Electronics and IT
Aalborg University Copenhagen
http://www.aau.dk

AALBORG UNIVERSITY
STUDENT REPORT

Title:
Exploration of interaction in AR

Theme:
Master thesis

Project Period:
Spring Semester 2020
MED10

Participant(s):
Jeppe Milling Korsholm

Supervisor(s):
Stefania Serafin

Copies: 1

Page Numbers: 27

Date of Completion:
May 28, 2020

Abstract:

In recent years, Augmented Reality(AR) has become more commercially available and is being applied to a width of areas, including museums. This has mainly been in the form of mobile phone powered AR, but Mixed Reality Headsets(MRH), allowing for more immersive AR has also become available. Research suggest that AR used in museums, improve the overall experience and peoples intention to revisit. The research suggest a strong link between the improved experience, and peoples feeling of presence, induced by AR. Assuming the more immersive MRH's can induce a stronger feeling of presence, a prototype for the device Magic Leap 1 was made. While the intention was to evaluate the prototype in a museum, this was not possible due to Covid-19, and focus was instead changed to explore how embodied interaction, possible with MRH has an effect on AR experience.

Exploration of interaction in AR

Project Report

Jeppe Milling Korsholm

Aalborg University
Electronics and IT

Contents

Chapter 1

Introduction

In recent years Augmented Reality (AR) has been applied to many areas, from enhancing shopping experiences[8] to education. It has been used to interact with virtual objects, for both educational and playful purposes, and can also be used in information providing systems, such as User Interfaces.

The ubiquity of smartphones, along with the release of ARkit[1] and ARCore[2], has meant the development of mobile AR applications has become a lot more accessible to many industries, including museums. Here AR can both be used to augment existing displays as well as allow virtual museums to be experienced from the comfort of one's own home.

AR advances, driven by developments in spatial computing, has also lead to the development of new mixed reality headsets such as Microsoft Hololens and Magic Leap 1, which allows embodied interaction to be combined with spatial audio and visuals.

While these devices are expensive, and not as widely used, their use in interactive displays are being tested, from creating new ways of shopping, to having museum experiences in the comfort of ones home. The latter example can be seen in Magic Leap's collaboration with the London Museum and Nexus Studios.[9] Microsoft's Hololens has been used on film sets, allowing actors to see the CGI worlds their characters inhabit, instead of just green screen sets[3].

In recent a survey[10] of the application of AR to museum experiences, the presence felt in AR experiences, is linked to enriched museum experiences. If what drives the improved experience is presence mediated via AR, then Mixed Reality Headsets (MRH), should deliver a more immersive experience, allowing

[1] https://developer.apple.com/augmented-reality/arkit/
[2] https://developers.google.com/ar/discover/
[3] https://picturethisconference.com/the-paradigm-shift-of-virtual-production

for a more natural and hands-free interaction. This project will explore the use of AR experiences in a museum in the following way:

- Examine how embodied interaction input available in a MRH compares, when used in basic interaction tasks.
- Following the first point, evaluate how an AR application featuring embodied interaction, affects the museum experience.

To explore these questions, a prototype has been made using the Magic Leap 1 Creator, focusing on the use of the following modalities: input controller, eye-gaze and mid-air hand gestures. The prototype features a navigable interface and interactive elements that could be used in a museum.

While the originally plan called for the prototype to be tested in a museum environment, due to the April-March 2020 covid-19 shutdown, this proved difficult, and the the main concern of the evaluation will instead focus on achieving the first mentioned exploration point.

Chapter 2

Analysis

In exploring how to answer the previously stated questions, we will first look at the technology that allows both mobiles and MRHs to place virtual objects properly in real scenes, thus facilitating augmented experiences.

2.1 SLAM

Simultaneous localization and mapping (SLAM), is a technology through which a computer system maps its surroundings, and from this mapping computes the systems location. SLAM thus allows a system, to locate itself in its surroundsings, using only data from it's sensors, without the need for any a priori data about the environment.[4] while SLAM technology has primarily been developped to support the navigation of autonomous vehicles, it is now also being applied to mixed reality headsets, where accurate SLAM enables virtual objects to seamlessly interact with the real world.

SLAM can be approached with an array of sensors, but monocular and stereo vision have been the most popular approaches[4]. Despite the popularity of vision based approaches, there are only limited, and often device-specific data-sets available for SLAM like tracking and reconstruction, when compared to other visual fields like image recognition and object detection.[2]

To solve this problem, Magic Leap developed a point tracking system for its MR headset, composed of two neural networks that could be trained on synthetic data, avoiding the before mentioned issue of limited datasets, as well as expensive ground truthing equipment. [2] Finally, Magic Leap's researchers managed to make the system run smoothly at 30+ frames on a single CPU, enabling the SLAM system to be built into a lightweight wearable device.

Having such systems in a wearable device, allows the user to see virtual objects interact with accurately mapped physical objects. This can add realism, such as when a virtual ball can bounce of a real table, or letting a virtual character be

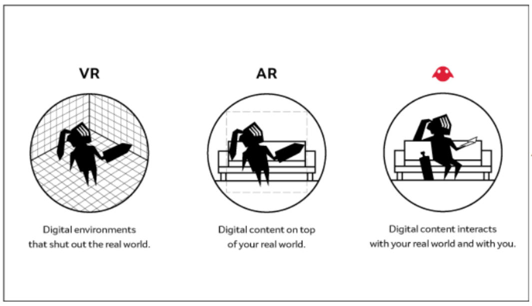

Figure 2.1: And illustration of what Spatial computing tries to offer. Source:`https://developer.magicleap.com/en-us/learn/guides/design-spatial-computing`

seated in a real physical chair, as seen later in figure 2.3.

How this added potential for realism, and what spatial computing can add to applications in a museum context will be explored below.

2.2 AR in a museum context

In a survey of museum visitors, researchers Jung et al.[10] found that social presence in VR and AR is a strong predictor for the several aspects of peoples' experience, as well as their intent to revisit.

AR may also prove of interest to museums as a storytelling tool, given its ability to provide a compelling narrative. After conducting semi-structured interviews with museums and heritage sites, researchers Ketchel et al. [11] , found museums were interested in how AR could help advance their use storytelling, by connecting people to the exhibited artefacts as well as facilitating the storytelling by following the visitors as they move through the exhibition.

The introduction of AR to a museum's experience may entice younger generations, who are not as interested in traditional museum visits, to visit more. This sentiment is posed by researchers Geronikolakis et al. [6], and in that work they further posit that interactive applications, through gamified elements, and the options to be physically involved in restoration work, could encourage people to return to museums more often.[6] Geronikolakis et al.'s research corroborates that presence and storytelling are important elements to consider when designing AR applications. [6]

Should the technology progress further, to the point where one cannot distinguish the virtual elements from the real, it would improve the users' immersion when using the application, making it advantageous over VR. The blending of the real and the virtual also allows applications to be tried on actual archaeological sites, where virtual layers can add interest to historical sites. [6]

2.2.1 Design input

When designing AR applications for use in museums or cultural sites, it appears that the most important factors are a strong feeling of immersion and a high image quality. The use of avatars and animation may also prove helpful in enhancing the entertainment aspect of the user's experience [10] The quality of the visual aspect is important, especially in regard to avatars, as low fidelity can make facial expressions hard to read, which as been shown to have a negative impact on the learning outcome from these experiences.[11]

While AR experiences could be built using both phones and MR headsets as platforms, it is possible to bypass the use of phones as a looking glass and just use MR headsets, so that the user is more naturally immersed in the MR experience. Geronikolaks et al. do mention limitations when using MR headsets (in their case Microsoft's Hololens) as the device has a limited processing power and Field of View (FOV), compared to the VR version they built for their own platform. [6] This is true for AR compared to VR, but does not compare the FOV of mobile AR and MRH's.

While MR headsets are more cumbersome and expensive than phone powered AR, they can provide embodied interaction input, which is not possible with the latter. This embodied interaction allows for greater immersion, an important design principle for for applications designed for the storytelling and knowledge conveying that museums engage in.

2.3 Embodied interaction and multi-modal interaction

This section covers the following:

1. Embodied interaction
2. How gaze and gesture can be combined for multi-modal interaction
3. Considerations for how to evaluate input modalities
4. User input, how it can be used for direct feedback and saved for post use evaluation

2.3.1 Embodied interaction

Embodied interaction can be seen as being in contrast to the Cartesian tradition, that views cognition as happening in the brain and separated from the body. In an embodied interaction, mind and body are seen as inseparable. [1]

The term was popularized in 2001 by Dourish [3], who emphasized interaction as not just what is being done when interfacing with a computer system, but also how it is done. He also mentions how interaction historically can be viewed as

> “a gradual expansion of the range of human skills and abilities that can be incorporated into interaction with computers “
>
> Dourish, 2001, chapter1, page 17

This view lends itself well to how Augmented Reality seems to merge the Virtual and the Real, and seeks to use similar ways to interact with both. Mixed reality headsets like the Microsoft Hololens and Magic Leap One, allow for a variety of embodied inputs which include gesture and eye-gaze tracking, as well as speech recognition and head pose, as can be seen in figure 2.2. While all of these inputs could be used in a prototype, the API available on the Magic Leap device for voice control, is not readily supported at the time of writing, and is as such not going to be this work’s focus. Instead, eye-gaze-driven interaction will be looked into first, and how it can be combined with gestures, to provide a natural way of interacting in AR.

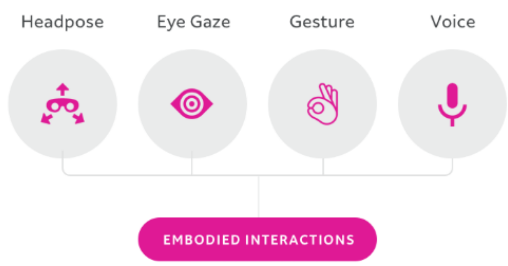

Figure 2.2: And illustration of the embodied input modalities available in the Magic Leap Source:`https://developer.magicleap.com/en-us/learn/guides/design-input-methods-overview`

2.3.2 Multimodal input using gaze and gesture

While there are many ways of capturing eye-gaze, the most used modern and unobtrusive method involves calculating the eye-gaze point based on captured images. And while eye-gaze input when used on a screen, is not as fast or accurate as using a mouse for selection tasks, it is sufficiently accurate, when targets are of a large enough size, and works well on larger displays [12]. However, if one uses only eye-gaze for interaction that requires selection, either dwelling time at a target or eye blinking is often used for interaction, which has inherent issues such as time consumption and involuntary blinking[12].

Some of the drawbacks of the pure eye-gaze interaction may be mitigated by combining it with other input modalities such as mid-air gestures. In a paper by Kim et al. [12], researchers found that when combining eye-gaze and hand-gestures, eye-gaze was well suited for global navigation on a display, with hand gestures for local navigation and selection, thus avoiding dwell time or involuntary blinking issues. Combining input like this in a complementary fashion seems ideal with a device like an MRH, that allows access to API for multiple input modalities, but it is also worth considering how long term use of such mixed systems would be. The setup for the test, described above by Kim et al., required the user's hand to hover over an infrared Leap Motion Sensor.[1] It is worth noting that while collecting

gesture data, researchers did not report any strain on the user for keeping their hand hovering, although the experiments were short in nature (all trials under 20 seconds). As MRHs can only capture hand gestures when they are held up within the field of view of the device, one can expect gesture interaction to be tiring if required for extended periods of time. That a system can both be more accurate, but potentially less enjoyable over time, makes it important to consider several aspects of the chosen input modality.

2.3.3 Evaluating modalities

When working with the MRH, then the goal in evaluating the input modalities would not be testing their accuracy per se, but would rather be to test their performance against each other, and evaluations of this kind would be of interest.

In a 2019 paper, Gentile et al., did a comparative evaluation of mid-air gestures and touch interaction, for use with interactive displays, the target group being people on the Autism Spectrum Disorder. For the comparison, they looked into several aspects: usability, effectiveness and enjoyment, finding that while touch interaction provided more effective actions with higher usability scores, the mid-air gestures provided higher scores for engagement and enjoyment. [5]

This method of evaluating several aspects of interaction modalities seems appealing, when looking at how people will interact in MR, and for further evaluating how and when different modalities best serve to enhance the MR experience.

2.3.4 User feedback using sensor input

Other than just using the MRH's sensor directly for interaction, the sensors can also be used for providing user feedback post interaction, based on captured user info during MR experiences.

Researchers Hartholt et al. demonstrate how a MRH can be combined with virtual humans, as a teaching application for young adults with Autism Spectrum Disorder (ASD), in this case for virtual job interviews. [7]

Some of the elements they touch upon in the project is the use of surface detection that allows virtual characters to interact with the real world, such as letting a virtual character sit in a real physical chair, by mapping the chairs surface, as

[1] `https://www.ultraleap.com/product/leap-motion-controller/`

seen in figure 2.3. They also use the eye-tracking capabilities provided by the Magic Leap Headset to provide user feedback as well as to improve the animated characters' immersive effect.

Figure 2.3: Figure showing virtual character interacting with the environment [7]

Allowing these virtual characters to interact dynamically with the environment through SLAM and eye-tracking, allows for more believable interaction.

In addition to believable interaction, they highlight how these devices allows the use of metrics like eye-gaze, blink and head-rotation both to provide feedback as well as better inform researchers during evaluation of user tests.

2.4 RelatedWorks

This section will look at some of the latest novel apps on MRH's , their use of interaction modalities and the possibilities they open up.

2.4.1 Holographic Museum Experiences

Augmented reality, has also been used to make museum pieces more accessible. Guided by Sir David Attenborough, the Museum Alive application leverages the Magic Leap to bring to life scientifically correct prehistoric animals in your living room. [13]

A similar project was made by Nexus Studio, which was commissioned by the BBC to make an AR application that would bring the art and artefacts covered in BBC's series Civilizations to life. The studio developed the application in collaboration with the British Museum, among others, whose artefacts were scanned, to then be brought to life again using a ARHMD. According to the creators of this project, this allowed people to get much closer to museum pieces and explore them from new angles [9].

Figure 2.4: Left: Interactive look at sarcophagus from [9]. Right: Interaction with natural history with [13]

While these applications use modern scanning techniques to get you close to realistic looking artefacts, they use the device's input controller, and head pose, as the primary drivers of interaction. This does let you inspect interesting artefacts up close, but does not show the artefacts or animals in their historic context.

This is attempted, or at least hinted at with the True AR Authoring tool, mentioned in 2.2. The images in figure 2.5 show how they are also incorporating an avatar, to help convey information to users. The right image also shows how gesture commands allow a user to interact with a model of real historical ruins. As the authors mention in their future works, it would be interesting to see how this kind of interaction could be used on an actual historical site. Perhaps gesture interaction could allow one to piece together virtual parts on top of existing ruins, to get a view of how sites of historical importance looked in their prime.

Figure 2.5: Left: A character, based on the scan of a real priest, gives information about a historical site, a model of which is shown on the table in front of him. Right: The objects on the table can be manipulated with hand gestures. Both images from [6]

Chapter 3

Design and Implementation

In this chapter the design and implementation of a prototype AR experience is explained.

The prototype is built using the Unity Engine [1] by implementing Magic Leap's plug-in for Unity. The prototype was then built to Magic Leap's Lumin platform, and loaded onto the Magic Leap 1.

A unity project template made available by Magic Leap[2], was used as the foundation for the prototype.

This chapter details the design considerations for navigation of interfaces and with virtual objects, and how to accommodate using multiple input modalities.

3.1 User input

This section will cover the following topics:

1. Considerations for handling UI in an AR setting
2. General input handling in AR
3. Setup of Input controller
4. Setup of gaze input
5. Setup of gesture input

3.1.1 UI considerations

As shown in figure 2.2, the Magic Leap headset allows several modalities to be used as an input. One of the first challenges to consider was how to navigate a

[1] `https://unity.com/products/core-platform`

[2] `https://github.com/magicleap/UnityTemplate`

UI, as users would have to be presented text or given options to choose between, which could be handled with traditional buttons and text boxes. For this, Unity's UI system was leveraged, and regular canvases with text and images were used, with the addition of a helping script made available from Magic Leap, that kept the World Space in front of the user via tracked head-pose. The script allowed the canvas to follow at a delay, using smoothed interpolation, compared to direct constant tracking, which detracts from the experience. Concretely, this means that the projected UI does not follow every single flicker of the user's head, but only follows broad movements in the user's field of view. This prevents the user from experiencing a constant shaking in the UI which his head's natural movements would induce.

For traditional desktop and phone prototypes, Unity has a built-in Standalone Input Module[3], which handles input events from a mouse and traditional controllers. This module could not immediately be translated to work with both eye-tracking, handgesture and controller input, so custom scripts had to be made.

These scripts utilize raycasts from Unity's built-in physics engine. This means that anything interactive, whether UI buttons or 3d instruments simply have to be bounded by a Unity physics Collider component, to be interactive with the proposed modalities.

3.1.2 Handling Input

For handling simple input, it was first determined that whatever input modality was chosen, the system should be able to differentiate whether a user was selecting and item or trying to activate it. First a user should be able to select a button or instrument, with a highlight or similar reaction showing the user it is an object of interest that can be interacted with. Secondly, the user should then be able to activate a highlighted object of interest.

For the actual selection of an object, we will look at it through the example of pressing a virtual button.

The system was set up so that all interactive buttons had a tag that identified them as such, and a collider as previously mentioned. When present in a scene with a UI, the system would try to raycast from the virtual scene-camera through a pointer such as a fingertip or eye-gaze fixation point, and if this ray hit an interactive button, it would light up as seen in figure [MAKE FIGURE SHOWING highlighting UI BUTTON]. Depending on the chosen input modality, activation could then happen in one of the following ways: - with the input controller registering a button press, - with pure eye tracking registering a blinking event, - with gestures recognizing the hand-pose chosen to represent activation.

[3] `https://docs.unity3d.com/2018.4/Documentation/Manual/script-StandaloneInputModule.html`

Input Controller

For controller input, the position and rotation of the controller was tracked. This position and direction was then used for raycasting, when the trigger button was pressed. As a visual aid, a thin cube, was put in the scene, showing the position and pointing direction of the controller. Code was also implemented, to exit the prototype if the controllers "home" button was tapped twice in quick succession.

Gaze Input

For gaze input, calibration of gaze is handled by the Lumin OS itself, so no calibration code was necessary. Preliminary testing using a scene with just a virtual cursor shown at the gaze point illustrated that while the system is quite accurate, targets should not be too small. While blink control can have issues as mentioned in 2.3, it was still put in the prototype, in order for mono-modal input to be evaluated. To remove the issue of uncontrolled blinking, the prototype was setup to activate for blinking only with the right eye.

```
void Update(){
      headlook = (MLEyes.FixationPoint - maincamera.transform.position
    ).normalized;

    if (MLEyes.RightEye.IsBlinking)
        {
          pressingButton = true;
        }
        else
        {
          pressingButton = false;
        }
        EyeRayPhysical();
    }...
public void EyeRayPhysical()
  {
    RaycastHit hit;
    if (Physics.Raycast(maincamera.transform.position, headlook, out
    hit))
    {
      if (MenuSystem.Instance.testCondition == MenuSystem.
    TestCondition.EyeGaze)
      {
        if (pressingButton && hit.transform.CompareTag("Button"))
        {
          hit.transform.gameObject.GetComponent<ARButton>().Pressed();
        }
        else if (hit.transform.CompareTag("Button"))
        {
          hit.transform.gameObject.GetComponent<ARButton>().Selected()
    ;
```

```
        }
      }
    }
  }
```

Listing 3.1: Eye RayCasting Code

Gesture Input

For gesture input, two different hand-poses were needed for selecting and activating buttons and objects. The Magic Leap comes with 8 predefined and recognized hand poses as seen in figure 3.1. Of these mappings, "pointing with the index finger" was chosen for selection, and "extending the thumb while pointing" was chosen for activating, as "pointing with index finger" was found by the author to be more relaxing than the activation pose when holding your hand up for extended periods of time. Note, this configuration is the opposite of what is used in the multi-modal study described in [12].

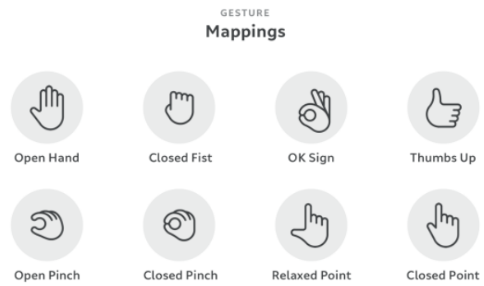

Figure 3.1: 8 predefined hand poses for the Magic Leap. Source: `https://developer.magicleap.com/en-us/learn/guides/design-gesture`

In the code these two hand-poses are described as 'Finger' for the selecting pose, and 'L' for the activation pose. For using gestures as primary input, the code below was used. It illustrates how a ray was cast with direction from the virtual camera, through the pointing index finger. If an interactive object was hit and the relevant pose was recognized, the AR Button script on the object would activate the desired function.

```
public void FingerCast()
  {
    Vector3 StartPos = MLHandTracking.Left.Index.Tip.Position;
    Vector3 fingerDir =(StartPos-Camera.main.transform.position ).
   normalized;

    RaycastHit hit;
    if (Physics.Raycast(Camera.main.transform.position,fingerDir, out
   hit))
```

```
    {
      if (pose == HandPoses.L && hit.transform.CompareTag("Button"))
      {
        hit.transform.gameObject.GetComponent<ARButton>().Pressed();
      }
      else if (pose == HandPoses.Finger && hit.transform.CompareTag("
    Button"))
      {
        hit.transform.gameObject.GetComponent<ARButton>().Selected();
      }
    }
  }
```

Listing 3.2: Gesture Raycasting code

This use of ray casting works well when interacting with virtual objects at a distance. And while it was possible to also interact with virtual objects closer to the user, it was not always as smooth. This is due to the user's different manner of interaction with close objects, as explained in the following example. Code was written to allow a user to pick up an object with a pinch gestures, which would then move the position of the object in relation to the position of the index and thumb tip. But if the user's natural reaction was to also turn their hand, expecting the object to rotate, then the device would loose the tracking points needed to recognize the gesture, and connection with the virtual object was lost. Therefore it was chosen for the time being, to keep interaction at a distance, while close object manipulation would wait until positional, rotational and scalable manipulation could be implemented, using different gestures or similar control means.

3.1.3 Head Pose

Another final input modality that is used in conjunction with all the aforementioned, is head-pose. When using the device, the virtual camera in the prototype follows the head of the user (both position and orientation). This allows the user to explore a scene freely, and can also be used to setup triggers for events when the virtual camera and thus the user, gets close to an object of interest.

3.2 Virtual Interactables

This section will describe the setup of the UI and virtual interactive objects, and how they were setup in Unity.

3.2.1 UI

While UI considerations are described above, this section provides further clarification.

For input testing's general ease of use, a UI was setup allowing users to see three images of instruments, with their names above. These three music buttons were flanked by arrow keys, allowing you to browse among 14 different instruments. The setup can be seen in figure [add UI picture]

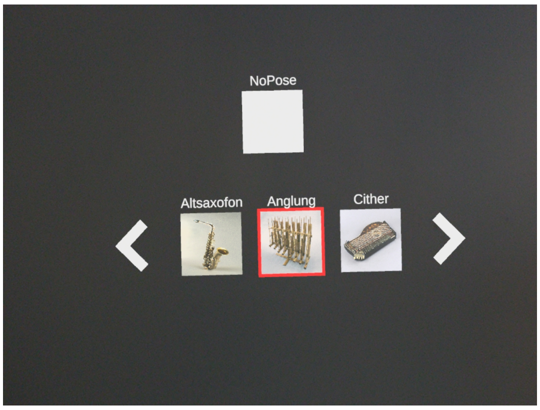

Figure 3.2: Image of the canvas as seen in the device. The red border on the middle image is due to being highlight via eye gaze. The empty square above, shows NoPose, as no hands and thus poses are in view of the device. Image taken with the device.

3.2.2 Virtual Objects

With SLAM technology, virtual objects can now do much more than simply being placed on top of real-world objects, unlike with a simpler UI. With a virtual representation of the Real, objects can lean on walls or stand on tables, provided the mapping is done properly. On top of this, the Magic Leap allows objects to be placed not in relation to the virtual camera's coordinates and thus the device, but according to what is called Persistent Frame Coordinates (PFC). [4].

PFCs anchor an object to the physical world via multiple known positions based on recognized geometry in the frame. This means that regardless of where the device is turned on, the object will be located in the same physical spot as it is not tied directly to the coordinatess of the virtual camera. This also allows multiple users on the same network to see the object positioned correctly and in the same place. The use of PFC can thus allow an author of an experience, to place virtual objects along a planned museum exhibition route, or anchor extra information nodes next to the physical installations and artefacts they relate to.

While any type of virtual object could be put into the prototype, the plan for the project was to interact with a musical museum. Therefore the example shown here will be a virtual instrument. The virtual instrument consists of a 3D model and an

[4]https://developer.magicleap.com/en-us/learn/guides/persistent-coordinate-frames

audio clip. These are linked to components for spatial audio, and a sphere collider that can trigger the music to start when the player gets close to the instrument. Over the instrument, the name is displayed as 3D text, always facing the user, an example can be seen in figure 3.3.

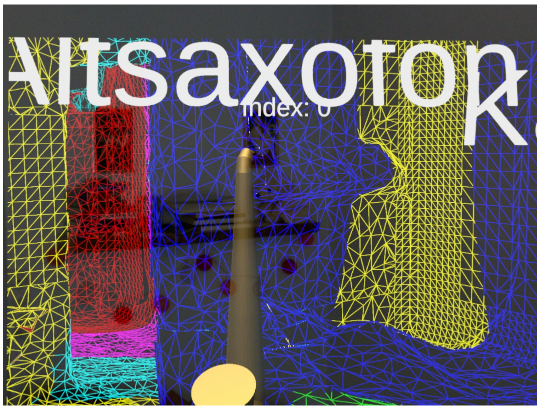

Figure 3.3: A virtual instrument placed in a physical location. The geometric grid behind it, shows the virtual mapping of the environment

An addition to be made, is a canvas that can roll out with additional information about the instrument, if the user chooses to activate it.

3.3 Tracking user data

While it is possible to monitor what is being shown on the device on a connected computer, this requires a constant connection and even then, a high frame rate of the playback is not guaranteed. If data is wanted on how and when the user interacts with elements in the system, it is therefore better to build data capture code into the device. This could be useful both for gathering metrics such as completion time and input method used for post-use analysis, as well as for direct user feedback during the experience.

For post-use analysis, data can be stored on the device, using the application path. It can be extracted from there, using the Magic Leap Companion software, namely the functionality called the Device Bridge.

For ease of analysis, any data related to input should be labeled so it can be stored in a css file. File entries should encode all necessary data such as input modality, time, and if for testing, should also include participant ID. For user feedback, it would depend on the given scenario.

Chapter 4

Evaluation

From chapter 2 we know there is a link between the quality of an experience, and the feeling of presence AR can help induce. While it is currently not possible to test at a museum, this chapter will detail how one could seek to evaluate elements of the prototype that have been made.

Elements of interest here are how embodied interaction, and range of input in general, can influence presence.

4.1 Influence of input

When evaluating the input modalities, their intuitiveness, effective use and usability are all aspects that can be used to assess their overall usefulness for the user experience. The preferred input modality can change based on what the user wants to do, as was suggested in the multi-modal interface of the analysis section.

This might indicate that a direct comparison of them would not work unless tested across multiple interaction tasks, but this project will mainly focus on basic exploration and navigation tasks, such as UI selection tasks. It would also be of interest to see how users intuitively try to interact, using the different input modalities. For this sort of test, the conditions to be tested would be gaze, gesture and input controller. Taking into account the weaknesses of gaze input only, a fourth condition mixing gaze and input controller is proposed.

The four conditions, will be compared on multiple criteria, such as intuitiveness, usability, effectiveness and enjoyment of the interaction. And while test cases are limited, this test would also serve as the outline for what should be included in further tests of different interaction tasks.

Intuitiveness

For intuitiveness, the participants will be presented with the prototype with little to no explanation of the workings, and will only be told what to do when stuck for prolonged periods of time. The participants should be encouraged to think out loud, so the researcher can note down if the prototype has any shortcomings.

The time taken to grasp the workings of the particular interaction condition is measured. As a participant will have to test four different versions of the prototype, they will be informed of the relevant input modality before the test starts.

To further clarify, in each condition the participants will be given a version of the prototype that starts with them reading an introduction on the UI. The UI will instruct the participants to look left and right before trying to press a button in front of them. This movement will start a timer. The timer is activated by the head-pose, not any of the input conditions, thus not biasing the users attempt.

Usability

The usability of the modalities will be measured using a System Usability Scale[1], filled out by the user after interacting with a version of the prototype. Also, the researcher will take notes of the users interaction while testing is in progress.

Effectiveness

For measuring the effectiveness of the different modalities, the prototype will log tasks completion time, for all instructed tasks.

Enjoyment

Post test, the participants will be asked to fill in a questionnaire regarding how enjoyable using the given modality was. In addition, the researcher will be taking note of observed behavior and any comments from the participants.

4.1.1 Test Setup

For the setup, all that is required is a space wide enough to move around comfortably. As the data is captured on the device itself, filming is not necessary, but the researcher should still be taking notes of the participant's behaviour.

4.1.2 Planned Procedure

As this part is influenced by the research by Gentile et al., it will follow a similar setup, with differences regarding the measure of how intuitive the input is.

[1] https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html

The test should be a with-in group study, with all participants experiencing all four conditions. Ideally, this would happen with time enough between to not have a high carry-over effect on conditions tested after the first, but if time is limited, the order should at least be randomized for the different participants, to counteract any carryover. Below, the procedure is outlined in steps:

1. Before the participant is given the device, the researcher ensures the prototype is running, that it is at the start screen and has a unique participant ID.

2. After being given a brief introduction to the device and the prototype, the participant puts on the device, assisted by the researcher to make sure it is worn properly.

3. On the start screen, the participant is given a keyword regarding the input for the test and can read instructions for how to start the test. This would be by looking up and down or side to side, in order to start the timer for intuition test using head pose instead of one of the input conditions tested.

4. Once the timer has begun,unbeknownst to the participant, the intro canvas is removed and a button is shown to the participant. The participant will now try and figure out how to press it. If 30 seconds passes without a button press, a researcher can give a hint.

5. Once the first button is pressed, the canvas seen in figure 3.2 is shown, and the participant will be given tasks relating to navigating the canvas.

6. Once navigation tasks are completed, they are free to keep using the app for as long as they find browsing the instruments interesting.

4.1.3 Data Analysis Considerations

When all the required data has been obtained, the conditions can be compared using one-way analysis of variance. In addition, the data should be inspected for any relations that can be explained based on researchers' notes of the experiment, and reasons for any outliers should be looked into.

Chapter 5

Discussion and Future works

As the evaluation described in the previous section has not been carried out, this section cannot discuss any results, but only how the evaluation could be carried out, and what other evaluations could be made concerning the topic. First a note concerning the test described in the evaluation section. While multi-modality is discussed in the analysis, the current evaluation seeks to compare them one on one. This works for an initial exploration of the system, but the findings of such an evaluations should inform the design of multi-modal interaction for further exploration.

If multi-modality is to be explored further, scenarios using gaze for global navigation between canvases and active areas of interest (including voice or gesture for selection) would be better. Other ways to expand the proposed test would be to task participants with interacting with virtual objects and manipulating them to some degree. They could also be made to explore environments to find objects of interest based on visual and audio cues. This level of interaction and the free exploration it affords would go well with exploring a museum installation. As with the test proposed in the evaluation section, it would be interesting to look into how such an interaction is best setup before being attached to an AR experience. The reason for this would be that once it is implemented, if people are new to AR and MRHs, the novelty factor might influence their answers to a degree where it is difficult to tell how well the actual features were implemented.

Following an evaluation of the usability of available input modalities in MRHs, a more fully fledged application could be tested out in a museum. Exploring a museum for contextual information about physical museum pieces, could then be tested. Thus the power of AR could be leveraged to augment an existing physical

display instead of only featuring virtual displays situated in a museum context. For this the participants could be surveyed in a manner similar to the evaluations mentioned in [10].

Mobile and headset AR comparison Should a study seek to assess whether embodied interaction has a significant impact on the AR experience, it would have to isolate this variable. This could be done by building a prototype that works on both platforms, with the only variable being whether input is embodied or phone based. A test prototype could be similar to the one made for this project. Making it in unity would ensure cross-platform capabilities, but due attention should be given to these platforms' interoperability.

While this test could be considered a proof of concept, it does not address the feasibility of a large scale use in museums. It is worth noting that at this point that phones are more widespread and applications for them easier to implement.

Use of avatars A future iteration of this work could also introduce an avatar. Works in the analysis suggest that having a human avatar in the experience increases the sense of presence, and having an avatar to convey information, would also be more natural than just a wall of text.

5.1 Conclusion

In conclusion, the study of all the aspects of the interplay between AR and a museum experience is a complicated project. To do it justice would require successive evaluations of its different aspect.

However, after evaluating possible input for interaction, building an application using a optimal combination of the available input is possible. This application could be tested in a museum context, not on completely virtual displays, but with virtual elements augmenting the physical artefacts present, with the goal of enhancing the museum experience, through the information gathered by visitors, and their likelihood for wanting to revisit. Finally, it would be interesting to see how an augmentation would work on a historical site, and how seeing a historical site in its prime, virtually and being able to interact with it, would affect peoples experience of history.

Bibliography

[1] Alissa N. Antle, Greg Corness, and Milena Droumeva. "What the body knows: Exploring the benefits of embodied metaphors in hybrid physical digital environments". In: *Interacting with Computers* 21.1-2 (2009). Publisher: Oxford University Press Oxford, UK, pp. 66–75.

[2] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. "Toward geometric deep SLAM". In: *arXiv preprint arXiv:1707.07410* (2017).

[3] Paul Dourish. *Where the action is*. MIT press Cambridge, 2001.

[4] H. Durrant-Whyte and T. Bailey. "Simultaneous localization and mapping: part I". In: *IEEE Robotics Automation Magazine* 13.2 (June 2006). Conference Name: IEEE Robotics Automation Magazine, pp. 99–110. ISSN: 1558-223X. DOI: 10.1109/MRA.2006.1638022.

[5] Vito Gentile et al. "Touch or touchless? evaluating usability of interactive displays for persons with autistic spectrum disorders". In: *Proceedings of the 8th ACM International Symposium on Pervasive Displays*. 2019, pp. 1–7.

[6] Efstratios Geronikolakis et al. "A True AR Authoring Tool for Interactive Virtual Museums". In: *Visual Computing for Cultural Heritage*. Springer, 2020, pp. 225–242.

[7] Arno Hartholt et al. "Virtual Humans in Augmented Reality: A First Step towards Real-World Embedded Virtual Roleplayers". en. In: *Proceedings of the 7th International Conference on Human-Agent Interaction*. Kyoto Japan: ACM, Sept. 2019, pp. 205–207. ISBN: 978-1-4503-6922-0. DOI: 10.1145/3349537.3352766. URL: https://dl.acm.org/doi/10.1145/3349537.3352766 (visited on 05/05/2020).

[8] Magic Leap inc. *H&M and Magic Leap Redefine the Customer Experience*. en-us. Library Catalog: www.magicleap.com. URL: https://www.magicleap.com/news/news/h-and-m-and-magic-leap-redefine-the-customer-experience (visited on 05/05/2020).

[9] Magic Leap inc. *Preserving history, with spatial computing*. en-us. URL: https://www.magicleap.com/news/partner-stories/preserving-history-with-spatial-computing (visited on 05/05/2020).

[10] Timothy Jung et al. "Effects of virtual reality and augmented reality on visitor experiences in museum". In: *Information and communication technologies in tourism 2016*. Springer, 2016, pp. 621–635.

[11] Sarah Ketchell, Winyu Chinthammit, and Ulrich Engelke. "Situated Storytelling with SLAM Enabled Augmented Reality". In: *The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*. 2019, pp. 1–9.

[12] Hansol Kim, Kun Ha Suh, and Eui Chul Lee. "Multi-modal user interface combining eye tracking and hand gesture recognition". en. In: *J Multimodal User Interfaces* 11.3 (Sept. 2017), pp. 241–250. ISSN: 1783-7677, 1783-8738. DOI: 10.1007/s12193-017-0242-2. URL: http://link.springer.com/10.1007/s12193-017-0242-2 (visited on 05/07/2020).

[13] Alchemy Immersive Magic Leap inc. *Museum Alive*. en-us. URL: https://world.magicleap.com/en-us/details/com.alchemyimmersive.museumaliverelease (visited on 05/05/2020).

Appendix A

Appendix A name

Here is the first appendix