



**Cartography M.Sc.**

**Master thesis**

# **Developing Gaze-based Map Interactions in Mixed Reality Devices**

Kurumbayeva Nargiz



2021

# **Developing Gaze-based Map Interactions in Mixed Reality Devices**

submitted for the academic degree of Master of Science (M.Sc.)  
conducted at the Department of Aerospace and Geodesy  
Technical University of Munich

Author: Nargiz, Kurumbayeva  
Study course: Cartography M.Sc.  
Supervisor: Dr.-Ing. Christian Murphy (TUM)  
Reviewer: M.Sc. Wangshu Wang (TUW)

Chair of the Thesis  
Assessment Board: Prof. Dr. Liqiu Meng

Date of submission: 18.11.2021

## **Statement of Authorship**

Herewith I declare that I am the sole author of the submitted Master's thesis entitled:

"Developing Gaze-based Map Interactions in Mixed Reality Devices"

I have fully referenced the ideas and work of others, whether published or unpublished. Literal or analogous citations are clearly marked as such.

Munich, 18.11.2021

Nargiz Kurumbayeva

## Acknowledgements

*I am grateful to be given such a life-changing opportunity to be a part of the Cartography Master's programme, and I would like to express my gratitude to all the people who participated in the very important stage of my life.*

*To Juliane Cron for always being the supporting and caring coordinator and friend,*

*To Christian Murphy for always encouraging and motivating during all the thesis stages,*

*To Wangshu Wang for useful feedback and interesting questions,*

*To my new CartoFamily for all the lovely and warm moments,*

*To my closest friends for being the most supportive and the coolest people,*

*And to my family for everything,*

*I say thank you!*

# Abstract

This research features the development and evaluation of the gaze-based cartographic interactions to facilitate the user-map experience in the Mixed Reality environment. The interactions are chosen from the fundamental cartographic, gaze-based, and Mixed Reality interactions. The implemented interactions are: retrieve, overlay, and rotate. The Mixed Reality application is assembled to present the gaze-based, gaze-aware, and conventional interfaces that feature the selected interactions. The gaze-based interface uses the gaze to control the interactions. The gaze-aware interface uses gaze and voice modalities, and the conventional interface is hands-controlled. The experiment and survey are conducted for the evaluation of the assembled interfaces. The performance and user experience of the implemented interactions and interfaces are evaluated.

Keywords: eye-tracking, gaze-based interactions, cartographic interactions, mixed reality

# Table of contents

Acknowledgements .....	i
Abstract.....	ii
Table of contents .....	iii
List of figures .....	v
List of tables .....	vii
List of abbreviations.....	viii
1. Introduction.....	1
1.1. Motivation and problem statement.....	1
1.2. Research identification .....	2
1.3. Thesis structure .....	3
2. Theoretical background.....	4
2.1. User-map interactions .....	4
2.1.1. Definition .....	4
2.1.2. Taxonomy.....	4
2.2. Eye-tracking.....	7
2.2.1. Eye and eye movements.....	7
2.2.2. Eye-tracking techniques and devices .....	10
2.2.3. Gaze interaction .....	14
2.2.4. Gaze-based interactions in cartography.....	18
2.3. Mixed reality .....	19
2.3.1. Definition .....	19
2.3.2. Interactions .....	20
2.3.3. MR devices .....	23
3. Methodology.....	25

3.1. Research approach .....	25
3.2. Case study .....	27
3.2.1. Selected interactions .....	27
3.2.2. Resources .....	27
3.2.3. Implementation .....	33
3.2.4. User study setup.....	43
3.3. Data analysis .....	45
4. Results and discussion .....	46
4.1. Performance.....	46
4.2. User experience .....	49
4.2.1. User Experience Questionnaire .....	49
4.2.2. Task load .....	52
4.3. Interface ranking .....	55
4.4. Further statements .....	56
4.5. Challenges and limitations .....	59
5. Conclusion and outlook.....	62
Bibliography .....	64
Appendices .....	73
Appendix 1: Pre-study questionnaire .....	73
Appendix 2: Post-study questionnaire .....	75

## List of figures

Figure 1: The eye.....	7
Figure 2: The human eye muscles.....	8
Figure 3: Relationship between saccades and smooth pursuits.....	9
Figure 4: EOG data collection .....	11
Figure 5: The scleral contact lens with embedded search coil and attached electromagnetic field frame for eye movement measuring. ....	11
Figure 6: Deduction of gaze direction.....	12
Figure 7: Eye-tracking setup .....	13
Figure 8: Remote head-boxed eye-tracking system .....	14
Figure 9: The EyeWrite gesture alphabet .....	16
Figure 10: Examples of eye movement recordings .....	17
Figure 11: The virtuality continuum.....	20
Figure 12: Visual-based modality and its contexts and interaction methods. ....	22
Figure 13: Audio-based, haptic-based, and sensor-based modalities, contexts, and interaction methods .....	23
Figure 14: Optical see-through and see-through video devices .....	24
Figure 15: The research approach structure. Abbr. ....	25
Figure 16: The city model for the retrieve and overlay interactions .....	28
Figure 17: The terrain model for the rotation interaction. ....	28
Figure 18: HoloLens 2, top view .....	29
Figure 19: HoloLens 2 cameras setup (front view).....	31
Figure 20: Software utilised for the application development.....	32
Figure 21: Schematic workflow of the application development .....	35
Figure 22: The root scene and the main menu .....	36



Figure 23: The gaze-based interface scene .....	37
Figure 24: Box collider for the interactions with the terrain .....	37
Figure 25: Overlay interaction: points of interest activated.....	38
Figure 26: Retrieve interaction triggered for the building .....	39
Figure 27: Scene for the eyes-voice controlled interactions.....	39
Figure 28:The second interface - overlay interaction: 1) activated hotels, .....	40
Figure 29: The scene of hands-controlled interactions.....	41
Figure 30 Hands-controlling interface .....	42
Figure 31 Hands-controlled interface: air-press, overlay.....	42
Figure 32 The retrieve interaction of the hands-controlled interface: using hand-ray .....	43
Figure 33: Violin-boxplot visualisation of the tasks' completion time in regards to interfaces .....	48
Figure 34: Visualisation of UEQa results' mean values.....	51
Figure 35: Task Load results: mental demand, physical demand, and frustration ....	53
Figure 36: Task load results: users evaluate their performance; the most, moderately, and the least successful .....	53
Figure 37: Task load results: the effort spent to accomplish the user's level of performance .....	54
Figure 38: Results of the Interface Ranking questionnaire .....	55
Figure 39: Eye-cursor responsiveness evaluated by participants without/with glasses or contact lenses:.....	56
Figure 40: Difficulties that users mentioned in the open-ended questions: .....	57
Figure 41: User suggestions from open-ended questions.....	59

## List of tables

Table 1: The description of cartographic interaction operator primitives .....	6
Table 2: Overview by Mollenbach (2013) of eye interactions .....	18
Table 3: The conceptual framework of the MR notions .....	20
Table 4: HoloLens 2 technical specifications. ....	29
Table 5: The three developed interfaces and their characteristics .....	34
Table 6: Overlay, retrieve, and rotate tasks' completion time in seconds.....	47

## List of abbreviations

AR	Augmented reality
AV	Augmented virtuality
MR	Mixed reality
NASA TLX	National Aeronautics and Space Administration Task Load Index questionnaire
UEQ	User Experience Questionnaire
VR	Virtual reality
POR	Point of regard
EOG	Electro-oculography
GDO	Graphic display object
UX	User experience
HCI	Human-computer interaction
3D	Three-dimensional
2D	Two-dimensional
OSM	Open Street Map
SoC	System on the chip
CPU	Central processing units
GPU	Graphical processing units
HPU	Holographic processing unit
SDK	Software development kit
MRTK	Mixed Reality Toolkit
IDE	Integrated development environment



# 1. Introduction

## 1.1. Motivation and problem statement

Eye-tracking is becoming pervasive in various research fields. Its implementation ranges from biomedical engineering (Cho & Kim, 2013) to smart television and wearables such as smartwatches and smart glasses (Mardanbegi et al., 2016). The eye-tracking technology allows recording the position and movements of a person's gaze for analysis and interaction purposes. In the cartographic field, passive analysis of eye-tracking data has been used for evaluating interactive map designs by their effectiveness and efficiency (Coltekin et al., 2010; STEINKE, 1987). Kiefer and Giannopoulos's (2012) visualisation of gaze history on mobile display for easier orientation on small display maps is a case of active analysis of gaze data. Gaze-based interactions have been implemented for zooming and panning in a natural and fast way (Stellmach & Dachsel, 2012). In only a few approaches, the interaction with cartographic interfaces has utilised gaze as an input (Giannopoulos et al., 2015, 2013). Previous research has shown that gaze-supported interaction can contribute to cartographic applications (Serrao, 2020).

Gaze-based interactions have been applied for different hardware, such as desktop, public displays (Khamis et al., 2017), virtual reality glasses, etc. Eye-tracking is entering the mass market. Microsoft's new mixed reality device, the HoloLens 2, has a built-in eye-tracker that allows access to single eye gaze rays at 30 Hz via an eye-tracking API. Schweigert et al. (2019) suggest that the combination of eye-gaze referencing and hand-pointing gesture ("EyePointing" method) can replace the mid-air tap gesture of HoloLens. However, research is missing on how alternative human-computer inputs like hand gestures in a mixed reality environment can be substituted by eye-tracking for user-map interactions. Its development can help with offering unusual possibilities for differently-abled people and proposing new opportunities in human-computer interaction (Piotrowski & Nowosielski, 2020). Implementing user-map interactions controlled by gaze can allow for an easy, natural, and fast way of interacting in MR devices.

## 1.2. Research identification

### *Research objectives*

This research has the main objective of developing gaze-based interactions to facilitate user-map interaction in the MR environment. The main objective consists of the following sub-objectives:

1. Identify and determine cartographic interactions for the gaze control in the MR environment
2. Assemble a subset of MR interfaces for the selected interactions
3. Evaluate the performance and user experience of assembled interfaces

### *Research questions*

1. How can users interact with the map in the MR environment?
  - a) What are the fundamental cartographic interactions?
  - b) How can eyes control interactions?
  - c) What are the fundamental interactions in the MR environment?
2. How are MR interfaces assembled for the selected gaze-based interactions?
  - a) What are the limitations of the developed gaze-based interfaces?
  - b) What are the challenges in the development of the selected gaze-based interfaces?
3. How effective are the implemented user-map interactions in MR?
  - a) What is the performance of the assembled interfaces?
  - b) What is the user experience with the implemented user-map interactions in the MR environment?

### *Delimitations*

The focus of this research is not the passive analysis of eye-tracking data. Instead, this research focuses on using eye-tracking data as an input for the interactions. Furthermore, this research does not have a goal of optimising existing gaze-based user-map interactions. Instead, gaze-based user-map interactions are developed for the MR environment and evaluated in this thesis.

### *Innovation*

The novelty of the work lies in generating user-map interactions for mixed reality devices with gaze-based control. The research will possibly help guide future gaze-

based interactions implementation and optimisation of gaze-based interactions within the domain of cartography.

### 1.3. Thesis structure

Chapter 1 states the motivation and identifies the research objectives and questions of developing gaze-based cartographic interactions for the mixed reality devices.

Chapter 2 familiarises the reader with the theoretical background of the research and addresses the research questions of the first objective. The subchapters reflect three main aspects of the work: cartographic interaction, eye-tracking, and MR.

Chapter 3 describes the methodology of developing and evaluating gaze-based user-map interactions for MR devices. The case study defines the implementation matter of the selected methodology.

Chapter 4 presents and discusses the findings of the user study for gaze-based user-map interactions in the MR environment. The questions of the second and third research objectives are answered here.

Chapter 5 concludes the paper by outlining the main results, the limitations, and future research of gaze-based map interactions in MR.

## 2. Theoretical background

### 2.1. User-map interactions

#### 2.1.1. Definition

Cartographic visualisation being a visualisation of information implicates representation as well as interaction (Roth, 2011; Yi et al., 2007). After Google maps' release in 2005, the accessibility and utilisation of interactive maps for expert and general use have drastically increased (Miller, 2006). A map is considered interactive if it can respond to the user's manipulation by changing itself in some way. This reciprocal action constitutes the user map interaction and possibly impacts the user's comprehension of the visualised phenomena (Roth, 2011). One example of a simple interaction would be the user clicking on some object on the map to retrieve additional information about it and the map responding with a small text window appearing near the object.

User map interactions, as an integral part of map use, influence the user experience and are supported by an interface (Yi et al., 2007). Before proceeding to the next parts of this research, we need to differentiate between interaction and interface. Interaction is a bilateral action of a human and a map mediated by a computing device, whereas the interface is an individual or a set of digital components that enables the user input, e.g., buttons, menu, and others.

#### 2.1.2. Taxonomy

Taxonomies of interaction presented below, regardless of having some common units, notably differ in granularity (Yi et al., 2007) and have diverse natures. Roth explains the diversity through Norman's stages of (inter)action (Norman, 1988, pp. 45–52; Roth, 2012). Norman (1988), in his book "The work of everyday things", presents the action model, which provides an insight into the performance of cartographic interaction. The action model consists of seven stages that describe the human's way of interacting with objects, both physical and virtual:

- 1) forming the goal
- 2) forming the intention
- 3) specifying the action
- 4) executing the action



- 5) perceiving the state of the system
- 6) interpreting the state of the system
- 7) evaluating the outcome

Existent taxonomies, according to Roth (2012), commonly abide by one of the following approaches: an objective-based approach, an operator-based approach, and an operand-based approach – each of which aligns with the interaction stage. Roth (2012) summarised them as follows:

Objective-based approaches correspond to Norman's forming intention stage and distinguish interactions according to the user "tasks" (Amar et al., 2005; Crampton, 2002; Zhou & Feiner, 1998), or, in alternative terminology, "operations" (MacEachren et al., 1999; Wehrend & Lewis, 1990), and "intentions" (Yi et al., 2007). Identify and compare are the most common primitives within objective taxonomies.

Operator-based approaches align with Norman's specifying action stage and categorise interactions according to interface functionality, or, in alternative terminology, visualisation "manipulations" (Buja et al., 1996) and "interaction techniques" (Keim, 2002). Some of the common operator primitives are focusing, linking, and brushing (Roth, 2012).

Operand-based approaches categorise interaction according to visualisation content: the type (Andrienko et al., 2003; Keim, 2002; Peuquet, 1994; Shneiderman, 2003) or the state (Chi, 2000; Chuah & Roth, 1996; Haber & McNabb, 1990) of information represented on the map.

Roth's operator taxonomy of cartographic interaction primitives (Roth, 2013) consists of working and enabling operators. User objectives are realised through interactions, which are defined as working operators: reexpress, arrange, sequence, symbolise, overlay, reproject, pan, zoom, filter, search, retrieve, and calculate. In comparison, enabling operators are necessary to support and allow for working operators. Enabling operators include import, export, save, edit, and annotate. Table 1 presents Roth's cartographic interaction operator primitives adapted by Tolochko (2016).

*Table 1: The description of cartographic interaction operator primitives adapted from Roth by Tolochko (Tolochko, 2016)*

	Interaction operator primitive	Description
<b>Work Operators</b>	Pan	Changes the geographic center of the map; adjusts the part of the map that is in the current view, since part of the map is off the screen
	Zoom	Changes the scale and/or resolution of the map; “zoom in” commonly refers to changing from a smaller to larger scale, while “zoom out” refers to changing from a larger to a smaller scale. Zoom can also describe a change in map detail without a change in map scale.
	Retrieve	Requests details about a particular map feature or features of interest, usually through direct manipulation (e.g. clicking on the feature).
	Filter	Identifies features or places on the map that meet one or several conditions, defined by the user. Can be confused with Search, see below for clarification.
	Search	Identifies a specific place or feature of interest on the map. Similar to Filter (see above), with the difference that Search identifies a specific feature, while Filter produces multiple results that match specific characteristic(s).
	Overlay	Adds or removes features in the current map view (e.g. toggle layer visibility).
	Reproject	Changes the map projection.
	Resymbolize	Changes the design of a map, but does not change the map type itself (e.g. a change in color scheme for a choropleth map, while still using the choropleth map type).
	Reexpress	Changes the map representation type (e.g. changes from choropleth to proportional symbol).
	Arrange	Changes the layout of different views in a linked visualization.
	Sequence	Creates a set of related maps that are placed in a particular order (e.g. small multiples showing change over time).
	Calculate	Computes new information about map features (e.g. calculates new statistics).
<b>Enabling Operators</b>	Import	Loads a new dataset or map to the current map view.
	Export	Pulls out geographic information or a map created by the map interface to be used in different map setting or interface.
	Save	Conserves the current state of the map, including its associated geographic information and/or the current system status.
	Edit	Alters the underlying geographic or attribute information of the map.
	Annotate	Allows the user to add text or graphics to the map interface.

Cartographic interaction has been defined by Roth (2012) as a user-map dialogue. This metaphor implies a human “asking” a question by initiating interaction (e.g., clicking a button) and a map “responding” by adding or changing content. In this research, the above-mentioned taxonomy of interaction primitives will be used to enable and confine the “dialogue” between user and map we use.

## 2.2. Eye-tracking

### 2.2.1. Eye and eye movements

To better understand eye movements and eye-tracking techniques, firstly, we need a brief survey of the eye composition and basic eye movements. Figure 1 schematically shows the eye by a section. The pupil lets the light in, and the image turned upside-down by the lens is projected onto the retina. The retina contains cells sensitive to light: cones that allow us to have colour vision and rods that support our vision during dim lighting. Cones are primarily concentrated in the area on the retina called the fovea. The cones concentration results in the full acuity only on the fovea, which takes less than 2 degrees of the visual field and provides a foveal vision. When we look at an object to see it sharply, the eye is moved in the direction of the light from the object, so the light falls exactly on the fovea. This process is called foveating. (Holmqvist et al., 2011)

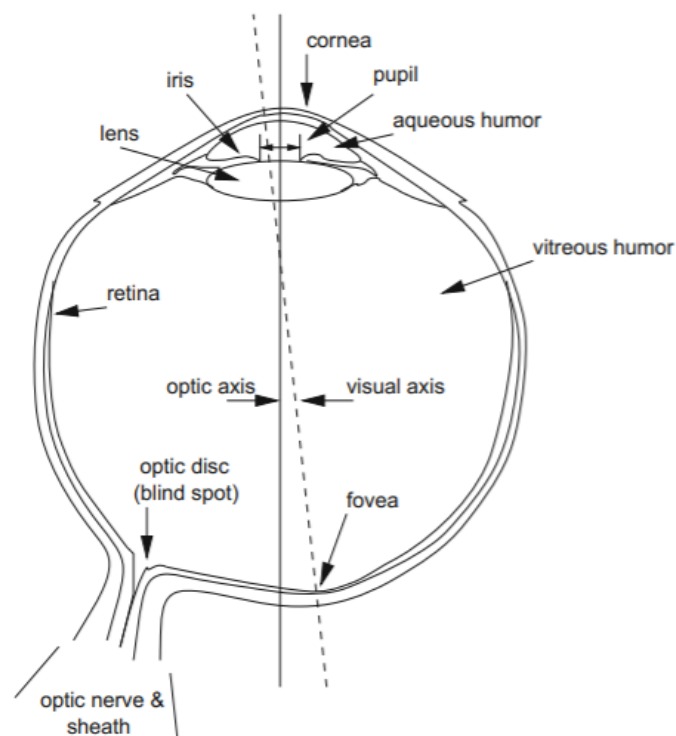


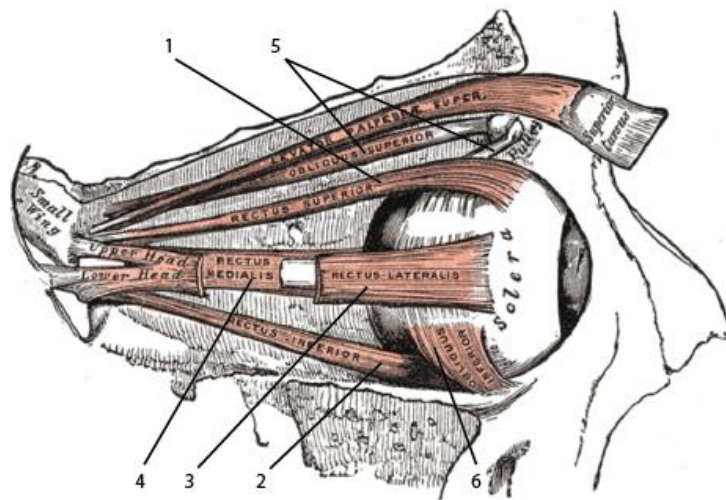
Figure 1: The eye.

Adapted by Duchowski & Duchowski (2017) from Cornsweet (2012)

The gaze pupil and cornea are essential to measuring eye movements (pupil-and-corneal reflection technique). The cornea is the outer layer of the eye that reflects light. Thus, the visible reflection on the eye is a corneal reflection (Holmqvist et al., 2011).

The corneal reflection and the techniques for eye tracking are described in more detail in the “Eye-tracking techniques and devices” subchapter below.

An eye uses three pairs of muscles to allow movements, the scheme of which you can see in Figure 2. The vertical movements are controlled by the superior (2) and inferior (3) recti (Gray, 1918). The medial (4) and lateral (5) recti control the horizontal movements (Gray, 1918). The twist movement (the torsional rotating) of the eye requires the superior (7) and inferior (8) obliques (Davson, 1990).

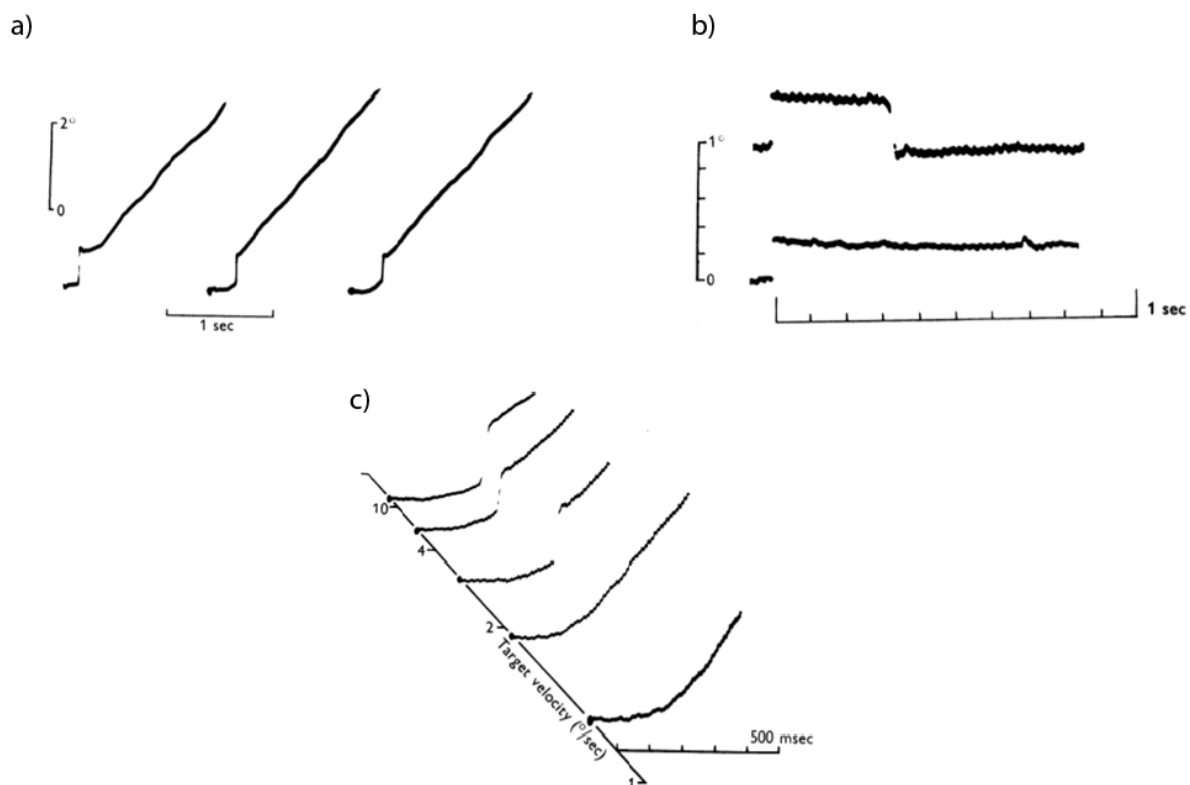


*Figure 2: The human eye muscles  
the pair (1)-(2) - up-down movements; the pair (3)-(4) - right-left movements; the pair (5)-(6) – the torsional rotation (twist).  
Adapted from (Gray, 1918)*

Eye movements have non-positional aspects to them, such as adaptation, accommodation, vergence and others. However, to obtain the information about the user’s visual attention, only positional eye movements are of primary importance, because assumably, they indicate evident visual attention (Duchowski & Duchowski, 2017). There are three types of eye movements considered to be the primary requirement for eye movement analysis: saccades, smooth pursuits, and fixations.

A saccade is a rapid movement of the eye from one location to another. These movements are performed when changing attention from one object to another. During the saccades, we are effectively blind as the duration of these transitions is from 10 to 100 milliseconds, which is insufficient for image rendering (Shebilske & Fisher, 1983).

While following a moving object, eyes make two types of movements: brisk movements – saccades, and smooth movements filling the intervals between the saccades, called smooth pursuits (WESTHEIMER, 1954). Rashbass (1961) conducted a number of experiments that defined the relationship between saccades and smooth pursuits. According to Rashbass (1961), smooth pursuits are performed when tracking the moving object with a relatively small displacement and uniform velocity. In Figure 3, the results from some of the experiments are demonstrated.



*Figure 3: Relationship between saccades and smooth pursuits.*

*Records of horizontal eye direction (x-axes – time; y-axes – displacement) as a response to: a) an object moving horizontally with uniform velocity: initial saccadic movement and change to a smooth pursuit; b) sudden displacement of the object: lower record - no response to the displacement within a threshold, the upper record – saccadic response to the displacement higher than the threshold; c) to a change in the object velocity: saccadic response to a velocity gain is the more significant magnitude of the saccade with the higher velocity.*

*Adapted from Rashbass's (1961)*

The first graph (a) shows the initial saccadic response to the onset of the moving object with uniform velocity followed by smooth pursuit. The linear relation between a moving object and the smooth component allows eyes to match the object's velocity. Graph (b) demonstrates how the saccade is evoked by the sudden lateral displacement of the target. However, displacements less than the threshold get no saccadic response. On the last graph (c), the variations of the target velocity gain result in different

responses. With the more significant change in velocity, the magnitude of the saccadic movement is more significant (Rashbass, 1961). That means the faster the target moves, the more distinguishable the saccadic movement is on the graph.

The foremost event to be reported during eye tracking is, interestingly, not the movement but the static state of the eye (Richardson & Spivey, 2004). The condition when the eye stays still or fixates on something is called fixation. However, the eye during fixation is not exclusively motionless because it performs three types of micro-movements: tremor (nystagmus), micro-saccades, and drifts (Martinez-Conde et al., 2004). Without these miniature movements (for example, if we artificially stabilise the image on the retina), within about a second, the scene becomes blank, and we don't see anything. These components fluctuate around the fixation point and are oversimplified as a noise for eye-tracking modelling purposes. (Duchowski & Duchowski, 2017)

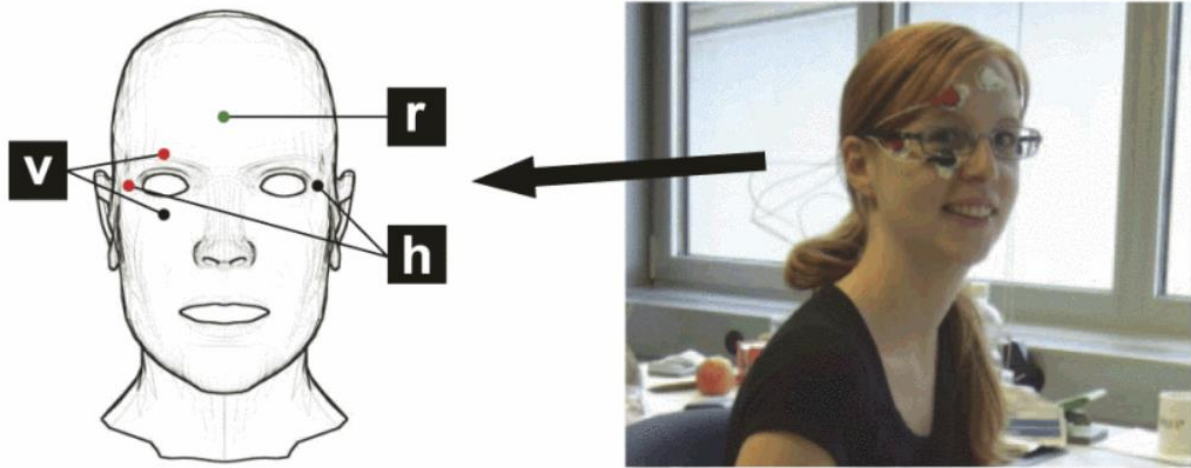
#### 2.2.2. Eye-tracking techniques and devices

Generally, eye trackers monitor eye movements by measuring either the eye position relative to the head or the eye orientation in the space - "point of regard (POR)" (Young & Sheena, 1975). The projection of POR defines the user gaze that is the desired output of eye-tracking. Different techniques to monitor eye movements form four categories and correspond to four generations of the apparatus required for measurements (Duchowski & Duchowski, 2017):

- 1) electro-oculography (EOG) and scleral contact lens/search coil (first-generation systems)
- 2) photo- and video-oculography (second generation systems)
- 3) analogue video-based measurement of pupil/corneal reflection (third generation systems)
- 4) digital video-based measurement of pupil/corneal reflection (fourth generation systems)

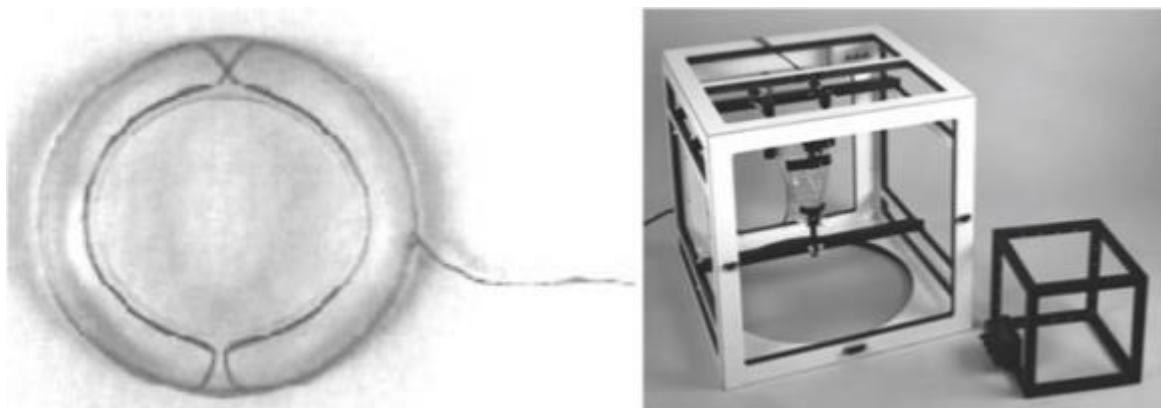
The early technique that was widely used in mid-70th is electro-oculography (Young & Sheena, 1975). The electro-oculography method records the difference in electric potential of the electrodes placed on the skin around the eye cavity. Vertical, horizontal, and reference electrodes are set as shown in Figure 4, recognising different eye movements relative to the head position; thus, generally, it does not provide POR

unless an additional head tracker is used to position the head (Duchowski & Duchowski, 2017). Figure 4 shows a subject for EOG testing from an eye movement analysis study by Bulling et al. (2011).



*Figure 4: EOG data collection  
horizontal (h), vertical (v), and reference (r) electrode placement (Bulling et al., 2011).*

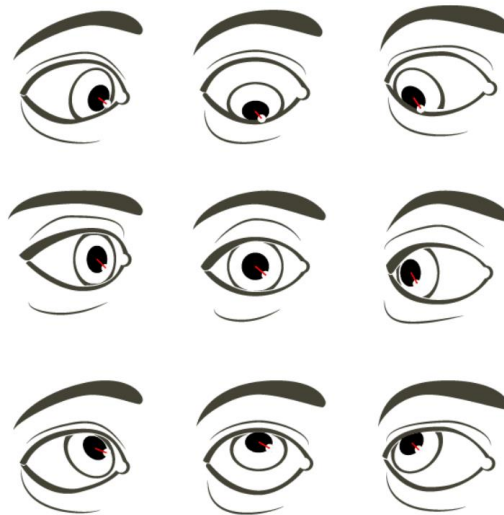
Another method that measures eye-in-head positions uses a scleral contact lens with an embedded search coil. The scleral contact lens is purposefully bigger than the standard contact lens to avoid slippage from the cornea. The search coil in the form of optical or mechanical devices (e.g., wire coil) attached to a stalk on the scleral contact lens is then measured for changes in the electromagnetic field to measure eye movements (Duchowski & Duchowski, 2017). The scleral contact lens with embedded search coil and attached electromagnetic field frames are shown in Figure 5. This method provides susceptible and precise measurements (Young & Sheena, 1975). However, it causes discomfort to the user because of its intrusive nature.



*Figure 5: The scleral contact lens with embedded search coil and attached electromagnetic field frame for eye movement measuring.*



Non-invasive methods are based on the measurement of eye features like a pupil and iris-sclera boundary (second generation) or direct light (usually infrared) reflection on the cornea (Figure 6). To differentiate between head movements and eye movements, two reference points are needed: the pupil centre and the corneal reflection. Crane (1994) names corneal reflections as the Purkinje reflections/images. Naturally, there are four reflections. The first Purkinje reflection is a corneal reflection and is the brightest one registered by video-based trackers. The positional difference between the first Purkinje reflection and the centre of the pupil stays relatively constant during slight head movements and changes during eye movements. This positional difference allows for the estimation of the gaze direction.



*Figure 6: Deduction of gaze direction.  
The first Purkinje reflection (small white circle) in close proximity to the pupil centre (What Is Eye Tracking and How Does It Work?, 2019)*

The second generation of eye trackers uses photo- or video-oculography techniques that require offline manual inspections of photo or video frames. They do not provide POR measurements because the eye measurements are relative to the head position. Whereas the third-generation systems measure POR coordinates using analogue video processors for combined pupil/corneal reflection and deliver the calculated gaze position in real-time. Another essential feature is self-calibration that excludes the necessity of tedious calibration routines to be performed by a researcher like for the older equipment (Duchowski & Duchowski, 2017).



The fourth generation of eye trackers has increased the accuracy of output, the usability of the device, and the speed of processing. This is achieved by utilising auto-focusing digital cameras and on-chip digital signal processors. The devices can be table-mounted and head-mounted, but typically, identical optics are built into these devices with differences in size (Duchowski & Duchowski, 2017).

Additionally, eye trackers can be categorised regarding their placement. Valtakari et al. (2021) referred to the relative placement of user and eye tracker as “eye-tracking setup”. They divided the eye-tracking setups into three groups: head-restricted setup, head-boxed setup, and head-free setup (Figure 7). Head-restricted and head-boxed setups both use remote eye trackers mounted on a table. The head-restricted setups use head stabilisers, such as chin rest or bite-bar, to constrain the user’s head movements (*Different Kinds of Eye Tracking Devices*, 2020). The reduction of head-movement artefacts results in more accurate data (Duchowski & Duchowski, 2017). However, it can limit the user’s natural interaction and level of comfort.

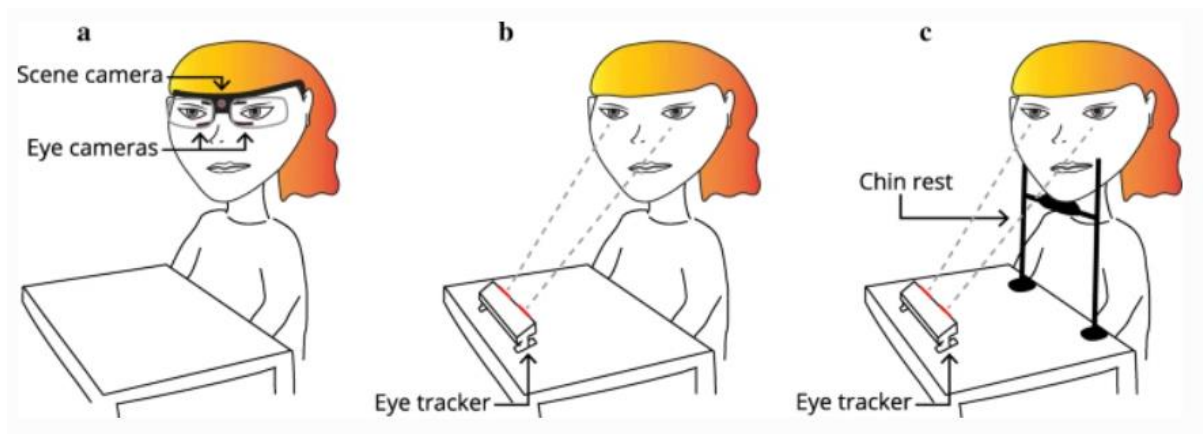


Figure 7: Eye-tracking setup  
a) head-free setup with a wearable eye-tracker; b) head-boxed setup with a remote eye tracker; c) head-restricted setup with a remote eye tracker and a chin rest (Valtakari et al., 2021).

The head-boxed setups (Figure 8) are named this way because the device tracks the user’s eyes within a functional working box. The head-free setup is mobile eye-tracking (*Different Kinds of Eye Tracking Devices*, 2020). It utilises head-mounted eye-trackers and allows for tracking user’s gaze not only on the displays but also in the user's environment. These eye-trackers require three cameras: two cameras with an unobstructed view of eyes and one camera facing the scene.

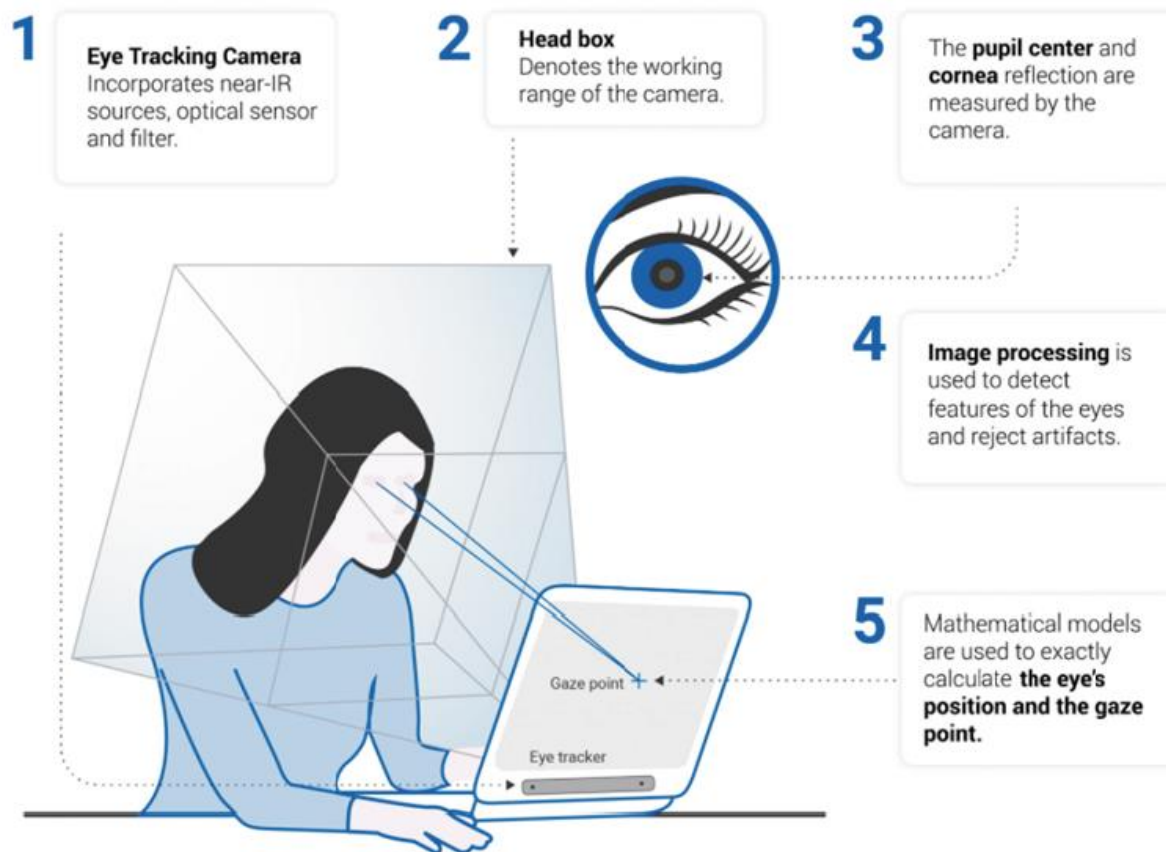


Figure 8: Remote head-boxed eye-tracking system  
(Different Kinds of Eye Tracking Devices, 2020).

Modern eye-trackers can be embedded in other kinds of technologies (for instance, MR and VR devices). In 2019, Microsoft released the second generation of MR glasses, the HoloLens2. The HoloLens2 is the head-mounted MR device with an integrated eye-tracker. The eye-tracker has two infrared cameras for the pupil and the corneal reflection technique. The eye-tracking API is used for accessing the gaze ray for each eye at 30 Hz.

### 2.2.3. Gaze interaction

Gaze interaction enables user control by using eye tracking to define the user's gaze position in relation to the object of interest (e.g., screen). Implementation of gaze input as the only mode of application needs that time, and ways of interactions are identified. When the application is multimodal, additional questions arise: which means of control to use and how to fuse them. (Møllenbach et al., 2013)

The mono-modal gaze interaction can be challenging due to eyes being always “active”. Jacob (1991) refers to it as the “Midas touch” problem, alluding to the king Midas from Greek mythology. The king wished to have the gold touch so that he could turn

anything he touched into gold. Warning him, God Dionysus granted this wish. Everything Midas touched was turned to gold, including food. The king nearly starved because he could neither eat nor drink. This can be an allegory to an unintentional activation of the gaze interaction when the user only wants to explore the scene. The overcoming of the Midas touch problem is essential for the mono-modal cases of gaze interaction. Almost similar to the river Pactolus washing the gift of gold away from Midas and lifting the gold touch, there are means to prevent the Midas touch problem in the gaze interactions.

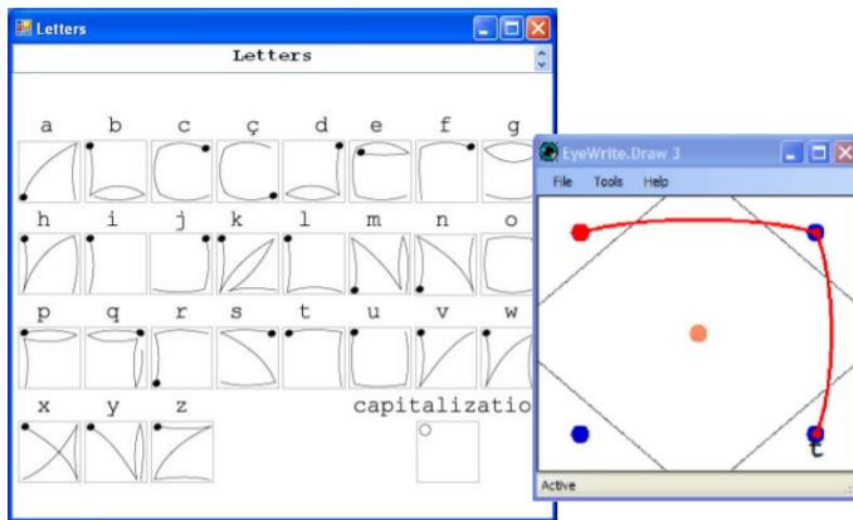
Gaze interaction principles were explained by Mollenbach et al. (2013) through types of eye movements:

- fixation based interaction
- saccade based interaction
- smooth pursuit based interaction

Fixation-based interaction is initiated when the user is looking at the target. Some interactions are conceived to be activated just by looking at the target, for example, to provide a responsive design. An example of a responsive design can be an auto-scrolling of a text when a user reads an article and is close to the end of the page, or an object replying to the user gaze by emitting light to provide confirmation. However, when the gaze interaction is intended to add or change the content in some way, it is difficult to just inspect the object due to Midas touch problem. This case can be addressed by implementing a dwell time to activate the interaction. The user needs to look at the target for a prolonged time to enable the interaction. Dwell duration is advised to be set in the range of 200ms -1000ms (Jacob, 1991; Majaranta et al., 2009), considering user abilities and preferences. While helping with the separation of the selection and inspection, the dwell time can limit the user's capacity to explore the interface and the speed to make selections (Møllenbach et al., 2013).

The main target group for the research into dwell-time-based gaze interaction has been people with motor-skill impairments. The goal has been to ensure the ability to use software for differently-abled people. Dwell time-based gaze interaction was implemented in various eye typing systems (Hansen & Hansen, 2006; Majaranta et al., 2009), environmental control systems (Shi et al., 2008), and computer gaming (Istance et al., 2009).

Saccade-based interaction achieves gaze interaction by utilising saccadic movements to form gaze gestures. For clarification purposes, it should be noted that saccades as strokes constitute gaze gestures. The usual saccade is a rapid motion between any two fixations, whereas the stroke is the movement between two intended fixations (Møllenbach et al., 2013). Gaze gesture is defined by Istance et al. (2010) as “a definable pattern of eye movements performed within a limited time, which may or may not be constrained to a particular range or area, which can be identified in real-time, and used to signify a particular command or intent”. Figure 9 shows the eye gestures, resembling the shape of a character, implemented in the EyeWrite system for text entry built by Wobrock (2008). The experiment showed that eye gesture interactions were slower than dwell-time typing. However, they resulted in higher accuracy (fewer mistakes). The speed of writing with eye gestures was considered to reach the speed of dwell-typing with practice.

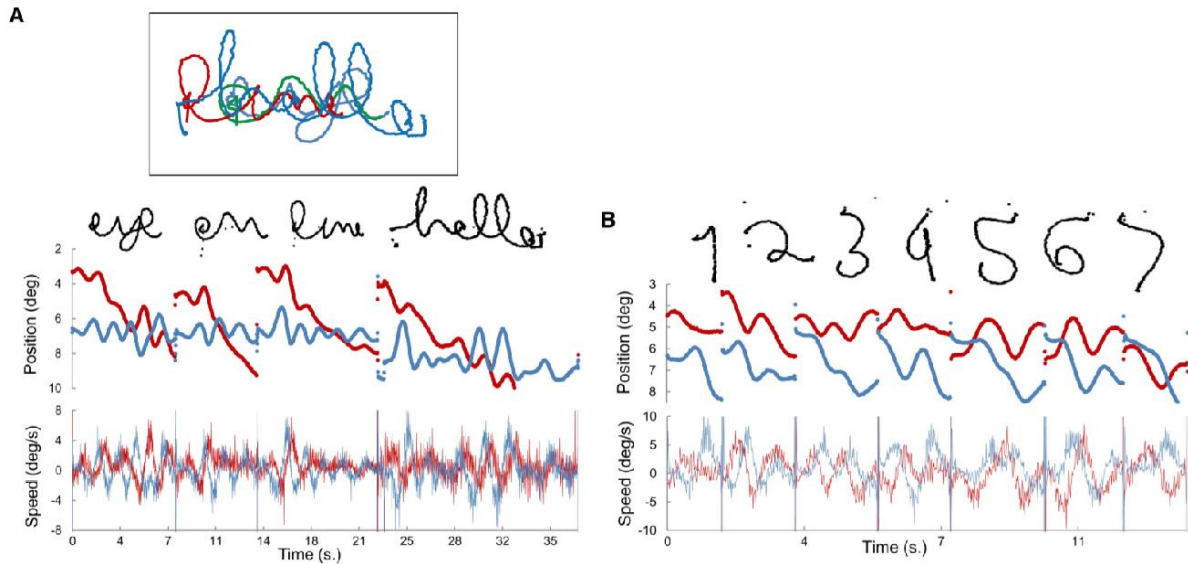


*Figure 9: The EyeWrite gesture alphabet writing a character "t" (Wobrock et al., 2008).*

Beyond based text entry systems (VisionKey (Kahn et al., 1999), pEYWrite (Huckauf & Urbina, 2008)), gaze gesture interfaces (Drewes & Schmidt, 2007), drawing applications (Heikkilä & Rähä, 2010), and computer gaming interfaces (Heikkilä & Rähä, 2010) were invented.

Smooth pursuit-based interaction suggests interacting while fixating upon a moving target (Møllenbach et al., 2013). Lorenceau (2012) developed a smooth pursuit-based

text input system, recording examples of which can be seen in Figure 10. The results were achieved with preliminary user training for gaining control over smooth pursuits. The interaction based on smooth pursuits is still a topic for future research (Møllenbach et al., 2013).



*Figure 10: Examples of eye movement recordings  
(A) for eye-generated words and (B) for eye-generated numbers. Top - raw recordings before and after segmentation and scaling; middle – vertical (blue) and horizontal (red) gaze positions over time; bottom – vertical and horizontal speeds over time. (Lorenceanu, 2012)*

Mollenbach (2013) characterised gaze interactions in regard to the eye movement type and displayed graphic objects. The graphic display object (GDO) is divided into static GDO (directs the user gaze and provides feedback), dynamic GDO (dynamically adapts and guides gaze), and no GDO (when no graphic object is related to the interaction). Table 2 provides an insight into the categorisation.

Table 2: Overview by Mollenbach (2013) of eye interactions with static/dynamic graphic display objects and without.

	Fixations	Saccades	Smooth Pursuits	Characteristics
No Graphic Display Object	–	Single and Complex Gaze Gestures	–	Does not use screen real estate Does not require a screen Does not provide feedback in the selection process Limited vocabulary
Static Graphic Display Object	Static Dwell Objects	Single and Complex Gaze Gestures	–	Requires screen real estate Requires screen Provides selection feedback Limitless Vocabulary
Dynamic Graphic Display Object	Dynamic Dwell Interfaces	Dynamically Defined Gaze Gestures	Moving Targets	Requires screen real estate Requires screen Provides feedback Limitless Vocabulary

#### 2.2.4. Gaze-based interactions in cartography

Eye-tracking applications in cartography can be generalised as map design evaluation, cognitive processes evaluation, explicit and implicit gaze-based interaction (Göbel et al., 2019). The first studies analysed eye-tracking data to evaluate designs of static maps (STEINKE, 1987). Later on, interactive maps and the effectiveness and efficiency of their interface designs were evaluated using eye-tracking data and usability metrics (Coltekin et al., 2009, 2010). The cognitive processes that support map use were evaluated in the spatial decision-making studies (Wiener et al., 2012), indoor wayfinding (Schuchard et al., 2006), and outdoor wayfinding (Bektaş & Çöltekin, 2011).

For interaction purposes, eye-tracking data is used as a real-time input (Jacob, 1991). Schmidt (2000) distinguishes explicit and implicit interactions. Explicit interaction involves a user's intention to manipulate. For example, to type a word with eyes, the user looks at the corresponding keys on the screen keyboard (Majaranta et al., 2009). During implicit interaction, a user does not intend to manipulate anything directly with the gaze. However, the information about the user gaze is registered by the device and used to adapt the content. Visualising a history of user gaze on a map is one example of implicit interaction (Giannopoulos et al., 2012). This gaze-based interaction concept ("GeoGazemarks") was created by Giannopoulos et al. (2012) to help users with orientation within a map on the small-screen devices (smartphones). Another example of implicit interaction is implemented in the "GeoFoveation" approach of Bektaş & Çöltekin (2011). In this approach, geospatial data is loaded only on the map area where the user is looking.

Explicit gaze-based interaction was implemented by many researchers for zooming and panning (Adams et al., 2008; Bates & Istance, 2002; Fono & Vertegaal, 2005, 2005; Hansen et al., 2008; Lankford, 2000). There are monomodal (Adams et al., 2008; Hansen et al., 2008; Lankford, 2000; Zhu et al., 2011) and multimodal (Bates & Istance, 2002; Fono & Vertegaal, 2005; Stellmach et al., 2011; Stellmach & Dachsel, 2012) examples of gaze-based zooming and panning. Monomodal gaze-based interaction can use the dwell-time technique (Lankford, 2000) and zoom/pan regions on the screen (Adams et al., 2008; Hansen et al., 2008). Within multimodal gaze-based interaction, gaze input can be supplemented with manual control (Bates & Istance, 2002), keyboard (Fono & Vertegaal, 2005), computer mouse (Adams et al., 2008), or a touch screen (Stellmach et al., 2011).

## 2.3. Mixed reality

### 2.3.1. Definition

In 1994, Milgram et al. developed a “virtuality continuum” and introduced the mixed reality concept while distinguishing it from virtual reality and augmented reality. The virtuality continuum (Figure 11) has two extrema: the real environment – reality, and the virtual environment- virtuality. Reality consists of solely real objects and the virtuality, in other words, virtual reality (VR) – of solely virtual objects (Milgram & Kishino, 1994). A very straightforward definition of mixed reality (MR) is the combination of reality and virtuality. This means that VR is not a part of MR because it fully immerses a user in a synthesised environment. All cases in between the real and the virtual environment (not including extrema) are considered MR, for instance, augmented reality and augmented virtuality (AV). Augmented reality (AR) is the case of MR, located close to the reality continuum, and it overlays the real world with virtual objects (Milgram & Kishino, 1994). An example of AR would be a phone app that places an animated penguin on top of the real table or “wearing” a face mask on to a real person (for instance, “AR lens” from Huawei). AV lies closer to the virtuality continuum, in between AR and VR. (Milgram & Kishino, 1994)



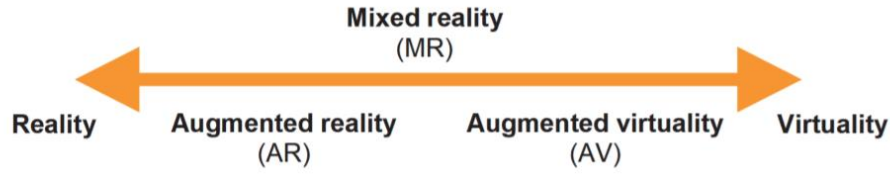


Figure 11: The virtuality continuum  
by Milgram et al. (1994) adapted by Sobota et al. (2020)

Speicher et al. (2019) extended the visual feature-oriented definition of MR (Milgram & Kishino, 1994). They identified the six most popular notions of MR through expert interviews and literature review. The first notion is the definition according to the virtuality continuum. The second notion is MR as a synonym for AR (interchangeable use of terms). The third notion defined MR as a collaboration of physically separated AR and VR. The fourth notion described MR as a combination of AR and VR. “Alignment of environments” is considered to be MR in the fifth notion. In this case, the environments are not necessarily AR and VR and are not physically separated. And the last notion describes MR as a “strong” AR. A conceptual framework (Table 3) was derived from organising the notions along seven dimensions: number of environments involved, number of users participating, the immersion level, the virtuality level, degree of interaction, input, and output (Milgram & Kishino, 1994).

Table 3: The conceptual framework of the MR notions  
classified along seven dimensions (Speicher et al., 2019)

Dimension value	# Environments		# Users		Level of Immersion			Level of Virtuality			Interaction		Input	Output
	one	many	one	many	not	partly	fully	not	partly	fully	implicit	explicit	any	any
1–Continuum	✓		✓		✓	✓	✓	✓	✓		✓		✓	✓
2–Synonym	✓		✓		✓	✓		✓			✓		✓	✓
3–Collaboration	✓	✓		✓		✓	✓	✓	✓		✓	✓	✓	✓
4–Combination	✓		✓			✓	✓	✓	✓		✓		✓	✓
5–Alignment		✓	✓		✓	✓	✓	✓	✓	✓	✓		✓	✓
6–Strong AR	✓		✓			✓		✓			✓	✓	✓	✓

The given table shows the differences between MR definitions. The definition which is used in this study responds to the continuum concept. According to the table, the MR as a continuum has one environment, one user, three levels of immersion, and full and part virtuality.

### 2.3.2. Interactions

Umbrella terms for the interactions, such as multimodal (Aladin et al., 2020; Fekri & Wanis, 2019), collaborative (Chalon & David, 2004; Wendrich, 2011), and tangible interactions (Couture et al., 2010; Lee et al., 2010), are widely used by the scientific



community in the MR field (Bekele et al., 2018). However, for the user experience (UX) and human-computer interaction (HCI) knowledge structuring, there is a need for a well-defined classification of interactions in MR (Pamparău & Vatavu, 2020). Papadopoulos et al. (2021) addressed this need and created a modality-based framework for classifying interactions in immersive realities.

The framework by Papadopoulos et al. (2021) consists of modalities, contexts, and interaction methods. Modality is a “human-computer communication channel” that allows for the input and output of information. Each modality corresponds to one of the human senses: visual-based, audio-based, haptic-based, smell-based, and taste-based modality (Papadopoulos et al., 2021). Papadopoulos et al. (2021) described the first three, defining all interactions that include visual, sound, and touch perception. The interactions not covered by the modalities mentioned above, but require sensors, are contained in the sensor-based modality (Papadopoulos et al., 2021). Depending on how many modalities are used, interactions can be unimodal and multimodal.

Context determines a specific case of a modality, for example, eye movement tracking functions within the gesture-based context of the visual-based modality (Fig.11). The eye movement from this context is an interaction method. An interaction method is utilised to perform an interaction task by a sequence of coordinated procedures (Papadopoulos et al., 2021). Bachmann et al. (2018) outlined four basic interaction tasks for three-dimensional interaction devices (such as MR devices): selection, manipulation, system control, and navigation. A user selects an object to execute an operation, for instance, to retrieve information. Manipulation implies the user adjusting any parameters of an object, such as rotation, scale, and others. The system’s state can be changed by system control interaction, for instance, using menu buttons. Navigation allows users to orientate in the MR environment (Papadopoulos et al., 2021).

The visual-based modality (Figure 12) includes interactions within the following contexts: gesture-based, surface-based, marker-based, and location-based context (Papadopoulos et al., 2021). A camera sensor captures and analyses state changes in these contexts. When interactions require body language recognition, the user interacts through gesture-based context. Surfaces of the real environment are considered when interacting through surface-based context (Papadopoulos et al.,

2021). Interactions can be activated by a camera recognising a marker assigned to the interaction. These markers can take different forms (Onime et al., 2020), e.g., a physical image sticker or an infrared marker (Nee et al., 2012). Within the location-based context, an interaction is provided based on the user's position. The user's location can be identified by the camera registering imagery around the user (Stricker et al., 2001) or by reading a QR code containing the position information (Evangelidis et al., 2021).

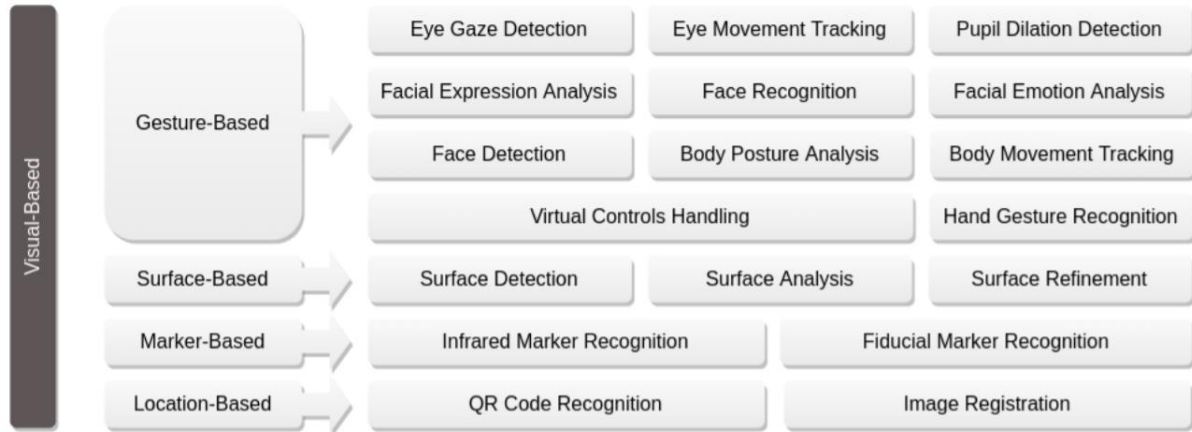


Figure 12: Visual-based modality and its contexts and interaction methods.  
by Papadopoulos et al. (2021)

The audio-based, haptic-based, and sensor-based modalities with related contexts and interaction methods are presented in Figure 13. The speech recognition method within a speech-based context is commonly used for executing tasks and mostly in multimodal systems (Billinghurst et al., 2009; Che Hashim et al., 2018; Hanifa et al., 2021). The haptic-based modality implies the tangible, graspable objects that need a touch to initiate an interaction, such as a joystick-controller (Hashimoto et al., 2011), a pen (Jiawei et al., 2010), and others. The head-gaze direction is detected by motion sensors in devices and operates within the sensor-based modality. This method is used in Microsoft's HoloLens devices as a fallback estimation when eye-tracking is not permitted but required for the interaction (sostel, n.d.).

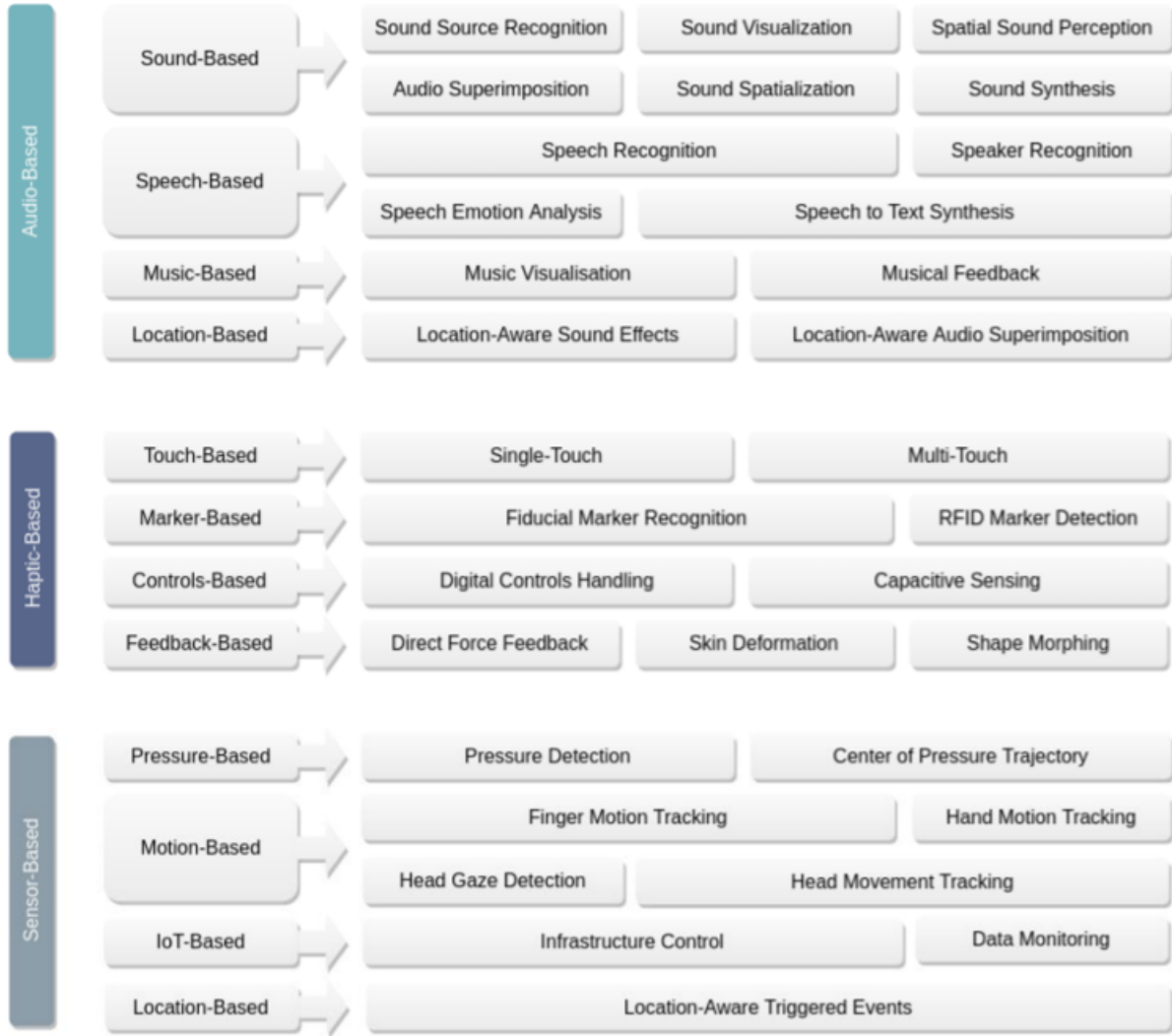


Figure 13: Audio-based, haptic-based, and sensor-based modalities, contexts, and interaction methods by Papadopoulos et al. (2021)

### 2.3.3. MR devices

Sobota et al. (2020) outlined different types of MR devices. There are two methods of reality-virtuality coexistence in these devices: optical see-through system and video see-through system. An optical see-through system (holographic) allows users to see the real world through transparent display and generates (virtual) objects on top of it (Figure 14 a). Whereas in the video see-through systems (immersive), the real world is represented on the non-translucent screen as captured by the camera sensor, and the virtual objects are added (Figure 14 b, c). In this case, the device blocks the view of a natural environment, similar to VR devices. Moreover, systems can be defined as marker systems that use the markers to recognise the scene and place objects or markerless systems that utilise GPS coordinates, image recognition, and other inputs instead of tags (Sobota et al., 2020).



*Figure 14: Optical see-through and see-through video devices  
 a) HoloLens 2 from Microsoft - holographic (qianw211, n.d.); b) HP Windows MR Headset – immersive (Mixed Reality Headset / Augmented Reality (AR) Headset | HP® Store, n.d.); c) Acer Windows MR Headset – immersive (view from above) (Windows Mixed Reality Headset, n.d.)*

While having defined general types of MR devices in the current sub-section, we describe the headset chosen for this study and its specifications in the case study section.

### 3. Methodology

#### 3.1. Research approach

This study combines different research methods to meet research objectives and answer the research questions of the thesis. The research design, which composes different research methods of this thesis, is mixed. It combines experimental and cross-sectional research designs (Bryman, 2016). The former entails a laboratory experiment, where the researcher can control the setup of the experiment. Latter includes questionnaires and is also known as survey design (Bryman, 2016). The schematic research approach is presented in Figure 15 and described in the following passages.

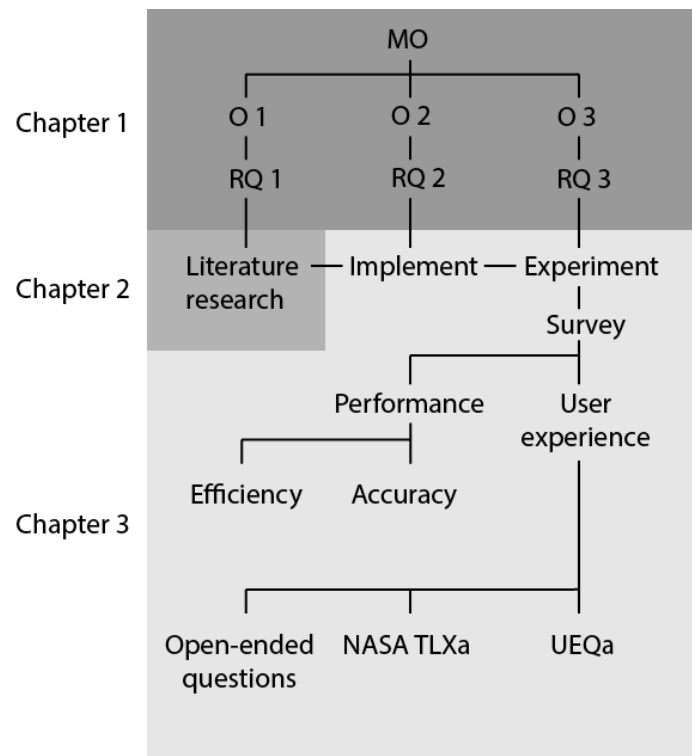


Figure 15: The research approach structure. Abbr.

MO – main objective, O – objective, RQ – research question, NASA TLXa – NASA TLX adapted, UEQa – UEQ adapted

The experiment implies different users interacting with maps in MR. Firstly, how the users interact with the maps in this study is defined by determining the interactions. Previously, in the second chapter (Theoretical background), we have identified the fundamental cartographic, gaze-based, and MR interactions. The gaze-based user-map interactions developed for this research are selected from these fundamental

interactions. Consequently, the map to be utilised in this user study corresponds to the needs of the selected gaze-based interactions.

The users experience the interactions chosen with the help of an application built on the MR device. The application has three kinds of interfaces to provide interactions. For evaluation purposes, along with the gaze-based interface, which is the focus of this study, conventional and gaze-aware (mixed) interfaces are also assembled. These interfaces differ in the modalities utilised. The gaze-based interface is unimodal and controlled explicitly by user-gaze. The gaze-aware interface is a multimodal interface that uses gaze and one other input. Finally, the conventional interface uses hands as the widely used concept of gesture-based modality for the MR environment (Papadopoulos et al., 2021).

This study evaluates the interfaces from the performance and user experience perspective. The task-based approach is used to gain data for the evaluation as one of the standard approaches (Lazar et al., 2017). For each of the implemented interactions, the tasks are composed, and they encourage users to interact. The efficiency is measured (task completion time) for every one of the interactions to assess the performance (Lazar et al., 2017).

Questionnaires used in this study are digital and have closed-ended questions (mainly ranking scales) and one open-ended one. They aim to collect data about the user experience, task load, and interface ranking related to the mixed reality experiment. The User Experience Questionnaire (UEQ) is adapted and used to evaluate the user experience in this user study. UEQ is the standardised subset of questions widely used by researchers for measuring user experience (Schrepp et al., 2014). The Task Load Index questionnaire (NASA TLX) is utilised to rate the user's cognitive load (Haklay, 2010). NASA TLX questions are adapted here for this research. The UEQ and NASA TLX are chosen for the user study because they can help answer the third research question. The interface ranking questions are composed to have an insight into user preferences. Finally, open-ended questions have a goal of considering any commentaries users provide after testing the application. These remarks can regard any aspects of the interfaces that are not covered by the previous questions.

## 3.2. Case study

### 3.2.1. Selected interactions

In this thesis, three interactions were chosen to be implemented and evaluated. They were chosen based on the fundamental interactions defined in second chapter. The results of Tolochko's (2016) study relate the retrieve and overlay interactions to the main interactions expected from web maps, along with the zoom, pan, search, and filter interactions. As previously defined, the retrieve interaction provides additional information about an object selected, and the overlay interaction adds layers of content to the map. Thus, the retrieve and overlay are selected for this research. In addition, the manipulation interaction is defined as one of the fundamental interactions tasks in the MR environment (Bachmann et al., 2018). Finally, the third interaction is manipulation, particularly rotation. The case study applies the fixation-based principle for gaze interactions. A user activates the rotation and the retrieve interaction by gazing at a target. Moreover, dwell-time is set for the overlay interaction.

### 3.2.2. Resources

#### *Data*

MR environment provides the opportunity to present holograms of three-dimensional (3D) objects. In addition, two-dimensional (2D) interfaces can be implemented, such as internet browser windows or other non-immersive applications. The browser window is presented in the same manner as on the laptop display, only placed in the MR environment. However, in this study, we focus on 3D maps. The selection of mapped areas did not entail any particular reasoning. The only criterium was for the site not to be an easily recognisable destination.

Two separate map models were created for the interactions to avoid information overloading. The following 3D models were created to correspond to the needs of the selected interactions:

- Retrieve and overlay – the city
- Rotate – the terrain

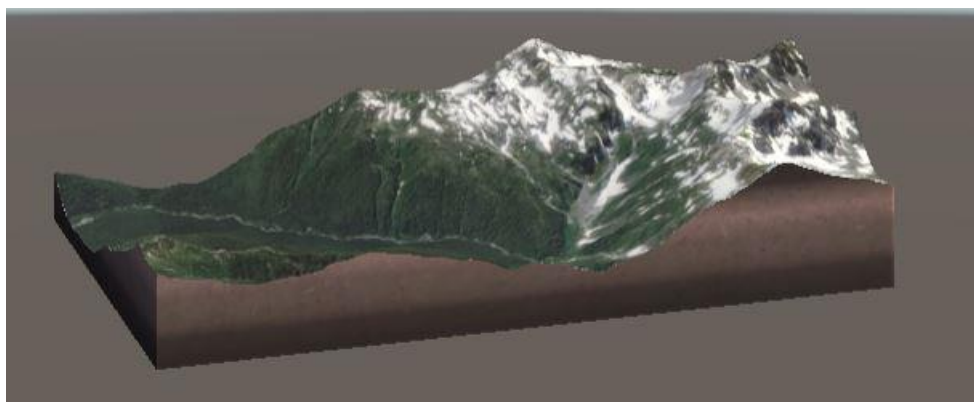
The city model shown in Figure 16 was created for the retrieve and overlay interactions. The model was generated using OSM (Open Street Map) data for the

Auckland city area in New Zealand. The names of the buildings can be retrieved. The overlay interactions can add points of interest (hotels, viewpoints, and others) to the map.



*Figure 16: The city model for the retrieve and overlay interactions (the points of interest are hidden).*

The second model is the terrain shown in Figure 17. The Google Satellite imagery and the SRTM (Shuttle Radar Topography Mission) elevation data of the mountain Wilder (Washington, USA) were used to generate the model. Users can rotate the terrain along the vertical axis.



*Figure 17: The terrain model for the rotation interaction.*

## **Hardware**

Different MR devices on the market have embedded eye-tracking technology, such as Oculus Quest 2, Lenovo Star Wars, Magic Leap One, Microsoft HoloLens 2, and others. Notably, the hosting university of the study provided the opportunity to use HoloLens 2. Therefore, the prioritisation of the device access impacted the device selection for the user study. In addition, some developers consider HoloLens 2 as the best MR headset ('HoloLens 2 Review', 2021).



The HoloLens 2 is the MR device and can also be identified as a head-free (mobile) eye-tracking system (Valtakari et al., 2021) due to the embedded eye-tracking technology. The main components of the device (Figure 18) are the visor and the headband (scooley, n.d.-b). The vizor contains the display and sensors, and the headband includes the processor part. Since many different users wear the device to test the interactions, the comfort aspect can be important. Indeed, HoloLens 2 has fitting elements that allow users to adjust the headset's position. In Figure 18, we can see two elements of the fitting system: wheel (1) and head strap (3) (scooley, n.d.-b). The wheel adjusts the headband's diameter, whereas the head strap regulates the vertical position of the headset. Hence, the user can control the headset's position on their head, no matter the head form or size. In addition, the rear pad and brow pad protect the user from overheating ('HoloLens 2 Review', 2021; *HoloLens 2—Overview, Features, and Specs* | Microsoft HoloLens, n.d.).

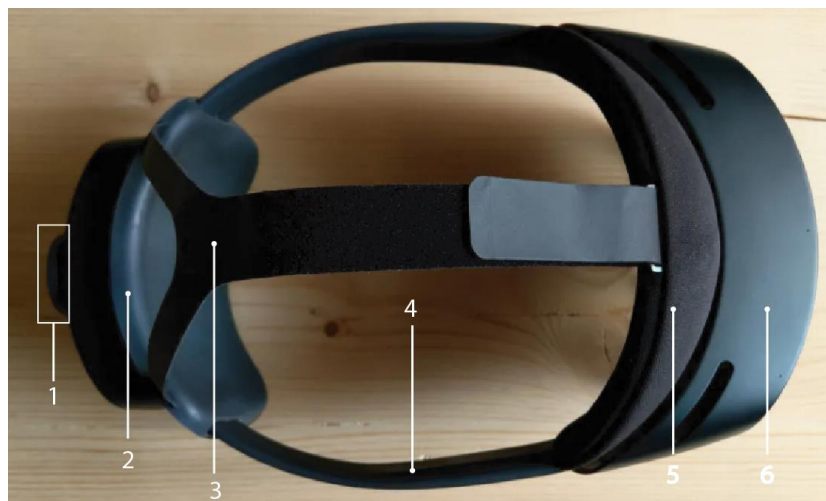


Figure 18: HoloLens 2, top view  
1) wheel; 3) head strap; 4) headband; 5) brow pad; 6) visor.  
The image adapted from ('HoloLens 2 Review', 2021).

The technical specifications of the headset are presented in Table 4. The overview covers four aspects: display, sensors and human understanding, and miscellaneous. The miscellaneous part offers other features, which are essential and only ungrouped in this context.

Table 4: HoloLens 2 technical specifications.  
Adapted from Microsoft (*HoloLens 2—Overview, Features, and Specs* | Microsoft HoloLens, n.d.)

Features	Characteristics
Display	

Optics	See-through holographic lenses
Field of view	Diagonally 52°, horizontally 43°, vertically 29°; ratio 3:2
Holographic density	>2.5k radiants (light points per radian)
Sensors and human understanding	
Hand tracking	Two-handed fully articulated model, direct manipulation
Eye tracking	Real-time tracking (30 Hz), two infrared cameras, eye tracking API
Head tracking	Four visible light cameras (stereo and periphery)
Voice	Voice commands, natural language recognition
Audio	Two speakers with built-in spatial sound
Miscellaneous	
Computing	System on the chip (CPU, GPU, and HPU)
Operating system	Windows Holographic
Battery	2-3 hours of active use
Weight	566g

The display is presented as see-through holographic lenses, thereby not isolating the user from the real world. In addition, the lenses are slightly dark to reduce the real environment brightness, thus facilitating the holograms. The users can see holograms within the 43 degrees of horizontal and 29 degrees of vertical view (3:2 ratio) (scooley, n.d.-b).

The visor of the HoloLens 2 has built-in cameras (Figure 19) to provide the tracking possibility, therefore facilitating human understanding. The device can track the hands' position, the eye gaze, and the head gaze. Usually, the head gaze is used as a fallback to the eye gaze. The head gaze defines where the user is looking by the direction of the head (sostel, n.d.).

HoloLens uses two infrared (Figure 19 d) cameras for the pupil/corneal reflection eye-tracking technique (Duchowski & Duchowski, 2017). The cameras are aimed at the user's eyes and can track the eye gaze in real-time. The eye-tracking application programming interface (API) gives access to the eye gaze origin and direction data.

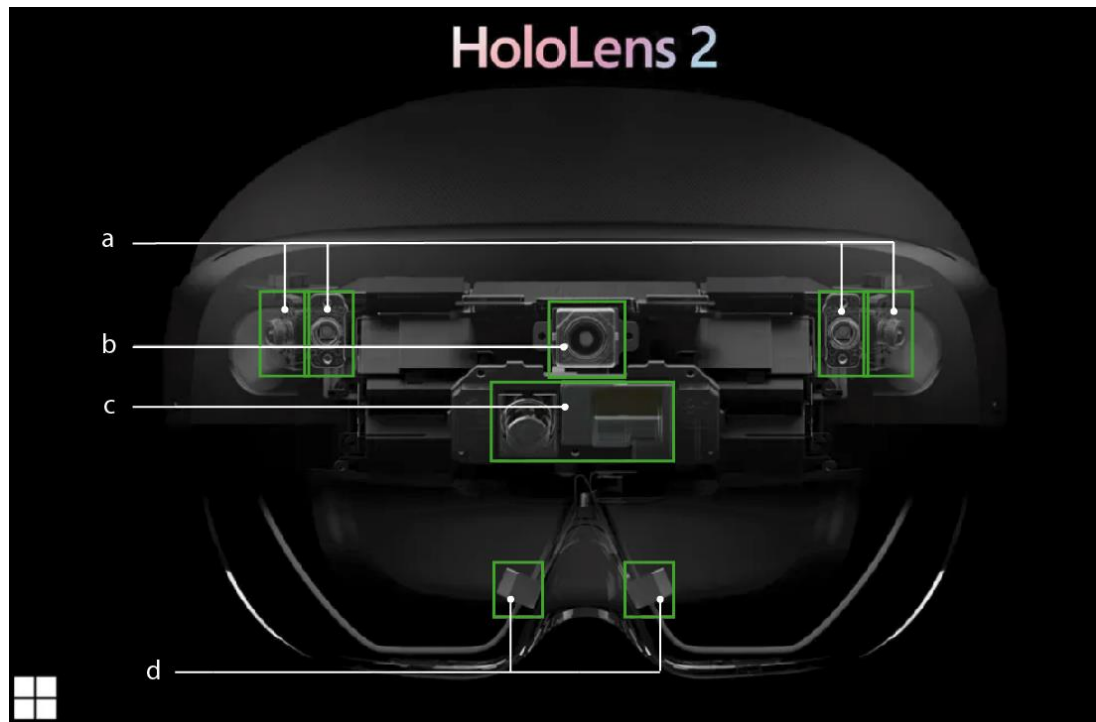


Figure 19: HoloLens 2 cameras setup (front view)

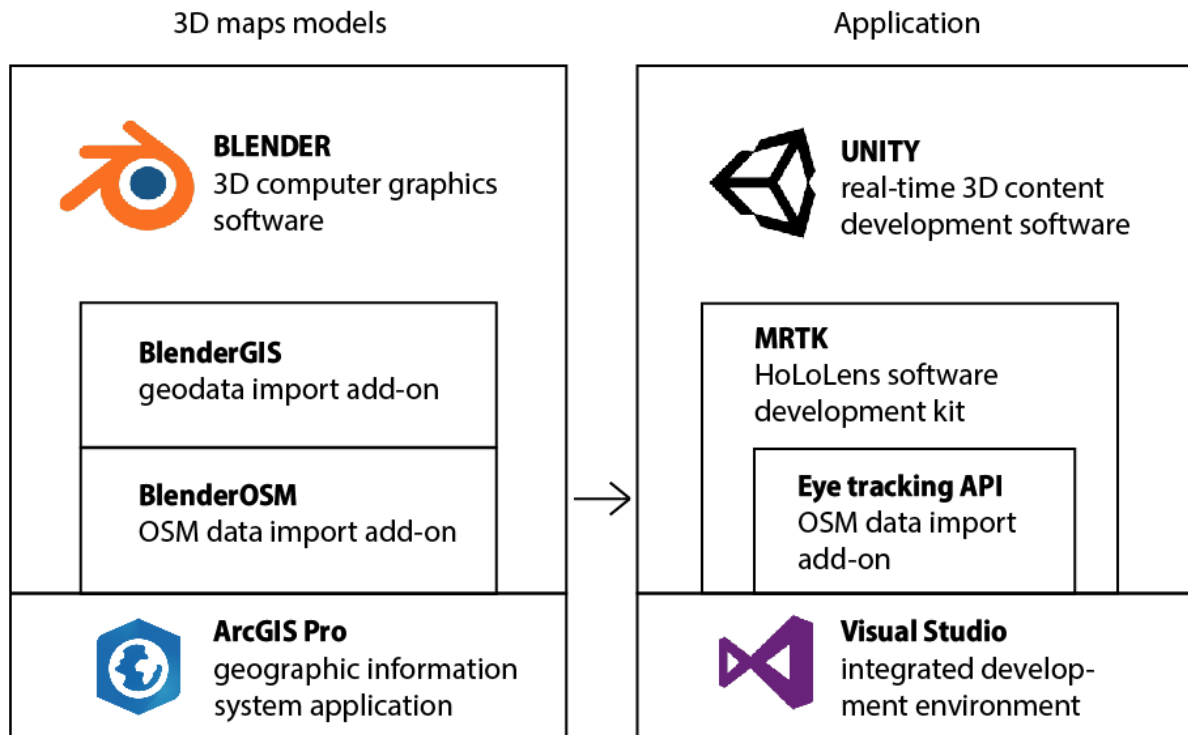
a) 4 head tracking cameras - stereo and periphery; b) RGB camera for photos/videos; c) depth camera – near and far range; d) 2 eye-tracking cameras – infrared. Adapter from Microsoft (scooley, n.d.-b)

Speech recognition is another fascinating feature of HoloLens 2. This feature and the microphones ensure that users can use voice commands to control the experience. In addition, two speakers on the headband above the ears produce spatial sound to immerse the user in the MR ('HoloLens 2 Review', 2021).

The processor has the system on the chip (SoC) that contains central, graphical processing units (CPU and GPU) and a separated holographic processing unit (HPU). HPU performs tracking computations, thus freeing CPU and GPU responsible for the MR experience. HoloLens operates on Windows Holographic Operating System, a Windows version that runs in the three-dimensional world (*HoloLens 2—Overview, Features, and Specs | Microsoft HoloLens*, n.d.).

### Software

In Figure 20 **Error! Reference source not found.**, one can see the primary software used to develop gaze-based user-map interactions for the HoloLens 2. The two subsets represent the maps' and application generation stages. The software utilised in the experiment and survey stages are defined in the User study setup sub-chapter.



*Figure 20: Software utilised for the application development*

The maps were created using the Blender software that is a 3d computer graphics software. Blender is a freely available open-source software founded by Blender Institute corporation (Foundation, n.d.). The geodata for the maps was imported using Blender add-ons. BlenderGIS was used for the terrain. The vector data for the city buildings were imported using BlenderOSM add-on, whereas points of interest for the city model were collected and exported using ArcGIS Pro (OSM import function).

The maps created were exported to a Unity environment for further application development. Unity is the development platform for creating real-time content that is games, 3D experiences, and others. Unity organisation provides students with a free license to develop and explore the possibilities of the software (Technologies, n.d.). It was used to create an application for the gaze-based, gaze-aware, and conventional interfaces of selected cartographic interactions.

Microsoft has released an open-source toolkit to help developers with the creation of new HoloLens applications. This software development kit (SDK) is named Mixed Reality Toolkit (MRTK) and can be imported to the Unity environment. Essential features of the MRTK are the project configurations, building tools, APIs, and samples. MRTK allows the developers to compile new applications faster by excluding

programming and designing the basic functionality and interface elements (*MRTK-Unity Developer Documentation - Mixed Reality Toolkit | Microsoft Docs*, n.d.).

Undoubtedly, we need to know where the user is looking to allow gaze-interaction. Hence, we are using the Eye-tracking API provided by HoloLens 2. API is the “messenger” between eye-tracking data and the application. Eye-tracking API gives access to the gaze origin and direction with a frequency of 30 Hz.

Finally, the developed app was installed on the HoloLens 2. The Visual Studio, the integrated development environment (IDE), was used for the instalment. It is a Microsoft software that allows running, editing, debugging, and deploying applications (*Visual Studio: IDE and Code Editor for Software Developers and Teams*, n.d.).

### *People*

The user study needs a minimum of twenty users. This is reasoned by the experiment and survey being conducted in the laboratory. Therefore, the required number of participants is relatively small. This study addresses people of any gender, age, experience, or professional background. Due to the design of HoloLens 2, users can wear the device along with prescription glasses. Thus, there is no restriction for people wearing spectacles to participate in the study. Moreover, a wider variety of the users' backgrounds is preferred.

#### 3.2.3. Implementation

The gaze-based user-map interactions were chosen after researching the related literature based on the existent fundamental cartographic, gaze-based, and MR interactions. The contents and presentation of the maps were defined regarding the selected interactions. Finally, the three interfaces with various control were determined to be compiled in one application. The workflow of developing the application was divided into five stages: 1) maps modelling, 2) root scene, 3) gaze-based interface, 4) gaze-aware interface and 5) conventional interface assembling. Figure 21 presents the schematic workflow followed to create the application.

### *Maps modelling*

The 3D models of maps were created in the Blender software (Figure 21, 1). The process of creating the terrain model for the selected area is outlined as follows:

1. importing the Google Satellite imagery and the SRTM elevation data using BlenderGIS add-on
2. extruding the imagery according to the elevation
3. adjusting the visualisation

The city model generating models is similar to the above mentioned but implies additional technical and design steps. The outline follows as:

1. importing the base map and buildings using BlenderOSM
2. extruding the buildings according to their height
3. importing the points of interest (POIs) to ArcGIS and exporting to Blender
4. adjusting the visualisation

To evaluate the selected gaze-based interactions of the gaze-based interface (1), other two interfaces are created: the interface that uses gaze but not entirely relies on it (2) and the interface that uses the typical MR mean of control (3). The modalities for these interfaces were defined as shown in Table 5.

*Table 5: The three developed interfaces and their characteristics*

Concept	Modality	Retrieve	Overlay	Rotate
1. Gaze-based	gaze	gaze	gaze and dwell	gaze
2. Gaze-aware	gaze and voice	gaze	voice command	gaze
3. Conventional	hands	hand-ray	air-press	near/far manipulation

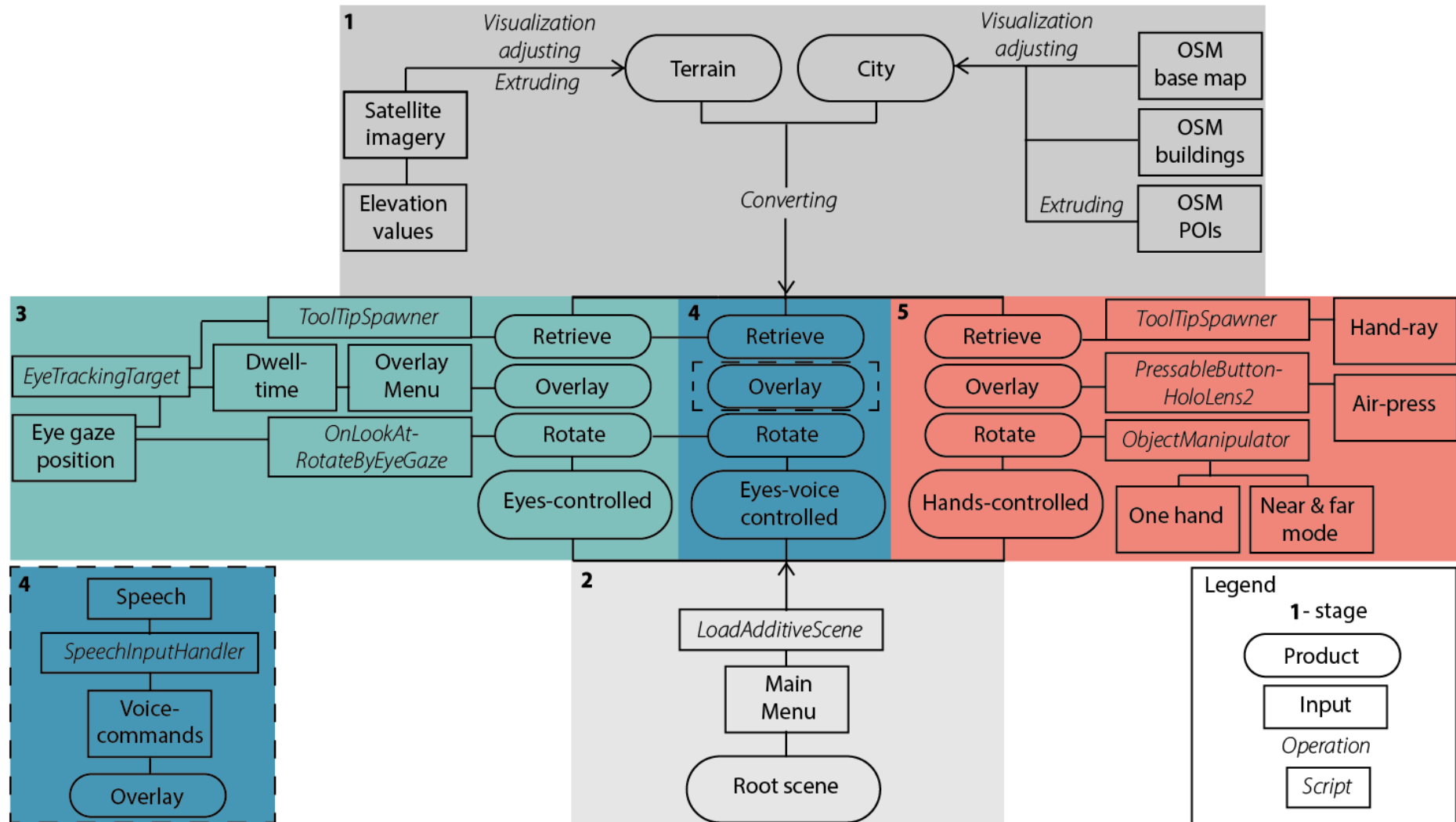
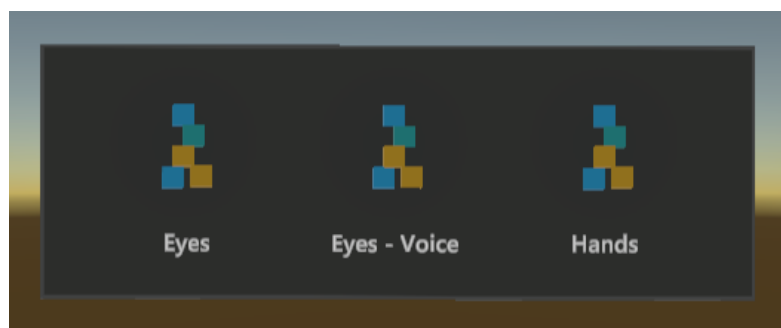


Figure 21: Schematic workflow of the application development  
 1) maps modelling, 2) the root scene compiling, 3) gaze-based interface, 4) gaze-aware, and 5) conventional interface assembling

The application was assembled for the Windows Universal platform that is HoloLens 2 suitable. There are four Unity scenes, the experience spaces, that compose the application. Each scene has an MRTK configuration profile, where the experience settings, input (modalities), diagnostics, extensions, and other settings can be defined and changed. In addition, MRTK has template assets of interface elements that follow the HoloLens applications' style, such as menus, buttons, etc.

## Root scene

The root scene (Figure 21, 2) is the main scene that opens other scenes. At the start of the application, only the root scene is initialised. It is presented as the main menu (Figure 22) slightly above the user's view field. Such a position is due to the activated interface scene being added to the experience space in the user's view field. For that, the menu uses the MRTK LoadAdditiveScene script. The menu has three buttons for the interfaces. The buttons are named after the modalities used for the interaction control: gaze-based interface - "Eyes"; gaze-aware interface - "Eye - Voice", conventional - "Hands". Initially, the root scene uses gaze and dwell on activating any interface. Afterwards, the modalities for the main menu are defined as the current scene because every scene has its configuration profile.



*Figure 22: The root scene and the main menu*

## Interface 1

The first interface is assembled for the gaze-based (eyes-controlled) interactions. The scene (Figure 23) contains the terrain and the city model, two panels with short instructions, and the floating overlay menu that follows the user.



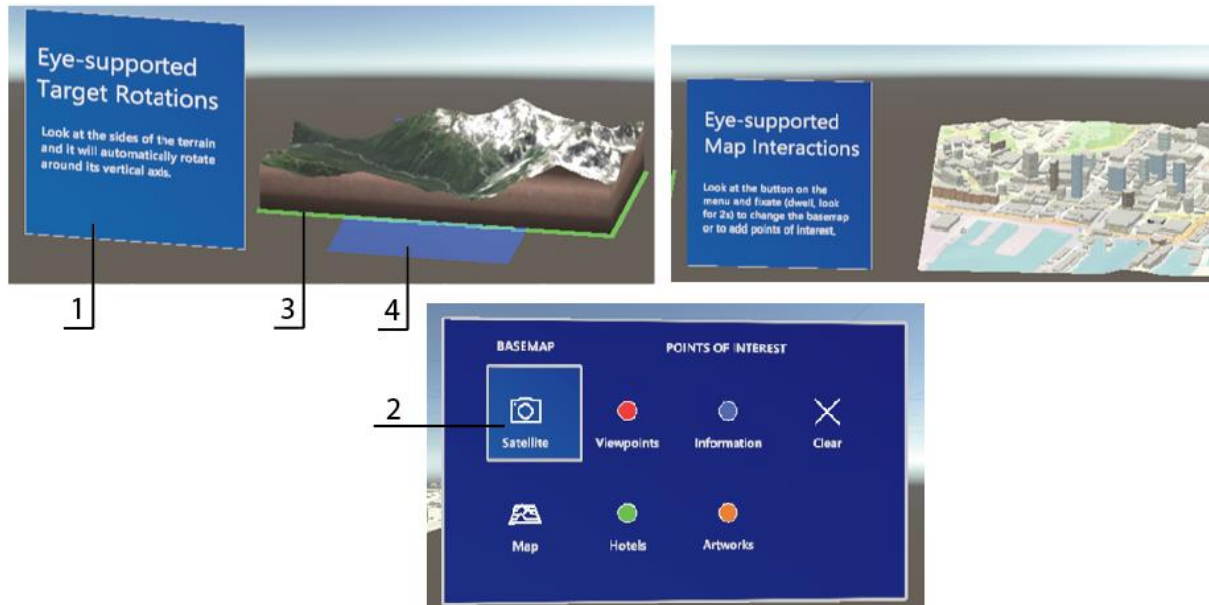


Figure 23: The gaze-based interface scene  
 1) instruction panel, 2) highlighted button - visual cue of the activated layer, 3) initial position of the terrain (green rectangle), 4) the Rotation task position (blue rectangle)

The rotation interaction of the first interface is activated by the user gaze. To be able to interact with objects in Unity, the objects need a collider. It is a 3D shape of the object that defines collisions and triggering of interactions. In this case, when the user is looking at the target, the gaze position is predicted to be on the target rather than going through the hologram. Figure 24 shows the terrain's collider in the form of the box. The box collider is chosen here for easier computations. The rotation interaction uses the MRTK EyeTrackingTarget script, thus defining that gazing at the terrain triggers an interaction. Further, the MRTK OnLookAtRotateByEyeGaze script rotates the model toward the user when the user looks close to the edges of the terrain.

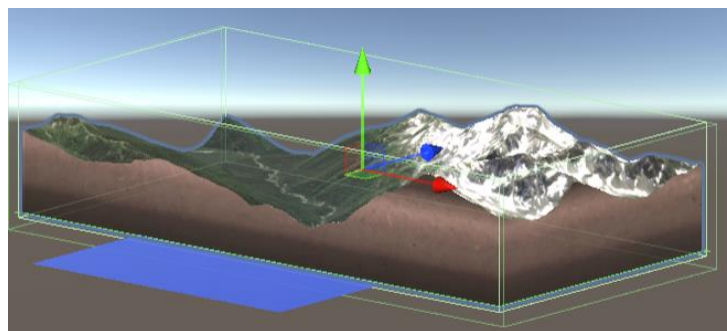


Figure 24: Box collider for the interactions with the terrain

The overlay interaction uses the menu (Figure 23) with buttons divided into two groups: base map and points of interest. Therefore, users can change the base map and add the points of interest. To do so, they need to gaze at the button and dwell for 2

seconds. The visual cues guide the user by slightly highlighting the button when the user is looking at it and outlining the button when it is activated (Figure 23, 2). In addition, the audial cue of the clicking sound is activated when some layers from the menu are added to the map. The interaction functions use the MRTK EyeTrackingTarget script's "OnDwell ()" function.



*Figure 25: Overlay interaction: points of interest activated*

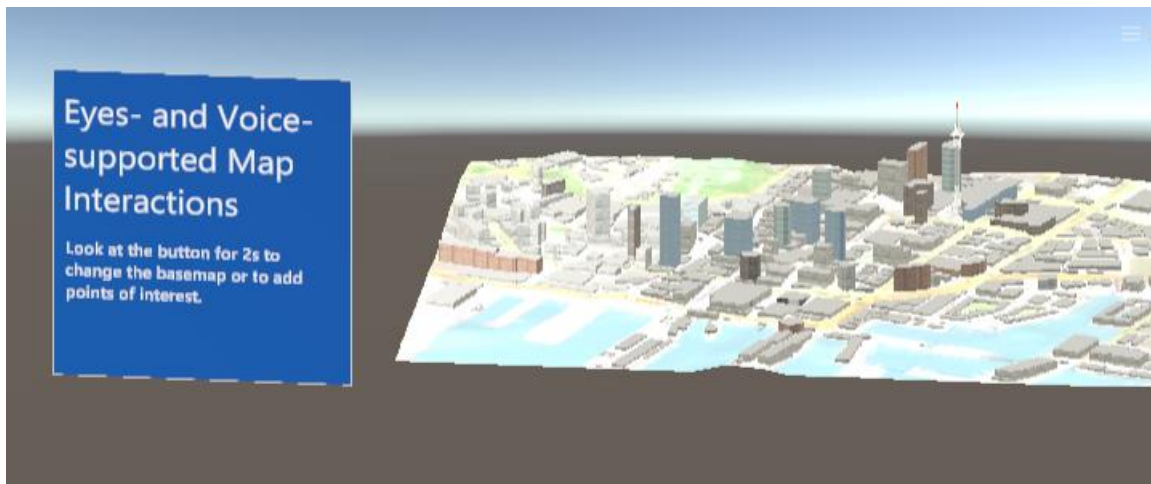
By looking at the buildings, the retrieve interaction is triggered. The building's name appears on the focus enter and vanishes on the focus exit. An example of the triggered retrieval is shown in Figure 26. Interestingly, the EyeTrackingTarget script is not utilised here because the MRTK ToolTipSpawner uses the "OnFocus()" function that accepts different kinds of focus no matter gaze or hand-ray. Only three buildings on the map have the retrieve function, with which the user is familiarised during the experiment.



*Figure 26: Retrieve interaction triggered for the building*

## Interface 2

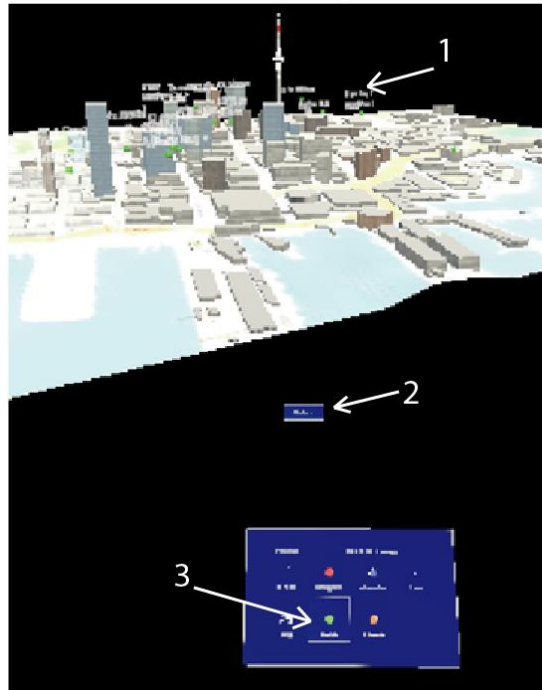
During developing the gaze-aware interface, some challenges were encountered, details of which are discussed in the Challenges and limitations sub-chapter of the Results and discussion chapter. The voice control does not apply to the rotate and retrieve interactions, whereas the overlay interaction is triggered by the voice commands. The final eyes-voice scene presents the city model, the instructions panel, and the Overlay menu (Figure 27). The featuring interactions are overlay and retrieve.



*Figure 27: Scene for the eyes-voice controlled interactions*

The names of the buttons on the Overlay menu correspond to the voice command that triggers the button activation. The user says the name of the button aloud, and the floating voice-command confirmation appears as the visual cue (Figure 28). The text

of the confirmation is the same as the spoken command. As in the previous interface, the button is highlighted when the user looks at it.

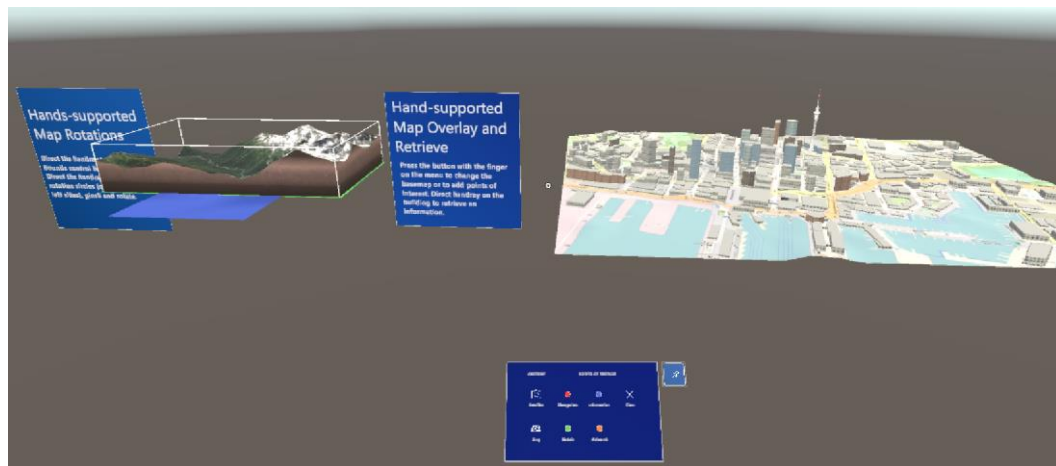


*Figure 28: The second interface - overlay interaction: 1) activated hotels, 2) floating visual cue of the voice-command, 3) highlighted button as the visual cue of activation*

The terrain model was excluded from the scene because the model and the rotation interaction were not changed during the second interface development. In addition, the retrieve interaction functions the same way as in the previous interface but is kept in the scene for evaluation purpose. Therefore, the interface does not feature the mixed-modality interactions but rather the mixed-modality interface.

### Interface 3

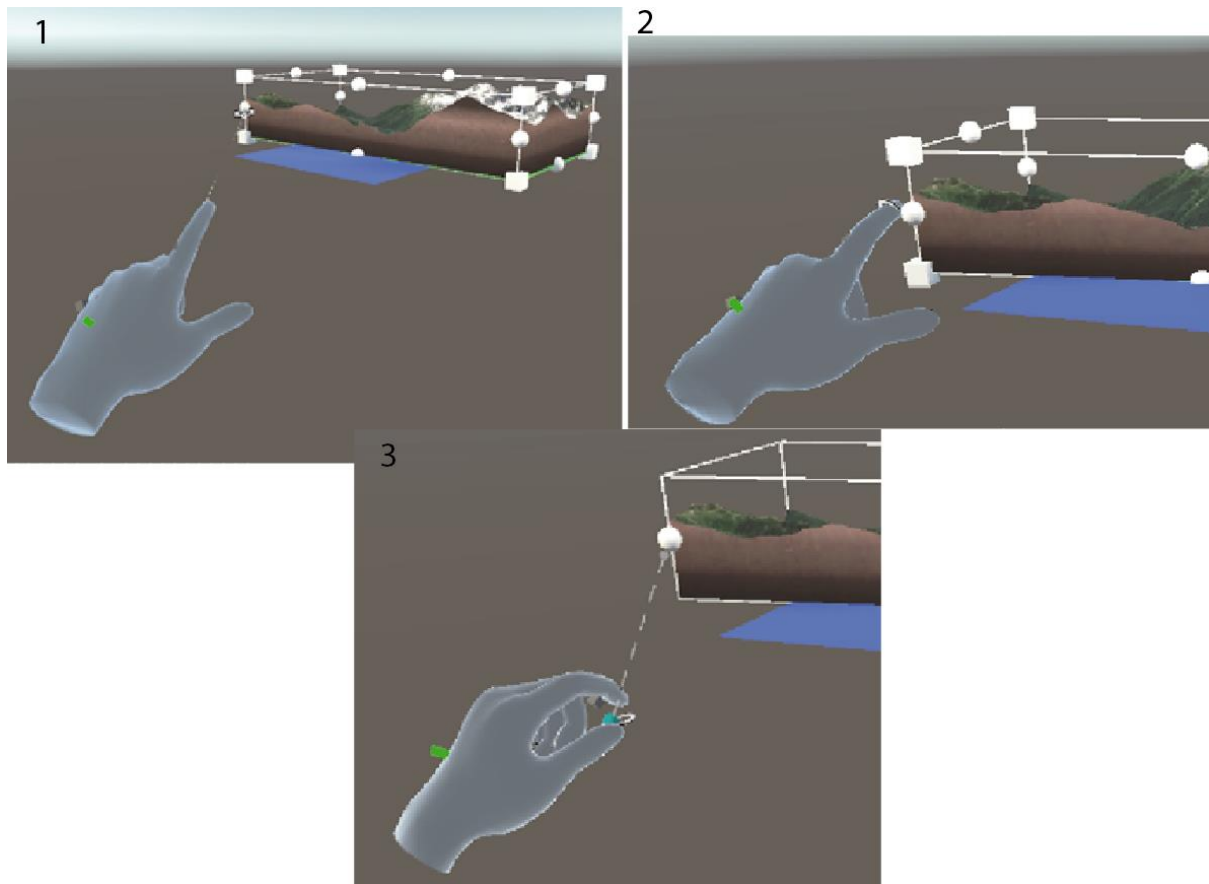
Finally, the third interface uses one hand (no matter right or left) to control the interactions. The scene contents (Figure 29) are the same as in the first interface, but the interactions and instructions are different. The interface features all three interactions due to already established hands interactions' design and functionality for the HoloLens 2. However, the terrain manipulations were not restricted to the horizontal rotations, but only the horizontal rotations were evaluated. That is discussed in the Challenges and limitations sub-chapter of the results.



*Figure 29: The scene of hands-controlled interactions*

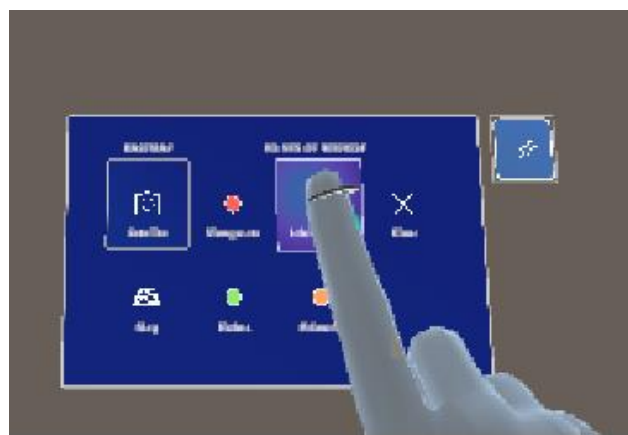
The rotation interaction has two modes: far and near interaction. The far interaction is used when the user is relatively far from the object. The far interaction casts the ray of the hand (hand-ray) to show the hand direction. The pointer (white circle) on the end of the ray defines where the hand-ray reaches. On the contrary, the near interaction activates when the user's hand is near the collider of the target. The ring on the index finger defines the triggering zone, and the hand-ray appears when the user is using the far interaction again.

The proximity of the pointer (no matter far or near interaction) to the terrain's collider activates the bounding box used for the object manipulations. The bounding box has the spherical and cubical elements that define the rotations and scaling, respectively. To rotate the object horizontally, the user needs to focus on one of the left/right sides' spheres, pinch and move the hand to the selected direction. When focusing on the spheres, a small arrows icon appears, indicating the opportunity of rotating. The pinch gesture is shown in Figure 30.



*Figure 30 Hands-controlling interface  
1) far interaction with hand-ray, 2) near interaction with the index finger ring, 3) pinch gesture*

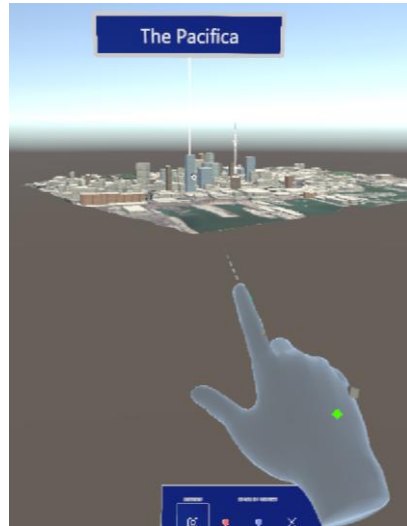
The overlay interaction uses the same Overlay menu. Additionally, the menu, by default, already has some animations added to the buttons. The user needs to press the hologram button from the Overlay menu as if there was a physical one. This gesture is defined in this work as air-press and is shown in Figure 31. The ring on the index finger indeed represents the near interaction.



*Figure 31 Hands-controlled interface: air-press, overlay*



Lastly, the retrieve interaction in the conventional interface uses the same script as the gaze-based interface to show the names (ToolTipSpawner). As mentioned previously, the script uses the focus position regardless of the input method. Again, the far interaction is used here, and the hand-ray defines the target (Figure 32).



*Figure 32 The retrieve interaction of the hands-controlled interface: using hand-ray*

#### 3.2.4. User study setup

##### *Experiment*

The experiment starts with the user filling the pre-study forms: health and hygienic rules forms, the informed consent (Appendix 1, section 1) and the background check questionnaire (Appendix 1, section 2). The background form checks the user's experience with maps and the MR devices and whether the user is wearing glasses/contact lenses.

The personal computer runs for the remote control (Microsoft HoloLens application) over the MR headset, facilitating user coordination. As the pre-requisite for the experiment, the user calibrates eyes with the HoloLens 2 eyes calibration application. The calibration verifies the user gaze to be utilised in gaze-based and gaze-aware interactions.

The application experience starts, and the user explores the first interface, reading the instructions and coordinated by the researcher. Indeed, user is allowed to walk around the map models, come closer or further at any stage of the experiment. Then the experiment proceeds with the tasks part, the first interface is reloaded to the initial state, and the recording of HoloLens 2 view begins. The view recording is controlled

from the computer and utilised in the evaluation stage for measuring task completion. The researcher announces the first task for the first interface and allows the user to begin with the “Start” command. The user indicates the accomplishing of the task by saying “Done”. Finally, the recording ends when all the tasks for the first interface are accomplished. The same procedure applies to the second and third interfaces.

The rotate, overlay, and retrieve tasks follow the same pattern for the three interfaces. The examples for the interaction tasks are:

- rotate the terrain to the task position (indicated by the blue rectangle below the terrain (Figure 23, 4))
- turn on the satellite view and add the hotels
- activate the name tag for the Sky Tower

The rotate task is the same for the gaze-based and conventional interface (the gaze-aware interface does not include the rotation interaction). The overlay task involves activating the layers from the Overlay menu and is changed for different interfaces. For example, the user is asked to add viewpoints or other POIs instead of the hotels. Any of the three buildings is chosen to retrieve the name tag for the last task.

Finally, the survey starts when the user is familiar with the three interfaces and accomplishes all the tasks. At this stage, the user fills the post-study questionnaires with the task load, user experience, interface ranking, and open-ended questions.

### *Survey*

The questionnaire forms are digital and filled in using the researcher's tablet. Thus, data collection is automated, and it is not needed to convert the data into digital format afterwards. Four questionnaires are used in this study: background check, task load, user experience, and interface ranking questionnaires.

The task load questionnaire is the adapted version of the NASA TLX with fewer questions. It evaluates the mental and physical load on the user while accomplishing the tasks.

The user experience questionnaire is adapted from the UEQ. The adaption targets reducing the time required for filling the form and excluding the confusing terms that



can seem similar to the user. The researcher strives to have an insight into the user's impression from the interfaces and interactions.

The interface ranking form allows users to formulate their interface preferences after the experiment. Additionally, it included open-ended questions regarding the experience challenges. Users can write about any difficulties they encounter and make suggestions for the interfaces' design and performance.

### 3.3. Data analysis

Data that has been collected in the user study were recordings of user task completion (qualitative) and questionnaires (quantitative and qualitative). The performance is evaluated using the task completion time. The user task completion time is measured from the HoloLens view recordings from the "Start" to "Done" command. The questionnaires were conducted by using an online tool. The obtained data was structured, filtered and analysed. The results of the open-ended questions were grouped by the common features and visualised.

## 4. Results and discussion

Twenty-four users took part in the experiment and survey. Unfortunately, two test records are not included in the evaluation due to technical issues during the user study: 1) the task accomplishing and 2) the survey forms were not recorded for these users. Therefore, twenty-two user test results are evaluated. The majority of participants (77%) use maps often, and the rest of them use maps sometimes. Moreover, 54 per cent of users are either an expert or experienced in working with geospatial information. However, only six people (27%) have had experience with an MR device once or several times, and one person uses it often. The majority of the assumptions described in the discussion paragraphs of the Performance, User experience, and Interface ranking sub-chapters can be supported by the Further statements and Challenges and limitations sub-chapters.

### 4.1. Performance

The performance evaluation is based on the task completion time. The extremums and the mean value of the users' task completion time are presented in Table 6. The observations can be formulated as follows:

- The overlay interaction tasks were accomplished the fastest using the gaze-aware interface. The average completion time with the gaze-aware interface is 13,68 seconds. The highest mean value (15,43s) and the maximum overlay task completion time (66,25s) was performed using the conventional interface.
- The retrieve interaction task has similar results for the gaze-based and conventional interfaces. However, the average retrieve task completion is faster when using the gaze-based interface.
- The average completion time for the rotation task significantly differs (10s) between the gaze-based and conventional interfaces. The gaze-based control, on average, offers faster target rotation (20,02s).

Table 6: Overlay, retrieve, and rotate tasks' completion time in seconds

		Overlay	Retrieve	Rotate		Overlay	Retrieve	Rotate		Overlay
Min	Gaze-based	7,28	4,48	6,9	Conventional	6,09	4,82	8,66	Gaze-aware	5,68
Max		39,33	24,32	54,62		66,25	38,12	85,66		23,56
Mean		14,27	11,33	20,02		15,43	12,38	30,34		13,68

The violin-boxplot visualisation (Figure 33) shows the distribution of the task completion time values. On the visualization the gaze-based interface is labelled as the eyes control, the gaze-aware interface as eyes and voice control, and the conventional interface as the hands control. The outliers and the “main body” of the violin give a better insight into the performance differences. The visualisation provides with the following observations:

- Even though the fastest results (min and mean) of the overlay task were performed using the gaze-aware interface, the majority of the gaze-aware observations have higher values than the majority of the gaze-based or conventional interfaces. However, the distribution is more consistent for the gaze-aware interface.
- The similarity in the retrieve task results from gaze-based and conventional interfaces is apparent. Both interfaces have outliers.
- The dispersion of the observations is prominent for the rotation task results.
- The hands-controlled interface has a more significant variation in the results of all the interaction tasks.

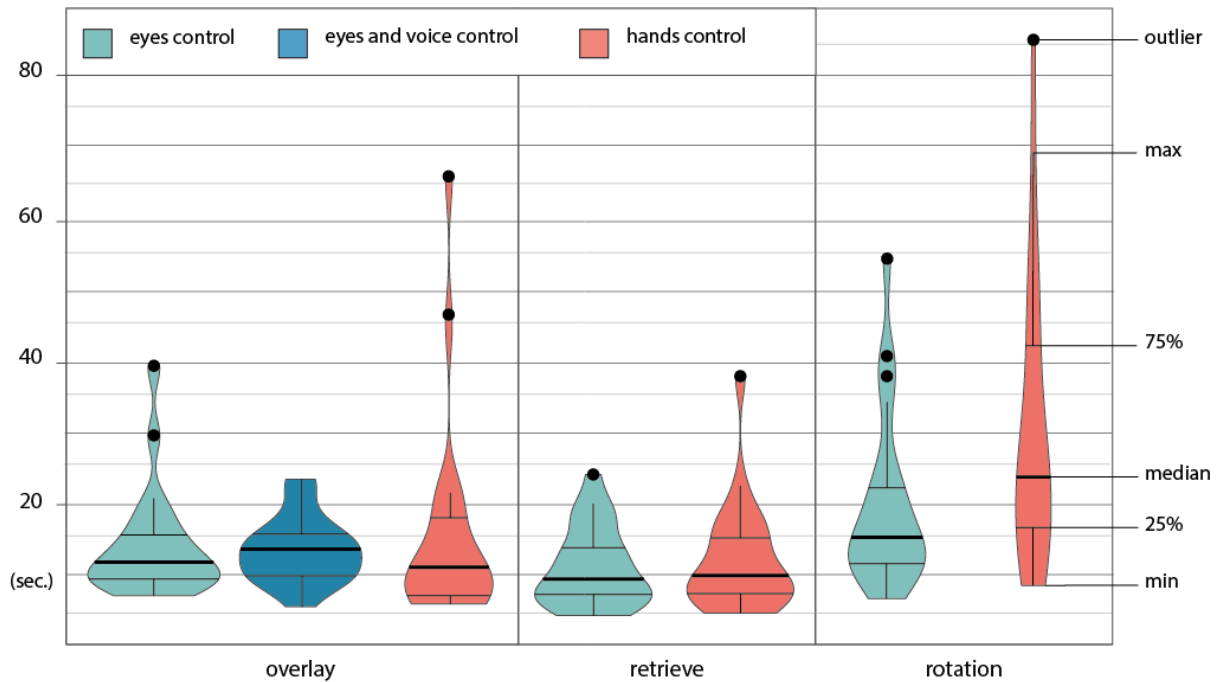


Figure 33: Violin-boxplot visualisation of the tasks' completion time in regards to interfaces

The eyes-control for the overlay interaction has given a middle result between gaze-aware and the conventional interface. The control method can not be called slow compared to others because most of the observations in the second quartile have lower values than the gaze-aware interface. The outliers can occur due to unintentional activation of other buttons, the Midas touch. The dwell time was used for the gaze-based interface to address the problem. However, the more individual approach to assigning the dwell time can be helpful due to the users' possible attention differences.

Using eyes-voice control for the overlay interaction has given the fastest and relatively consistent results. The consistency of the eyes-voice controlled results can occur due to the time needed for the voice-command processing. The gaze-aware interface's high task completion time values can arise if the user needs to repeat the voice command. The repetition can be required if the user says the voice command in a sentence or without pauses.

The longest time to perform the overlay tasks using hands can be due to the air-press gesture. During the exploration stage, the gesture was explained to the users and tested. Nevertheless, some users confused the press gesture with the tap. Another aspect of it can be the orientation in the MR environment. For example, some users pressed the buttons with their hands being under the overlay menu hologram.

The gaze-based retrieve interaction shows similar results with the conventional retrieve interaction, possibly because both methods use pointing and not gestures. The high task completion values can occur due to the city model and gaze/hand-ray pointer occlusion issues. The objects in the MR environment have colliders to give the physics experience (collision, occlusion, etc.). Usually, the colliders are around the same size as their objects for more realistic physics. The HoloLens 2 predicts the gaze position within 1,5 degrees in visual angle around the target. That means the gaze prediction area increases with the distance between the user and the target (sostel, n.d.). However, colliders can be small regarding the distance between the user and the target. Therefore, it can be challenging to point at a small target. A similar scenario applies to the hand-ray pointer, whose radius increases with the distance (scooley, n.d.-a).

The eyes-controlled rotation presents fast task completion relatively to the hands-controlled performance. The outliers and high values possibly occur due to the user not being prepared to stop the rotation. The rotation happens when the user is looking at the sides of the model. Therefore, to stop the rotation, the user needs to either look at the centre of the model or look away. That is possibly not the intuitive way to stop the interaction as the user is looking at the model to ensure its correct position. However, this interaction can be more relevant for the map responsive design. For example, the map model can rotate towards the user to provide a better view of the side to which the user is reaching.

The hands-controlled rotation shows dispersed task results. The manipulation gesture can be confusing for people who do not use the HoloLens 2 often. The rotation gesture involves pointing, pinching, and moving. Users can unintentionally mix these steps, which leads to unsuccessful attempts. For example, some users pinpoint the target while holding the hand in the closed pinch state, which is not the correct gesture performance. These gestures are the default for the manipulation tasks in the HoloLens and imply the learning curve (scooley, n.d.-a).

## 4.2. User experience

### 4.2.1. User Experience Questionnaire

The user experience questions from the post-study survey allowed users to assign points to the interfaces regarding the different qualities. Figure 34 shows the mean

values for the results of the UEQa. All the mean values are higher than 3.5 points. The more extensive shape area visualised for the interface implies an overall better experience. The user experience questionnaire identifies the following main facts:

- All the interfaces are exciting to the users.
- The eyes and voice control interface is evaluated as the most enjoyable, practical, understandable, and easy to learn. Thus, it is rated superior to other interfaces by almost all the features except for being less inventive than the gaze-based interface.
- Users choose the gaze-based interface as the most inventive. The rest of the qualities are rated moderately relatively to the other interfaces, except where the excitement is slightly lower.
- The hands-controlled interface has the lowest evaluation relative to the other two interfaces. It is inferior to other interfaces in almost all characteristics, except for being highly exciting as the gaze-aware interface and less practical as the gaze-based interface.
- The overall experience of all three interfaces is positive as the average values are higher than the median value (3).

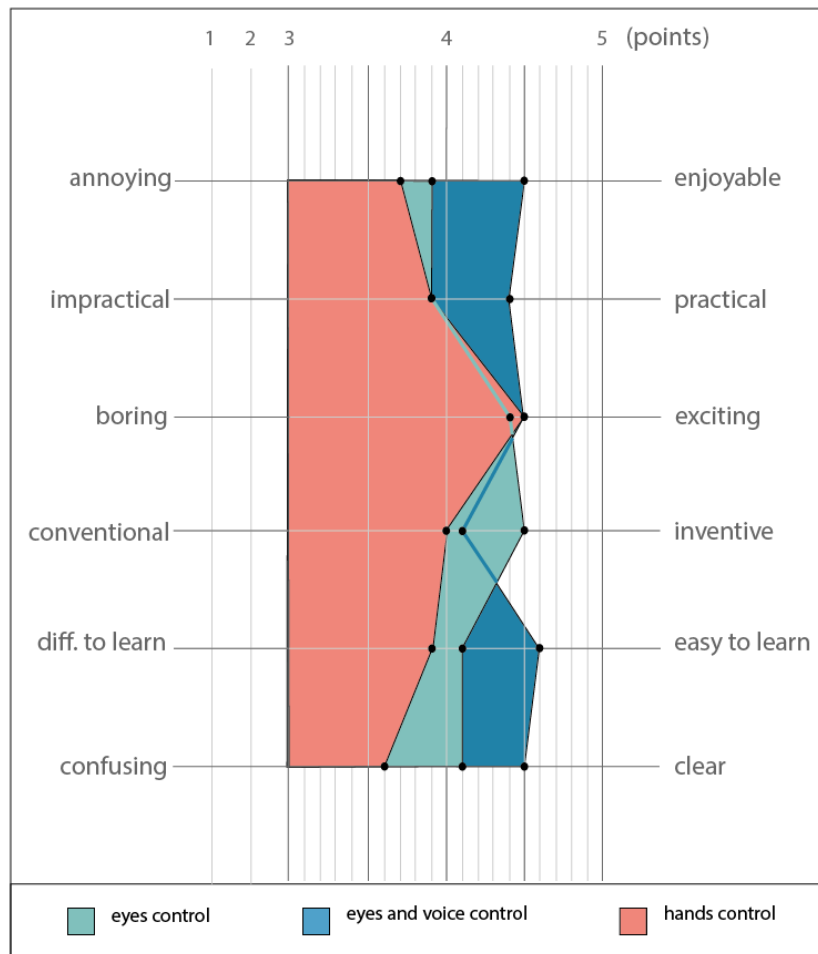


Figure 34: Visualisation of UEQa results' mean values

As most of the participants were using the MR device for the first time, they were excited to test the interfaces. The difference in the average excitement rate is only 0,1 point. However, it is worth mentioning that different activity preferences can influence the degree of excitement. For example, the person who prefers a passive state to an active state would probably like the gaze-based interactions more and on the opposite.

During the experiment, the gaze-aware interface was introduced as the interface that uses two modalities and has two interactions: overlay and retrieve. However, as the retrieve interaction is the same with the gaze-based interface, only the overlay task is given. Nonetheless, some users possibly evaluated the gaze-aware interface regarding the voice-controlled overlay interaction, but not in combination with the eyes-controlled retrieve interaction. Therefore, part of the gaze-aware interface experience results can be the rating of the voice-controlled overlay interaction.

In the current technologies such as smartphones, personal computers, personal assistants (Alexa), etc. the speech recognition is commonly utilised. Additionally, the

hands-control is frequently used for MR devices. These factors can influence users' perception of the "inventive" quality. Therefore, the gaze-based interface can be rated as the most inventive. However, overall the gaze-based interface rating is moderate in comparison to the others.

The user interface control with hands has received the lowest rates relative to the gaze-based and gaze-aware interfaces. Again, the gesture learning process can drive this. Some users can find it challenging to learn hand gestures. Therefore, they can rate the interface as the most difficult to learn and understand within the three interfaces. In addition, the annoyance finds its place in the ranking.

#### 4.2.2. Task load

The task load in this study is characterised by mental and physical demand, frustration (Figure 35), effort (Figure 37), and how the users evaluate their performance (Figure 36). The following figures present the percentage of "the least/the most/moderate" answers to the questions (Appendix 2) per interface and interaction. The prominent observations from visualisations are:

- The hands-controlled interface is rated as the most mentally and physically demanding interface and the eyes-voice interface - as the least.
- The frustration rating is not as apparent as the demand results. However, the hands-controlled interface being the most frustrating is still prominent. The eyes-voice interface is rated as the least frustrating.
- Again, the gaze-based interface is rated moderately in all three aspects.



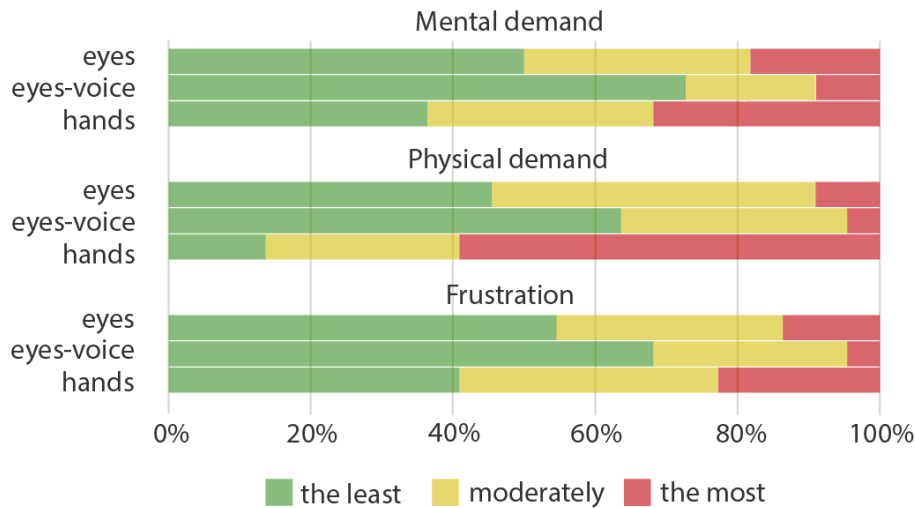


Figure 35: Task Load results: mental demand, physical demand, and frustration

- The majority of the users consider their hands-controlled overlay performance as the most successful and the eye-controlled overlay performance as the least successful.
- The retrieve task controlled by hands is considered to be the most successful.
- The rotate task completion, on the opposite, is considered to be the most successful when using eyes-control.

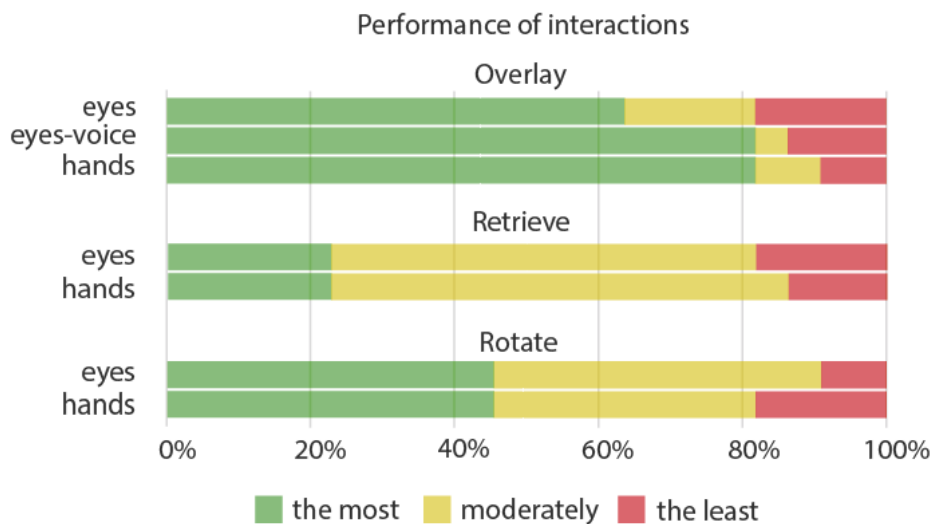


Figure 36: Task load results: users evaluate their performance; the most, moderately, and the least successful

- The level of overlay task performance mentioned above was less hard to achieve with the gaze-aware interface. Eyes- and hands-control seem to require more effort, but on the same level for both
- The rotation required less effort when interacting with the eyes.

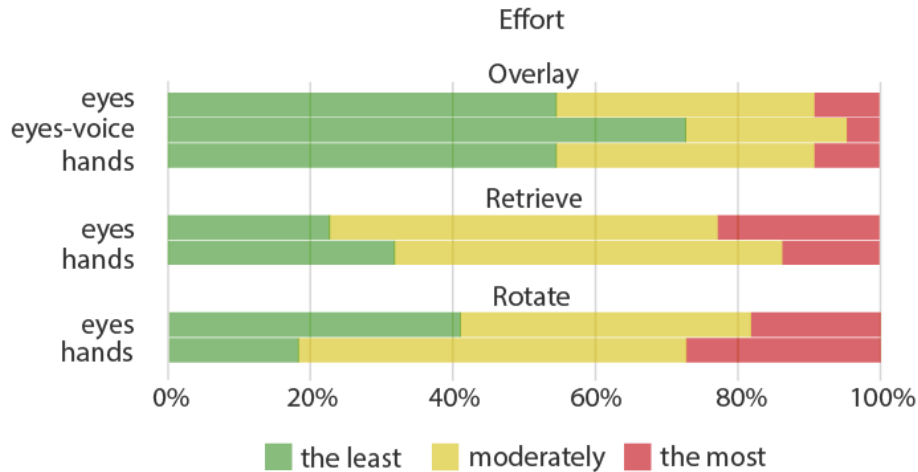


Figure 37: Task load results: the effort spent to accomplish the user's level of performance

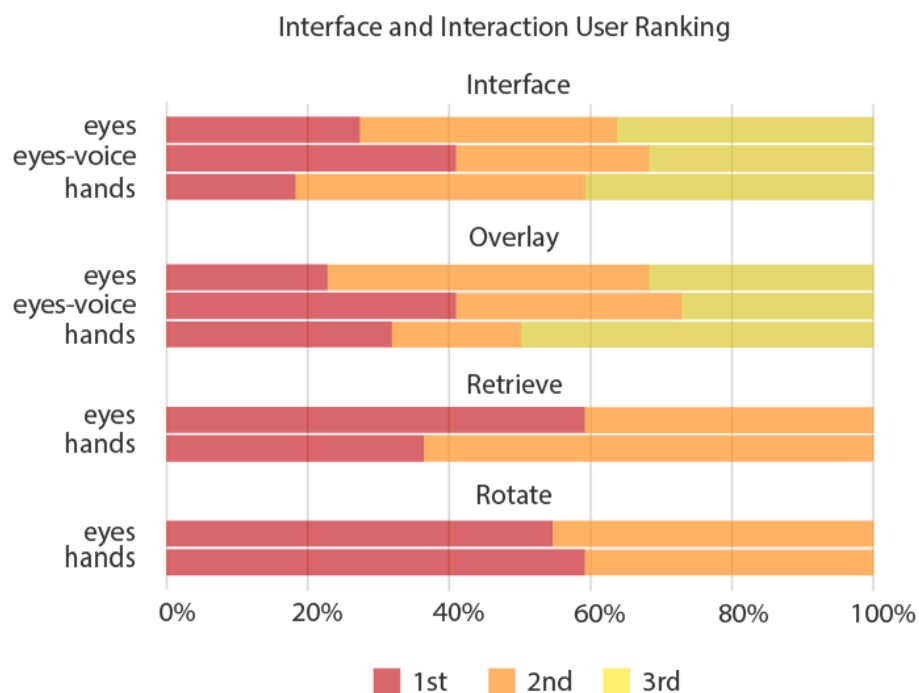
The high mental and physical demand for the hands-controlled interface is noted by users. Similarly to the UEQa, these results can address the learning curve of the hands-controlled interactions. That is supported by the relatively high effort users spent accomplishing the rotate task with the relatively low performance. The overlay interaction performance is considered the most successful using hands as controls but requiring more effort than the gaze-aware interface. The results of the TLXa also present the frustration for the hands-controlled interface that is probably based on mental and physical demand.

The overall task load results for the gaze-aware interface are positive, requiring the relatively lowest levels of demands and effort. The overlay performance is considered moderately successful regarding the hands-interactions. These results can be based on less necessity to learn how to perform voice commands. Additionally, as mentioned before, this evaluation is possibly done by users for the voice-controlled overlay interaction and generalised to the interface results.

The gaze-based interface is rated as requiring a moderate level of mental/physical demand and frustration. The performance success is considered as the lowest within the three interfaces with the comparatively low effort for the rotation and high effort for the retrieve tasks. This can be explained by the users' effort to focus on the small building colliders and the gaze slightly offset position due to the distance.

### 4.3. Interface ranking

Overall, users have ranked their preferences regarding the interface and interactions and awarded the first, second, and third places. Figure 38 presents the results of interface and interaction ranking. The gaze-aware interface is the favourite of the three interfaces, winning in the overlay interaction preference competition. The second preferred interface is the gaze-based interface, being ranked the first choice for the retrieve interaction. The third place takes the hands-controlled interface, nonetheless winning the rotate interaction competition by five percent.



*Figure 38: Results of the Interface Ranking questionnaire  
1st, 2nd , and 3rd places awarding; x-axis – percentage of participant answers; 100% - answers of 22 users*

The preferred option for the overlay interaction can be explained by less effort that voice commands require and the overall user experience results. The gaze-based interface being favourite for the retrieve interaction can be supported by less effort, physical/mental demands, and other TLXa and UEQa results described above. The excitement of using hands can probably support the rotation choice. However, the enthusiasm alone does not seem to be enough to overweight the task load of the hands

interactions. That leaves the third place in the users' interface ranking to the conventional interface.

#### 4.4. Further statements

Thirteen out of twenty-two participants wore glasses and contact lenses (1 user). The post-study survey starts with the question asking to evaluate the responsiveness of the eye-cursor (eye-pointer). Before the experiment majority of the users had asked if they could wear glasses while wearing the HoloLens 2. Figure 39 shows the evaluated cursor performance regarding wearing the glasses.

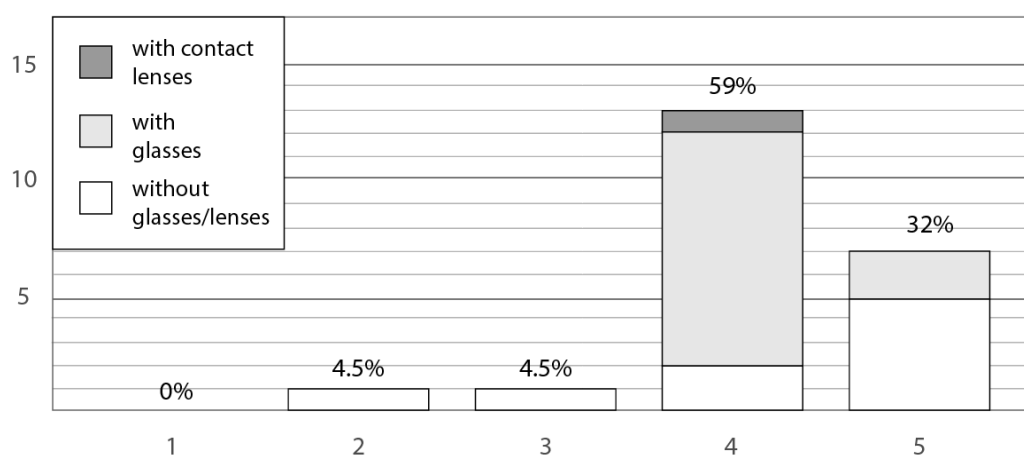


Figure 39: Eye-cursor responsiveness evaluated by participants without/with glasses or contact lenses: x-axis - the eye cursor responsiveness in points; y-axis - number of participants

83% of the users wearing glasses voted for the four points responsiveness. However, the lowest ratings (2-3 points) belong to users not wearing glasses or contact lenses. Therefore, a well-supported conclusion about the influence of wearing glasses or contact lenses on the MR experience cannot be done. However, the general overview can state that most users ranked the eye-cursor responsiveness as good or very good (4-5 points). Notably, the HoloLens 2 documentation mentions the possibility of using the device while wearing prescription glasses (scooley, n.d.-b).

The survey's open-ended questions allow users to define the difficulties they encounter and the suggestions for interactions and interfaces. Figure 40 and Figure 41 present the word clouds of all the mentioned user difficulties and suggestions for all the interfaces. The ratio of the terms in the word cloud defines the frequency of them appearing in the results.

Thirty-six per cent of the users have reported difficulties with the gaze-based interface. In contrast, the conventional interface received difficulties reports from seventy-two per cent of users. Additionally, only nine per cent of the users reported difficulties with the gaze-aware interface. The most referred features that implied the challenges are the hand-ray control, pinch, focus, bounding box, and feedback.



*Figure 40: Difficulties that users mentioned in the open-ended questions:  
red shades – hands- interface; green shades – eyes- interface; blue shades - eyes-voice interface*

The described difficulties with the handray control imply the map occlusion, target size, and handray offset issues. The occlusion issue happens due to the model limitations. The city model has the colliders only for the retrieve interaction objects (three buildings). The colliders give the physics experience (collision, occlusion, etc.) in the MR. That means the other objects do not present an obstruct for the pointers/rays. Therefore, the pointers/hand-rays go through those objects extending until meeting any obstruct (e.g. the wall). This might be confusing to the users. Additionally, the relatively small size of the target regarding the distance can make the pointing challenging due to the increasing size of the pointer with the distance (scooley, n.d.-a). In the experiments, the buildings and the rotation spheres can be the small targets. Finally, a notable handray offset is mentioned by several users.

The pinch hand gesture for the rotation interaction was described as requiring time to understand how it works and working from time to time. In addition, the bounding box issue was addressed to the hands' interactions but related to the terrain model

limitations. Thus, the users were confused about which elements of the bounding box to pinch for the rotation interaction due to the bounding box elements not being restricted to only rotation.

The gaze-based difficulties imply mostly the focus issue. Again, it can be challenging for users to focus on the buildings to retrieve the names due to the small size. Additionally, the possible offset in gaze position (sostel, n.d.) can affect the user experience.

There are other difficulties mentioned by some users related to the gaze-based interface: menu, feedback, calibration, and brightness. The overlay menu follows the user relative to the head position. Thus, the menu can sometimes cross the maps. That can be addressed by improving the collision aspect of the interfaces. The feedback (visual and audial cues) were designed for the gaze-based interface. However, an occasional technical interruption (the root of this issue is not defined) happened in the work of the application that led to the audio and visual cues not being activated. Therefore, that was reported in the results. Moreover, not only the technical application issues can be worth mentioning, but the design of them. For example, the button pressing in the HoloLens has a predesigned feedbacks set, including animation. In contrast, the designed interaction feedback for the gaze-based and gaze-aware interfaces in this user study is not as perceptible. Accordingly, the gaze-aware interface received a similar feedback concern. The last difficulty reported for the gaze-aware interface noted that the user's voice was not recognised. The voice recognition, in some cases, worked intermittently. That can be due to the voice commands being a part of a sentence or said in a low voice. Additionally, one can expect a quicker response from the interface, hence saying the same voice command several times in a row, causing the extra load for the system.

The second word cloud features the multi-modality as the most often mentioned suggestion. Users suggest combining all three modalities to improve usability. For example, adding hands-control to the eyes-voice interface and using hands for the overlay and eyes-voice for the other interactions. One user noted the preference of combining hands and voice modalities, supporting it with the following sentence:

*“I think using gaze is quite nice, but maybe people are not used to this type of interaction yet. The test results might be biased with their previous experience and their personalities ( if one likes to try new experience or not).”*



*Figure 41: User suggestions from open-ended questions*

The other words imply the design suggestions, such as “using larger colliders” for the retrieve task, improving the labels’ (hotels, name tags) visualisation, using a more extensive scale of the map, and applying a better menu placement. Another critical commentary suggests elongating the dwell-time.

#### 4.5. Challenges and limitations

The city map model has colliders for the retrieve interaction objects. Only three prominent buildings on the map were chosen to test the retrieve interaction and not to overload the map. The rest of the city map objects, including base maps, do not have the colliders. That can limit the user experience of this interaction due to the pointers not having enough obstructions and going through the objects that do not have colliders. Moreover, due to following the user, the overlay menu can cross the map when the user is near the model.

The main challenge in the gaze-aware interface development was combining the gaze and voice modality for the interactions. The final gaze-aware interface has the retrieve and the overlay interactions, the retrieve being gaze-controlled and the overlay being voice-controlled. The rotation interaction is not included in the gaze-aware interface because no changes are applied to the terrain model. Therefore, that can present a limitation for the comparative evaluation of the three interfaces. Additionally, the visual and audial feedback design for the gaze-based and gaze-aware interfaces is not

as detailed as the default feedbacks of the conventional (hands-controlled) interface. That can influence the user experience results.

The conventional (hands-controlled) interface evaluation limitation is that interactions with the terrain are not restricted to the horizontal rotation but also features vertical rotation and scaling. During the survey, users evaluate only the horizontal rotation. However, the additional elements of the terrain bounding box (spheres for the vertical rotation and cubes for the scaling) can lead to user confusion and complicate the task.

Finally, the application has its limitations: the experience works correctly only in the particular order of the interfaces (1st, 2nd, and 3rd). Thus, the experiment follows the same order of the interfaces for every user. That can influence the results of the interfaces' evaluation.





## 5. Conclusion and outlook

This research achieves the main objective of developing gaze-based user-map interactions in three steps:

Firstly, identifying cartographic interactions for the gaze control in the MR environment and determining overlay, retrieve, and rotate interactions for the implementation;

Secondly, assembling gaze-based, gaze-aware, and conventional MR interfaces for the selected interactions. The gaze-aware interface uses gaze and voice control, and the conventional interface uses hands control;

Finally, evaluating the performance and user experience of the built interfaces and answering the research questions.

The first research question is answered by defining the fundamental cartographic, gaze-based, and MR interaction in the Theoretical background chapter.

The second research question is answered by analysing the development process and the open-ended questions from the survey:

*How are MR interfaces assembled for the selected gaze-based interactions?*

a) *What are the limitations of the developed gaze-based interfaces?*

The limitations of the developed interfaces are the incomplete maps' design, the gaze-aware interface using gaze and voice modalities separately, instead of combined, and the constant order of the interfaces during the experiment.

b) *What are the challenges in the development of the selected gaze-based interfaces?*

The challenge encountered while developing the subset of interfaces for the gaze-based cartographic interactions evaluation was the assembling of the gaze-aware interface with combined gaze and voice modalities.

The third research question is answered by analysing the results of the experiment and survey:

*How effective are the implemented user-map interactions in MR?*

*c) What is the performance of the assembled interfaces?*

The performance of the gaze-based interface is defined by the overlay, retrieve, and rotate tasks completion time. The gaze-based overlay interaction has a higher performance than the conventional hands-controlled overlay interaction; however, the lower performance compared to the interaction with the voice modality. The gaze-based retrieve interaction has the fastest task completion time compared to the hands-controlled interface. Again, the gaze-based rotation allows for faster task completion than when using hands.

*d) What is the user experience with the implemented user-map interactions in the MR environment?*

The gaze-based interface is considered the most inventive interface within the three assembled interfaces. However, it is evaluated as more enjoyable, easier to learn, and less confusing than the conventional hands-controlled interface. Nonetheless, the gaze-based interface is inferior to the gaze-aware interface in the same qualities. Moreover, the gaze-based interface is evaluated as requiring more mental and physical demand and effort than the gaze-aware interface; however, less than the conventional interface requires.

Future research can be directed to improving the usability of gaze-based user-map interactions. The multimodality, as suggested by the users, can be explored for gaze-based interactions. The limitations and challenges mentioned in this research can be addressed to improve the overall experience of gaze-based cartographic interactions.

## Bibliography

- Adams, N., Witkowski, M., & Spence, R. (2008). The inspection of very large images by eye-gaze control. *Proceedings of the Working Conference on Advanced Visual Interfaces*, 111–118.
- Aladin, M. Y. F., Ismail, A. W., Ismail, N. A., & Rahim, M. S. M. (2020). Object selection and scaling using multimodal interaction in mixed reality. *IOP Conference Series: Materials Science and Engineering*, 979(1), 012004.
- Amar, R., Eagan, J., & Stasko, J. (2005). Low-level components of analytic activity in information visualization. *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, 111–117.
- Andrienko, N., Andrienko, G., & Gatalsky, P. (2003). Exploratory spatio-temporal visualization: An analytical review. *Journal of Visual Languages & Computing*, 14(6), 503–541.
- Bachmann, D., Weichert, F., & Rinkenauer, G. (2018). Review of three-dimensional human-computer interaction with focus on the leap motion controller. *Sensors*, 18(7), 2194.
- Bates, R., & Istance, H. (2002). Zooming interfaces! Enhancing the performance of eye controlled pointing devices. *Proceedings of the Fifth International ACM Conference on Assistive Technologies*, 119–126.
- Bekele, M. K., Pierdicca, R., Frontoni, E., Malinverni, E. S., & Gain, J. (2018). A survey of augmented, virtual, and mixed reality for cultural heritage. *Journal on Computing and Cultural Heritage (JOCCH)*, 11(2), 1–36.
- Bektaş, K., & Çöltekin, A. (2011). An approach to modeling spatial perception for geovisualization. *Procedia-Social and Behavioral Sciences*, 21, 53–62.
- Billinghurst, M., Kato, H., & Myojin, S. (2009). Advanced interaction techniques for augmented reality applications. *International Conference on Virtual and Mixed Reality*, 13–22.
- Bryman, A. (2016). *Social research methods*. Oxford university press.
- Buja, A., Cook, D., & Swayne, D. F. (1996). Interactive high-dimensional data visualization. *Journal of Computational and Graphical Statistics*, 5(1), 78–99.
- Bulling, A., Ward, J. A., Gellersen, H., & Tröster, G. (2011). Eye Movement Analysis for Activity Recognition Using Electrooculography. *IEEE Transactions on*

- Pattern Analysis and Machine Intelligence*, 33(4), 741–753.  
<https://doi.org/10.1109/TPAMI.2010.86>
- Chalon, R., & David, B. T. (2004). IRVO: An Architectural Model for Collaborative Interaction in Mixed Reality Environments. *MIXER*.
- Che Hashim, N., Abd Majid, N. A., Arshad, H., & Khalid Obeidy, W. (2018). User satisfaction for an augmented reality application to support productive vocabulary using speech recognition. *Advances in Multimedia*, 2018.
- Chi, E. H. (2000). A taxonomy of visualization techniques using the data state reference model. *IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings*, 69–75.
- Cho, D.-C., & Kim, W.-Y. (2013). Long-Range Gaze Tracking System for Large Movements. *IEEE Transactions on Bio-Medical Engineering*, 60.  
<https://doi.org/10.1109/TBME.2013.2266413>
- Chuah, M. C., & Roth, S. F. (1996). On the semantics of interactive visualizations. *Proceedings IEEE Symposium on Information Visualization'96*, 29–36.
- Coltekin, A., Fabrikant, S., & Lacayo-Emery, M. (2010). Exploring the efficiency of users' visual analytics strategies based on sequence analysis of eye movement recordings. *International Journal of Geographical Information Science*.  
<https://doi.org/10.1080/13658816.2010.511718>
- Coltekin, A., Heil, B., Garlandini, S., & Fabrikant, S. (2009). Evaluating the Effectiveness of Interactive Map Interface Designs: A Case Study Integrating Usability Metrics with Eye-Movement Analysis. *Cartography and Geographic Information Science*, 36, 5–17. <https://doi.org/10.1559/152304009787340197>
- Cornsweet, T. (2012). *Visual Perception*. Academic Press.
- Couture, N., Rivière, G., & Reuter, P. (2010). Tangible interaction in mixed reality systems. In *The Engineering of Mixed Reality Systems* (pp. 101–120). Springer.
- Crampton, J. W. (2002). Interactivity types in geographic visualization. *Cartography and Geographic Information Science*, 29(2), 85–98.
- Davson, H. (1990). *Physiology of the Eye*. Macmillan International Higher Education.
- Different Kinds of Eye Tracking Devices*. (2020, June 12). Bitbrain.  
<https://www.bitbrain.com/blog/eye-tracking-devices>
- Drewes, H., & Schmidt, A. (2007). Interacting with the computer using gaze gestures. *IFIP Conference on Human-Computer Interaction*, 475–488.

- Duchowski, A. T., & Duchowski, A. T. (2017). *Eye tracking methodology: Theory and practice*. Springer.
- Evangelidis, K., Papadopoulos, T., & Sylaiou, S. (2021). Mixed Reality: A Reconsideration Based on Mixed Objects and Geospatial Modalities. *Applied Sciences*, 11(5), 2417.
- Fekri, A., & Wanis, I. (2019). A review on multimodal interaction in Mixed Reality Environment. *IOP Conference Series: Materials Science and Engineering*, 551, 012049. <https://doi.org/10.1088/1757-899X/551/1/012049>
- Fono, D., & Vertegaal, R. (2005). EyeWindows: Evaluation of eye-controlled zooming windows for focus selection. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 151–160.
- Foundation, B. (n.d.). About. *Blender.Org*. Retrieved 13 November 2021, from <https://www.blender.org/about/>
- Giannopoulos, I., Kiefer, P., & Raubal, M. (2015). GazeNav: Gaze-based pedestrian navigation. *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 337–346.
- Giannopoulos, I., Kiefer, P., & Raubal, M. (2012). GeoGazemarks: Providing gaze history for the orientation on small display maps. *Proceedings of the 14th ACM International Conference on Multimodal Interaction*, 165–172.
- Giannopoulos, I., Kiefer, P., & Raubal, M. (2013). Mobile outdoor gaze-based geohci. *Geographic Human-Computer Interaction, Workshop at CHI 2013*, 12–13.
- Göbel, F., Kiefer, P., & Martin, R. (2019). FeaturEyeTrack: Automatic matching of eye tracking data with map features on interactive maps. *GeoInformatica*, 23. <https://doi.org/10.1007/s10707-019-00344-3>
- Gray, H. (1918). Anatomy of the human body. *Annals of Surgery*, 68(5), 564–566.
- Haber, R. B., & McNabb, D. A. (1990). Visualization idioms: A conceptual model for scientific visualization systems. *Visualization in Scientific Computing*, 74, 93.
- Haklay, M. M. (2010). *Interacting with geospatial technologies*. John Wiley & Sons.
- Hanifa, R. M., Isa, K., & Mohamad, S. (2021). A review on speaker recognition: Technology and challenges. *Computers & Electrical Engineering*, 90, 107005.
- Hansen, D. W., & Hansen, J. P. (2006). Eye typing with common cameras. *Proceedings of the 2006 Symposium on Eye Tracking Research & Applications*, 55–55.

- Hansen, D. W., Skovsgaard, H. H., Hansen, J. P., & Møllenbach, E. (2008). Noise tolerant selection by gaze-controlled pan and zoom in 3D. *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications*, 205–212.
- Hashimoto, S., Ishida, A., Inami, M., & Igarashi, T. (2011). Touchme: An augmented reality based remote robot manipulation. *The 21st International Conference on Artificial Reality and Telexistence, Proceedings of ICAT2011*, 2.
- Heikkilä, H., & Rähkä, K.-J. (2010). Speed and accuracy of gaze gestures. *Journal of Eye Movement Research*, 3(2).
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Weijer, J. van de. (2011). *Eye Tracking: A comprehensive guide to methods and measures*. OUP Oxford.
- HoloLens 2 Review: The Best In Class For AR, But... (2021, August 12). *The Ghost Howls*. <https://skarredghost.com/2021/08/12/hololens-2-review/>
- HoloLens 2—Overview, Features, and Specs | Microsoft HoloLens. (n.d.). Retrieved 12 November 2021, from <https://www.microsoft.com/en-us/hololens/hardware>
- Huckauf, A., & Urbina, M. H. (2008). Gazing with pEYES: Towards a universal input for various applications. *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications*, 51–54.
- Istance, H., Hyrskykari, A., Immonen, L., Mansikkamaa, S., & Vickers, S. (2010). Designing gaze gestures for gaming: An investigation of performance. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, 323–330.
- Istance, H., Vickers, S., & Hyrskykari, A. (2009). Gaze-based interaction with massively multiplayer on-line games. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems* (pp. 4381–4386).
- Jacob, R. J. (1991). The use of eye movements in human-computer interaction techniques: What you look at is what you get. *ACM Transactions on Information Systems (TOIS)*, 9(2), 152–169.
- Jiawei, W., Li, Y., Tao, L., & Yuan, Y. (2010). Three-dimensional interactive pen based on augmented reality. *2010 International Conference on Image Analysis and Signal Processing*, 7–11.

- Kahn, D. A., Heynen, J., & Snuggs, G. L. (1999). Eye-controlled computing: The VisionKey experience. *Proceedings of the Fourteenth International Conference on Technology and Persons with Disabilities (CSUN'99)*.
- Keim, D. A. (2002). Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, 8(1), 1–8.
- Khamis, M., Hösl, A., Klimczak, A., Reiss, M., Alt, F., & Bulling, A. (2017). *EyeScout: Active Eye Tracking for Position and Movement Independent Gaze Interaction with Large Public Displays*. <https://doi.org/10.1145/3126594.3126630>
- Kiefer, P., & Giannopoulos, I. (2012). Gaze map matching: Mapping eye tracking data to geographic vector features. In *GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems* (p. 368). <https://doi.org/10.1145/2424321.2424367>
- Lankford, C. (2000). Effective eye-gaze input into windows. *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, 23–27.
- Lazar, J., Feng, J. H., & Hochheiser, H. (2017). *Research methods in human-computer interaction*. Morgan Kaufmann.
- Lee, J. Y., Rhee, G. W., & Seo, D. W. (2010). Hand gesture-based tangible interactions for manipulating virtual objects in a mixed reality environment. *The International Journal of Advanced Manufacturing Technology*, 51(9–12), 1069–1082. <https://doi.org/10.1007/s00170-010-2671-x>
- Lorenceanu, J. (2012). Cursive Writing with Smooth Pursuit Eye Movements. *Current Biology*, 22(16), 1506–1509. <https://doi.org/10.1016/j.cub.2012.06.026>
- MacEachren, A. M., Wachowicz, M., Edsall, R., Haug, D., & Masters, R. (1999). Constructing knowledge from multivariate spatiotemporal data: Integrating geographical visualization with knowledge discovery in database methods. *International Journal of Geographical Information Science*, 13(4), 311–334.
- Majaranta, P., Ahola, U.-K., & Špakov, O. (2009). Fast gaze typing with an adjustable dwell time. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 357–360.
- Mardanbegi, D., Khamis, M., Jalaliniya, S., & Majaranta, P. (2016). *6th international workshop on pervasive eye tracking and mobile eye-based interaction* (p. 1655). <https://doi.org/10.1145/2968219.2968333>



- Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, 5(3), 229–240.
- Milgram, P., & Kishino, F. (1994). A Taxonomy of Mixed Reality Visual Displays. *IEICE Trans. Information Systems*, E77-D, no. 12, 1321–1329.
- Miller, C. C. (2006). A beast in the field: The Google Maps mashup as GIS/2. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 41(3), 187–199.
- Mixed Reality Headset / Augmented Reality (AR) Headset | HP® Store*. (n.d.). Retrieved 19 October 2021, from <https://www.hp.com/us-en/shop/cv/mixed-reality-headset>
- Møllenbach, E., Hansen, J. P., & Lillholm, M. (2013). Eye Movements in Gaze Interaction. *Journal of Eye Movement Research*, 6(2), Article 2. <https://doi.org/10.16910/jemr.6.2.1>
- MRTK-Unity Developer Documentation—Mixed Reality Toolkit | Microsoft Docs*. (n.d.). Retrieved 19 November 2021, from <https://docs.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/?view=mrtkunity-2021-05>
- Nee, A. Y., Ong, S. K., Chryssolouris, G., & Mourtzis, D. (2012). Augmented reality applications in design and manufacturing. *CIRP Annals*, 61(2), 657–679.
- Norman, D. A. (1988). *The Design of Everyday Things*. New York City, NY, USA: Doubleday.
- Onime, C., Uhomobhi, J., Wang, H., & Santachiara, M. (2020). A reclassification of markers for mixed reality environments. *The International Journal of Information and Learning Technology*.
- Pamparău, C., & Vatavu, R.-D. (2020). A Research Agenda Is Needed for Designing for the User Experience of Augmented and Mixed Reality: A Position Paper. *19th International Conference on Mobile and Ubiquitous Multimedia*, 323–325.
- Papadopoulos, T., Evangelidis, K., Kaskalis, T., Evangelidis, G., & Sylaiou, S. (2021). Interactions in Augmented and Mixed Reality: An Overview. *Applied Sciences*, 11, 8752. <https://doi.org/10.3390/app11188752>
- Peuquet, D. J. (1994). It's about time: A conceptual framework for the representation of temporal dynamics in geographic information systems. *Annals of the Association of American Geographers*, 84(3), 441–461.

- Piotrowski, P., & Nowosielski, A. (2020). *Gaze-Based Interaction for VR Environments* (pp. 41–48). [https://doi.org/10.1007/978-3-030-31254-1\\_6](https://doi.org/10.1007/978-3-030-31254-1_6)
- qianw211. (n.d.). *What is Mixed Reality? - Mixed Reality*. Retrieved 19 October 2021, from <https://docs.microsoft.com/en-us/windows/mixed-reality/discover/mixed-reality>
- Rashbass, C. (1961). The relationship between saccadic and smooth tracking eye movements. *The Journal of Physiology*, 159(2), 326–338. <https://doi.org/10.1113/jphysiol.1961.sp006811>
- Richardson, D. C., & Spivey, M. J. (2004). Eye tracking: Characteristics and methods. *Encyclopedia of Biomaterials and Biomedical Engineering*, 3, 1028–1042.
- Roth, R. E. (2011). *Interacting with Maps: The science and practice of cartographic interaction*. The Pennsylvania State University.
- Roth, R. E. (2012). Cartographic interaction primitives: Framework and synthesis. *The Cartographic Journal*, 49(4), 376–395.
- Roth, R. E. (2013). An Empirically-Derived Taxonomy of Interaction Primitives for Interactive Cartography and Geovisualization. *IEEE Transactions on Visualization and Computer Graphics*, 19(12), 2356–2365. <https://doi.org/10.1109/TVCG.2013.130>
- Schmidt, A. (2000). Implicit human computer interaction through context. *Personal Technologies*, 4(2), 191–199.
- Schrepp, M., Hinderks, A., & Thomaschewski, J. (2014). Applying the user experience questionnaire (UEQ) in different evaluation scenarios. *International Conference of Design, User Experience, and Usability*, 383–392.
- Schuchard, R. A., Connell, B. R., & Griffiths, P. (2006). An environmental investigation of wayfinding in a nursing home. *Proceedings of the 2006 Symposium on Eye Tracking Research & Applications*, 33–33.
- Schweigert, R., Schwind, V., & Mayer, S. (2019). *EyePointing: A Gaze-Based Selection Technique*. 719–723. <https://doi.org/10.1145/3340764.3344897>
- scooley. (n.d.-a). *Getting around HoloLens 2*. Retrieved 18 November 2021, from <https://docs.microsoft.com/en-us/hololens/hololens2-basic-usage>
- scooley. (n.d.-b). *HoloLens 2 hardware*. Retrieved 12 November 2021, from <https://docs.microsoft.com/en-us/hololens/hololens2-hardware>
- Shebilske, W. L., & Fisher, D. F. (1983). *Understanding extended discourse through the eyes: How and why*. Lawrence Erlbaum Associates: Hillsdale, NJ.

- Shi, F., Gale, A., & Mollenbach, E. (2008). Eye, Me and the Environment. *International Conference on Computers for Handicapped Persons*, 1030–1033.
- Shneiderman, B. (2003). The eyes have it: A task by data type taxonomy for information visualizations. In *The craft of information visualization* (pp. 364–371). Elsevier.
- Sobota, B., Korecko, S., Hudák, M., & Sivý, M. (2020). *Mixed Reality: A Known Unknown*. <https://doi.org/10.5772/intechopen.92827>
- sostel. (n.d.). *Eye tracking—Mixed Reality*. Retrieved 19 October 2021, from <https://docs.microsoft.com/en-us/windows/mixed-reality/design/eye-tracking>
- Speicher, M., Hall, B. D., & Nebeling, M. (2019). What is Mixed Reality? *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–15. <https://doi.org/10.1145/3290605.3300767>
- STEINKE, T. (1987). Eye Movement Studies In Cartography And Related Fields. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 24, 40–73. <https://doi.org/10.3138/J166-635U-7R56-X2L1>
- Stellmach, S., & Dachzelt, R. (2012). *Investigating gaze-supported multimodal pan and zoom*. <https://doi.org/10.1145/2168556.2168636>
- Stellmach, S., Stober, S., Nürnberger, A., & Dachzelt, R. (2011). Designing gaze-supported multimodal interactions for the exploration of large image collections. *Proceedings of the 1st Conference on Novel Gaze-Controlled Applications*, 1–8.
- Stricker, D., Karigiannis, J., Christou, I. T., Gleue, T., & Ioannidis, N. (2001). Augmented reality for visitors of cultural heritage sites. *Proc. of Int. Conf. on Cultural and Scientific Aspects of Experimental Media Spaces*, 89–93.
- Technologies, U. (n.d.). *Unity Real-Time Development Platform | 3D, 2D VR & AR Engine*. Retrieved 13 November 2021, from <https://unity.com/>
- Tolochko, R. C. (2016). *Contemporary professional practices in interactive web map design* [PhD Thesis].
- Valtakari, N. V., Hooge, I. T. C., Viktorsson, C., Nyström, P., Falck-Ytter, T., & Hessels, R. S. (2021). Eye tracking in human interaction: Possibilities and limitations. *Behavior Research Methods*, 53(4), 1592–1608. <https://doi.org/10.3758/s13428-020-01517-x>

- Visual Studio: IDE and Code Editor for Software Developers and Teams.* (n.d.). Retrieved 13 November 2021, from <https://visualstudio.microsoft.com/>
- Wehrend, S., & Lewis, C. (1990). A problem-oriented classification of visualization techniques. *Proceedings of the First IEEE Conference on Visualization: Visualization90*, 139–143.
- Wendrich, R. E. (2011). A Novel Approach for Collaborative Interaction With Mixed Reality in Value Engineering. *World Conference on Innovative Virtual Reality*, 44328, 103–111.
- WESTHEIMER, G. (1954). EYE MOVEMENT RESPONSES TO A HORIZONTALLY MOVING VISUAL STIMULUS. *A.M.A. Archives of Ophthalmology*, 52(6), 932–941. <https://doi.org/10.1001/archopht.1954.00920050938013>
- What is Eye Tracking and How Does it Work?* (2019, April 1). Imotions Publish. <https://imotions.com/blog/eye-tracking-work/>
- Wiener, J. M., Hölscher, C., Büchner, S., & Konieczny, L. (2012). Gaze behaviour during space perception and spatial decision making. *Psychological Research*, 76(6), 713–729.
- Windows Mixed Reality Headset.* (n.d.). Acer. Retrieved 19 October 2021, from <https://www.acer.com/ac/en/US/content/series/wmr>
- Wobbrock, J. O., Rubinstein, J., Sawyer, M. W., & Duchowski, A. T. (2008). Longitudinal evaluation of discrete consecutive gaze gestures for text entry. *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications*, 11–18.
- Yi, J. S., ah Kang, Y., Stasko, J., & Jacko, J. A. (2007). Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 13(6), 1224–1231.
- Young, L. R., & Sheena, D. (1975). Survey of eye movement recording methods. *Behavior Research Methods & Instrumentation*, 7(5), 397–429.
- Zhou, M. X., & Feiner, S. K. (1998). Visual task characterization for automated visual discourse synthesis. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 392–399.
- Zhu, D., Gedeon, T., & Taylor, K. (2011). “Moving to the centre”: A gaze-driven remote camera control for teleoperation. *Interacting with Computers*, 23(1), 85–95.

# Appendices

## Appendix 1: Pre-study questionnaire

### Section 1: Informed consent

Purpose of the research study: to compare the performance and user experience of the three Mixed Reality interfaces: eyes-controlled, eyes- and voice-controlled, hands-controlled.

What you will do in the study: take the pre-study survey; use the app wearing the HoloLens 2 and perform several tasks; take the post-study survey.

Time required: estimated time is 40 minutes. Buffer is 20 minutes. The appointment is booked for an hour.

Confidentiality: The HoloLens view will be streamed to the researcher's laptop for the successful coordination. Short videos will be recorded during the app experience for the performance evaluation purpose (what user is seeing at the moment, not the face). Videos will be deleted after evaluation of all participants' videos. All information that you give in the study will be handled anonymously.

Right to withdraw from the study: You have the right to withdraw from the study at any time. Your data will be cleared from the research.

1. I have read the "Informed Consent" and voluntarily agree to participate in the research study.
  - ☐ Yes

### Section 2: Background check

2. How often do you use Mixed Reality or Virtual Reality devices? (e.g. HoloLens, Oculus Rift, etc.)
  - ☐ haven't used it yet
  - ☐ one or several times
  - ☐ often
3. How often do you use maps?
  - ☐ haven't used

- sometimes
  - often
4. What level of proficiency do you have in working with geospatial information?
- I have never worked with it
  - I'm a novice
  - I'm experienced in it
  - I'm an expert
5. What is your vision situation?
- I don't need to wear glasses or contact lenses
  - I'm wearing glasses (right now)
  - I'm wearing contact lenses (right now)
  - My vision is not very good, but I'm not wearing glasses or contact lenses (right now)

## Appendix 2: Post-study questionnaire

1. How would you rank the responsiveness of the eye cursor?

	1	2	3	4	5	
poor	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	very good

### Section 1: Task load

2. How mentally demanding were the interactions? (mental demand)

	the most	moderately	the least
Eyes interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eyes-voice interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hands interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

3. How physically demanding were the interactions? (physical demand)

	the most	moderately	the least
Eyes interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eyes-voice interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hands interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

4. How successful were you in accomplishing the Rotation tasks?  
(performance)

	the most	moderately	the least
Eyes interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hands interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

5. How successful were you in accomplishing the Retrieval tasks?  
(performance: building name tag)

	the most	moderately	the least
Eyes interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hands interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

6. How successful were you in accomplishing the Overlay tasks? (performance: change basemap, add points)

	the most	moderately	the least
Eyes interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eyes-voice interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hands interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

7. How hard did you have to work to accomplish your level of performance for the Rotation interaction? (effort)

	the most	moderately	the least
Eyes interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hands interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

8. How hard did you have to work to accomplish your level of performance for the Retrieval interaction? (effort)

	the most	moderately	the least
Eyes interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hands interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

9. How hard did you have to work to accomplish your level of performance for the Overlay interaction? (effort)



	the most	moderately	the least
Eyes interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eyes-voice interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hands interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

10. How insecure, discouraged, irritated, stressed, and annoyed were you?  
(frustration)

	the most	moderately	the least
Eyes interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hands interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## Section 2: User experience

Please, circle the corresponding number for each interface.

	1 - annoying	2	3	4	5 - enjoyable
Eyes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eyes-voice	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hands	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

11.

	1 - impractical	2	3	4	5 - practical
Eyes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eyes-voice	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hands	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

12.

	1 - boring	2	3	4	5 - exciting
Eyes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eyes-voice	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hands	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

13.

	1 - conventional	2	3	4	5 - inventive
Eyes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eyes-voice	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hands	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

14.

	1 - diff. to learn	2	3	4	5 - easy to learn
Eyes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eyes-voice	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hands	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

15.

	1 - confusing	2	3	4	5 - clear
Eyes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eyes-voice	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hands	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

16.

### Section 3: Interface ranking

#### 17. Overall

	1	2	3
Eyes Interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eyes-Voice Interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hands Interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

#### 18. Overlay Interaction (adding POIs, changing the basemap)

	1	2	3
Eyes Interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eyes-Voice Interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hands Interface	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

19. Retrieve interaction (retrieving the building name)

	1	2
Eyes Interface	<input type="radio"/>	<input type="radio"/>
Hands Interface	<input type="radio"/>	<input type="radio"/>

20. Rotation Interaction (rotating the terrain)

	1	2
Eyes Interface	<input type="radio"/>	<input type="radio"/>
Hands Interface	<input type="radio"/>	<input type="radio"/>

21. Have you encountered any difficulties or issues while using the Eyes Interface? If yes, please, elaborate.

☐ No

☐ Other: \_\_\_\_\_

22. Have you encountered any difficulties or issues while using the Eyes-Voice Interface? If yes, please, elaborate.

☐ No

☐ Other: \_\_\_\_\_

23. Have you encountered any difficulties or issues while using the Hands Interface? If yes, please, elaborate.

☐ No

☐ Other: \_\_\_\_\_

24. Please, write the suggestions for any of the interfaces or for the user-study here.

---

---

---

---

---