any questions the participant may have are answered.

The experiment and its data gathered were approved by the faculty's ethics committee[1] and all participants signed an informed consent form.

4.2.5 Pilot testing

Two pilot tests are performed to ensure a smooth procedure, each using one participant. Once conducted and their feedback processed, the main tests are performed with the remaining 35 participants.

First Pilot

The first pilot test is conducted using a participant with similar technological and usability background and affinity. Feedback was provided and discussed during the procedure, pausing when needed. This allowed for discussing each part in depth on its use and reasoning. The following aspects were changed after the first pilot test, some of these are already mentioned before.

Initially, the participant's task was putting on and off their jacket/coat. In this pilot test, it was determined this could possibly lead to different results between participants, and it was deemed not long enough: people can put their own jacket/coat on and off very quickly. Therefore, the tasks were changed to a more controlled and time intensive activity: putting on a buttoned shirt and tying a tie. Additionally, these tasks also better fit the context of a mirror.

As described in the methodology chapter, the screen fades to black when the user gets too close, preventing multiple issues and oddities. A problem arose with the fade to black, as black is the default state of any screen that is turned off. When the screen turns black, from a UX point-of-view, it does not clearly communicate whether this is by design or it is not working properly. Besides this, the screen turning black essentially turned it in to a mirror with its reflections, which could potentially influence results. The fade is therefore changed to a light grey, solving both issues and being easier on the eyes than f.e. white.

Another observation was the impact of clothing colour on the visibility of the stain. Instead of the stain using a fixed opacity, it should change depending on the brightness of the clothing. An automatic system for this was attempted, not achieving consistent results. Instead, the researcher manually labelled each participant's upper clothing as either 'light' or 'dark' and changed the opacity of the stain accordingly. The clothing brightness label was saved for each participant so its possible impact can be analysed later.

Finally, various interview and questionnaire questions were rephrased, reducing interpretation errors, to yield more consistent results. A few questions were also removed for being too similar to another questions, or for adding no further insights. The question *"What made this not feel like a real mirror?"* was added to ensure participants would provided insightful commentary on the prototype.

Second Pilot

The first pilot test saw many changes made to the procedure and prototype, a second pilot test is thus valuable. This test was performed similar to a real test, using a person with a different background from the researcher. Feedback was only provided after the procedure, so that the general flow (and time) of the procedure could also be tested.

Comparatively speaking, the feedback was minor and more positive. The phrasing of the briefing was adjusted to remove potential bias and suggestiveness. The digital mirror

[1] https://www.utwente.nl/en/eemcs/research/ethics/

was also moved to a more optimal location in the room, allowing for more freedom of movement. Additionally, the setup was made cleaner with less visible 'machinery', by hiding the wiring better and removing any distractions.

The stain was also changed: it was found to be too obvious, due to its resolution being too high in comparison to the video feed. It was therefore reduced in resolution and a blur effect was applied, until it roughly matched the video feed. It was also found too large according to the pilot tester, and thus too noticeable. The stain was reduced in size, resulting in a more subtle effect. The final implementation can be seen in Figure 4.1, alongside the initial implementation.

After feedback was integrated, the prototype and procedure proved ready for final testing.

4.2.6 Participant tests

The effect of pilot testing was noticeable as the participant tests ran smoothly. Participants were scheduled in 30 minute slots, allowing enough time to perform the test and reset. Resetting the system could be completed within a minute. This quick testing meant that the majority of the 35 total participants could be tested in the first few days. Most tests fell within a couple minutes deviation of the 15-minute mark, as essentially zero issue plagues testing. The system's robust design proved extremely reliable, never crashing or delaying the tests.

Seemingly due to the short nature of the test, the instructions were well received, with most participants understanding the assignment. While all participants were eager to try the digital mirror, there was a noticeable difference in participant behaviour: some stood still, wondering what to do and sometimes needed encouragement from the researcher, contrasted by others trying to test the limits of the system.

Chapter 5

Results

Each interview audio recording was transcribed and anonymized, after which responses were summarized and categorized based on the predefined interview questions. This data was then coded into recurring themes and topics to enhance its quantitative interpretability, reduce complexity, and minimize bias. The same technique was applied to the observations and participants remarks made during the digital mirror interaction phase. An overview of questionnaire responses can be found in Appendix A.

All statistical tests performed used a p value of 0.05 to test for significance.

5.1 Participant impressions and behaviours

This section covers the most important findings, according to the participants, as well as emerging ones. These have no overarching theme, the next sections will cover specific topics. Findings are not categorised per data modality or related research question, but rather by topic, to provide holistic impressions.

5.1.1 Virtual stain

When asked directly during the interview, about participants' initial reaction the virtual stain, 17 out of 35 participants stated the stain might be physically there on their person ('believers'), 15 participants stated it was fake ('sceptics'), and 3 participants did not notice the spot at all ('unaware'). From the believers, not everyone interpreted the stain as a stain. A few participants though it was something else, like a spot of light, but believed it was physically on their person nonetheless.

The above-mentioned statistics were self-reported. For another, possibly more objective, perspective, observations were made of participant behaviour regarding the virtual stain, during the interaction with the digital mirror. Four main actions were observed, and their occurrences counted. The unaware participants were excluded, as they exhibited no behavioural cues towards the virtual stain. In total, out of 35 participants ...

... **22** performed at least one of the actions in this list.
... **19** touched the spot where the stain was projected on their body, using their hands.
... **5** looked down at the spot where the stain was projected on their body.
... **5** moved towards the mirror to look closely at the stain.
... **3** (consciously) waved their hands in front of the projected stain.

Figure 5.1 below visualises the results mentioned above, allowing alternative data interpretation. The figure shows that all participants that reacted, used their hands, as 19 tried to touch the stain, while 3 only waved their hands in front of it. Furthermore, 6 participants exhibited two of the observed behavioural cues, while 2 participants demonstrated three.

The behavioural cues mentioned above were obviously detectable, as they are deliberate towards the stain. Others cues were observed that might be in relation to the stain, such as rotating or moving the lower body to if the stain would be stationary or not. A few of participants confirmed this, but these cues were too ambiguous to allow for accurate labelling without major assumptions.

Comparing the self-reported spot reaction to the observed cues shows that the amount of reactions is greater in the observed group: 22 compared to 17. From the believers, all but one participant exhibited behavioural cues, in addition to 6 from the sceptics.

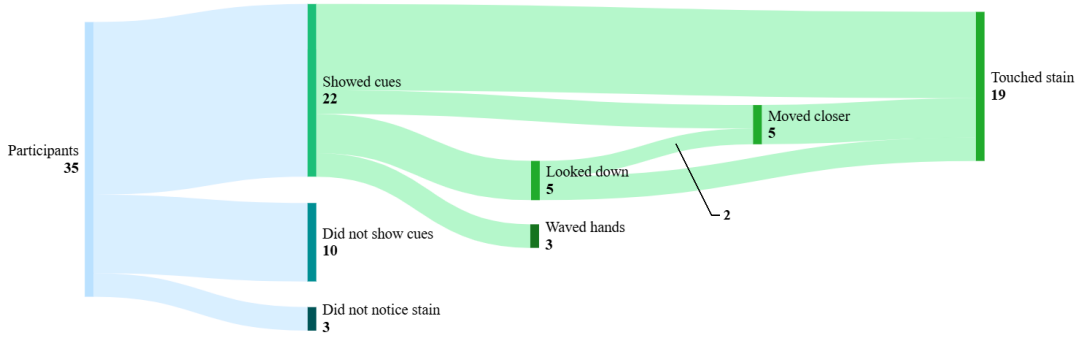


FIGURE 5.1: Overview of participants' behavioural cues towards the virtual stain. Not in order.

Examining which behavioural cues took precedence, combined with their corresponding reaction times, shows that from the participants that performed more than one cue (n=8), 6 tried to touch the stain first, while 2 first moved closer to the mirror. Looking down, although occurring five times, was never the first action. All participants who showed behavioural cues were divided in three groups based on their reaction time: quick (<10s, n=5), short (<1min, n=10) and long (>1min, n=7), see Appendix C. In all cases, the reaction time did not appear to vary significantly based on which behavioural cue occurred first. When multiple actions occurred, it always happened in quick succession (<5s).

No demographics seem to influence the believability of the virtual. Having a technical background, was presumed to negatively impact the believability of the stain. This correlation was tested using a Chi^2 test. The test found no statistically significant correlation ($p = 0.71$) between the two variables.

At the interview question, *"When did you notice the spot? What was your first reaction?"*, most participants from the sceptics reported their interpretation of the stain. Roughly half of this group assumed the stain was somehow related to the pose tracking, some thought it was a visual artifact, while others thought it was a reflection, or something on the lens of the camera. Four participants mentioned they 'expected something' AR related before starting the experiment, of which three self-reported the stain to be fake, and three exhibited the behavioural cues of interest. Two additional participants mentioned that were certain they had no stain on their shirt, therefore not reacting.

5.1.2 Digital mirror

The first question of the interview, *"What did you think of the mirror?"*, is an open question by design. People tend to start with what they found most important first, in this case most participants described the experience along the lines of 'fun'. The word fun, and related words such as engaging, cool, interesting, were the word 22 participants used, 26 out of 35 mentioned it in total. Many participants expressed an eagerness to trying out the system and testing its capabilities. Besides fun, participants reported the system felt mirror-like, with some small downsides, mostly notably the delay and camera angle (elaborated in the next section). Almost all participants mentioned the delay during the experiment and/or answered it at *"What made this not feel like a real mirror?"*, although many stated it did not bother them or got in the way of the experience, much. Three participants mentioned the 'wobble' could be nauseating/disorienting. This effect was worsened for smaller movements where participants did not expect the reflection to change, resulting in the effect feeling too strong. P18 even remarked that "it felt like I had to plant my feet firmly on the ground, because the movement didn't feel entirely intuitive".

The interaction with the digital mirror was quite simple, since it was intended to mimic a real one. Although the system was introduced to the participants as a 'digital mirror', many expected more features to be present, akin a 'smart mirror', presumably due to the experimental setting and technical University environment. Because of this, some participants tried touching the screen (n=3), looking for gesture controls (n≥4), or started talked to the system (n=2), assuming there would be more interaction possibilities. This caused a feeling of disappointed for some participants, with P35 stating the system is "playful, but limited".

Besides these points, resolution was also a popular topic. Most participants did not consider it immersion breaking, but a higher resolution was expected, especially since a real mirror provides an 'infinite' resolution.

5.1.3 Camera angle

Many participants noticed the angle of the camera, which a few participants further adding how it affected their experience. One participant stated it was strange they could not look themselves in the eyes, thus not wanting to use the mirror up close. Others mentioned that it noticeably affected their body proportions, with their legs noticeably longer, especially when closer to the mirror. One participant mentioned that these incorrect proportions could be detrimental to the mirror experience, as it "reduced self-confidence" (P16). Another participant stated they "had less recognition of their own body" (P21), compared to real mirrors. Some participants suggested moving the camera to eye height could solve these problems.

5.1.4 'Incorrect reflection'

A slight majority of participants (n=19) stated during the interview that the reflection was not accurate compared to a real mirror, mostly reporting that it felt like it 'followed them'. The general consensus in this group was that the mirror showed the opposite to what should be shown (horizontally). Many participants had difficulty explaining their rationale, and 4 participants even corrected themselves during their answer, as they noticed they were mistaken.

This phenomena is reflected in the scores from the questionnaire between these two groups, shown in Table 5.1 (the questionnaire was filled prior to the interview, therefore

the four participants who changed their mind are represented by the 'incorrect' group). While not statically significant, an observable difference between the two groups exists. The question *"The digital mirror showed me what a real mirror would show me."* expectedly shows a slightly lower score in the group that stated the reflection was incorrect. Interestingly, this effect is also observed to negatively impact the score on usability of the mirror, extending to the overall score of the experience.

	Reflection consensus		
	Incorrect (n=19)	**Correct** (n=16)	**All**
"I enjoyed using the DM"	4.9	5.4	5.1
"The experience with the DM was engaging"	5.2	5.3	5.2
"The DM felt intuitive to use"	4.9	5.2	5.0
"The DM showed what a real mirror would"	4.5	5.2	4.8

TABLE 5.1: Questionnaire scores for a subset of questions, using a 7 point Likert scale (1 = "Strongly disagree", 4 = "Neutral", 7 = "Strongly agree"). Digital mirror is abbreviated to DM for brevity.

5.2 User experience design

The following findings specifically discuss the user experience (and by extension user interface) of the digital mirror, discussing communication clarity and its easy to use.

5.2.1 Usability

As the system mimics an existing object that participants are accustomed to using, the experience is expected to be intuitive, though slightly reduced due to technical limitations, most notably latency. The questionnaire directly addressing this, *"The digital mirror felt intuitive to use."* scored an average rating of 5.0 on a 7-point Likert scale. This score is expected to correlate with *"The digital mirror showed me what a real mirror would show me"*, as a system is typically experienced as more intuitive if it aligns with users' expectations. Removing two outliers (P32 & P34), a Pearson correlation test ($r = 0.40$, $p = 0.02$) shows a statistically significant, moderately positive, correlation between the answers to the two questions.

The two outliers both contain a 4-point difference between the answers of the two questions (where the mean is 1.3, median of 1), each the other way around. These outliers being the only participants that had opposing answers to the two questions. These participants share no specific demographic or answer pattern, only both reporting in the interview that the reflection was not correct. These outliers are therefore most presumed a mistake or misinterpretation of (one of) the questions. Answers to other questions were relatively in line with similar respondents.

The usability of the system can also be evaluated by examining the answers to *"What do you mostly use a mirror for?"*, specifically its follow-up question *"Could you use this digital mirror for that use case?"*. Checking outfits (n=27), hair (n=21), and other personal hygiene related activities were reported as the primary uses of mirrors by all participants. These activities require a 'medium' distance from the mirror, and most participants indicated that the digital mirror would suffice for these activities. Some participants mentioning that it would outperform an actual mirror, as they thought it showed a larger image than

a normal mirror. Closer activities, such as shaving, brushing teeth and doing makeup were also mentioned as usual mirror activities. These take place at a much closer distance, as detail becomes more important. For these activities, most participants indicated the digital mirror would not be suited, mainly due to: the strange perspective caused by the camera angle, the delay, and the system not working at close distance (fade to grey).

5.2.2 Calibration process

On average, participants required 11.4 seconds to be guided to the correct calibration spot, following the instructions on screen. This showed a high variability: a standard deviation of 7.3 seconds, inflated by some large outliers on either side.

Some very low values contribute to this relatively high standard deviation: 6 measurement were in the 2-4 second range. In these cases, it was observed that the participant (almost) immediately stood in the correct spot, skipping any further instructions.

Higher measurements were observed to mostly be the result of a low threshold around the desired spot, ± 25 cm, which meant overshooting the spot happened often. Combined with the delay in calibration screen switching, caused by smoothing of the distance values, many participants expressed difficulty finding the exact spot to stand.

The interview did not contain questions directly addressing the calibration process, thus only few participants mentioned the it. P16 made an interesting comment, namely that they were hesitant to move around after the calibration, since they were instructed to stand still, after which no further instructions were given. Some more participants felt this way, this being observed and also mentioned in passing during the interaction with the mirror. P20, who expressed interest into the inner workings of the system, mentioned that the calibration process was "simple, but works well". They also praised its clean and minimalistic aesthetic, as well its intuitive design, specifically the small animations.

5.2.3 Usage quantified

Participants' mirror usage can by quantified by combining three variables: time spent in front of the mirror; the movement tracked by the system, where increased movement indicates curiosity; and engagement and observational data quantifying the use the mirror during tasks. These three variables can be normalized and combined to form a usage value.

While this value would be interesting to test against other variables, such as rated usability from the questionnaire, testing the three sub-variables against each other in a Spearman correlation test, reveals that they share little statistical correlation. The results of this test, found in Appendix B, indicate that combining all three of these would yield values nearly indistinguishable from random noise. Noteworthy is a moderate negative correlation (ρ = -0.38) between 'observed use for tasks' and 'movement over time'. Indicating that participants who used the mirror more during tasks generally moved less, as they were possibly more focused on the mirror image.

5.3 Technical feasibility

The following findings are directly related to the technical aspects of the digital mirror, used in assessing each parts of implementation. This will contain user test data in combination with extra performance tests of the system.

5.3.1 Virtual stain implementation

The opacity of the spot was adjusted based on participants' clothing brightness, using two presets: 'light' and 'dark'. To assess whether this implementation affected the believability of the stain, a Chi^2 test was conducted examining the relationship between clothing colour and participants' reactions. The test found no statistically significant correlation ($p = 0.31$), indicating the implementation caused no significant impact.

The occurrence of behavioural cues in Table 5.2, again showed no significant difference. Both shades of clothing contain similar reaction time distributions, with the light clothing overall showing less reactions, though shorter on average, specifically containing less long reaction.

	No reaction	**Quick**	**Short**	**Long**
Dark Clothing	6 (30%)	4 (20%)	6 (30%)	4 (20%)
Light Clothing	7 (47%)	3 (20%)	4 (27%)	1 (07%)

TABLE 5.2: Reaction time by clothing shade. Quick = <10s, short = <1min, long = >1min.

From the group with no reaction, three participants stated to never have noticed the stain: one was not wearing their glasses (self-reported), the other two of wore white clothing. While the stain was objectively harder to see on white clothing, participants also wore a darker shirt as part of the tasks, providing another observation opportunity. No other shared characteristic among these participants could be found.

The behavioural cues to the virtual stain all lasted only a few seconds. Most participants tried to confirm the stain's existence using the observed behavioural cues mentioned before. Some also moved their body checking whether the stain would remain in place. Despite all efforts to properly place the stain, most of these participants mentioned during the experiment and/or in the interview that this effect was quickly negated when moving or rotating, especially when quick movements were applied. Some participants expressed that the believability could be improved by placing the stain higher on their body, or changing it to f.e. a pimple on their head.

5.3.2 Latency

As mentioned before, most participants noticed, but were unbothered by, the delay between action and reflection. For proper analysis, a quantitively measurement was conducted. End-to-end system latency was measured using two methods, first as described in Section 3.3.2, which, averaging 10 measurements, resulted in a latency value of 0.50 seconds. For the second test, the screen was filmed with the Unity feed open, resulting in twice the system delay, making small fluctuation less prominent. Tested was repeated 10 times, averaging 1.02 seconds, validating the roughly 500 ms measurements from the first test. This total system latency includes the 133 ms of artificially added delay from running in coupled mode, which, interestingly, seemed an imperceptible difference to the 8 participants who tested the latency in Section 3.7.4.

Further dissection of the system latency was attempted, however, too many factors influenced the results. Without complicated analysis or specialised tools there was no proper method available to measure latency, especially between programs.

Despite the relatively high latency, the system ran smooth, providing a consistent frame rate of over 60 in the Unity editor. The latency only delayed the image, it did not bottleneck the frame rate.

5.3.3 User position

To quantify the accuracy of the system's estimated user position, a test was conducted. Figure 5.2 shows a top-down view of user position measurements. Markers (shown in green) were placed on the floor, with the user then standing on these markers. The markers were placed in a grid of 1 meter cells with respect to the mirror, within the camera's FOV. The mean measurement for each marker is highlighted in red, and a line is drawn to the corresponding marker.

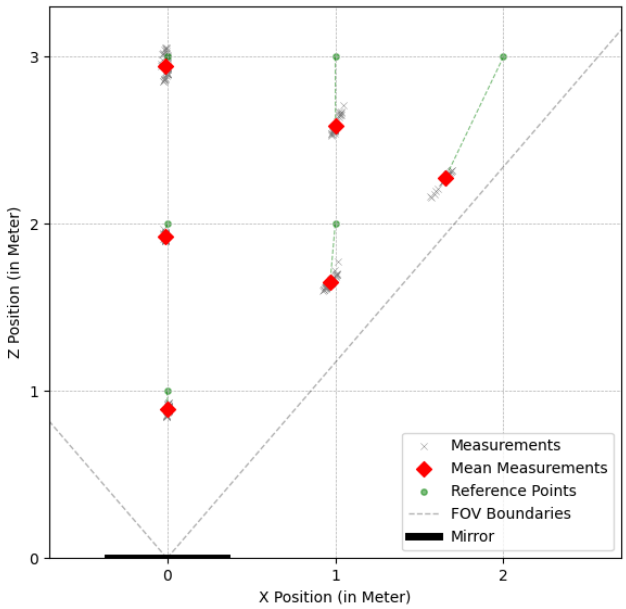


FIGURE 5.2: Estimated user position accuracy test (top-down). Measurement were taken standing still on markers highlighted in green. Results are similar for negative x-values, mirrored over the y-axis.

The results show that the markers straight in front of the mirror (x=0) provide values close to their reference point. All three points have a similar z offset of ∼8 cm, which is also presumed present at the other points, though not clearly visible. This offset can be explained. Standing at the calibration point marker (0,2), the distance sensor takes the reference distance measurement. Though the user is standing exactly in the middle of the marker, their body 'sticks out' on all sides, resulting in distance sensor offset (8 cm in this case). Since the eyes of the user stick out a similar amount from the middle of the body, this provides the correct position for the perspective calculations. This is thus a flaw in the test, not the implementation.

Figure 5.2 further shows that measurements are more spread out at longer distances. This spread occurred predominantly in a line to the camera, not side-to-side, indicating the distance scaling is the cause. The landmarks that are used for scaling are close together on the camera feed at longer distance, which results in small changes causing relatively bigger fluctuations in distance calculations.

Chapter 6

Discussion

This chapter discusses the results in relation to the literature presented in the background chapter, aiming to provide a deeper understanding of the findings. The structure is mostly similar to the previous chapter to facilitate easy cross-referencing.

6.1 Participant impression and behaviour

6.1.1 Virtual stain

The reaction to the stain was captured using two methods. In the first method, 17 out of 35 (49%) participants self-reported the stain as real: the 'believers'. The 'sceptics' (n=15) thought the stain was not real. To clarify: 'real' in this context means physically present on the person. Not all participants perceived the augmentation as a stain, especially participants from the sceptics group. If a person perceived it as a stain, they have already placed it in context of their environment and are thus presumed to believe it.

The second method for capturing reactions was observing behavioural cues. Here, 22 out of 35 (63%) participants reacted, as shown in Figure 5.1. This number is higher than the self-reported number, as all but one of the participants from the believers also exhibited behavioural cues, in addition to 6 out of the 15 participants from the sceptics. The behavioural cues of interest were all actions that directly aimed to confirm the existence of the stain. The high participation from the believers makes sense given this context.

The 6 participants from the sceptics group warrant more discussion, as they reported the stain was fake, yet showed behavioural cues towards it: 4 participants touched the stain, 2 others waved their arms in front of it. While it can be assumed participants usually answer truthfully during an experiment, it is possible that responses were influenced by self-presentation bias. The slight deception introduced by the stain may have caused psychological discomfort, or cognitive dissonance [6], leading participants to reduce this discomfort by reinterpreting the stain as fake.

The relatively high rate of behavioural reactions to the virtual stain is a sign of a strong self-association effect. As the stain is projected directly onto the participant's self-associated mirror image, it seemingly inherited that association (similar to [11]), causing automatic capturing of the participants' attention, and prioritising actions towards it.

The act of touching the stain may suggest participants perceived it as real, however, this cannot be stated with certainty. Since the system is digital, and participants became aware of being tracked during the calibration process, several participants assumed the stain was interactive. In such cases, touching the stain does not necessarily indicate that it was perceived as real. Another action, waving the arms in front of the stain, inherently implies

a critical approach towards the stain, as it tests whether or not the stain occludes when something is placed in front. For these participants, it is resumed they indeed perceived the stain as fake, assuming they self-reported that.

It is therefore expected that actual the amount of participants thinking the stain was real is closer to the self reported number than the observed number. However, although 6 participants from the sceptics stated the stain was fake, the act of trying to interact with it does signal they perceived it as 'on their person'. Self-relevance was thus low, but a sense of self-association was present. The stimulus, in this case the stain, seemingly triggered this association and was therefore incorporate into the 'self' [26].

Some participants either expressed or showed signs of suspicion towards the mirror. Mostly seeking its capabilities were beyond a standard mirror. This potentially influenced the believability of the stain, though not quantifiable enough to analyse properly.

In summary, getting 63% of participants to react to the stain shows a promising future for augmented reality mirrors. With a more subtle effect and refined implementation this number is expected to increase.

6.1.2 Reaction time

A noteworthy result is the variability seen in reaction times to the virtual stain. While some participants reacted immediately, which can potentially be explained by the SPE [8, 9] triggering, others took much longer or failed to notice the stain altogether (n=3). This could be caused by 'inattentional blindness', where stimuli can be overlooked when attention is focused elsewhere, as shown by Simons and Chabris [95] in their famous 'Gorillas in our midst' study. However, the stimulus in question was designed to trigger self-association, which directly opposes this blindness. In the context of this experiment, participants may have hyper-focused on their mirror image, technical limitations, or exploring the possibilities of the mirror, resulting in them overlooking the stain initially. Wood and Simons [96] found that people who fail to notice something immediately, may never notice it, possibly explaining the 3 participants in the unaware group.

6.1.3 Digital mirror

This research tested the prototype in isolation, without multiple conditions or control groups. It was assumed that a mirror is so commonly used that participants have clear expectations. This appeared somewhat of an oversight, as 19 out of 35 (54%) participants stated the mirror did not show the correct horizontal reflection, with many stating the reflection 'followed them'. While the system does follow (track) the user, the reflection always showed the other side of the room, similar to a real mirror.

This highlights that, in the absence of a physical reference mirror, many individuals either do not understand or have never critically reflected on how a mirror functions. Most are expected unable to answer why a mirror image is horizontally flipped, but not vertically. However, familiarity is so high with mirrors, that any alteration to its workings immediately signals something is 'off'.

While the reflections produced by the digital mirror were not perfectly identical to those of a real mirror, it is improbable this explains the participants' misinterpretations. The more probable cause is the (significant) latency introduced by the system, seemingly broke the natural puppeteering effect that a mirror provides [17]. Additionally, with a delay of approximately 500 ms, compared to the practically 0 ms response time of a real mirror, participants may have interpreted the delayed spatial feedback as incorrect spatial feedback [97]. In physical mirrors, the reflection moves with the user, making the movement

imperceptible, while with the digital mirror, the added delay emphasizes this movement, creating the 'off' feeling.

These two reasons suggest that, were the image indeed horizontally reversed and shown to the user, a similar sized group would again point out it is incorrect, for the same reasons.

6.1.4 Camera angle

A (single) static camera-based approach is unable to replicate the reflections of a real mirror perfectly. The placement of the camera is crucial, and while many participants expressed concerns regarding their skewed body proportions, placing the camera on top appears to have been the most appropriate choice. The test described in Section 3.3.2, showed a overwhelming preference for this approach. Alternative camera placement is thus expected to negatively affect the experience further.

Nonetheless, the result are affected by these incorrect body proportions, best described by P16: "it reduced self-confidence" and P21: "I had less recognition of my own body". This reduction of resemblance negatively impacted the self-association participants had with their mirror image. However, this effect should be limited, since the distortions were mostly affecting the legs of the participants, which take less priority in self-association compared to the head and torso area [20, 25, 26], which were largely unaffected.

The statements of these participants directly compare the system to a real mirror, where self-representation is perfect. Additionally, the fact that participants noticed the image distortions suggests they maintained a strong sense of self-association with the digital mirror, creating sensitivity to imperfections. This strong sense of self-association was expected in the context of a digital mirror, given the high resemblance and immersion [19], and the matching movement [17].

Future iterations could attempt to reduce these distortions, by f.e. warping the image. This will not solve the angled image problem, but should promote higher self-association.

6.2 User experience design

6.2.1 Calibration process

The calibration process generally worked well, with participants taking an average of 11.4 seconds to complete it, though with considerable variability (SD = 7.3s), due to some extreme times. Quick calibrations (<4 seconds) were typically due to participants directly stepping near the correct spot, while longer times were mostly caused by a strict ±25 cm threshold, combined with a slight screen-switching delay, causing some to overshoot, which is consistent with the literature [61, 62].

While not directly addressed in the interview, some participants expressed confusion about whether they were expected to remain in the calibrated position. P16, for example, hesitated to move due to the lack of further instructions, and the system previously instructing them to stand still. Others, such as P20, found the process simple and effective.

Overall, the calibration fulfilled its purpose an can be improved using a prompt clarifying participants they are free to move afterwards. In the long term, reducing latency could lower some of the higher calibration times, in the short term, increasing the calibration threshold could achieve a similar effect.

6.2.2 Usage quantified

Usage was expected to be quantifiable, and while individual usage related variables could be formed, such as total distance travelled, these variables did not correlate. This resulted in no correlation between engagement scores and mirror usage, opposing the findings of Bianchi-Berthouze [98], who found that more movement caused higher scores of engagement. This discrepancy in results can possibly be attributed to the high latency of the system.

6.2.3 Interactions

While interactions were outside the scope of this research, the user tests showed that interactive elements are often expected of digital systems. In other, future, research, added interaction possibilities could prove beneficial. For example, most participants stated that projection outfits on themselves would be a practical use case for this system. Gesture recognition [72], touchscreens [71], hover-based inputs [76, 77], and speech commands or proximity-based inputs [78] are useful additions for controlling such an application.

6.3 Technical feasibility

6.3.1 Virtual stain implementation

Comparing the visibility of the virtual stain between lighter and darker clothing, showed it was objectively more difficult to notice on lighter colours. However, the results do not reflect this, as brightness of the clothing the stain was projected on, did not significantly influence the reactions or reaction times. The workaround solution of changing the opacity of the stain manually can thus be considered adequate.

An interesting alternative application of the stain for future research could be to start at 0% opacity, slowly increasing over time. Reaction time might be less valuable in this scenario as different people and different colour clothing would yield inconsistent results. However, it might trick more participants into thinking it is real, as the contrast between the stain and image will start much less obvious.

The stain did not occlude when moving something in front of it, such as hands. For most participants who thought the stain was real at first, this was how they confirmed it was not. Proper occlusion implementation could prevent this, though it was out of the scope of this project. The sceptics mostly looked at the stain's reaction to movement and rotation to confirm its existence. The stain did not stick perfectly to the user, the effect worsened with quicker movements. Running the video in coupled mode helped tremendously, however, if the effect is expected to stay believable for longer, this implementation needs improvement. Using machine vision to achieve this is presumably the best option.

For the context of this research, the implementation was found sufficient. A decent portion of participants believed the stain, providing interesting results. Placing the spot higher, closer to the face [20, 25, 26], has potential to improve the results, some participants mentioned this. Additionally, the stain was found slightly too obvious visually, thus a more subtle stain might yield better results. This can be achieved by reducing its size and opacity, and exploring different shapes.

6.3.2 Latency

The total system latency reached around 500 ms, far exceeding the preferred 100 ms set up in the requirements as proposed by [63, 64]. While a lower latency is certainly achievable with optimisations, displaying a camera feed in Windows takes up 90 ms minimum using the quickest camera tested (see Figure 3.3), the 100 ms figure was thus never achievable. Theoretically it is possible to get lower with specialised hardware, for example, a video feed streamed from an FPV drone to its goggles can achieve a latency as low as 30 ms. Even if this reduction was applied, it is doubtful the total system latency could stay below the 100 ms.

The output frame rate was a consistent 60 fps, far exceeding the 25 fps minimum required for a seamless experience [66]. The fps was also stable, with no apparent frame time spikes, which was also preferred [65]. The frame rate can thus be considered a success.

6.3.3 User position

The results in Figure 3.6, show estimating the user position was reasonably accurate, greatly worsening at greater horizontal angles. This was not implemented properly, as accurate horizontal angles were considered not important. Looking into a mirror from an angle is uncommon, and people are unlikely to notice discrepancies, especially minor ones. None of the participants mentioned the position, or horizontal angles being inaccurate. If accurate angles are important, this can cause problems, in other cases this is thus a non-issue.

Using head-based scaling for distance measurement provides accuracy within a few percent, when the user is positioned directly in front of the camera. The landmarks values that comprise the face are relatively close together, at distance, the differences between them results in rapid changes in the scaling. Some participants found this 'wobble' to be distracting or even nauseating. This can be solved, either using more advanced smoothing algorithms, or using another way of scaling, f.e. continuously using the distance sensor to cross-check the scaling.

Chapter 7

Conclusion

This research set out to explore the behavioural and perceptual responses of participants to a digital mirror system, specifically focusing on the believability and behavioural response of virtual augmentations, and the practical feasibility of a mirror with physically accurate reflections. The findings, while promising in certain aspects, also highlight several critical limitations.

The virtual stain successfully provoked (noticeable) reactions, with 63% of participants displaying behavioural cues, suggesting that augmentations placed on the self-associated mirror image capture attention well. Assuming most participants self-reported their belief truthfully, many of the behavioural cues may have stemmed not from actual belief, but from curiosity or expectations about interactivity. Still, even sceptical participants exhibited self-relevancy towards the stain, suggesting the stain's successfully integrated into the participants' self-representation.

The system's latency was evidently the biggest limiting factor. At 500 ms, it far exceeded the requirements, and distanced itself from an actual mirror experience. This delay was presumably the main contributor to participants' sense of unease, and the misinterpretation of the mirror's reflections. Additionally, camera placement introduced further image distortions, which some participants explicitly mentioned as reducing body recognition and self-confidence.

This experiment demonstrated that an augmented reality digital mirror can engage the user, evoke meaningful behaviours, and trigger self-perception processes, even with the aforementioned, severe, technical limitations. This suggests that the concept holds value for self-perception research, and has improvement potential with further iterations, particularly regarding latency, image distortion, and augmentation refinement.

Summery per research question

mRQ1: To what extent can an augmented reality mirror influence individuals' self-perception and body image?

The digital mirror prototype successfully triggered self-perception related behaviours in the majority of participants. Even sceptical participants engaged with it, indicating some degree of self-association in most participants. However, its influence was limited, as many participants quickly detected the illusion due to technical shortcomings. Better implementation, less provocative augmentations could improve this effect on self-perception, and by extension body image in a powerful way.

mRQ2: How well can a digital mirror reproduce physically accurate reflections and meet user expectations?

From a technical standpoint, the reflections were reasonably accurate given the inherent physical limitations of a static camera-based approach. However, the system struggled to convey this to the participants, mostly due to the high latency and image distortions. This resulted in over half of the participants (54%) questioning the basic reflection behaviour of the mirror.

sRQ1: What types of self-perception research are enabled by a digital mirror?

This system, shows promise for research into attention, self-prioritization effects, self-association, and potentially body image manipulation. However, meaningful research involving subtle self-perception processes will require a more refined system, with lower latency and higher visual fidelity.

sRQ2: How do individuals perceive the digital mirror's realism and behaviour compared to a real mirror?

The latency, skewed image proportions, and the resolution made the system noticeably digital. Some misinterpreted these limitation as incorrect mirror behaviour, showing that participants' understanding of real mirrors isn't always accurate, but they notice when something is off. Realism was generally seen as acceptable, as most participant stated they could use this system in place of a real mirror.

sRQ3: What design guidelines can be set up for creating digital mirrors?

Due to the novelty of the mirror, many pitfalls were encountered and described in the design process. Below is an overview of all recommendations for future researchers aiming to (re)create a physically accurate augmented reality digital mirror system.

- **Minimize latency** — The literature found, and this research confirmed, that a low latency is critical for a high-immersion system. It should be lowered systemically and computational processes optimised.
- **Camera placement** — If similar limitations are in place, a top-mounted camera is recommended. Distortions should be minimised if this approach is used. Ideally, multiple cameras would be used, fusing the images.
- **Get physical context** — The system requires real-time physical properties to produce accurate and consistent results. The ToF sensor and calibration process for distance both worked great for this system.
- **Perspective calculations** — A rules-based approach should be used to emulate the behaviour of a mirror. Inherent physical limitations prevent exact simulation, which is why the 'logical' ('faithful') approach will not work.
- **Augmentations** — When using a self-developed solution, prioritize subtlety in augmentations, since technical imperfections will be quickly noticed and undermine the illusion. A machine vision approach is recommended for better tracking quality.

Bibliography

[1] Christopher J. Wilson and Alessandro Soranzo. The use of virtual reality in psychology: A case study in visual perception. *Computational and Mathematical Methods in Medicine*, 2015:1–7, 2015.

[2] Daryl J. Bem. Self-perception theory11development of self-perception theory was supported primarily by a grant from the national science foundation (gs 1452) awarded to the author during his tenure at carnegie-mellon university. volume 6 of *Advances in Experimental Social Psychology*, pages 1–62. Academic Press, 1972.

[3] Dunser Andreas, Grasset Raphael, and Farrant Hamish. *Towards Immersive and Adaptive Augmented Reality Exposure Treatment*. 2011.

[4] Bogdan Wojciszke. Morality and competence in person- and self-perception. *European Review of Social Psychology*, 16:155–188, 1 2005.

[5] Daryl J. Bem. Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, 74:183–200, 1967.

[6] Elliot Aronson. *The Theory of Cognitive Dissonance: A Current Perspective*, pages 1–34. 1969.

[7] Mark P. Zanna, James M. Olson, and Russell H. Fazio. Self-perception and attitude-behavior consistency. *Personality and Social Psychology Bulletin*, 7:252–256, 6 1981.

[8] John A. Bargh. Attention and automaticity in the processing of self-relevant information. *Journal of Personality and Social Psychology*, 43:425–436, 9 1982.

[9] Theodore Alexopoulos, Dominique Muller, François Ric, and Christian Marendaz. I, me, mine: Automatic attentional capture by self-related stimuli. *European Journal of Social Psychology*, 42:770–779, 10 2012.

[10] C. Whan Park and V. Parker Lessig. Familiarity and its impact on consumer decision biases and heuristics. *Journal of Consumer Research*, 8:223, 9 1981.

[11] Gabriela Orellana-Corrales, Christina Matschke, Sarah Schäfer, and Ann Katrin Wesslein. Does an experimentally induced self-association elicit affective self-prioritisation? *Quarterly Journal of Experimental Psychology*, 6 2022.

[12] Matthew Botvinick and Jonathan Cohen. Rubber hands 'feel' touch that eyes see. *Nature*, 391:756–756, 2 1998.

[13] Mel Slater. Towards a digital body: The virtual arm illusion. *Frontiers in Human Neuroscience*, 2, 2008.

[14] Mel Slater, Bernhard Spanlang, Maria V. Sanchez-Vives, and Olaf Blanke. First person experience of body transfer in virtual reality. *PLoS ONE*, 5:e10564, 5 2010.

[15] Angelo Maravita and Atsushi Iriki. Tools for the body (schema). *Trends in Cognitive Sciences*, 8:79–86, 2 2004.

[16] Sarah Schäfer, Ann Katrin Wesslein, Charles Spence, Dirk Wentura, and Christian Frings. Self-prioritization in vision, audition, and touch. *Experimental Brain Research*, 234:2141–2150, 8 2016.

[17] Emma Westecott. The player character as performing object. 2009.

[18] Marnix S. van Gisbergen, Ilay Sensagir, and Joey Relouw. *How Real Do You See Yourself in VR? The Effect of User-Avatar Resemblance on Virtual Reality Experiences and Behaviour*, pages 401–409. 2020.

[19] Maximilian A. Friehs, Sarah Schäfer, and Christian Frings. The (gami)fictional egocenter: Projecting the location of the self into an avatar. *Frontiers in Psychology*, 13, 7 2022.

[20] Maximilian A. Friehs, Martin Dechant, Sarah Schäfer, and Regan L. Mandryk. More than skin deep: about the influence of self-relevant avatars on inhibitory control. *Cognitive Research: Principles and Implications*, 7, 12 2022.

[21] Selen Turkay and Charles K. Kinzer. *The Effects of Avatar-Based Customization on Player Identification*, pages 247–272. IGI Global, 2015.

[22] Eleonora H. Customization, emotional bonds and identification with the player character: A study into the effects of text-based gameplay. 2016.

[23] Hyunjin Kang and Hye Kyung Kim. My avatar and the affirmed self: Psychological and persuasive implications of avatar customization. *Computers in Human Behavior*, 112:106446, 11 2020.

[24] Rabindra A. Ratan and Michael Dawson. When mii is me. *Communication Research*, 43:1065–1093, 12 2016.

[25] Jakub Limanowski and Heiko Hecht. Where do we stand on locating the self? *Psychology*, 02:312–317, 2011.

[26] S. Schäfer, Dirk Wentura, Marcel Pauly, and Christian Frings. The natural egocenter: An experimental account of locating the self. *Consciousness and Cognition*, 74, 9 2019.

[27] Mark Billinghurst, Adrian Clark, and Gun Lee. A survey of augmented reality. *Foundations and Trends® in Human–Computer Interaction*, 8:73–272, 2015.

[28] Jun Rekimoto and Katashi Nagao. The world through the computer. pages 29–36. ACM, 12 1995.

[29] Philipp A. Rauschnabel, Reto Felix, Chris Hinsch, Hamza Shahab, and Florian Alt. What is xr? towards a framework for augmented and virtual reality. *Computers in Human Behavior*, 133:107289, 8 2022.

[30] J.M. Zheng, K.W. Chan, and I. Gibson. Virtual reality. *IEEE Potentials*, 17:20–23, 1998.

[31] T.S. Mujber, T. Szecsi, and M.S.J. Hashmi. Virtual reality applications in manufacturing process simulation. *Journal of Materials Processing Technology*, 155-156:1834–1838, 11 2004.

[32] John Estrada, Sidike Paheding, Xiaoli Yang, and Quamar Niyaz. Deep-learning-incorporated augmented reality application for engineering lab training. *Applied Sciences*, 12:5159, 5 2022.

[33] Mika Yasuoka, Marko Zivko, Hiroshi Ishiguro, Yuichiro Yoshikawa, and Kazuki Sakai. *Effects of Digital Avatar on Perceived Social Presence and Co-presence in Business Meetings Between the Managers and Their Co-workers*, pages 83–97. 2022.

[34] Carolina Cruz-Neira, Daniel J. Sandin, and Thomas A. DeFanti. Surround-screen projection-based virtual reality. pages 135–142. ACM, 9 1993.

[35] Lizhou Cao, Chao Peng, and Jeffrey T. Hansberger. Usability and engagement study for a serious virtual reality game of lunar exploration missions. *Informatics*, 6:44, 10 2019.

[36] Devon Allcoat and Adrian von Mühlenen. Learning in virtual reality: Effects on performance, emotion and engagement. *Research in Learning Technology*, 26, 11 2018.

[37] Charles Xueyang Lin, Chaiwoo Lee, Dennis Lally, and Joseph F. Coughlin. Impact of virtual reality (vr) experience on older adults' well-being. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10927 LNCS:89–100, 2018.

[38] Domna Banakou, Sameer Kishore, and Mel Slater. Virtually being einstein results in an improvement in cognitive task performance and a decrease in age bias. *Frontiers in Psychology*, 9, 6 2018.

[39] Karin A. Buetler, Joaquin Penalver-Andres, Özhan Özen, Luca Ferriroli, René M. Müri, Dario Cazzoli, and Laura Marchal-Crespo. "tricking the brain" using immersive virtual reality: Modifying the self-perception over embodied avatar influences motor cortical excitability and action initiation. *Frontiers in Human Neuroscience*, 15, 2 2022.

[40] A Garcia-Palacios, H Hoffman, A Carlin, T.A Furness, and C Botella. Virtual reality in the treatment of spider phobia: a controlled study. *Behaviour Research and Therapy*, 40:983–993, 9 2002.

[41] Jimmy Bush. Viability of virtual reality exposure therapy as a treatment alternative. *Computers in Human Behavior*, 24:1032–1040, 5 2008.

[42] Mark B. Powers and Paul M.G. Emmelkamp. Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *Journal of Anxiety Disorders*, 22:561–569, 4 2008.

[43] Ronald T. Azuma. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, 6:355–385, 8 1997.

[44] Silhouettes from real objects enable realistic interactions with a virtual human in mobile augmented reality. *Applied Sciences*, 11:2763, 3 2021.

[45] Helena Roeber, John Bacus, and Carlo Tomasi. Typing in thin air. page 712. ACM Press, 2003.

[46] Jiann-Der Lee, Hae-Kuang Wu, and Chieh-Tsai Wu. A projection-based ar system to display brain angiography via stereo vision. pages 130–131. IEEE, 10 2018.

[47] Gun A. Lee, Andreas Dunser, Seungwon Kim, and Mark Billinghurst. Cityviewar: A mobile outdoor ar application for city visualization. pages 57–64. IEEE, 11 2012.

[48] Oliver Kutter, André Aichert, Christoph Bichlmeier, Jörg Traub, Sandro Michael Heining, Ben Ockert, Ekkehard Euler, and Nassir Navab. Real-time volume rendering for high quality visualization in augmented reality. *International Workshop on Augmented environments for Medical Imaging including Augmented Reality in Computer-aided Surgery (AMI-ARCS 2008)*, 2008.

[49] Hanhoon Park and Jong-Il Park. Invisible marker–based augmented reality. *International Journal of Human-Computer Interaction*, 26:829–848, 8 2010.

[50] Cristina Botella, Juani Bretón-López, Soledad Quero, Rosa Baños, and Azucena García-Palacios. Treating cockroach phobia with augmented reality. *Behavior Therapy*, 41:401–413, 9 2010.

[51] Carlos Suso-Ribera, Javier Fernández-Álvarez, Azucena García-Palacios, Hunter G. Hoffman, Juani Bretón-López, Rosa M. Baños, Soledad Quero, and Cristina Botella. Virtual reality, augmented reality, and in vivo exposure therapy: A preliminary comparison of treatment efficacy in small animal phobia. *Cyberpsychology, Behavior, and Social Networking*, 22:31–38, 1 2019.

[52] Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, and Andrew Fitzgibbon. Kinectfusion: Real-time 3d reconstruction and interaction using a moving depth camera*. pages 559–568. ACM, 10 2011.

[53] Clara Isakowitsch. How augmented reality beauty filters can affect self-perception. *Communications in Computer and Information Science*, 1662 CCIS:239–250, 2023.

[54] Rebecca Fribourg, Etienne Peillard, and Rachel McDonnell. Mirror, mirror on my phone: Investigating dimensions of self-face perception induced by augmented reality filters. pages 470–478. IEEE, 10 2021.

[55] Marie Luisa Fiedler, Mario Botsch, Carolin Wienrich, and Marc Erich Latoschik. Self-similarity beats motor control in augmented reality body weight perception. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–11, 2025.

[56] Harish Kumar and Madhushree Nanda Agarwal. Filtering the reality: Exploring the dark and bright sides of augmented reality–based filters on social media. *Australian Journal of Management*, 50:152–172, 2 2025.

[57] Susruthi Rajanala, Mayra B. C. Maymone, and Neelam A. Vashi. Selfies—living in the era of filtered photographs. *JAMA Facial Plastic Surgery*, 20:443–444, 11 2018.

[58] Oliver Baus and StÃ©phane Bouchard. Moving from virtual reality exposure-based therapy to augmented reality exposure-based therapy: A review. *Frontiers in Human Neuroscience*, 8, 3 2014.

[59] L. Barbieri, F. Bruno, F. Cosco, and M. Muzzupappa. Effects of device obtrusion and tool-hand misalignment on user performance and stiffness perception in visuo-haptic mixed reality. *International Journal of Human-Computer Studies*, 72:846–859, 12 2014.

[60] Gayani Samaraweera, Alex Perdomo, and John Quarles. Applying latency to half of a self-avatar's body to change real walking patterns. pages 89–96. IEEE, 3 2015.

[61] Daniel Rakita, Bilge Mutlu, and Michael Gleicher. Effects of onset latency and robot speed delays on mimicry-control teleoperation. pages 519–527. ACM, 3 2020.

[62] T.B. Sheridan and W.R. Ferrell. Remote manipulative control with transmission delay. *IEEE Transactions on Human Factors in Electronics*, HFE-4:25–29, 9 1963.

[63] Robert B. Miller. Response time in man-computer conversational transactions. page 267. ACM Press, 1968.

[64] Mark Claypool and Kajal Claypool. Latency and player actions in online games. *Communications of the ACM*, 49:40–45, 11 2006.

[65] Roland Thomaschke and Carola Haering. Predictivity of system delays shortens human response time. *International Journal of Human-Computer Studies*, 72:358–365, 3 2014.

[66] Magdalena Kanabus, Elzbieta Szelag, Ewa Rojek, and Ernst Pöppel. Temporal order judgement for auditory and visual stimuli. *Acta Neurobiologiae Experimentalis*, 62:263–270, 12 2002.

[67] A multi-layered display with water drops. *ACM Trans. Graph*, 29:76, 2010.

[68] I'm in the game: Embodied puppet interface improves avatar control. *Proceedings of the 5th International Conference on Tangible Embedded and Embodied Interaction, TEI'11*, pages 129–136, 2011.

[69] Sean Follmer, Daniel Leithinger, Alex Olwal, Akimitsu Hogge, and Hiroshi Ishii. inform: Dynamic physical affordances and constraints through shape and object actuation. 2013.

[70] Miri Kim and Kim Cheeyong. Augmented reality fashion apparel simulation using a magic mirror. *International Journal of Smart Home*, 9:169–178, 2015.

[71] Ana Javornik, Yvonne Rogers, Delia Gander, and Ana Moutinho. Magicface: Stepping into character through an augmented reality mirror. pages 4838–4849. ACM, 5 2017.

[72] Tobias Blum, Valerie Kleeberger, Christoph Bichlmeier, and Nassir Navab. mirracle: An augmented reality magic mirror system for anatomy education. pages 115–116. IEEE, 3 2012.

[73] Marc Erich Latoschik, Jean Luc Lugriny, and Daniel Rothz. Fakemi: A fake mirror system for avatar embodiment studies. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST*, 02-04-November-2016:73–76, 11 2016.

[74] Chontira Nimcharoen, Stefanie Zollmann, Jonny Collins, and Holger Regenbrecht. Is that me? - embodiment and body perception with an augmented reality mirror. *Adjunct Proceedings - 2018 IEEE International Symposium on Mixed and Augmented Reality, ISMAR-Adjunct 2018*, pages 158–163, 7 2018.

[75] L Bharath, S Shashank, V S Nageli, Sangeeta Shrivastava, and S Rakshit. Tracking method for human computer interaction using wii remote. pages 133–137. IEEE, 12 2010.

[76] Fraser Anderson, Tovi Grossman, Justin Matejka, and George Fitzmaurice. Youmove: enhancing movement training with an augmented reality mirror. pages 311–320. ACM, 10 2013.

[77] Gun A. Lee, Hye Sun Park, and Mark Billinghurst. Optical-reflection type 3d augmented reality mirrors. pages 1–2. ACM, 11 2019.

[78] Perttu Hämäläinen. Interactive video mirrors for sports training. pages 199–202. ACM, 10 2004.

[79] Theodore H. Mita, Marshall Dermer, and Jeffrey Knight. Reversed facial images and the mere-exposure hypothesis. *Journal of Personality and Social Psychology*, 35:597–601, 8 1977.

[80] Robert B. Zajonc. Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, 9:1–27, 1968.

[81] N. J. Block. Why do mirrors reverse right/left but not up/down. *Journal of Philosophy*, 71:259–277, 1974.

[82] K.A. Vogel. The design space for interactive pose tracking as an in-action training tool for climbing, January 2025.

[83] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43:172–186, 1 2021.

[84] Nadine B. Sarter. Multimodal information presentation: Design guidance and research challenges. *International Journal of Industrial Ergonomics*, 36:439–445, 5 2006.

[85] Christopher D. Wickens, Justin G. Hollands, Simon Banbury, and Raja Parasuraman. *Engineering Psychology and Human Performance*. Routledge, 2015.

[86] Yvonne Rogers, Helen Sharp, and Jenny Preece. *Interaction Design: Beyond Human-Computer Interaction*. John Wiley & Sons, 2011.

[87] Ángela Di Serio, María Blanca Ibáñez, and Carlos Delgado Kloos. Impact of an augmented reality system on students' motivation for a visual art course. *Computers Education*, 68:586–596, 10 2013.

[88] Sindre Rolstad, John Adler, and Anna Rydén. Response burden and questionnaire length: Is shorter better? a review and meta-analysis. *Value in Health*, 14:1101–1108, 12 2011.

[89] M. Galesic and M. Bosnjak. Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*, 73:349–360, 6 2009.

[90] John Brooke. *SUS – a quick and dirty usability scale*, pages 189–194. 01 1996.

[91] Kraig Finstad. The usability metric for user experience. *Interacting with Computers*, 22:323–327, 9 2010.

[92] Bettina Laugwitz, Theo Held, and Martin Schrepp. *Construction and Evaluation of a User Experience Questionnaire*, pages 63–76. 2008.

[93] Monique Hennink and Bonnie N. Kaiser. Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social Science Medicine*, 292:114523, 1 2022.

[94] Antony Bryant and Kathy Charmaz. Grounded theory in historical perspective: An epistemological account. *The SAGE handbook of grounded theory*, pages 31–57, 2007.

[95] Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, 28:1059–1074, 9 1999.

[96] Katherine Wood and Daniel J. Simons. Now or never: noticing occurs early in sustained inattentional blindness. *Royal Society Open Science*, 6:191333, 11 2019.

[97] Monica Gori, Maria Bianca Amadeo, and Claudio Campus. Temporal cues trick the visual and auditory cortices mimicking spatial cues in blind individuals. *Human Brain Mapping*, 41:2077–2091, 6 2020.

[98] Nadia Bianchi-Berthouze. Understanding the role of body movement in player engagement. *Human–Computer Interaction*, 28:40–75, 1 2013.

Appendix A

User tests

A Questionnaire

First, the age and gender of the person were asked. Then the following questions used a 7-point Likert scale, from 'strongly disagree' to 'strongly agree'.

1. "I enjoyed using the digital mirror."
2. "The digital mirror felt intuitive to use."
3. "The experience with the digital mirror was engaging."
4. "The digital mirror produced high quality images."
5. "I could perform the given tasks well."
6. "I could use the digital mirror to aid in the given tasks."
7. "The digital mirror showed me what a real mirror would show me."
8. "The digital mirror felt like a real mirror."

The final question used a 9-point scale for more precision. Underneath it is an extra description, since the wording here is important and the results are only valid if everyone's understanding is the same.

- "How would you describe/rate the experience of using the digital mirror, between a static camera and a real mirror?"

  A 'static camera' example would be your own webcam feed in a video call, or taking a selfie on a phone.

B Interview

Below are the questions of the semi-structured interview. Possible further indents beneath a question indicates a follow-up question, if the participant had not satisfied this already. Sometimes follow-up questions were asked that are not on this list.

1. "What did you think of the mirror?"
2. "When did you notice the spot? What was your first reaction?"
    - "Why did you (not) look down? Or used your hands?"
    - "Did you (shortly) believe the spot was real?"
3. "Can you think of another use case for augmentations in mirrors?"
4. "What do you mostly use a mirror for?"
    - "Could you use this digital mirror for that use case?"
5. "What made this not feel like a real mirror?"
    - "What did you think of it from a technical point-of-view?"
6. "Do you have any other remarks?"

Appendix B

Statistics

A Questionnaire responses

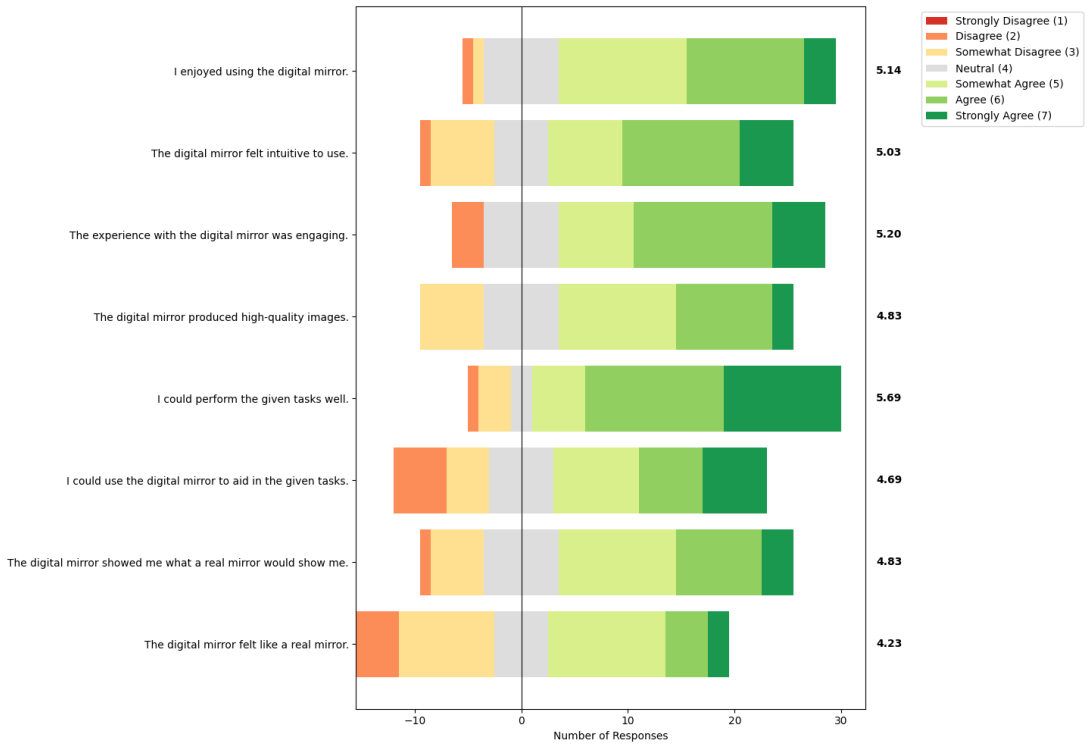


FIGURE B.1: Likert scale responses to questionnaire.

B Usage variables correlation test

	Observed use for tasks	**Total usage time**	**Movement over time**
Observed use for tasks	1.00000	0.00014	-0.38305
Total usage time	0.00014	1.00000	-0.15938
Movement over time	-0.38305	-0.15938	1.00000

TABLE B.1: Spearman correlation coefficients between the three usage variables.

C Behavioural cues reaction time

Action	<10s	<1min	>1min
Touch	4	9	4
Inspect	1	0	1
Wave	0	1	2

TABLE B.2: Distribution of reaction times of initial behavioural cues.

Appendix C

AI disclaimer

Portions of this thesis were supported by the use of AI-based tools, to improve the research quality, increase the scope of the project and decrease workload. Any AI generate content was always critically reviewed first by the researcher, verifying information and removing bias.

A Tools and uses

The following AI have been utilised, with their use cases explained.

- **ChatGPT** has been used extensively for the technical side of this thesis, such as programming and brainstorming technical solutions, as well as Latex formatting. It was sparingly used for writing, only sometimes aiding in the restructuring and condensing of sentences.
- **Google MediaPipe** uses the BlazePose ML model to generate its pose estimation.
- **Google Gemini** local live transcribe was used during the interview to save time later on. The generated files proved too inaccurate, therefore they were not used.