Department of Music

Norwegian University of Science and Technology

Department of Musicology

University of Oslo

Author
Simon Rønsholm Sandvik

Supervisor
Ph.D. Andreas Bergsland
Department of Music, NTNU

Co-Supervisor
Ph.D. Anders-Petter Andersson
Department of Design, NTNU

External Partner
Rockheim
National Museum of Popular Music
Trondheim, Norway

Gesture-based interaction for Interactive Museum Exhibits

Master's Thesis

Master's programme in Music, Communication and Technology

Trondheim
May 2021

ABSTRACT

This thesis looks at research literature and identified principles useful for gesture based interaction Museum Exhibit design. It also the reports on the design of a prototype exhibit. It showcases the use of dynamic gestures and spatial positioning as a means of interactively controlling the exhibit, made possible using the Azure Kinect within the Unity game engine.

Keywords:
Motion control, Gesture recognition, Interactive exhibit, Mixed Reality, User Experience, Museum, Azure Kinect, Unity

Accompanying Blog

A short summary of this thesis and a demo can be found by visiting the MCT blog with the following link:

https://mct-master.github.io/master-thesis/2021/06/20/simonrs-gestures.html

Acknowledgements

I extend my sincerest thanks to Arve Guldbrantsen who has facilitated and aiding testing, procured hardware for me to use. I would also like to thank my supervisors Andreas Bergsland and Anders-Petter Andersson who have devoted their time to help me through this thesis. I also extend my heartfelt thanks to my fiancée who have endured me during this stressing period.

Table of Contents

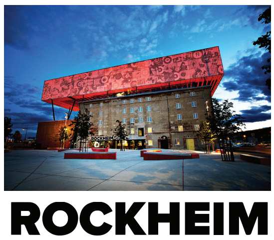
INTRODUCTION

1

1.1 Background

1.1.1 Deciding upon thesis (Introduce topic and motivation)

1. FIGURE:
A view of Rockheim



Rockheim, the national museum of popular music in Norway, is a museum with interactive exhibits of Norwegian popular music from the fifties to the present day. The museum has also been my employer for the last two years, where I have worked as a guide. Upon deciding my thesis, I had a sit-down with *Arve Gulbrandtsen*, head of IT, and where we decided he would be a liaison with Rockheim as an external partner for my thesis.

The focus would be to investigate how movement can be used in exhibits, what to account for with the use of gesture-based interactivity and explore the technical possibilities and limits of this. Exploring this topic, the thesis identifies theory and key-issues on this topic and documents the design of a prototype as a case study. The prototype is not intended for deployment, but as a preliminary study and framework for a potential full-scale exhibit. The case envision replacing the current flagship exhibit currently installed at Rockheim, "The Honorary Wall", and build content, location and intended function upon this exhibits.

The Honorary Wall

One of the flagship exhibits is under consideration of receiving an upgrade or replacement. The exhibit is known as "The Honorary Wall" and features a selection of some of the most prominent Norwegian artists. Split into six sections, the canvas display portraits of different artists in rotation, one for each decade. Visitors can step onto marked areas on the ground and use movement to break the artist's image currently on display, thereafter a music video of the artist will play. This exhibit has been present at Rockheim since it was opened in 2010 and is the first thing visitors see when entering the main exhibition floor. The exhibit was made in TouchDesigner by an external Trondheim-based company, Ablemagic. A camera is present above each marked area. The cameras imagery is superimposed upon the artists image, but pixels within a threshold of change is visible in a monochrome green to display the users movement. The image is split into pieces, likely using Delaunay triangulation. Changes in pixels within the triangles make them fall and gives the appearance of shattering glass.

TouchDesigner
is a node-based brogramming environment used by visual artists.

Delaunay triangulation
is an algorithm for splitting an area into triangles.

While the exhibit has been a success, judged upon visitor feedback, it is also flawed in many ways. Since its creation, we have entered two new decades, which are not present in the exhibit. Due to little or no visual room to spare on screen, simply adding the decades is not an option. The exhibit is made in what can now be considered an outdated version of TouchDesigner. This makes it challenging to update content, particularly for museum curators and technicians in-house. Any museum must renew its contents and exhibits to remain relevant and encourage repeat visitors. While creating a fully-fledged exhibit is beyond the scope of this thesis, I decided to provide a prototype to showcase some of the newer technologies available and identify its challenges and possibilities with a focus on spatial movement and gestural control.

1.1.2 Computer Vision and Gesture Recognition

Technological advancements have significantly increased communication bandwidth between human computer interaction. Specifically sensor technology and AI have opened several new viable modalities, such as real time gesture analysis and voice recognition. Computer vision has made considerable advancements over the last decade as AR, VR and MR become standard in various fields and technology. An increasing number of players in the game stems from the mobile and application industry. Augmented reality has become a central feature in daily communication through applications such as *Facebook* and *Snapchat*. According to Snapchat, 180m engages daily with AR, no less than three out of four of their 238m daily users in 2020[1]. Beside facial recognition having had a large focus, there is now much attention towards a spatial perception in computer vision to place virtual objects in a captured scene accurately. To effectively do this, there is a need for depth perception. *Microsoft Hololens* relies heavily on this to let users pin applications and virtual tasks to physical walls and furniture.

AI
Artificial Intelligence

Modality
A channel of communication between human and computer

AR, VR, MR
Augmented Reality
Virtual Reality
Mixed Reality

1 *Betsy (2021)*

Time-of-Flight
A sensor used for measuring depth

ToF cameras have made their debut in the cellphone industry entering the 22nd century. In 2021 most flagship and high-end cellphones have this feature included in their array of sensors. These sensors can accurately retrieve depth information, but the technology is also being challenged by AI algorithms that can perceive depth from 2D imagery. Facebook is researching an approach to continuously estimate depth from video through ML and steadily come closer to achieving this in real-time[2]. ML and AI go hand in hand with computer vision, pushing the boundaries and reducing the gap between a humans and computers ability to identify imagery, spatiality and objects. Another significant player is the automobile industry, which relies upon accurate visual recognition to enable safe and capable self-driving vehicles. Not only do they use this technology for navigation, but research is being put into gesture recognition to control the interior interface, enabling them to remain attentive to surrounding traffic while interacting with the car's interface[3].

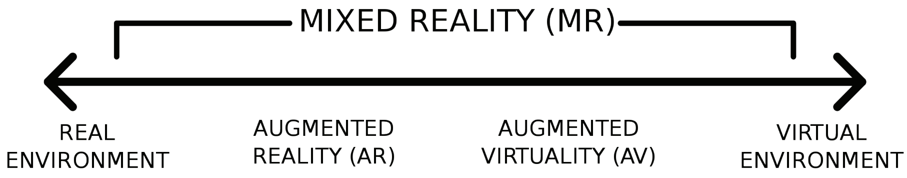
ML
Machine Learning

2 *Luo et al. (2020)*

3 *Alpern, Minardo (2003)*

The gaming industry is at the forefront, chasing immersive gameplay where accuracy and response time are among the main contributors. As gaming has embraced this new technology, VR sees trends of exponential growth[4]. This puts a responsibility on significant game engines such as Unity and Unreal Engine to make this new technology both responsive and approachable. For this thesis, I chose to use Unity as the engine solution for creating the prototype.

4 *Maida (2016)*



2. FIGURE: *The Reality-Viritual Continuum*

1.1.3 Museum Trends

Staying Relevant

A quick google search of museum trends from recent years shows that MR has already become one of the most prominent approaches to the design of museum exhibits. Museums rely on attracting visitors, and "blockbuster" exhibits can be a great marketing tool. Museums are allocating more funds on singular exhibits of greater value.[5] In museums where displaying a *Picasso* or something of similar fame is not an option, they can create these blockbuster experiences by creating innovative exhibits that provide a novel and out of the ordinary experience through immersive interaction with museum content. Many museums have doubled down on smaller displays and

5 *Carliner (2001)*

exhibits in favour of larger ones facilitating several users. While being more costly they allow for more work and technology to be put into single exhibits that are novel and are easily marketable through unveiling and events. By focusing the museum real-estate, exhibits facilitate social interaction between visitors and are often guide-friendly for larger audiences.

The museums have had to widen their field, and entertainment is becoming an increasingly relevant factor in attracting visitors.[6] Mixed Reality can be used as engaging tool to provide interactivity with otherwise static exhibits. The increase in adoption of digital technologies provide additional help museum showcase digital renditions of their collections, which otherwise would require physical space. Museums often have far greater collections than locale can afford to display.

6 *Trajkova et al. (2020)*

The increased use of technology enables and require attention to design and usability, prompting museums to engage programmers, game developers, visual designers or artists to aid in the developing new exhibits. At the heart of these, we often see aesthetically pleasing projections controlled by innovative interactive interfaces or solutions relying on sensor inputs. This has transformed the traditional museum experience, and the way exhibits are designed, introducing terms such as; user-centric, embodied interaction and affordance, previously foreign to the museum field. Studies show interactive exhibits to increase both attraction power and average time spent at exhibits[7]. The success of exhibits is often measured attraction and the time which it can occupy a visitor. This is a slight shift from the earlier notions of how well visitors absorbed information. There seems to be a greater focus on the overall visitor experience, and many museums today are designed with the entire exposition in mind[8] through a holistic approach to provide visitors a cohesive experience.

Embodied interaction
Interaction making use of the body in a natural or significant way

Affordance
The sum of actions an object *can* give.

7 *Vaz et al. (2018)*

8 *Carliner (2001)*

Museums cross-inspire each other and a move towards digital presentation of content regardless of being art, science, history or culturally focused. Digital exhibits offer a way to showcase the old in new formats, that can attract the modern audience. This have shift have made museums technological interactive galleries, where science, history and artworks are displayed in extended fashion through digital mediums. These exhibitions have becoming increasingly elaborate and engaging, and exhibits can in some cases be considered works of art themselves. Museums having the function of presenting content with a heritage value, see technology as a way to merge the old and new and engage younger visitors with little to no prior relation to content, and give new perspective to those with relation to the material.

1.2 The Purpose and Scope of this Thesis

This thesis aims to provide an overview of approaches to exhibit design with gestural interaction, to provide theory and a practical exemplified prototype made with new and approachable technology and software solutions that do not require specialized knowledge of computer vision or ML expertise. The prototype is designed for windows architecture using the Unity engine and Azure Kinect DK sensor.

1.3 Research Impact

While there exist a good deal of research on the field, the current rate of technological development demand a continuity to stay relevant. This thesis makes use of recently released technological solutions such as the Azure Kinect DK(2020), Unity HDRP(2020). While previous versions do exist, few examples of practical documentation of design exist. This thesis aim to provide an overview of a design process, and reflect on its viability, pros and cons to identify challenges, advantages and work-flow utilizing this technology.

Provides groundwork to utilizing gesture control for a potential exhibit.

1.4 Relevance to Field of Study

This thesis have a relevance to several of the program courses. While its relation to music is indirect, it remains relevant through extensive use of sensors, data processing and media playback. Another aspect is the theoretical discussions around exhibit artistry, design, interaction and feedback. Human-Computer Interaction and Gesture-based interactions are a primary focus of this thesis.

1.5 Research question

What design principles can we use to create successful exhibits with gesture recognition?

What issues can be identified and solved through the development of a touch-less gesture-based exhibit prototype using accessible technology?

Method

2

3.1 Research

There are four key aspects to the research which have been approached flexibly.

- Identifying key literature that provide insight into the different aspects of designing gesture-based interactive museum exhibits.

- Analysing literature research and studies for data, specifically important findings and important discussions and reflections and key concepts presented in the literature.
In a qualitative approach the descriptive secondary data collected will be used for summarization

- Summarize and present the finding and provide discussions around the identified concepts to create a theoretical text that may aid the addressing design principals, conceptualization and design of gesture-based interactive Museum Exhibits.
There are a wide range of existing literature on the subject matter, there is a gap in the collection of the available key texts and key findings.

- Conducting applied research by designing Design a working Museum exhibit prototype.

One of the issues faced when addressing literature on technology is the rate at which it might be outdated. Recognizing this and focusing on literature that give theory, and studies over implementation will help in making sure what is presented is current, relevant and broadly applicable.

3.2 Design Methodology

The tools used for this thesis were the Azure Kinect DK, Azure Kinect API, Azure Kinect Body Tracking API, Unity HDRP 2021, "Examples for Kinect" Unity asset pack.

The design process will follow an agile methodology.
My knowledge of the Unity environment prior to this thesis have been limited. Allowing for a flexible approach is necessary as it allows me to address different features, issues and bugs as my knowledge have improved.

Development of a prototype of a gesture-based Museum Exhibit, using accessible available technology.

Finding are based on empirical data gained throughout development and testing.

The tools used for this thesis were the Azure Kinect DK, Azure Kinect API, Azure Kinect Body Tracking API, Unity HDRP 2021, "Examples for Kinect" Unity asset pack.

These tools were chosen as they are gaining relevance in the design of museum exhibits. They have a relatively low level skill requirement compared to other Gesture Recognition approaches. The aim was to develop making use of accessible technology and identify how viable this is as an approach, based on the design principal findings during the basic research phase.

One of the weaknesses of that the methodology poses

THEORY AND CONTEXT

2

4.1 The Museum

4.1.1 Role of the Museum

A museum space offers several consideration to take into account when designing exhibits. First is the need to be aware of the museum role in society. A museum is traditionally a site to preserve heritage values. In this role they are collectors and caretakers but heritage value extends beyond just physical objects, into context, ideas, movements etc. Their second role are as academics, researchers and educators to the public. Thirdly is the museums have obtained a new-found role as centers of entertainment, a role necessary to survive in the modern age. Museums have to compete against other cultural arenas such as, cinema, sports etc. this have pushed them into the territory of an entertainment arena[9]. This paradigm have by some been seen as a of "desacralization of the museum institution". More so there is a worry that the increase adoption of the entertainment role and new technologies may lead to a "Guggenheim effect" where the appeal of the technological medium overshadow the content and information it provides.[10]

As Museums have entered the digital era, research reflects this with a significant increase in studies on museum and digital technology.[11]

4.1.2 Exhibit Designs in the Museum Environment

There is a large trend for adopting embodied technologies in museums.[12] These have opened up new opportunities such as enhanced storytelling[13], virtual tours[14] and virtual hands on artefact and heritage site explorations.[15] The adoption of new and novel technologies are seen as an important factor in attracting mass audience.[16]

9 *Carliner (2001)*

10 *Carrozzino, Bergamasco (2010)*

11 *Shah, Ghazali (2018)*

12 Trajkova et al. (2020)

13 *Hammady et al. (2020)*

14 *Pollalis et al. (2017), Dam et al. (2013), Manghisi et al. (2018)*

15 *Bekele et al. (2018), Garzotto, Rizzo (2007)*

16 *Carrozzino, Bergamasco (2010)*

Exhibit design have become increasingly multidisciplinary, involving developers, designers, curators and AV-expertise. Participatory design involving all stakeholders have become important due limited funds and the extensive costs of new exhibits. Museums are often economically limited in-house to keep this expertise on payroll and thus often purchase exhibits from external contractors. There is also a push towards "in the wild" design, allowing for in-situ iterations to improve upon features and identify flaws as users often find unintended new and compelling ways to use and break the interactive aspects of exhibits. [17] One of the issues faced with the adoption of new interactive methods are how users respond and make use of it, underlining the need for extensive testing.

AV
Audio/Video

17 *Rogers (2011)*

One of the main issues faced in a museum setting is to effectivly engage the audience. Thus it is imporant that exhibits are able to attract, and facilitate rapid interaction. The time visitors devote to an exhibit are often limited, as nearby things may compete for attention.[18] Many exhibits have limited interactability as a design choice, due to the need for throughput of visitors to avoid queues.[19] Other exhibits may warrant visitors to engage for longer periods of time. How visitors approach an exhibit can be largely influenced by their age, e.g. older visitors will have a greater focus on nostalgia while younger visitors will likely have a greater interest in the enjoyment they can gain through interactivity and technology. Embodied exhibits have been found to be able to attract and engage visitors of all ages compared to non-embodied.

18 *Serrell (2010)*

19 *Hornecker, Stifter (2006)*

A museum is considered a public space. Exhibits should therefore be accommodating to a variety of people through accessible design and use of space. The bigger/more important the exhibit, the more people it should also accommodate in parallel to avoid queue and promote shared experiences. The use of space should also account for guided tours, where each exhibit can provide room for learning.

When creating exhibits for museums it is a valid question to ask if our exhibit is a tool or toy, and what is the target audience is.
Museums are in most cases state sponsored or owned, and there are often laws in place to ensure that the institution is accessible to and relevant for different demographics. While it might be favourable that an exhibit have appeal to all ages, it is not in all cases a necessity if there are other exhibit options that are available. Schools tend to frequent museums, and having exhibit options that target younger audiences, e.g. gamified, may be a good

addition. But we should also be aware that these may not have the same appeal to older age groups who are may be more reluctant to participate. Museums also have a defined role as an institution that provides education. If exhibits are overly gamified or focused on visual appearance, effects or other "gimmicks" that have an attractive power, it reduces attention to the curated content.

To engage the audience an exhibit must account for a variety of obstacles. If exhibits demand a lot of cognitive load, or put a high demand on a users skill, visitors will pay less attention.[23] In public settings social aspects must also be considered when designing gestural control. When engaging in embodied exhibits, participants are actively taking on a role that can be observed by bystanders.[24] Gesture-based interaction can put a visitor in the spotlight and social factors may function as a limiter due to the users need of upholding his or hers standards of social behaviour.[25] Many may feel reluctant in putting themselves in a situation that is potentially attention grabbing and may have social or psychological consequences. Social embarrassment is found to be a major deterrent in participation.

Worts[26] provide us with the Critical Assessment Framework (CAF), as a way of measuring the cultural value a museum provides. While these can provide us with some insight and measures onto the content in an exhibit, we need to account for several other aspects when measuring success of an exhibit, such as usability heuristics[25] that provide guiding principles in design.

Measuring the success of exhibits can be challenging but common criteria in addressing this is the ability to educate, provide specific content, prolonged engagement and level of engagement.[28] Some criteria are easily quantifiable, such as the attraction power, and time spent, and information transfer rate. Information transfer rate depends on a collection of aspects such as user planning time, time per input, system response time, and correcting error and unintended interactions.[27] Time allocated on a task can also be an indicator or predictor on the amount of education takes place.[28]

When creating exhibits the local must also be considered, such as spacing, adjecent exhibits and ambiance. Museums are often dimmed in lighting to create an ambiance. In exhibit design one should also consider not disrupting the existing ambiance of neighbouring exhibits with excessive light and sound leakage, as external inputs are disruptive to an immersive experience. Important is also hiding technology outside of where we want to focus visitors attention. Computers and wires are unappealing and cooling solutions for technology are also additively noisy, generate temperatures

20 *Carliner (2001)*

21 *Serrell (2010)*

22 *Walter et al. (2013)*

23 *Muller et al.(2010)*

24 *Worts (2006)*

25 *Nielsen, Molich (1990)*

26 *Hornecker, Stifter (2006)*

27 *Keates, Robinson (1998)*

28 *Serrell (2010)*

unsuited for the museum space. Whenever possible placeing back-end solutions in remote server rooms etc. Hiding sensors, monitors, wires etc. are important to maintain a natural environment that is technologically non-intrusive.

4.2 Gestures

4.2.1 Understanding and Defining

Oxford Learners Dictionary	1. *A movement that you make with your hands, your head or your face to show a particular meaning*
	2. *Something that you do or say to show a particular feeling or intention*
Miriam Webster	1. *A movement usually of the body or limbs that expresses or emphasizes an idea, sentiment, or attitude*
	2. *The use of motions of the limbs or body as a means of expression*
	3. *Something said or done by way of formality or courtesy, as a symbol or token, or for its effect on the attitudes of others*

1. Table: *Definition of the word "Gesture" from Oxford Learners Dictionary and Miriam Webster*

Gestures and Communication

One of the first things that may come to mind when we talk about gestures, are the simultaneous movements we conduct during speech, known as gesticulation. When we communicate in personal interaction, the communicative message is the gestalt of how and what we convey. You may say something, but your facial expression may implicate a contradictory falsehood to your statement. We normally understand communication in complementarity, but also contextualized to previous knowledge and physical or abstract points of referral. Gestures are in this way context driven, as it may vary its semantic content. It also imply that gestures performed out of context are likely not recognizable and will lose their meaning. How we utilize and perform gestures may vary greatly between individuals, making it a idiosyncratic form of communication.

Idiosyncratic
Varying by the individual

Inter-human communication relies on a complex form of communication where complementary modalities of speech, gaze, and body language form the basis of our relaying information. Furthermore minute differences in the dramaturgy in body language, intonation, visual focus, and complex use of metaphors, abstractions and sarcasm, may entirely shift our understanding of the information. The sum of our communication is often redundant, reinforcing a recipient understanding of the content, that may be ambiguous. An important aspect of human communication is the turn-based dialogue. We alternate in sharing information respond and reflect on the reactions we receive. If information does not get across we can further elaborate to get the meaning across. Replacing the receiving end with a computer interface, may create a difficult situation, where there is no such natural feedback. A way of relieving this may be to immersify the user through embodied interfaces or make use of personified agents to adress or embody to create some natural form of feedback.[29]

Personified Agent
Viritual objects that maintain some human properties, such as avatars.

29 *Shneiderman, Maes (1997)*

Human communication is flexible and fluent, and aiming to recreate this communication with computers may be the ultimate goal increasing general accessibility as everyone possess this skill set. Approaching a more natural human social communication is thus a long-term goal in HCI[10].

30 *Thorisson (1996)*

Gestures are an important complementary tool to speech, these are sometimes referred to as co-verbal gestures.[31] In dialectic conversation, gesticulation is helpful in both conveying information, adding dramaturgy and increasing our rhetoric capabilities. It also provide various cues to our state of mind, displaying our emotions and enabling others to empathize with us. Not only is it effective in conveying information, but also improve comprehension and memory recall.[32] Gestures in day to day social interactions are often rapid, continuous and free-form and relies heavily on abstractions, metaphors, symbolism and deictic references. The effectiveness of gestures are dependant on our ability to perform and understand visual representations through movement, that subsequently demands similar conceptions of the gesticulated imagery between the performer and the recipient.

31 *Karam, Schraefel (2005)*

32 *Chee So et al. (2011)*

However if we remove speech from the equation and the true power of gestures can reveal itself, as in a game of charades, or through the storytelling of pantomimes. This alternative way is vital means of communication to those with hearing or speech impairments through the form of sign language.

Movement with information

Let us take a second to think about what gestures are. It is the expression of the human through movement, through the use of pose, arms, face and eyes. Even static gestures have a to and from, so let us use the temporal term movement. Movement constitutes our every interaction with the world, it is the medium between thought and action, the temporal evolution of our body, our physicality and our spatial presence. Some movements we use to communicate, others to navigate and some in response.
While not all human movement necessarily gestures, the difference between them in HCI have become increasingly ambiguent.

In the attempt of explaining gestures and movement in more detail i will provide a, arguably refutable, personal interpretation and non-conventional classification of different movements in a simplified manner. Thereafter I will address some of the more common taxonomy used by researchers.

Semantokinetic: We use our hands to create metaphors or abstractions that exemplify the contents of our communication. These movements/gestures are a language, and as all other languages although with fewer rules, differ from their cultural origin e.g. beckoning sign is commonly perform palm down in parts of Asia and finger counting is often reversed starting from the pinky. By putting communicative gestures into a system and reducing idiosyncrasy, you have what is referred to as Manual Communication, which sign language is an example of. We can term these movements as semantokinetic, gestures with a representative meaning or semantic value. These kind of movements have descriptive and pantomimic qualities through extensive use of expressive metaphors, abstractions and points of referrals. How you conduct these gestures are comparative to rhetoric in speech, the efficiency at which you communicate and convey meaning, e.g. Italians are often known for their excessive use of body language complementary to speech. Not only does this require some performing skill from the user performing the gesture but also requires the recipient to have shared experiences or conceptions to understand metaphoric referrals. Semantokinetic gestures are in human computer interaction fundamental building blocks for *discrete gestures*.

Pathokinetic: We also use movement to externalize our emotions. We clench our fists when we are angry and smile when happy. These are also gestures that can be both conscious or nervous responses. These are universal emotional responses that is also a form of communication we can interpret

HCI
Human-Computer Interaction

Beckoning sign
A "come here" gesture

Manual Communication
Non-verbal systems of communication.

Semantic
The meanings of words, phrases or systems.

and empathize with. Whether as simple as a smile or complex as a dancer's routine, it can evoke an emotional response among recipients. We can consider these as movements as Pathokinetic.

Logokinetic: We use movement to manipulate the world around us, interact with it and spatially displace and orient ourselves within. While we often do not refer these movements as gestures, we could term them as Logokinetic movements, which are based on our reasoning and understanding of the world. In human computer interaction, it is one of the fundamental building blocks of *continuous gestures*.

Taxonomy

Kurtenbach and Hulteen (1990) gave the following description of the term gesture:

> *A gesture is a motion of the body that contains information. Waving goodbye is a gesture. Pressing a key on a keyboard is not a gesture because the motion of a finger on its way to hitting a key is neither observed nor significant. All that matters is which key was pressed.*[33]

33 *Buxton, Billinghurst (2018)*

The term gesture have seen a widened use by the adoption of the term within in the field of HCI. GBI is a way of using gestures as a input method to achieve interactivity between user and computer. The way it have been used varies greatly, and have received much attention being considered a highly natural form of interaction.[34]

GBI
Gesture-based Interaction

34 *Wexelblat (1995)*

It is often used to describe both static positions as well as dynamic movement, and can encompass individual parts or the entire body. It is also used, often confusingly, in conjunction with the similar term pose. As gesture became standard use as the term for finger-interaction with mobile touch-screens, this further causes issues as it can be interpreted both tactile and not. Traditional definitions does not fully encompass the increased use, being a source of confusion. There have been several attempts at creating a taxonomy.[35] Defining suitable categorization and classification is important to create a baseline for research and discussion. Thus finding and creating descriptive subsets of gestures help us understand commonalities and apply previous knowledge in both discussion and research.

Taxonomy
The science of naming, describing and classifying

35 *Vafaei (2013), Vogiatzidakis et al. (2018), Karam, Schraefel (2005)*

Among the most prominent researchers who have contributed to this is Adam Kendon[36] and David McNeill[37] from a linguistic viewpoint. Their

36 *Kendon (2004)*
37 *McNeill (1994)*

taxonomies have been used as templates for Francis Quek[38] and Alan Wexelblat[39], who have been central in developing an HCI specific taxonomy.

Describing interaction gestures it is also important to relay some information about the context, such as style, application domain and input technology and interactive response. This have not always been the case in studies and been a source of critique.[40]

Standardization is critical in the understanding and reproducing of gestures. This is challenging in HCI where there are continuously emerge new ways of using gestures and movements as modalities for interactivity. Having not yet come to agreement, there have been a wide variety of terms used to address and describe different gestures. Below I will provide a short overview of some of the more commonly used ones, and some of the definitions this thesis will adhere to. Please note that some of these are not mutually exclusive and some may be overlapping as they stem from multiple taxonomic proposals by the aforementioned researchers.

Pose is one of the terms used interchangeably with gesture. Arguably pose is more often understood as whole-body over peripheral movement and static over dynamic. This thesis will refer to pose as a whole-body, static position.

Gestures are separated into two descriptive labels, static or dynamic. This thesis will primarily focus on empty/free-hand gestures, conducted in mid-air. In cases where the term is used to refer to whole body or other body parts this will be specified.

Gesture can further be subdivided into Complex gestures, when performed in parallel. E.g. involving multiple body parts, such as both hands. When performing gestures in a series the term Composite gesture will be used. Dividing gestures by their temporally interactive nature, we can separate them into either Discrete gestures, ones that trigger an event upon completion, or Continuous gestures, which continuously feed information to a given interaction. Discrete gestures rely on largely on semantokinetic movements while Continuous gestures rely on primarily logokinetic movement.

Deictic gestures are ones that rely on referral to objects in the domain context, e.g. pointing which can be both static or dynamic.

Manipulative gestures are a used to spatially displace, rotate and morph objects. They are inherently continuous and interactively reliant on a recipient/referential object to manipulate.

Semaphoric gestures can be simplified in description as a signal. They are communicative and refer to specific definitions in a "stylized dictionary", in other words, we have a common accepted conception into the specific

38 *Quek et al. (2002)*
39 *Wexelblat (1997)*
40 *Karam, Schraefel (2005)*

meaning of the gesture. E.g. Thumbs up or OK gesture. They can also be dynamic, e.g. a beckoning sign. Semaphoric gestures are large part of speech accompanied gesticulation. They are representational; we understand them by their Symbolic, Iconic meaning. Thus sometimes referred to as symbolic gestures and iconic gestures. These gestures can also take on different view-points; object-centric when depicting something external or character-centric when impersonating. Semaphoric gestures also include mimetic instances where a gesture is trailing or drawing imagery and shape, comparable to onomatopoeic words.

Onomatopoeia
A word that sounds like what it represents

Gesticulation is gesture that accompany speech, Often in a spontaneous manner.

Metaphoric Gestures are gesticulation that represent abstract ideas and concepts, commonly gestured as a point or expanding or deforming of shapes in some way.

Beat gestures and baton gestures are rhythmic in nature.

Pantomimic gesturing is storytelling through use of sequential gesturing to convey information.

The temporal domain

When understanding gestures as an input, we also need to see them in their temporal evolution.[41] All gestures have different temporal phases; beginning, during and end. A discrete gesture will start reading when recognized as a potential gesture. Varying upon the system, this may be when reaching a certain position spatially(hard-coded), or when a movement is elicited that the computer may recognize as potential(ML).

41 *Walter et al. (2013)*

Temporal
Referring to time

A computer have the additional phase where it is continuously looking for potential gestures to be elicited. In cases where gestures are hard-coded the middle-phase is in essence waiting for a threshold to be fulfilled. In the case of hard coded gestures, this is the arrival at the threshold of a spatially defined relation, while in ML approaches this threshold is determined by the algorithms perceived certainty of a gesture. The more similarities the movement share with a pre-trained gesture, the more the certainty will increase.

Arriving at the threshold we have reached the final phase, where the gesture ends, interactivity happens, and the algorithm may return to looking for new gestural inputs.

In continuous gestures, the beginning and end phase of a gesture is commonly triggered by the fulfilment of a secondary task. There have to be a command that prompts the interactivity to take place. When this command is elicited, the application will respond interactively until another secondary task is complete, issuing the command to stop responding interactivity to the continuous input or the reading of it altogether.

4.2.2 Gesture-based Interaction

Historical overview

The first attempts at gesture control was in 1977 when researchers developed the Sayre Glove at MIT Media Lab.[42] Among the first free-hand mid-air GBI, was the "*Put That There*"[43] also developed at MIT in 1979. A seated user would by pointing and voice commands place shapes at designated places.

42 *Premaratne (2014)*
43 *Bolt et al. (1982)*

During the late 80's and early 90's several new interaction styles appeared, such as 3D mouse with sensor gloves[44], Neural Network used to recognize sign language at high accuracy[45], and Object manipulation with two-handed gestures.[46]

44 *Weimer, Ganapathy (1992)*

45 *Murakami, Taguchi (1991)*

46 *Bolt, Herranz (1992)*

In 2006 the Wii console was released by Nintendo. The console have sold more than 100 million[47] units worldwide, and was for many their first experience of using GBI, with the hand-held Wii Remote.

47 *https://en.wikipedia.org/wiki/Wii*

The same year Microsoft released the Xbox Live Vision(2006), a simple camera that would be the predecessor of a significant piece of hardware for GBI; The Xbox Kinect (2010).[48] The Kinect have seen a large number of use cases providing a accessible solution through their affordable hardware, relatively simple API, and Body tracking implementations.

48 *https://en.wikipedia.org/wiki/Kinect*

The gaming industry's interest in GBI have been one of the largest contributors to technological development. And may be even more so today as gestures are considered an immersive interaction method that is highly suitable for use in VR and HMD environments.

In 2002 the movie Minority Report (Steven Spielberg), a science fiction feature film that was ground breaking in its depiction of the gesture interface using a glove. While we might not be at the technological stage presented in the science fiction movie, there is an increasing tendency of adopting gestures as a method of interaction.[50]

3. FIGURE: *Poster image for Minority report.*

> *I had a great time creating the future on 'Minority Report' and it's a future that is coming true faster than any of us thought it would.*[121]

50 *Lee et al. (2012)*
49 *Huffpost (2013)*

Today, wearable sensors may provide more accurate data but is less common as external camera/sensors have been favoured. These allow us to track gestures free-hand as perceptual input[51], providing instantaneous interaction potential.

Perceptual Input
Term used for gestural input with no wearable/held sensors.

51 Karam, Schraefel(2005)

GBI have become common in mobile, smart watches and TV's, but also see an increased use in all sorts of fields, where it among other usecases, have been used to complete secondary tasks while driving[52], assisting in operating rooms[53] and in military applications[55].

52 Alpern, Minardo (2003)

53 Wipfli et al. (2016)

55 Sathiyanarayanan et al. (2014)

> *Novel technologies are developed to augment people, places, and settings, without necessarily designing them for specific user needs. Instead of developing solutions that fit with existing practices, there is a move toward experimenting with new technological possibilities that can change and even disrupt behavior.*[54]

54 Yvonne Rogers (2011)

Advantages of Gestures

The approach presents several advantages in specific scenarios, but often at a cost. Gestures can provide quick access to an interface, as the need for proximity is reduced, and no intermediary tools are required. It frees the user from physical contact, which have been timely relevant during the Covid-19 outbreak. A no contact solution is hygienic as well as reducing issues of wear and tear. It is a user-centric approach that can provide an increased level of immersion, embodiment and engagement.

GBI is often used in the shift towards *Natural User Interface* (NUI), a design principle of effectively hiding the technology and interface, and in turn increase focus on the content.
In a museum setting, making use of this kind of technology can make an exhibit a novelty as currently few have exposure to this technology.[56]

56 Aigner et al. (2012)

While losing tactility, GBI maintains a level of sensorial experience through embodyment. It is immersive, and may provide a sense of presense for the user in a VE. Gestural input can augment humans in VE and reduces the gap between user and technology, providing a feeling of the technology acting as an extension of the user through embodiment. [57]

VE
Viritual Environment

57 Lee et al. (2012)

Remediation Theory[58], presents the opposing concepts of hypermediancy and immediancy. Respectively complete immersion on one hand and the opposing state of awareness in acting upon a medium. Hypermediancy

58 Michelis, Resatsche (2007)

is by extention a synonym for immersion. Immersion is a recurring term in both exhibit design and HCI.

The complex nature of gestures can make designers think more carefully about simplifying interactions. When simplicity or a level of natural interaction is achieved, it can be easily understood even by non-specialized users.[59] GBI is still by many considered a novel form which may generate interest from many.

59 *Carrozzino, Bergamasco (2010)*

> *Yet it may be argued that interaction fluidity on par with human interaction would be extremely beneficial when interacting with machines, since most of the people in the world are experts in this style of communication.* [131]

131 *Thórisson (1996)*

Disadvantages

The major drawback is often the limits of the technology and the challenges it presents in the design phase. Users must be in the field of view of sensors, which can be affected by lighting conditions, the users clothing, and *occlusion*. Additionally gestures must be performed in a readable speed and with sufficient precision for the algorithm to recognize the movement. Each interaction gesture must also feel intuitive and natural within range of motion. Gestures are more physically demanding over other interaction methods, and can over longer periods cause fatigue and repetative strain or in worst case, injury. Some researchers have pointed out the importance of using ergonomic gestures. They made note that gestures should be in neutral extensions and performed in a relaxed manner, avoiding repetition, static positions and stopping the flow of blood.[60]

Occlusion
Blockage of the view

60 *Nielsen et al. (2003)*

One issue presented with GBI is that a gesture-vocabulary might be challenging to perform and memorize. Furthermore they can pose a high cognitive load, and are prone to unintended interactions.[61]

61 *Keates, Robinson (1998), Wexelblat (2008)*

Contrary to intention, GBI are in cases a trade-off reducing ease of use and speed, compared to traditional interaction methods.

Some system also rely on invasive wearable or held technology, which removes one of the primary advantages free-hand gestures provides, quick access.

The experience of GBI is highly dependant on the system and implementation. Using ML algorithms can make the system computationally heavy, and there may be several instances of less than optimal feature detection, as sensors are reliant on lighting, proper noise filtering etc. Algorithms may also struggle when sequenced gestures are performed, identifying the gesture as one.[62]

It is valid to consider if it is better for a scenario over alternatives such as speech recognition.

62 *Wittorf, Jakobsen (2016)*

Choosing the best Gestures

Technology is currently limited in its understanding of the sum of our communication, and to design gestures for interactive purposes, distinctively identifiable gestures that computers can effectively and reliably identify is critical.

Gestures should also have inherent communicative value in and of itself that naturally correlate to the action performed, as it aids the user's understanding of causality and help them identify gestures based on expectancy. Secondarily these must be **proprioceptially** distinctive so as easy for users to perform and repeat. Thirdly we must consider what affordance movements and gestures give to mapping and the intended referent.

Priorioception
Perception of self movement and spatial position.

Interactions based upon real world interaction is easier to both identify, learn and use.

There are no established standard for how or what gestures we have to use. The need is highly dependant on the application domain and some gestures are more relatable for certain interaction outputs. Some of the criteria such as guessability, learnability, ease-of-performance, memorability and reliability, can help us identifying what is considered a good gesture.[63] However there might also be cases where designers must sacrifice gesture suitability, freedom and complexity for ease of recognition.[64]

4. FIGURE: *Different ways of performing a swiping gesture*

63 *Vogiatzidakis, Koutsabasis (2018)*

64 *Morris et al. (2010)*

Finding gestures that are easily guessed, can be important as there is an expectancy of instant success today when dealing with technology. By working on a many-to-one mapping of input gestures to output interactions we can increase the overall guessability for rapid successful interaction.[65] When we identify a successful gesture interaction, this will likely affect follow-up gestures by repetition for affirmation of causality or exploring in opposing or different directionalties. It may be beneficial in this way to create gestures in pairs, of opposing directions.

65 *Wobbrock et al. (2005)*

Elicitation Studies

A common approach on finding gestures to use in an application follow four general steps[66]:

66 Nielsen et al. (2003)

- Find the interactions
- Gather gestures from users
- Define a gesture vocabulary
- Benchmark the vocabulary

Gesture elicitation studies have been the method of choice in obtaining user preference and expectancy of gestures. It is a variation of the "guessability" method was designed by Wobbrock[67]. In these studies users are shown a *referent* after which they are asked to, while thinking out loud, suggest movements and gestures that fit the interaction. Variations of these include the Wizard of Oz method[68] where the referent interacts when a user perform a gesture, while someone else elicit the command without the participants knowledge, giving them a false sense of autonomous control. Gestures elicitation studies are usually conducted prior to implementation through coding gestures or training gesture recognition algorithms on the identified vocabulary. Agreement/consensus, how many used the same gesture, is the primary measure. However gestures are also often measured in Memorability, Ease-of-Execution, Ease-of-Conception and Enjoyability.

67 Villarreal-Navaraez et al. (2020)

68 Harwood (2018)

Referent -A feature of the user interface that can be controlledindependently using a command

Command -A signal that actuates the execution of a function in the user interface.

> *We are working to understand not why people perform (or fail to perform) certain gestures but rather how we could make gesticulation understandable to computers as an input mode.*[71]

69 Wexelblat(1995)

These studies attempt to obtain "good" gestures, and provide a kinetic corpus or syntax that may aid in the design of interactions and push towards a standardization. A set of gestures used for an application are commonly known as a gesture vocabulary. Elicitation studies can also be reversed to an identification study by showing participants a different set of gestures and ask for a participants opinion on those. A bottom-up approach of finding gestures suitable for a referent seem to be the most common.

Certain notable findings and commonalities have been found from various elicitation studies. According to one study, users expect that their arms are the main contributor to interaction and for them to be independent of pose[70]. The same study also note a close relation to physicality, where size and distance of virtual interactive objects can affect the size of movements. It suggests that we tend to base our gestures on our inherent knowledge of physically manipulating the real world. On this we can make the assumption that users will initially base the affordances of virtual objects upon preconceived notions through their morphological resemblance with real objects. Facilitating this may provide intuitive interaction, but poses significant design challenges.

70 *Wittorf, Jakobsen (2016)*

Another study found a higher consensus in gestures than speech, in an elicitation study finding gestures suitable for TV web browsing.[71] Research conducted on surface gestures found participant preference to be ones that were created by end-user agreement in elicitation studies, over developer designed ones which were noted as possibly being more "physically and conceptually complex".[72]

71 *Morris (2012)*

72 *Morris et al. (2010)*

A review[73] study looked at 47 different papers that conducted different elicitation studies from 2011-2018. The review makes a note that many of elicitation studies, providing tables of gestures, however seem to lack of detailed description / sufficient information, for designers and developers to replicate and make use of. This make it hard to account for different variables, which may affect end results and findings.

73 *Vogiatzidakis,Koutsabasis(2018)*

Wittorf and Jakobsen[74] made several observations of interest in a study.

74 *Wittorf, Jakobsen (2016)*

• Most gestures are preferred are the initial gesture produced. (69%)
• Gestures tend to have a static pose(53%) and path(50%) while conducted dynamically.
• A majority of gestures are physical in nature (55%), emulating manipulation of real world objects.
• Gestures are more often than not world dependent (47%) or object centric (34%).
• There was a fair split between Continuous (56%) and Discrete(44%)
• Most gestures used only the dominant hand (69%)

Their research also found indications that the hand-pose was less important than overall directionality and expression, the exception was when hand-pose had an explicit meaning important to the gesture.

They have found what can be argued as a case of digital convergence,

Digital convergance
A tendency where different technologies to become similar, unified or integrated.

where mid air gestures are often larger variations of touch-based gestures. This may be explained as they have an existing notion of certain movements applying to certain referents. This may also be why there is a prevalence of swipes, push, zoom drag, in various elicitation studies as they can be interpreted as larger variants of common touch gestures.

This may allow for a transfer of already known skills to a new form of interaction. The translation of smaller touch gestures, benefit by building upon already familiar movements, and potentially reducing the learning-curve.

The study also found gestures to be variable by display size, where larger displays prompt larger and more physically based gestures. The study also found larger gestures to be dependant on size of virtual objects manipulated.

A different study[75] suggest there have emerged a convention of discrete gestures. The study, looked at all top downloaded (250) applications making use of Kinect, Leap Motion and Myo Bracelet. They were able to identify 15 gestures as frequently recurring. Pointing, waving and swiping were the most used, respectively.

75 *Theil, Hwang (2015)*

Koutsabasis and Domouzis[76] conducted a study that share a notable similarity to the prototype developed for this thesis. They developed an application for browsing band albums, making use of gesture recognition with the Kinect sensor v.1.8 on a 42" display. They had a gallery style browsing with a next/previous interaction and a select/deselect interaction. During elicitation studies (Wizard of Oz) they found swiping to be "a strong user preference" (22/24) however in testing they found a sideways hand extension gesture to be superior to swiping and a wheel/rotating gesture in gallery style browsing. Outperforming in both speed, workload, and perceived usability. They also noted finding some confusion between right and left navigation in swipe and wheel gestures. Additionally they found holding a gesture to be a preferred action for fast forwarding.

76 *Koutsabasis, Domouzis (2016)*

Natural Interaction

> *The command set commonly seen is small, often unnatural, and usually restrictive, leading one to wonder what benefit the user gains by using this mode and learning the new gesture command language.*[132]

132 Wexelblat (1995)

Alan Wexelblat[77] have provided several critical remarks on the excessive use of discrete gestures. He argues that this reduces gestures to something that can be achieved just as easily with the press of the button. He underlined the importance of providing users freedom of movement through the use continuous gestures over discrete, as this is a more natural extension of real world interaction. Wexelblat notes that in gesticulation, which he describes as the most natural form of gesturing, people tend to vary gestures while speaking about the same thing. Gesticulation and semaphoric gestures are difficult to use as inputs in a natural way due to the variety.

77 Wexelblat (1995) Wexelblat (1997)

Discrete gesture have their place, but should be variable in execution in a many to one mapping.

> *Introducing system-required pauses significantly disrupts the natural flow of interactions, but makes life easier for the system builders. If the input stream is not artificially segmented in some manner then designers must deal with segmentation issues themselves. These are not easy issues, as we have no solid guidelines on how segmentation should be done, but they need to be addressed.*[78]

78 Wexelblat (1997)

VR games have come a long way in achieving more natural interactions, however these systems often provide sensors that are far superior in hand-tracking accuracy by held controllers. In these systems there is also a prevalence of first person viewpoints, where direct manipulation is proprioceptially and spatially easier to understand.

Generally there is also a need for explicit commands, in which case continuous gestures are not suited. However we may recognize his critique as a suggestion to design applications more akin to virtual environments where physical and where continuous gestures become natural, and users can leverage from their spatial reasoning and knowledge of object manipulation.

Studies highlight the benefit in providing alternative movements for identical actions in a many-to-one mapping, which may give users a more natural interaction[79]. Research have found users to favour a possibility of varying gestures, for triggering the same interaction[80]. This increased feeling of choice and autonomy, even for trivial variations, and functioned as a strong motivator of extended use. This show us a way to approach discrete gestures and maintain a level of naturalness by allowing for variety of gestures for single referents in a many-to-one mapping. While redundancy through a variety in gestures to elicit the same interaction may be favourable, it also poses some potential problems. Aiming to create natural interaction in this way is challenging for designers and will substantially increases complexity of systems. One issue that may present itself is conflicting movements. While users might easily perceive differences between movements, similarity in kinetic motion between gestures may trigger referents contrary to intention. Furthermore, Sensors and algorithms may not have sufficient accuracy of input *modalities*, e.g. finger-tracking, to derive the minor differences presented in a gesture.

To exemplify this, a wave could be interpreted as a swipe, or a push as a pointing motion. Thus designers should be careful in this approach if there is also a large number of referents. It should also be considered whether a large amount of referents is necessary as this can put a cognitive load on participants.[81]

Several researchers[82] have pointed out that freedom of expression is important in achieving natural interaction. Increasing degrees-of-freedom have also shown to improve rate of interaction.[90] Varying interactions between discrete, continuous, semaphoric, deictic and manipulative will overall add to a more natural interaction. E.g. deictic pointing is a gesture that might requires both finger tracking and accuracy.

Some sensors and systems have better accuracy than other thus not all kind of gestures can apply to all systems. The challenges of designing with gesture interactions show the importance of designers putting interactions into the preliminary stages of development.[84]

79 *Heinrichs, Carpendale (2011)*

80 *Wittorf, Jakobsen (2016)*

81 *Villarreal-Navaraez et al. (2020)*

82 *Lee et al. (2012)*

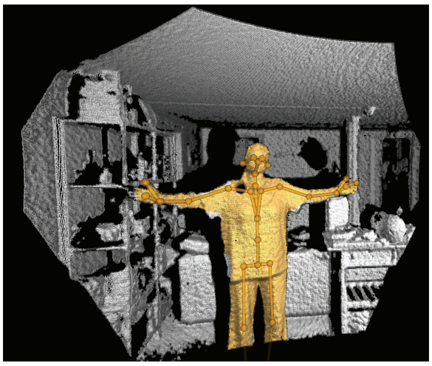
83 *Keates, Robinson (1998)*

84 *Wittorf, Jakobsen (2016)*

4.2.3 Available Technology

Technical Approaches

There are multiple ways to obtain data suitable for gesture recognition. The technology that have been around longest base themselves upon wearables with embedded sensors such as accelerometers and gyroscopes. With technological advancement we are now able to extract data from external sensors such as cameras. This allows for a quick access to the interface and sensors can be hidden, reinforcing calm technology principles. There are a number of inputs usable for gesture recognition. Stereoscopic camera solution have historically been favoured but is today largely replaced by depth sensors, such as Time-of-Flight. ToF sensors output bursts of point-based directional energy-waves*(infrared, laser etc.). Constant speed of the wave allows for temporal measuring of the reflections from which distance is obtained. These are often used in combination with Infrared and or RGB cameras to give a multiple modes of input. New developments in neural networks have also increased AI ability to perceive depth from RGB images and video. While an array of sensors can provide the most information, monoscopic approaches are highly cost effective and will likely be more common in the future. New developments are also being made using WIFI signals to retrieve gesture interaction.[85]



5. FIGURE: *A pointmap and body tracking for the Kinect DK.*

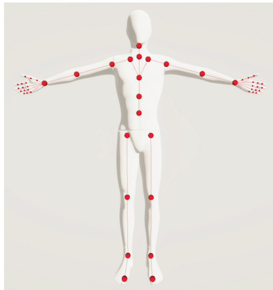
85 Venkatnarayan et al. (2021)

This thesis makes use of several techniques to achieve the interaction. It uses **Motion Detection** through sensor and camera technology. The incoming data temporarily stored, **Motion Capture***in cases where finger and facial movements are included it is often referred to as performance capture. . This data is continually computed in real-time and simplified into **Kinematic Chain**, consisting of joint position, angle and constraint modalities through a technique known as **Inverse Kinematics**. This data is used for **Match-Moving**, where the data is used to mirror movements through a virtual avatar. Secondarily the data is also used for **Gesture Recognition**, the process of identifying specific movements which are triggers for interactive control.

Essential in most systems today is the use of ML to identify people in raw images. This is done achieved with variations of deep neural networks and convolutional neural networks, to produce simpler modalities to work with such as segmentation, outlines and joint/skeleton key-points etc. ML can further be utilized to estimate gestures from this data.

Segmentation
Seperating objects or humans by finding boundaries in the image

There are two approaches to recognizing gestures. One is with the use of a Machine Learning algorithm that have been trained on a dataset. For this approach you need a dataset which corresponds to your inputs. Creating a dataset is a way to ensure this, but requires a significant workload and expertise. It is possible to make use of pre-made datasets or pre-trained algorithms, which however need to be compatible with your sensor setup. Using machine learning also adds a level of computation and may result in a loss of responsiveness depending on complexity and hardware. This essentially results in simultaneously running several layers of ML algorithms, as this is already needed to derive modal complementarity such as segmentation, outlines or key-points from sensor input data. Achieving a high level of accuracy, reliability and input resolution for a system to work on idiosymtomatic gestures may require a lot of work to attain. The system might also need to be training data of various people of different heights etc. Free-form gestures may also be hard to identify due to lack of uniform performance of a gesture between different individuals.[86]

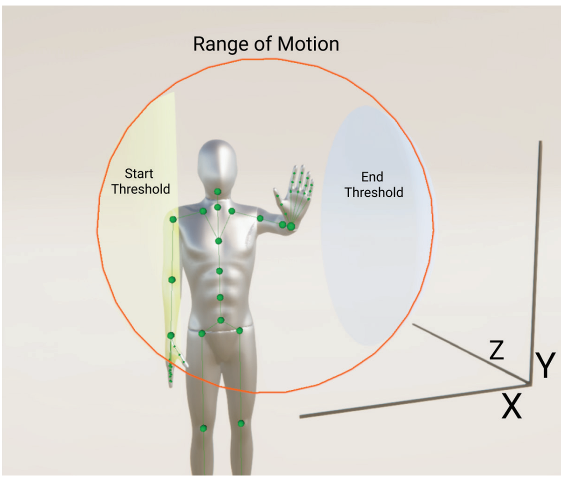


6. FIGURE: *Common keypoints retrieved from inverse kinematics.*

Input Resolution
How reliable interactions can be recognized

86 *Lee et al. (2012)*

Another approach is using key-points or outlines and hard code event thresholds such as a start and end position which must be completed within a temporal window. This method is limited, and may be tedious but it provide a way to create gesture interaction without prior collected data



7. FIGURE: *Thresholds of a swipe gesture*

necessary for training models. To create these spatial thresholds we can use the relation of key-points with conditions like RightHandX > LeftShoulderX.

While movement can often be recognized by pixel changes over time. To increase the computers ability to understand the sensor data we need algorithms to simplify and transform raw data into workable formats for further processing. In computer vision, machine learning is key to segment out a human from raw image/video. ML Algorithms often resulting in data such as veritices that can make out a 3D rendition. This is a computationally demanding approach but provides the most data. Simpler modalities are 2D outlines, and the more common skeleton based, where joints in 3D space are calculated.

The performance/system requirements of a gesture recognition system is highly variable, spanning from web-camera to high-end solutions providing large variety in input resolutions. Furthermore the different algorithms may have vastly different computational needs.[87]

87 *Karam, Schraefel (2005)*

While there are several arguments for the use of natural gestures, there are technical issues with finger tracking, which in many cases are an essential feature needed for "natural gestures", e.g. grab and drag. Considering finger tracking can add a upwards to 4 extra skeletal joint modalities per finger, there is also a significant added computation requirement. In many cases this results in a trade-off where multi-user tracking is disabled. There are few if none technologies that offer both, concurrently. Solution for collaborative gesture interaction with such a high level of requirement, could potentially be achieved through networked gesture interaction.

When working with cameras and ToF we are limited within a FoV. Some systems add additional camera at different angles to create an array which both increase the FoV as well as provide more accurate data from redundancy and increased viewing angle. Occlusion is often a common problem in gesture recognition, and more so if there are multiple users.

FoV
Field of View

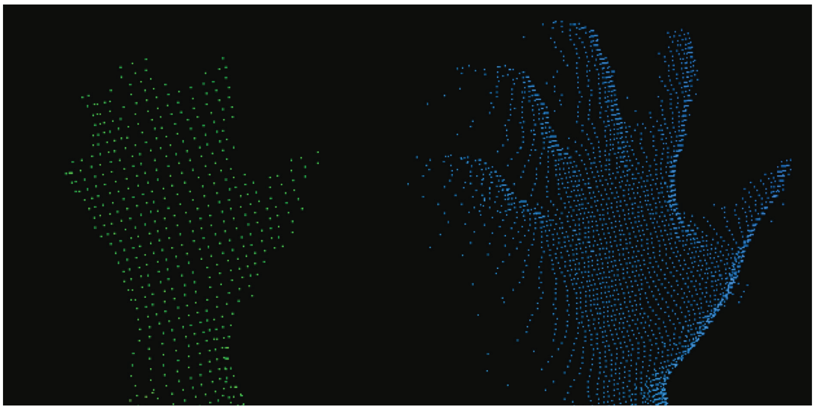
Azure Kinect

The Kinect, Kinect 2, and the newest Azure Kinect have been a popular hardware for both researchers and enthusiasts with a wide variety of uses.[88] The sensor used in the prototype for this thesis is the Microsoft Azure Kinect DK. The unit contains a 12 mega-pixel RGB camera, IR-Camera, and a 1 mega-pixel 30 fps Time of Flight unit, a 6 microphone array, gyroscope and accelerometer. The unit relies on the Azure Kinect API to retrieve sensor data through C/C#. Additionally there is a complementary Body Tracking

88 *Lee et al. (2012)*

API to retrieve kinematic analysis. It provides a total of 32 skeletal joint. While the body tracking API does provide a joint for index and thumb, there is no proper finger tracking.

The system specification states that the depth sensor works up to 4.2 meters. The range is primarily dependant on the reflectivity and lighting conditions resulting in reduced or increased range. At longer distances the point resolution significantly drops. Below you can see a figure showing difference in point resolution at 1.5m and 2.5m.



8. FIGURE: *Difference in resolution with the Kinect DK.*

In testing the Body Tracking SDK the sensors seems to significantly lose confidence in index and thumb joints at distances around 2m.

The older Kinect v2 allowed four states for each tracked hand to be identified: unknown, open, closed, and lasso. This have unfortunately not yet been implemented in the Body Tracking API for the newer Azure Kinect DK.

4.2.4 Interfaces and Interactions

Interfaces

Appropriating a User-Interface for gesture recognition and avatars it will likely have a significant impact. The far most common design approach these last decades have been the command-line interaction method and the Windows, Icons, Menus, and Points(WIMP). Most of us are so engrained in this interface style that it may be hard to detach us from it.

In most cases this is not an effective solution when designing for gesture interaction, unless making use of a pointer controlled e.g. by a pointing gesture. It may be argued that using a pointing approach is not taking advantage of what gesture recognition affords in the first place. A new interaction method should in principle have a user-interface specifically for the type of input modality. New input configurations have set way for the post-WIMP era, where natural user interfaces are becoming more common. Natural User Interfaces(NUI) is a termed as a interface where intuitive and natural actions are used as a means of operation commands, whereby the resulting interfaces are often minimalistic or hidden as direct commands are more common. NUI lets us both lower the cognitive load and improve user experience by less competing visual input and more provided screen real-estate. In this sense we can say that NUI favours virtual environments over traditional user interfaces.[89]

89 *Manghisi et al. (2018)*

Using a 2D environment is not necessarily inferior, as it may provide a sensible middle ground from its simplicity. It can reduce the needed technical modalities, and computational needs. With increasing complexity in interaction however, the reduced spatiality is a potential hindrance in direct manipulation and naturalness. Attempt at utilizing a 2D environment may quickly result in a WIMP inspired interface. 3D environments, provide additional degrees-of-freedom introducing extra directional and rotational axis. It is a more natural environment for movement, at the cost of design and interaction complexity. It also benefit from the full range of input modalities as depth information is likely to be present in any application where gesture recognition is present.

Navigation and Interaction

When an exhibit have more content than the display real-estate afford, a need for navigation emerges. In 2D environments, WIMP have been the most utilized method, where hierarchy is a key function. This hierarchical layering of windows which can be opened, scrolled within, moved and collapsed, significantly increases available real-estate. Navigation is an integral part of interacting with any interface, and there are several ways to achieve it. A user does not necessarily need to have explicit control over navigation, as it also can be automated temporally or conditionally. E.g. in a VE the camera may move through a scene automatically. Navigational interaction with a pointer and key-based solution such as click and drag or arrow-keys/WASD make navigation on computers fairly simple. Moving into a 3D environment, the most natural form is to change and displace the camera viewing angles within that environment. This is essentially the same as horizontal and

Camera
The point of origin a virtual environment is displayed

WASD
Using the WASD keys to move. The most common way of movement in computer games.

vertical scrolling in a WIMP system with an additional axis and Euler angle rotations. Controlling a camera with GBI is challenging from a mapping perspective. One study[90] aimed at tackling this proposed three different solutions all which use a torso rotation to rotate the viewing angle of the camera;

90 *Dam et al. (2013)*

• Virtual Foot DPad - Moving one foot to either direction will move the user in that direction.
• Dial DPads - Touchable buttons in the VE to navigate.
• Virtual Circle - a spatial circle around a centerpoint, moving in around the centre emulate a joystick navigational style.

VR-games usually provide a joystick on the controller for movement, and in cases where there is no use of a joystick the point and teleport, similar to google maps interaction are a common solution. VR have the benefit of free rotation contrary to external display where we need focal focus oriented towards the display.

Click and drag is also a common way of interacting in traditional mediums, however using this method with gestures can be potentially fatigue-prone. Another consideration is the need for excessive smoothing, as joint recognition at times provide data jitter, which may be disorienting for a user if it causes jagged camera movements. One elicitation study[91] did show thata drag gesture had the highest agreement for camera rotation.

91 *Manghisi et al. (2018)*

Any kind of navigation within a 3D environment, gives us some questions to consider when designing: Do we need to rotate the camera? Do we need to move within the VE? Is our viewpoint in first or third person?

Head tracking is also seen used to rotate the camera, popular among flight simulator enthusiasts by using TrackIR or similar hardware. This solution lets the user rotate the camera by head rotation with the limits of keeping eye focus on the screen. This is however not a generally suitable solution for exhibits where multiple participants may be present.

Navigation can also be content based in the form of page-flipping, gallery-browsing etc. In these cases discrete gestures, or virtual interactible buttons or objects are solutions to navigate. This becomes increasingly complex if navigation is hierarchical, adding the need of additional gestures, that do not conflict with existing ones. Elicitation studies show that swiping is the most commonly proposed action for this, however one study also found that a lifting a arm sidewards was a gesture that outperformed swiping in both speed, workload, and perceived usability.[92]

92 *Koutsabasis, Domouzis (2016)*

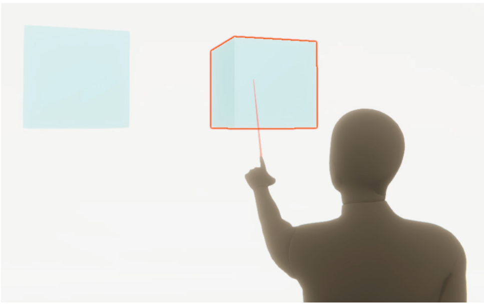
Alternatively virtual interactible buttons or objects can be solutions to for content oriented navigations. In content-based navigation zooming can also be considered as a means of interaction. Elicitation studies show that a two handed drag motion in opposing directions to be have the highest agreement.[93]

93 *Wittorf, Jakobsen (2016)*

A often necessary interaction is to select or confirm. A study conducted found a point and dwell gesture was most common as the way of selecting objects when no hints were provided. Seeing a significant preference of 90% over 4% who performed a push, 3,5% who performed a grip and 1% performing a wave.[94] They also made observation that people tend to explore before selection. On average 2.5/6 before selection, averaging 13 seconds after stepping in front of the display.

94 *Walter et al. (2014)*

Pointing gestures usually make use of ray-casting from the persons fingers or hands in relation to elbows, shoulders and eyes to obtain point of impact. While being one of the more intuitive gestures, pointing also is reliant on accuracy, and will limit user ability to perform other gestures while maintaining this accuracy. Prolonged pointing such as in cases where ray-casting controls a pointer may also be tiring. A common approach for point and select is the use of temporal thresholds as a triggering mechanism, point and dwell/hover.[95]



9. FIGURE: *Ray-casting a pointing gesture.*

95 *Manghisi et al (2018)*

It may be argued that using a pointing gesture aimed at controlling a pointer is to not take advantage of technologies opening for a post-*WIMP* era especially when the interface is heavily UI-based. Pointing is however one of the most common interaction methods.[96]

96 *Moeslund et. al (2001)*

A potential issue with pointing can be false positives. One study[97] where a pointing system was used in a public display found 1/4 gestures to be unintended. A possible explanation is the use of pointing being one of the more common gestures in social interaction to provide attention to a situation or object. While there are museum exhibits that have had good success with the use of pointing gestures[98], Pointing may be ill-suited in a museum or public context, as the main social interactions to direct attention.

97 *Walter et al. (2014)*

98 *Malerczyk (2004)*

Elicitation studies focusing on discrete gestures for selection found a click or push gesture to be the most common. These are to some degree similar in movement, however, variations show that a click also can be angled downwards.[99]

99 *Wittorf, Jakobsen (2016)*

It is possible to make use of a hierarchy of gestures, dependent on various conditions. E.g. When something is selected, a new set of gestures can be recognized or existing ones dynamically change their referent.
This can reduce the number of needed gestures This is a way of clutching with limitless possibilities in conditions. Clutching can also be spatially determined, e.g. standing on a virtual platform. The use of spatial positioning as a method of interaction is termed as proxemics.[100] Proxemics can both be discrete, by utilizing designated area or continuously mapped to an interaction. Proxemic interaction and spatial positioning is easy to combine with gesture interaction without adding a lot of complexity. One study[101] compared different interaction styles found walking over platforms to be score higher on a Likert scale than both, touching and gesturing.

Clutching
Make the output conditionally dependant

100 *Ballendat et al. (2010)*

101 *Campos, Dória (2009)*

Continuous gesture are often a form of direct manipulation. Direct manipulation is an deictic form of interaction where referent objects are acted upon in the virtual environment by incremental actions. Direct manipulation offers the user immediate and continuous feedback. E.g. Picking up a ball and tossing it. Direct manipulation consists of spatially relocating objects, rotating, scaling or otherwise morphing, grabbing or applying force to. While the referent response may be a commit action, the continuous gesture itself is considered reversible, in contrast of discrete gestures which you fully commit to as there is no interaction until the gesture is complete.

Continuous gestures are in some ways easier to design for virtual environments when making use of game or physics engines. Direct manipulation do not necessarily rely on gesture recognition, but instead makes use of modalities provided from inverse kinematics. The recognized key-points and their relation can be directly tied to the interaction output. It is however in most cases favourable to have some condition to disable or enable a continuous gesture, such as a discrete gesture. E.g. discrete grab, continuous carry, discrete release. In particular the emergence of VR application and gaming, there have been a significant improvement in continuous gestures, specifically in the form of direct manipulation. Using physics engines provide a simple and effective way to manipulate objects.

One of the main issues we face with direct manipulation is the need for good finger tracking. Proximity to sensors are vital for a accurate reading. Wearables and VR controllers lets us place sensors close, resulting in precision tracking. This becomes a challenge when using external sensors and increasingly so with further distances.

Designing for continuous gestures may make use of virtual objects in skeumorphic models that allow users to build upon familiar skills of real world manipulation. Building upon object resemblance and manipulative knowledge, these is are easily learn-able actions. While direct manipulation provide a less initial barriers, it might not effectively support content exploration without extensive user-labour, compared to a UI, which is a potentially speedier approach. A possible solution is the merging of UI into a 3D virtual environment. In this fashion menus may be graspable and movable objects. Selection, or a discrete gesture can open marker menus for a quick access and removal of additional interactions. When adding UI elements it is important to consider their properties such as if they should be affected by physical properties, be temporary, transparent or movable.

10. FIGURE: *Example of a marker menu.*

Other styles of intermediate solutions are abstractions in the form of a surrogate object[102] acting as a mediator to a domain object. E.g. a small tangible ball can control a distant planet. Interaction upon the surrogate will indirectly affect the domain object it represents by inheriting properties from the surrogate. Surrogates with a metaphoric value and skeumorphic similarity can strengthen the visual relation to the domain object. In this way surrogates provide an intermediate layer of control, to change the manipulative behaviour of different gestural interactions, and can provide a intuitive solution for continuous gestures. It can be an effective way of control, when there is a spatial gap and spatial movement is restricted.

102 Beaudouin-Lafon (2000)

These form of abstractions can is also present in instrument interaction[103], where different tools can be utilized to achieve different tasks. E.g. picking up a shovel, can let you dig. Picking up another tool may afford entirely new referents. These are a natural extension of how continuous gestures and manipulation can be utilized.

103 Beaudouin-Lafon (2000)

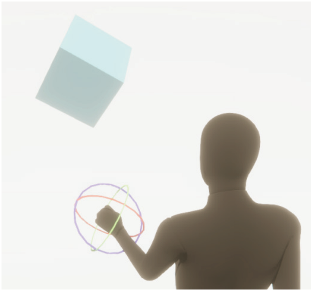
Magic lenses[104] are a different form of intermediate instrument that alter visual properties or modifying the scene when looked through, possibly revealing priorly hidden objects.

104 Bier et al. (1993)

Another form of intermediate control is the use of widgets, e.g. separate ones for spatial displacement, rotation and scaling. These kinds of widgets are common in 3D software. Adobe conducted a study where participants commented "being able to define a particular axis of rotation is useful" and "being able to use one hand as a pivot point and the other to define the axis".[105] Adding these kinds of intermediate referents could aid in effective direct manipulation.

11. FIGURE: *Example of a magic lens*

105 *Arora et al.* (2019)



12. FIGURE: *Example of a widget-based interaction*

Continuous gestures have the possibility of being fatigue inducing, if the user is required to keep their arms in a position for a prolonged duration.

Aiming for natural direct manipulation may reduce entry barriers among demographies such as disabled or elderly, who often posses less technological aptitude.
With direct manipulation may also quickly lead to the need for using both hands for complex gestures.

WASD - Using the WASD keys to move. The most common way of movement in computer games.

Camera Viritual Camera System - The viewpoint within a Viritual Environment displayed to the user

Discoverability and Feedback

One study[106] found that users tend to stay less than 20 minutes at any exhibit. The study also found the majority to explore less than half of exhibits. Arguably this point may vary on the size of the museum and its exhibits as well as its layout. The study also noted that the exit have a strong attraction and that adjacent exhibit garner more visitors.

106 *Serrell* (2010)

VE Viritual Environment

Making sure people discover interactability is one of the largest issues faced in public displays. Facilitating the initial gesture to be discovered without explicit guidance, can generate a sense of achievement, and induce curiosity, for further exploration. Users interacting on public displays are more often than not first time users. Thus having a vocabulary of gestures that are quickly identifiable is challenging task[107] but imperative, as those are unsuccessful, will not maintain interest. Studies have found it critical that successful interaction must be near immidiate[108]. Users will not linger, if they do not understand an exhibit. If visitors are not successful they might

107 *Trajkova et al.* (2020)

108 *Walter et al.* (2013)

assume the exhibit is limited or broken and move on, often referred to as "affordance blindness"[109]

There are three key factors to account for:

Attraction power - The ability to grab attention to the display

Curiosity - How effectively we can make the user interested and curious to explore the display

Revelation - How likely is it that a user will discover an interaction.

Attracting visitors can be achieved through a multitude of ways. Some proposed models are the Bayesian surprise and behavioural urgency. Making someone aware of the presence of public display is not necessarily a given without attaining focal or sensory attention.[110]

Changes in luminosity, dramatic movements, or auditory features can grab attention quickly but can also be pervasive.Avatars have been shown to be an effective attention grabber when mirroring user movement.[111]

The honey-pot effect[112], is a often mentioned term with exhibits in public settings. It explains that groups tend congregate by moving towards other groups, an important factor in drawing in bigger crowds to an exhibit with people already interacting. This also increases the chances of other people discovering the interactability of an exhibit.[113]

In a museum context the initial discovery of exhibits are less challenging than other public areas, as visitors are more often there with the intention of exploring. Museums also tend to provide additional information with signs etc. to make the users aware of exhibits. However revealation of possible interactions and their input is a challenge that must be addressed.

Feedback is important for any user to reduce the time of initial interaction. Various signifiers[114] can be implemented to increase the percieved affordances given by an exhibit. Icons, labels or text all add to a users visual stimuli and cognitive load. Reducing additional information to a bare minimum limits sensory overload and provides a more natural environment and lets users focal attention remain on the avatar which provides continuous feedback or on the interactible content. If there is information overload this may also have an off-putting effect on visitors not prepared for the cognitive requirements. Overcrowding the interface and decision-making can cause feature fatigue.

109 *Trajkova et al. (2020)*

110 *Muller et. al (2010)*

111 *Trajkova et al. (2020)*

112 *Brignull, Rogers (2003)*

113 *Trajkova et al. (2020)*

114 *Shuhaili et al. (2017)*

One study[115] suggests three different strategies in how information can be relayed to the user:

• Spatial division - Splitting the screen where information is accessible
• Temporal division - temporary interruption revealing information
• Integration - The use of visual queues within the application.

They also suggest different"integration cues" that may provide guidance to users, such as a voodoo user, temporarily taking control of your mirror image, adding a fake user avatar, or reachable point guides.

Spatial and temporal division are more intrusive than integration, which may be subtle and more creatively implemented. Visual guides may be a suitable tool but still should not reveal the interaction. Rather we can apply choice architecture and nudge theory to prompt the user to move in the directions of functionality or physically engage and explore affordances. E.g. the use of particle or visual effects in a heat map method (the closer you move towards point of gesture start/end the more effects), providing subtle guidance while avoiding commandeering users. Other ways guiding effects can be through shadow-guides or tracing. A study by Adobe[116], found control of particle systems to be enjoyable. Similarly lighting or other dynamic effects can be used to elicit these responses among users. While flat out revealing the interaction is unfavourable, providing the user with a goal can be an acceptable middle-ground. Making use of a Initial mode where clues are provided and removing them upon entering the exhibit area may also be a constructive way suggest initial movement. This could further be enhanced in a gamification, which may provide a high level of enjoyment, engagement and achievement, but often at the cost of educational content or as an explorative tool.

At some point we need to balance between tool or toy[117]. Museums have a defined role as an institution that provides education. If exhibits are overly gamified or focused on visual appearance, effects or other "gimmicks" that have an attractive power, it reduces attention to the curated content.

For a gestural way of providing essential information could be the use of a "I don't know" shrug icon appearing if a user have not been able to interact within a set time window. Adding dynamic elements may prolong the amount of time a user is willing to stay at the exhibit without successful interaction. In designing windows 8, UX-Designers found visual cues prompting curiosity to be an important contributor to exploring/discovering gestures.[118] Clues and hints to yet unperceived affordances are often referred to as signifiers.[119] Good signifiers are essential for the user experience in discovering and navigating interactive systems.

Incremental and continuous feedback is essential to the user experi-

115 *Freeman et al. (2009)*

116 *Arora et al. (2019)*

117 *Muller et al. (2010)*

118 *Hofmeester, Wolfe (2012)*

119 *Shuhaili et. al (2017)*

ence as it provides clues on interactivity, causality, and may be informative and suggestive in how they interact. Good feedback will result in less errors and let users achieve goals more rapidly. Feedback can also boost curiosity by making use of ambiguous and attention grabbing effects. Tutorials and explicit directions may reduce a persons feeling of achievement in discovery, the amount in which it boost curiosity and formation of goals as the user may wait for further directions. In essense providing too much guidance can remove senses of autonomy from the user and in turn reducing the level of enjoyment. Thus finding a middle-ground solution vital where enough information is given for the user to identify interactability, without revealing too much.

Due to the variation in different gesture application and the fact that few people are daily exposed to it, we should design under the principle that every user is a novice. We can also assume that first time users act spontaneously as they do also not have an a goal or command in mind without knowing what interactions are offered.

Incremental feedback allows a user be less prone to error and simultaneously reduce the time they need to complete a task. Users rely on continuous feedback when using gestures for a good interaction and this may substantially affect interface layout and design decisions.

> “People often make surprisingly good decisions using simple, “fast and frugal” heuristics. These are rules of thumb that ignore most of the available information.[133]”

133 *Yvonne Rogers (2011)*

Display and Layout

Display positioning and real-estate are important considerations. Positioning of the screen and a visitors freedom of movement may affect our decision-making both in choosing suitable gesture but also in how users perform them. Large gestures require a lot of space, may cause issues when combined with proxemics by being potential hazards for adjacent visitors and a potential for injury if dramatic. More so large gestures should be limited if the exhibit is a multi-user experience or interaction.

Design should aim to decrease the amount of secondary tasks to achieve a goal. Simplicity is at the core of what good gesture interaction is.

Large display provide more communication potential and quickly becomes a requirement if it is a multi-user exhibit.[120]

120 *Shuhaili et. al (2017)*

4.2.5 Multi-User Environment

Creating a exhibit allowing multiple users to engage simultaneously can be beneficial to the experience, but several present design challenges. Most museum visitors are visiting in a group, and prefer a social shared experience. Designing exhibits that allow for multiple simultaneous participants, will increase the level of social engagement among visitors.

Designing cooperative exhibits can also engage bystanders by prompting them into collaborative action.[121]

121 *Lee et al. (2012)*

If large exhibits are only designed for individual participation, it is also a potential source of queues. Social engagement in exhibits may also potentially promote learning. If one is able to discover an interaction the others can replicate the movements. Multi-user experiences are a trade-off as it adds a lot to complexity, both in hardware requirements but also in conceptual design and execution. A result of this is often a limited set of available interactions.[122] An aspect researchers have made note of is that causality become increasingly difficult to keep track of whit a higher number of participants.

122 *Carrozzino, Bergamasco (2010)*

Another problem we are faced with how to solve the interactive aspects. Do every participant have their own interactible elements, or do they share control over the same interactions. Allowing all participants full control of the same elements have the inherent issue of simultaneous contradictory gesture commands that may give way to confusion and unexpected results. Allowing for shared control rely on social factors for a fluid interaction. If there is lack of communication and planning by participants, there may be confusion as to who “is in charge”. Large display exhibits, may also host participants with no previous acquaintance to each other which may cause reluctance to engage in cooperative or competitive behaviour. Children often tend to engage in a frivolous manner, and it can be beneficial to account for this. A solution to this may be to create a the ability to pass on control, and in this way create a pseudo-multi-user experience. E.g. Closest person, or a give/take control gesture.

Another consideration, is whether the exhibit should be collaborative or competitive.[123] Cooperative interaction may require users to work towards a common goal, E.g. by gamifying, or dividing control parameters.

123 *Yvonne Rogers (2011)*

Motivation can be gained from collaboration between participation if there exist opportunity for the individual to influence other participants.[124] This can be a strong argument for competitive design, but this also may push the exhibit towards gamification.

124 Muller et al. (2010)

In case of shared interactive elements, it could be necessary to impose a conditional control, or clutching. Some ways to approach this is by having proximal zones for navigational interactivity, or object that can be passed along tied to a specific referent. This way you maintain the social aspect if multi-user avatars are present, but maintain a individual control that appears as pseudo-shared level of control.

Implementing limits like is suboptimal, and conceptual ideas of how competition or cooperation can be maximized can benefit the experience. Letting users work together to achieve goals is a highly social way of interaction. Also important, is letting every users actions have visible effect to increase causality. Designing for multi-user interaction must also take notice of location and the size of gestures to avoid participants physically obstructing or occluding another. Engaged users will expectedly have their focal attention to the screen and may be unaware of adjacent visitors. Similarly volatile gestures should be avoided reducing chance of accidental injury.

4.2.6 Avatars

Advantages of Avatars

The use of a performance-driven Avatar with synchronous mirror-symmetrical movements serves to provide continuous feedback. Studies show an improved spatial perception when able to see their entire body[125]. The avatar functions as a visual aid to proprioception.[126] By directly imposing the user within the virtual environment, it is potentially a major contributor to embodiment and immersion. This kind of reflection is also a natural attention grabber, and an effective complement or substitute for visual or audible effects to provide feedback. A study found that making users part of the display was a very big attention grabber.[127]

125 Steed et al. (2016)

126 Wang et al. (2020)

127 Schönböck et al. (2008)

By identifying mirrored movements from an avatar can act as a natural motivator to elicit movement and exploration of affordances which in turn will increase the likelyhood of revealing interactions.

Avatars creates an extension of physicality and is an effective way to suggest direct manipulation with virtual objects in proximity. This approach also directly builds upon Virtual Reality technology, which is becoming increasingly common, and users may inherit knowledge of perceived affordances and interaction models from familiar VR implementations. While a first-person perspective is more common in VR, research show that user's prefer a third-person perspective for spatial tasks of whole body movement.[128]

Avatars provide additional focal points within an interaction, but the extensive use of this in gaming, suggest this is not an issue of concern. People have a innate self-awareness of spatiality and peripheral vision is likely sufficient if focused content is distanced from the avatar.

Using avatars, one must decide avatar relation to the user, which may, face the user or share orientation with the user. A shared orientation is technically a simpler choice, as sensors in most cases are expected to be positioned in front of the user to avoid occlusion of gestures. This creates an extra step of rotating the avatar to face the content and mirror the axis of movement. Research show that a first person perspective provide a lower adoption time and better close object manipulation. Third person perspective is preferred when doing whole-body movements, such as walking as it provide a superior spatial overview.[129] A study also reveal that avatars should share orientation with the user as opposed to facing them, as this is a source of disorientation and is experienced as more pervasive.[130]

Using a third-person perspective avatar effectively disallows the use of pointing as a continuous interaction as it creates a directional offset between avatar and user.

By spatially incorporating the user, with the virtual world, a person will base perceived affordances of objects upon how they can be manipulated in the physical world. There exists expectancy that virtual interaction adhere to preconceived notions of object affordances and base their manipulation and navigations upon these. Exploring the virtual world by physical manipulation can be expected from the user. This will then continuously reveal more perceived affordances. This may argue for having interactible objects within reach for direct manipulation. The size and positioning also become an important consideration as it can occlude a lot of content, especially in multi-user environments. Smaller avatars, however, can make manipulation more challenging, as it can also result in distancing from objects or causing a significant difference in size between user and content. In these cases surrogate interaction styles is a potential solution.

A problem that may occur with avatars is poor tracking, miscalibration and jittery input data. This may create necessities of implementing

128 *Wang et. al (2020)*

129 *Salamin et al. (2006)*

130 *Wang et al. (2020)*

various filtering to avoid sudden unintended avatar movements. Filters to reduce noise and sudden loss of data are likely a necessity without avatars as they also can be a source of false gestures inputs.

The use of avatars allow for virtual spatial zones as a proxemic interaction. Furthermore it can also allow for virtual barriers to guide and restrict user movement within sensor FoV.

Avatars could also potentially alleviate social reluctance, as bystanders will quickly be able to reason the users intent behind awkward movements. The avatar may also act as a dominant focal point over the performing user.

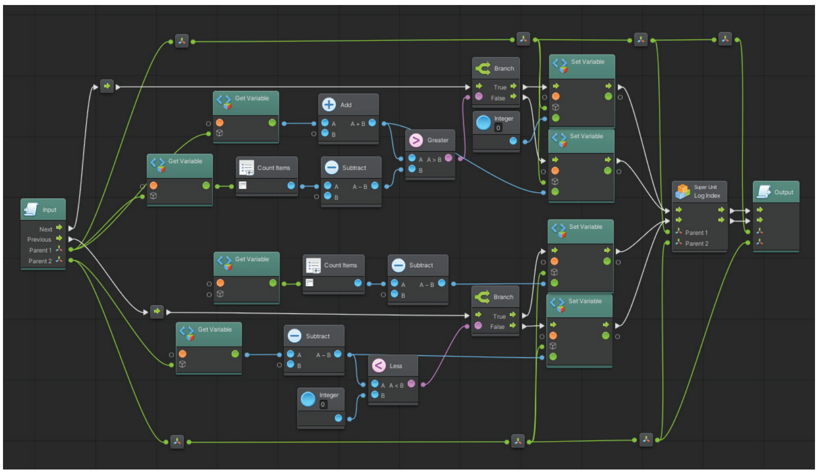
> “*Theories about human-computer interactions that were derived from lab-based research often do not map onto the messy human-computer interactions in the real world.*
> *Likewise, it has proven difficult to say with any confidence the extent to which a system or particular interface function can be mapped back to a theory. Typically, theories end up as high-level design implications, guidelines, or principles in interaction design.*”
>
> *Yvonne Rogers (2011)*

Design

4

5.1 Choice of Software

Unity 2020 is the engine used to develop the prototype. Unity is a game engine free of charge outside of commercial purposes. It also have a selection of assets available through their asset store. Among the important ones I have made use of is the "Azure Kinect Examples", made specifically for the two latest versions of kinect. This asset makes sensor data from the kinect API accessible through unity, which otherwise would require extensive C/C# knowledge. It also includes code examples and implementations of gesture recognition. This have served as a framework to build the gesture recognition on. I have also made use of the Bolt Library, a visual coding tool-kit Unity offers. This is a node based way of programming, where each functional algorithm that is available for unity is represented as nodes with input and outputs. Wires can be drawn between this to program functions. Essentially anything you can write in code could be replaced with visual nodes. This provide a simple way into developing applications through unity, without the need of C/C# proficiency and attention to syntax. The ease of quickly assembling application functionality make this a great asset for prototyping.



13. Figure: *The Bolt environment for Unity.*

5.2 Layout

The application have a limited amount of virtual objects. The camera faces a platform in the lower portion of the display area. This is the platform upon which avatars are positioned if a person enter the sensors field of view. These avatars are instantiated only when a user is tracked within the sensor field of view. On the edges of the numbers depicting each decade with a corresponding platform. The artists are represented by images as cover-art of vinyl records in a carousel.
3D models were created using blender. I chose to have a dark muted back-

14. FIGURE: *The layout of the proto-type exhibit.*

ground as this allows for a better focus on the content. In contrast I chose is white background for the platform, as it will easily contrast the avatars.

5.3 Functionality

The carousel and the different LP / artists objects are instantiate at the start of the game. This adds significant complexity as everything is put into the scene by code, contrary to the most common approach of working with objects in unity which is to place them as objects existing at the start of the application. The reason behind this approach is that it allows to dynamically add content into the carousel, e.g. by adding an extra artist in each decade. This allow for further dynamic entries through adding files to a folder that is read. This gives the opportunity for people who are non-familiar with the application and its framework to add content. Having first hand experience in working at a museum, this have presented itself as an issues, as adding content can time-consuming and may require the original developers to aid in updating the content.

15. FIGURE: *An avatar within the prototype.*

When a user walks into the Kinect sensors field of view, an avatar is instantiated and positioned at the platform. This inherits its movement from the by relating the skeletal joints to corresponding skeletal rig joints of the 3D model. This results in a mirrored movement of the user.

The model itself is not visible but instead a particle system engulfs the 3D mesh of the avatar. In this way the avatar provide anonymity and is androgynous to be suitable for both genders.

These particles can also provide an added trailing effect, which can be visually stimulating, but more importantly, be a strong proponent for elicitation of movement. The solution is a multi-user solution that dynamically adds and removes avatars as people enter and leave the field of view.

There are 8 discrete gestures that have functionality:

1.Swiping in a leftwards motion with the right hand
2.Swiping in a rightwards motion with the right hand
3. Swiping in a leftwards motion the left hand
4. Swiping in a rightwards motion with the right hand
5. Pushing with the right hand
6. Pushing with the left hand
7. Raising the right hand
8. Raising the left hand

There are three referents meaning half of the gestures are redundant but rather only allow for a variation in how the gesture is performed.

The 4 referents for the and the mappings are:

Rotate the carousel clockwise - Swipe
1. Swiping in a leftwards motion with the right hand
2. Swiping in a leftwards motion with the left hand

Rotate the carousel counter-clockwise
1. Swiping in a rightwards motion with the right hand
2. Swiping in a rightwards motion with the left hand

Starting a music video in full screen
1. Pushing with right hand
2. Pushing with left hand

Take control of navigation
1. Raise right hand
2. Raise left hand

The referent for the pushing gesture, will dynamically change to stopping the music video if repeated during the playback. In the case where the video plays its full length the video stops and the view returns to the scene and the referent for the push gesture changes back to starting a video.

There are also 8 discrete proxemic zones each mapped to a separate decade.

By stepping inside one of these zones, the display carousel will be replaced with carousel for the chosen decade, and the numbers signifying the respective decade currently showing will light up.

The zones each have a collider, an invisible object, When a user-avatar's collider overlap the referent is acts out.

There are several reasons I chose to utilize discrete gestures. Firstly is the lack of any hand state tracking by the Kinect. Secondly, Despite the vinyl models being three dimensional the primary content are images. Images are two dimensional and therefore there is no added benefit of any kind of direct manipulation.

There are various animations involved in changing the decade and browsing between the different artists. The animations make use of a technique called tween-animation. The animations involve resleeving the vinyl into its cover, rotating the carousel and unsleeving the vinyl of the artist now centred. Simultaneously with rotating the carousel all the vinyl's and their cover are rotated so as to face the user continue to face the camera.

Each decade carousel is positioned at different heights, outside the view

By walking on these circles the user will control the current decade on display.

These avatars are instantiated only when a user is tracked within the sensor field of view.

Each decade have its own carousel located at different heights, but is by the engine considered one object. By changing the decade the object will move on the y-axis corresponding to the selected decade. Moving animations are done through the use of tween-animations technique. Virtual objects are moved by generating "key-frames", simply to and from state. This can be done with a single movement or as consecutive or parallel animations. By providing the to and from states, as well as a duration and curve, smooth movements and transitions are computed.

The front-most vinyl is the currently selected artist. The user can browse through the different artists by eliciting a swipe gesture rotating the carousel. By performing a push gesture, a video of the selected artist will play, which may be stopped by repeating the push gesture.

5 Testing

One apparent issue with the proxemic interaction is the sensors narrow FoV. To combat this the sensor can be placed at further distance, this however can be a cause of drop-outs due to the limited range of the sensor. A natural solution could be to add additional Kinect sensors. This provides several advantages such as increasing the accuracy causing less jitter, decreasing occlusion and increasing the FoV. Adding more sensors do also pose other issues that would need to be addressed, among one is the need for separate computers for each and the transfer of data between those. Addressing this however is been beyond the scope of this thesis.

User testing

The current Corona outbreak made user-testing of this exhibit prototype outside the time-scope available, as the most restrictive guidelines were in effect in the Trondheim municipality. A outbreak on the museum have also forced the museum to close during planned testing period. Upon reopening a second unfortunate situation was the failure of an exhibit GPU at Rockheim, making the only spare GPU, which was initially used for testing for this prototype had to be used as a replacement as there is no way purchase GPU's due to the current microchip shortage.

As a user-test became a non option, I will report on issues and finding during the initial testing, who were conducted with two separate people at three different times.

One test was conducted at home with my fiancée on a computer monitor and the other two was conducted on-site at Rockheim at the "Honorary Wall" exhibit location, with the Head of IT as a tester.

One issue was immediately apparent, when I asked my fiancée to perform a swipe gesture. The way she performed it was outside of the constricted area. While I conducted swipes at shoulder hight, she performed it at waist hight and in a very horizontal motion. After several attempts she succeeded with guidance. Upon understanding the necessary motion she also noted that it did not work if performed in the opposing way with the same hand.

After these finding I increased the constrictions.

Testing the application at Rockheim, The same issues repeated itself, when he performed an overhead swipe, as the initial gesture. This was solved by further increasing the viable area for the gesture.

Upon enabling for overhead swiping, the participant commented on how it he perceived it as both natural and responsive.

Among the other finding was that there were a significant amount of false positives for the push gesture. The swiping motion would at time move far enough forward to prompt the push referent before reaching its destination. This was solved by increasing the length at which you had to push forward.

Another finding was on the hardware side of the Kinect DK. It requires a high-bandwidth USB-C cable. The provided cable is only two meters which was constricting in how different sensor heights could be tested. This can have implications on providing a solution for a remote computer, which is favourable in a museum setting due to noise and the need to hide technology. USB have a notable drop in signal strength over long distances, This is probably solvable by utilizing an usb to ethernet adapter to send the information over ethernet. This was not tested for.

Another issue that became apparent was sudden drop-outs when standing at angles and rotations difficult for the Body Tracking API to detect necessary features for joint skeleton data. This caused the Avatar to disappear and re-appear sporadically. If the current user dropping out was in control of navigation, the control would transfer back to the remaining user. As of yet I have yet to find a solution to this issue.

During the second testing I also became aware of an issue occurring when performing a swipe. In some cases, the returning of the used hand to a resting position could elicit the referent of the opposing direction, thus immediately browsing back/forward. This was fairly easily solvable by adding a small delay between each gesture interaction allowed.

6.1 Further Development

There are various implementations that can improve this prototype. Firstly is to add signifiers that will increase initial revelation and discoverability of gestures and the interactive aspect of the exhibit.
Currently there are is nothing to suggest what the user should do, except the for the perceived affordances given by the user avatar. It is possible the initial gestures would be guessable but this has not been tested.

More consideration should be put into identifying a use of continuous gestures, not related to direct manipulation.

It would also be natural to add the name of the current artist in front.

REFERENCES

A

Aigner et al. (2012)
Roland Aigner, Daniel Wigdor, Hrvoje Benko, Michael Haller, David Lindlbauer, Alexandra Ion,Shengdong Zhao, Jeffrey Tzu Kwan Valino Koh
2012
Understanding Mid-Air Hand Gestures: A Study of Human Preferences in Usage of Gesture Types for HCI
https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/Gesture-sTR-20121107-RoA.pdf

Alpern, Minardo (2003)
Micah Alpern, Katie Minardo
2003
Developing a car gesture interface for use as a secondary task
https://www.researchgate.net/publication/221519818_Developing_a_car_gesture_interface_for_use_as_a_secondary_task

Arora et al. (2019)
Rahul Arora, Rubaiat Habib Kazi, Danny M. Kaufman, Wilmot Li, Karan Singh
2019
MagicalHands: Mid-Air Hand Gestures for Animating in VR
https://dl.acm.org/doi/10.1145/3332165.3347942

B

Ballendat et al. (2010)
Ballendat, T., Marquardt, N., and Greenberg, S.
2010
Proxemic interaction: Designing for a proximity and orientation-aware environment.
https://dl.acm.org/doi/10.1145/1936652.1936676

Beaudouin-Lafon (2000)
Michel Beaudouin-Lafon
2000
Instrumental Interaction: An Interaction Model for Designing Post-WIMP User Interfaces
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.332.4880&rep=rep1&type=pdf

Bekele et al. (2018)
Mafkereseb Bekele, Roberto Pierdicca, Emanuele Frontoni, Eva Savina Malinverni
2018
A Survey of Augmented, Virtual, and Mixed Reality for Cultural Heritage
https://www.researchgate.net/publication/323975580_A_Survey_of_Augmented_Virtual_and_Mixed_Reality_for_Cultural_Heritage

Betsy (2021)
Betsy Frank
2021
SNAP INC. Q1 2021 TRANSCRIPT
https://s25.q4cdn.com/442043304/files/doc_financials/2021/q1/Snap-Inc.-Q1-2021-Transcript.pdf

Bier et al. (1993)
Eric A. Bier, Maureen C. Stone, Ken Pier, William Buxton, Tony D. DeRose
1993
Toolglass and Magic Lenses: The See-Through Interface
https://www.billbuxton.com/tgml93.html

Bitgood (1994)
Stephen Bitgood
1994
Designing Effective Exhibits: Criteria for Success,Exhibit Design Approaches, and Research Stat-egies
https://www.informalscience.org/sites/default/files/Designing_Effective_Exhibits.pdf

Bolt, Herranz (1992)
Richard A. Bolt, Edward Herranz
1992
Two-handed gesture in multi-modal natural dialog
https://dl.acm.org/doi/abs/10.1145/142621.142623

Bolt et al. (1982)
Richard Bolt, Chris Schmandt, Eric A. Hulteen
1982
Put That There
https://www.media.mit.edu/publications/put-that-there-voice-and-gesture-at-the-graphics-in-terface/

Brignull, Rogers (2003)
Harry Brignull, Yvonne Rogers
2003
Enticing People to Interact with Large Public Displays in Public Spaces
https://rauterberg.employee.id.tue.nl/conferences/INTERACT2003/INTERACT2003-p17.pdf

Buxton, Billinghurst (2018)
Bill Buxton, Mark Billinghurst
2018
Gesture based Interaction
http://www.billbuxton.com/input14.Gesture.pdf

C

Campos, Dória (2009)
Pedro Campos, André Dória
2009
Interactivity for Museums: Designing and Comparing Sensor-Based Installations
https://www.researchgate.net/publication/221053740_Interactivity_for_Museums_Designing_and_Comparing_Sensor-Based_Installations

Carliner (2001)
Saul Carliner
2001
Modeling Information forThree-dimensional Space:Lessons Learned fromMuseum Exhibit De-sign
http://core.ecu.edu/engl/henzeb/7701s06/ftp/ethnog2.pdf

Carrozzino, Bergamasco (2010)
Marcello Carrozzino, Massimo Bergamasco
2010
Beyond Viritual Museums: Experiencing immersive virtual reality in real musems
https://www.sciencedirect.com/science/article/pii/S1296207410000543

Chee So et al. (2011)
Wing Chee So , Colin Sim Chen-Hui & Julie Low Wei-Shan
2011
Mnemonic effect of iconic gesture and beat gesture in adults and children: Is meaning in gesture important for memory recall?
https://www.tandfonline.com/doi/full/10.1080/01690965.2011.573220?src=recsys

D

Dam et al. (2013)
P. Dam, P. Braz, A. Raposo
2013
A study of navigation and selection techniques in virtual environments using Microsoft Kinect
https://www.researchgate.net/publication/292784043_A_Study_of_Navigation_and_Selection_Techniques_in_Virtual_Environments_Using_Microsoft_KinectR

Freeman et al. (2009)
Dustin Freeman, Hrvoje Benko, Meredith Ringel Morris, Daniel Wigdor
2009
ShadowGuides: Visualizations for In-Situ Learning of Multi-Touch and Whole-Hand Gestures
https://dl.acm.org/doi/10.1145/1731903.1731935

G

Garzotto, Rizzo (2007)
Franca Garzotto, Francesca Rizzo
2007
Interaction paradigms in technology-enhanced social spaces: A case study in museums
https://www.researchgate.net/publication/221234257_Interaction_paradigms_in_technology-enhanced_social_spaces_A_case_study_in_museums

H

Hammady et al. (2020)
Ramy Hammady, Minhua Ma, Carl Strathearn
2020
Ambient Information Visualisation and Visitors' Technology Acceptance of Mixed Reality in Museums
https://dl.acm.org/doi/10.1145/3359590

Harwood (2018)
Fiona Harwood
2018
Wizard of Oz testing – a method of testing a system that does not yet exist.
https://www.simpleusability.com/inspiration/2018/08/wizard-of-oz-testing-a-method-of-testing-a-system-that-does-not-yet-exist/

Heinrichs, Carpendale (2011)
Hinrichs, U., & Carpendale, S.
2011
Gestures in the wild: Studying multi-touch gesture sequences on interactive tabletop exhibits.
https://doi.org/10.1145/1978942.1979391

Hofmeester, Wolfe (2012)
Kay Hofmeester, Jennifer Wolfe
2012
Self-revealing gestures: teaching new touch interactions in windows 8
https://dl.acm.org/doi/10.1145/2212776.2212855

Hornecker, Stifter (2006)
Eva Hornecker, Matthias Stifter
2006
Learning from interactive museum installations about interaction design for public settings
https://www.researchgate.net/publication/221332107_Learning_from_interactive_museum_installations_about_interaction_design_for_public_settings

Huffpost (2013)
N/A
2013
Steven Spielberg's 'Robopocalypse' Delayed Indefinitely
https://www.huffpost.com/entry/steven-spielberg-robopocalypse_n_2443894

K

Karam, Schraefel (2005)
Maria Karam and m.c. schraefel
2005
A taxonomy of Gestures in Human Computer Interaction
https://eprints.soton.ac.uk/261149/1/GestureTaxonomyJuly21.pdf

Keates, Robinson (1998)
Simeon Keates, Peter Robinson
1998
The use of gestures in multimodal input
https://dl.acm.org/doi/10.1145/274497.274505

Kendon (2004)
Adam Kendon
2004
Gesture Visible action as Utterance
https://books.google.no/books/about/Gesture.html?id=hDXnnzmDkOkC&redir_esc=y

Koutsabasis, Domouzis (2016)
Panayiotis Koutsabasis, Chris K. Domouzis
2016
Mid-Air Browsing and Selection in Image Collections
https://dl.acm.org/doi/10.1145/2909132.2909248

L

Lee et al. (2012)
Bongshin Lee, Petra Isenberg, Nathalie Henry Riche, Sheelagh Carpendale
2012
Beyond Mouse and Keyboard: Expanding Design Considerations for Information Visualization Interactions
https://www.researchgate.net/publication/260582987_Beyond_Mouse_and_Keyboard_Expanding_Design_Considerations_for_Information_Visualization_Interactions

Luo et al. (2020)
Xuan Luo, Jia-Bin Huang, Richard Szeliski, Kevin Matzen, Johannes Kopf
2020
Consistent Video Depth Estimation
https://research.fb.com/wp-content/uploads/2021/02/Consistent-Video-Depth-Estimation.pdf

M

Maida (2016)
Jesse Maida
2016
Global Virtual Reality (VR) Content Market to Experience Exponential Growth Through 2020, With a CAGR of Approximately 128%: Technavio
https://www.businesswire.com/news/home/20161107006321/en/Global-Virtual-Reality-VR-Content-Market-to-Experience-Exponential-Growth-Through-2020-With-a-CAGR-of-Approximately-128-Technavio.

Malerczyk (2004)
Cornelius Malerczyk
2004
Interactive Museum Exhibit Using Pointing Gesture Recognition
https://www.researchgate.net/publication/221546379_Interactive_Museum_Exhibit_Using_Pointing_Gesture_Recognition

Manghisi et al. (2018)
Vito M. Manghisi, Antonio E. Uva, Michele Fiorentino, Michele Gattullo,Antonio Boccaccio, Giuseppe Monno
2018
Enhancing user engagement through the user centric design of a mid-air gesture-based interface for the navigation of virtual-tours in cultural heritage expositions
https://www.sciencedirect.com/science/article/pii/S1296207417307549

Michelis, Resatsche (2007)
Michelis, D., & Resatsch, F.
2007
Alice Through the Interface Electronic Mirrors as Human-Computer-Interface.
https://doi.org/10.1007/978-3-540-71025-7_7

Moeslund et. al (2001)
Thomas B. Moeslund, Moritz Störring, Erik Granum
2001
A Natural Interface to a Virtual Environment through Computer Vision-Estimated Pointing Gestures
https://dl.acm.org/doi/10.5555/647592.728876

Morris et al. (2010)
Morris, Wobbrock, Wilson,
2010
Understanding users' preferences for surface gestures
https://dl.acm.org/doi/pdf/10.5555/1839214.1839260

Muller et al. (2010)
Muller, J., Alt, F., Michelis, D., and Schmidt, A.
2010
Requirements and design space for interactive public displays.
https://dl.acm.org/doi/10.1145/1873951.1874203

Murakami, Taguchi (1991)
Kouichi Murakami, Hitomi Taguchi
1991
Gesture recognition using recurrent neural networks
https://dl.acm.org/doi/10.1145/108844.108900

N

Nielsen, Molich (1990)
Jakob Nielsen, Rolf Molich
1990
Heuristic evaluation of user interfaces
https://dl.acm.org/doi/10.1145/97243.97281

Nielsen et al. (2003)
M. Nielsen, M. Störring, T. Moeslund, E. Granum
2003
A procedure for developing intuitive and ergonomic gesture interfaces for HCI
https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.62.4970&rep=rep1&type=pdf

P

Pollalis et al. (2017)
Christina Pollalis, Whitney Fahnbulleh, Jordan Tynes, Orit Shaer
2017
HoloMuse: Enhancing Engagement with Archaeological Artifacts through Gesture-Based Interaction with Holograms
https://dl.acm.org/doi/10.1145/3024969.3025094

Pollalis et al. (2017)
Christina Pollalis, Whitney Fahnbulleh, Jordan Tynes, Orit Shaer
2017
HoloMuse: Enhancing Engagement with Archaeological Artifacts through Gesture-Based Interaction with Holograms
https://dl.acm.org/doi/10.1145/3024969.3025094

Premaratne (2014)
Premaratne, P.
2014
Human Computer Interaction Using Hand Gestures.
https://doi.org/10.1007/978-981-4585-69-9

Q

Quek et al. (2002)
Francis Quek, David McNeill, Robert Bryll, Susan Duncan, Xin-Feng Ma, Cemil Kirbas, Karl E. Mccullough, Rashid Ansari
2002
Multimodal Human Discourse: Gesture and Speech
https://www.researchgate.net/publication/220286223_Multimodal_human_discourse_gesture_and_speech

R

Rogers (2011)
Yvonne Rogers
2011
Interaction design gone wild: striving for wild theory
https://dl.acm.org/doi/10.1145/1978822.1978834

S

Salamin et al. (2006)
Patrick Salamin, Daniel Thalmann, Frédéric Vexo
2006
The benefits of third-person perspective in virtual and augmented reality?
https://dl.acm.org/doi/10.1145/1180495.1180502

Sathiyanarayanan et al. (2014)
Mithileysh Sathiyanarayanan, Syed Azharuddin, Santhosh Kumar, Gibran Khan
2014
GESTURE CONTROLLED ROBOT FOR MILITARY PURPOSE
https://www.researchgate.net/publication/265425730_GESTURE_CONTROLLED_ROBOT_FOR_MILITARY_PURPOSE

Schönböck et al. (2008)
Johannes Schönböck, Florian König, Gabriele Kotsis, Dominik Gruber, Emre Zaim, Albrecht Schmidt
2008
MirrorBoard - An Interactive Billboard.
https://www.researchgate.net/publication/221439471_MirrorBoard_-_An_Interactive_Billboard

Serrell (2010)
Beverly Serrell
2010
Paying Attention: The Duration and Allocation of Visitors' Time in Museum Exhibitions
https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2151-6952.1997.tb01292.x

Shah, Ghazali (2018)
Nurul Fathihin Mohd Noor Shah, Masitah Ghazali
2018
A Systematic Review on Digital Technology for Enhancing User Experience in Museums
https://link.springer.com/chapter/10.1007/978-981-13-1628-9_4

Shneiderman, Maes (1997)
Ben Shneiderman, Pattie Maes
Direct Manipulation vs Interface Agents
1997
https://www.lri.fr/~mbl/ENS/FONDIHM/2013/papers/ShneidermanMaes-Interactions97.pdf

Shuhaili et al. (2017)
Fasihah Mohammad Shuhaili, Suziah Sulaiman, Mahamad Saipunidzam, Aliza Sarlan
2017
Visual Object Interface Signifier of Museum Application for Large Display
https://link.springer.com/chapter/10.1007/978-3-319-70010-6_54

Steed et al. (2016)
Steed, A., Pan, Y., Zisch, F., & Steptoe, W.
2016
The impact of a self-avatar on cognitive load in immersive virtual reality.
https://doi.org/10.1109/VR.2016.7504689

T

Theil, Hwang (2015)
Arthur Theil, Faustina Hwang
2015
An analysis of mid-air gestures used across three platforms
https://www.researchgate.net/publication/280134172_An_analysis_of_mid-air_gestures_used_across_three_platforms

Thórisson (1996)
Kristinn R. Thórisson
1996
Face-To-Face Communication With Computer Agents
https://www.researchgate.net/publication/2683031_Face-To-Face_Communication_With_Computer_Agents

Trajkova et al. (2020)
Milka Trajkova, A'aeshah Alhakamy, Francesco Cafaro, Rashmi Mallappa, Sreekanth R. Kankara
2020
Move Your Body: Engaging Museum Visitors with Human-Data Interaction
https://dl.acm.org/doi/abs/10.1145/3313831.3376186

V

Vafaei (2013)
Fereydoon Vafaei
2013
Taxonomy of Gestures in Human Computer Interaction
https://www.researchgate.net/publication/321058123_TAXONOMY_OF_GESTURES_IN_HUMAN_COMPUTER_INTERACTION

Vaz et al. (2018)
Roberto Vaz, Paula Odete Fernandes, Ana Cecilia Nascimento Rocha Veiga
2018
Interactive Technologies in Museums: How Digital Installations and Media Are Enhancing the Visitors' Experience
https://www.researchgate.net/publication/321490808_Interactive_Technologies_in_Museums_How_Digital_Installations_and_Media_Are_Enhancing_the_Visitors%27_Experience

Venkatnarayan et al. (2021)
Venkatnarayan, R. H., Mahmood, S., & Shahzad, M.
2021
WiFi based Multi-User Gesture Recognition
https://doi.org/10.1109/TMC.2019.2954891

Villarreal-Navaraez et al. (2020)
Santiago Villarreal-Narvaez, Jean Vanderdonckt, Radu-Daniel Vatavu, Jacob O. Wobbrock
2020
A Systematic Review of Gesture Elicitation Studies:What Can We Learn from 216 Studies?
https://faculty.washington.edu/wobbrock/pubs/dis-20.01.pdf

Vogiatzidakis, Koutsabasis (2018)
Vogiatzidakis, P., & Koutsabasis, P.
2018
Gesture Elicitation Studies for Mid-Air Interaction: A Review.
https://doi.org/10.3390/mti2040065

W

Walter et al. (2013)
Robert Walter, Gilles Bailly, Joerg Mueller
2013
StrikeAPose: Revealing Mid-Air Gestures on Public Displays
https://dl.acm.org/doi/10.1145/2470654.2470774

Walter et al. (2014)
Robert Walter, Gilles Bailly, Nina Valkanova, Jörg Müller
2014
Cuenesics: using mid-air gestures to select items on interactive public displays
https://dl.acm.org/doi/10.1145/2628363.2628368

Wang et al. (2020)
Tzu-Yang Wang, Yuji Sato, Mai Otsuki, Hideaki Kuzuoka, Yusuke Suzuki
2020
Effect of Body Representation Level of an Avatar onQuality of AR-Based Remote Instruction
https://www.mdpi.com/2414-4088/4/1/3

Wexelblat (1995)
Alan Wexelblat
1995
An Approach to Natural Gesture in Virtual Environments.
https://www.researchgate.net/publication/220286385_An_Approach_to_Natural_Gesture_in_Virtual_Environments

Wexelblat (1997)
Alan Wexelblat
1997
Research Challenges in Gesture: Open Issues and Unsolved Problems
https://link.springer.com/content/pdf/10.1007/BFb0052984.pdf

Wipfli et al. (2016)
Rolf Wipfli, Victor Dubois-Ferrière, Sylvain Budry, Pierre Hoffmeyer, Christian Lovis
2016
Gesture-Controlled Image Management for Operating Room: A Randomized Crossover Study to Compare Interaction Using Gestures, Mouse, and Third Person Relaying
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4833285/

Wittorf, Jakobsen (2016)
Wittorf, M. L., & Jakobsen, M. R.
2016
Eliciting Mid-Air Gestures for Wall-Display Interaction.
https://doi.org/10.1145/2971485.2971503

Worts (2006)
Douglas Worts
2006
Measuring Museum Meaning: A Critical Assessment Framework
https://www.jstor.org/stable/40283905?seq=1#metadata_info_tab_contents

LIST OF FIGURES