Mobile Augmented Reality as a Control Mode for Real-time Music Systems

Robert Clouth

MASTERS THESIS UPF / 2013

Master in Sound and Music Computing

Superviser: Sergi Jordà
Departament of Information and Communication Technologies
Universitat Pompeu Fabra, Barcelona

UNIVERSITAT POMPEU FABRA

Contents

Abstract

Mobile augmented reality has only been made possible in the last few years, owing to the increasing processing power, more responsive hardware sensors and improved cameras of modern mobile computers. As a result of this, the mobile app markets have started filling up with augmented reality applications, ranging from games to navigation assistants. Similarly, the increasing power of mobile devices combined with their multitude of interaction inputs has made them extremely suitable for music purposes – real-time synthesis, control and editing – and we are beginning to see more and more music creation apps on the market. However, as of yet there have been no attempts to mix these two ideas – augmented reality and music creation – in neither commercial products nor research. As such, an augmented reality real-time music system has been implemented inspired by the tangible table-top real-time music system, the Reactable with the aim of investigating the effects of an augmented-reality control mode on the enjoyment, intuitiveness of interaction, immersion and ease-of-use of a real-time music system. A series of subjective user tests were conducted and the results show that using an augmented reality control mode over a more traditional top-down screen-centric control mode improves the perceived enjoyment and engagement of the music experience, both of which are essential properties for such systems.

Acknowledgements

I'd like to thank Dr. Sergi Jorda and Sebastian Mealla for their help and inspiration during this thesis. Also I'd like to thank my fellow course mates, because it was by talking to them over a coffee and a cigarette that some of the most challenging problems of the thesis were solved. I am also very grateful to my family for both their emotional support and their cold hard cash, because without those I wouldn't have been able to do this.

1. Introduction

We live in a physical three-dimensional spatial universe populated tangible objects and scenes that can be seen, heard, touched, smelled and tasted. Since its humble origins in a chemical soup or as a stowaway on a stray meteorite that impacted the early earth, life has existed fully within the confines of this physicality. Billions of years of evolution later, we arrive at the modern-day human. Every step along this labyrinthine path of mutations and selections has been guided by the laws of the physics, and our perception is no exception. The way we perceive the physical world has been baked into us - the way objects move in space, the way they feel in our hands, the way perspective changes as we move our heads all seem so completely natural that we take them for granted. Then with the dawn of computing, the birth of the byte and the prelude of the pixel, we began to peer into a new realm of reality, that of the unreal: the virtual. All of a sudden we found ourselves dealing with on a daily basis the intangible bits of the digital realm. The internet came and people began flooding the world with data, and now 20 years later, we are drowning in 0s and 1s. We have got *used* to dealing with seas of digital information and tools have been created to help, but we can't overrule in just half a century 4.5 billion years of evolution that tells us digital is unnatural. Interfaces were developed in an attempt to close the increasing gap between us and the information, but while they made dealing with it faster, the gap didn't close. A new approach was needed to try to bring the real and the virtual back together: augmented reality (AR).

1.1 Augmented Reality

The term "augmented reality" was first coined by Professor Tom Caudell in the 1990s while working in Boeing's Computer Services' Adaptive Neural Systems Research and Development project in Seattle [1]. He applied the term to a head-mounted digital display that guided workers through assembling electrical wires in aircrafts. The early definition of augmented reality, then, was an intersection between virtual and physical reality where digital visuals are blended in to the real world to enhance our perceptions.

Information in the 2010s is highly contextual; it's relevant right here, right now and its importance decreases rapidly with increased distance in both space and time from the event. AR closes the gap between us and this spatial temporal data by overlaying it over reality, obtaining the flexibility and power of the unreal with the intuitiveness and naturalness of the real. It shouldn't be seen as the yielding of virtuality to reality, but as a symbiotic relationship in which the two inhabit the same space.

More formally, AR is a technology which allows the seamless overlaying of virtual computer-generated objects over the physical world in real-time. AR has been defined by Milgram [2] as existing on a continuum of real to virtual environments, and is thus a subset of mixed reality (figure 1). At one end is reality, and at the other end is completely virtuality (VR).

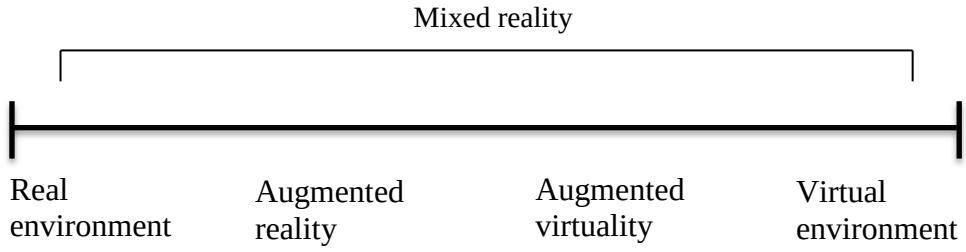

Figure 1 - The reality-virtuality continuum (adapted from Milgram [1])

AR differs from *augmented virtuality* in that the surrounding environment is real and populated with virtual objects, rather than a virtual environment being supplemented with real objects.
Perhaps a more formal definition provided by Azuma [3] states that an AR system has the following properties:

- Combines real and virtual objects in a real environment
- Runs interactively and in real-time
- Registers (aligns) the real and virtual objects with each other

Notice that the above definition does not mention the *sense* that is used to perceive the augmentations, AR can also be applied to other senses such as hearing [4].

With the miniaturisation of computers both in terms of size and price, it was only natural that they became more portable and ubiquitous, allowing the possibility of mobile AR. However, it has only been in the last few years that the processing and graphical power of mobile computing devices has reached a sufficient level for handling the high demands of a full AR system and so the field is still in relative infancy.

1.2 Mobile Music Applications

Advancing simultaneously both parallel and perpendicularly to the field of augmented-reality was that of digital music technology. *Music technology* once meant choosing the right wood and picking the right shape and dimensions of the body of a violin to get the best sound, in the 60s it meant synthesisers, in the 90s it meant desktop computers and VSTs, and in 2013 it means *mobile music applications*. Since the boom of mobile computing in the last few years and the resulting explosion of the mobile application development market, we have begun to see hundreds of music applications hit the various app stores, as people are beginning to realise the potential in using mobile devices as portable synthesisers [5], controllers [6] and even complete DAWs [7].

1.3 Motivation

Since the recent coming-of-age of mobile AR, we are beginning to see AR applications and games appearing in the Google and Apple app stores. These are typically either apps to aid navigation or to provide geo-location based information, or single player games that make limited use of AR. Similarly there are hundreds of musical applications on the market. However, there are currently no music applications that use AR as a control mode and no studies have been conducted investigating its uses as a musical interaction paradigm, and for

these reasons much research remains to be done in this area. The aim of this thesis is explore in detail the effect of an AR-based control mode on the intuitiveness, immersion and usability of a real-time music system.

2. State of the Art

The state of the art will be split in to two key sections:

- Enabling technologies – reviewing the technologies necessary to create a convincing AR experience
- Interaction techniques – reviewing modes of interaction made possible with AR

2.1 Enabling technologies

The following technologies are essential for a compelling AR experience:

- Display – necessary for viewing the augmentations
- Tracking methods – to overlay the virtual objects correctly the position of the device must be known

In this section these technologies will be examined with the main focus being on those specifically relating to mobile computing.

2.1.1 Displays

The display technologies can be grouped into three categories: head-mounted, projective and handheld.

Head-mounted displays (HMDs) are mounted on the users head and are either optical see-through (OST), or video see-through (VST). Optical see-through are those that allow to wearer to view the real world unimpeded with their natural eyes, and the virtual augmentations are overlaid via a holographic optical element or a half-silvered mirror. The advantage of this is that the clarity and response rate of the real are superior, however as a result of this the virtual overlay must be more responsive to maintain the seamlessness essential for a convincing experience. Google Glass is an example of an optical see-through HMD. Video see-through HMDs provide the user with a video view of the real world over which the virtual graphics are displayed. With VST HMDs, synchronising the real and the virtual is easier due to the added control over the real imagery with image-processing techniques.

Figure 2 – Video see-through HMD (left); optical see-through HMD (right)

Projective displays project the virtual information directly onto the real object with either a head-mounted or room-mounted projector. The advantage of projective displays is that they are less intrusive than HWDs since they don't require the user to wear any sort of device. One problem with this type of display however is that is cameras and projectors are difficult to operate simultaneously due to differing lighting requirements [8].

Handheld displays are typically flat-screen TFT displays with an attached camera, which act as a window through to the virtual space. Handheld displays are highly suitable for AR applications due to their mobility, low intrusiveness, low cost and ubiquity, owing to the boom of the mobile computing industry over the last decade. Early mobile AR prototypes like MARS [9] were often based on notebooks or custom hardware situated in bulky backpack, but in 2004 Möhring et al. [10] first presented a self-contained AR system running on a cell phone. The increased mobility and ease-of-use came with drawbacks – namely the low resolution of the camera input and the low computing power of the device. However with the vastly superior hardware and computing power of modern tablet PCs and mobile phones, these problems are reduced, making these devices particularly suitable for a mobile AR platform. A disadvantage of handheld screens over HMDs is that the immersion is reduced due to act of having to use a device in your hands in order to explore the space.

2.1.2. Tracking Methods

In order to accurately overlay the virtual information over the real, the position of the viewer must be known. In this section we describe some of the most effective state-of-the-art tracking techniques that are suitable for use on mobile devices, focusing on sensor-based techniques and vision-based techniques.

Sensor-Based Tracking Techniques

Sensor-based tracking techniques utilize various sensors such as acoustic, magnetic, inertial, GPS and optical sensors in order to localize the position of the viewer. Global positioning satellite (GPS) systems are commonly used for outdoor mobile device tracking, but due to the requirement of an active link to the satellites these don't function well indoors. For this reason, indoor positioning (IP) has become a field in itself focusing on the use of wireless technologies in order to localize the device. Apart from the sensor type (or combination) used the other variable that changes is the position estimation technique, which could be triangulation (positioning by using the geometric properties of triangles and the known positions of three reference elements), fingerprinting (positioning by comparing to pre-measured location related data) or proximity (positioning by detecting which reference point the sensor is closest to). Commercial mobile devices like smart phones and table PCs have limited sensors so effort has been put into performing IP with the sensors that are readily available on such devices, such as Bluetooth and WLAN. The problem with these techniques is that due to the complexity of indoor environments that contain various influencing factors, the accuracy of these systems is typically low, in the order of 1-3m [11].

The development sensor-based techniques slowed towards the end of the millennia due to their extensive use in virtual reality research [12], so most current research in this field focuses on combining different types of sensors for ubiquitous tracking, for example on how to smoothly handle the cross-over between different sensing techniques with dynamic data fusion [13]. A more recent example of sensor data-fusion attempts to combine accelerometer-based

dead-reckoning with Wi-Fi signal fingerprinting showing an improvement in localization accuracy compared to using each method separately [14].

Vision-Based Tracking Methods

Vision-based tracking is a significantly more active area of research than sensor-based techniques with approximately 80% of tracking papers submitted to the ISMAR conference describing computer-vision methods [12].

Vision-based methods use image processing of camera images to determine the camera pose relative to real-world objects. Many of the modern techniques are either feature-based or model-based.

Feature-based techniques first attempt to find a correspondence between features in the 2D image and their world frame coordinates in 3D space. This is typically done by exploiting prior knowledge of the geometry of the scene and camera lens. The estimated 3D coordinates are then projected into the detected 2D features to determine the pose of the camera [15].

Regarding the prior knowledge of the geometry, this can be in the form of fiduciary tracking markers with a known size and shape, for example with the popular ARToolkit system first published by Kato et al. in 1999 which utilized square markers [16]. By 2002, progress had slowed with general marker systems and much of the work moved over to tracking natural features. *Natural features* refers to points, lines, edges or textures that might be present in the camera image. A typical natural feature tracking system will initialise the camera pose using known features then continuously update the pose by detecting natural features in the image so that a marker need not remain in shot. One example of a natural feature tracker was proposed by Vacchetti L. et al. which combined tracking detected interest points in the image with an edge detection algorithm for improved robustness.

More recently, the focus of vision-based tracking methods has moved onto model-based tracking. Model-based systems track the camera by attempting to fit a known 3D model to the camera image. This 3D model can either be a hand-created model (e.g. in [17]) or generated at runtime through a process called simultaneous tracking and mapping.

Simultaneous tracking and mapping (SLAM) is a technique by which a system can simultaneously build up a 3D model of an unknown environment and track its current position. While this technique was initially conceived in the late 80s and later extended in 1991 by Lenard J. et al. [18], the first implementations primarily used non-optical sensors, and it wasn't until 2005 with the reduced cost of cameras and computers that the focus of SLAM research switched over to visual sensors [19]. SLAM research before this time was contained within the field of robot navigation, and then in 2007 Georg Klein and David Murray saw the potential of these techniques for use in AR and published their seminal paper: Parallel Tracking and Mapping for Small ARWorkspaces [20]. The system which is described in this paper first requires the map to be initialized with a stereo pair. This involves taking two photos a small translation apart which are then analysed to find matching pairs of regions in each image. The disparity between the pairs is then used to estimate to estimate the 3D position of the point. Once this initial map has been generated, it is expanded dynamically as the user moves the camera to reveal new areas of the environment. This map is used to estimate the pose in a method similar to the model-based methods described earlier.

Two years later the same group adapted the algorithm for use on an iPhone, most of the changes being to account for its significantly reduced processing power, the reduced field-of-view, the reduced frame rate and resolution of capture and the rolling shutter [21]. Although the

robustness and speed of mapping was much worse than the PC version, the system still displayed impressive tracking abilities and showed that tracking in a previously unknown environment is possible on current smartphones.

2.2 Interaction Techniques

As important as developing the technology itself is creating interaction methods that allow the user to experience and interact with virtual objects in a natural and intuitive way. While many interesting approaches have been explored since the beginnings of AR, it is only relatively recently that mobile devices such as smartphones and tablets have had the processing power to support immersive AR systems, and as a result the field of mobile AR interactive is very much a maturing area. Mobile AR presents added problems due to the reduced screen space, the limited resolution, CPU power and the fact that one hand is required to hold the device [22]. This section will first detail some novel approaches to AR interaction in general with the focus on the approaches that are applicable to mobile AR, and then later document some of the recent attempts to tackle some of the issues specific to mobile AR. Finally recent studies relating to collaborative mobile AR will be reviewed.

2.2.1 Tangible AR

From the first months of life, people spend their entire lives interacting with objects in the real world, and their handling and manipulation comes naturally to us. Tangible AR tries to take advantage of this to ease the interaction with virtual objects in the scene by coupling them with real-world physical counterparts. AR, with its mix of the real and the virtual makes this possible and it presents a potentially very intuitive interaction method owing to the familiarity and the natural properties of real-world objects. The idea of the tangible user interface (TUI) was introduced by Ishii H. and Ullmer B. in 1997, in which they proposed a paradigm they named "Tangible Bits" through which user can manipulate digital information by coupling it with physical objects facilitating a more natural interaction with the virtual domain [23]. Tangible AR is the fusion between this approach and augmented reality – an attempt to combine this intuitive interaction we have with physical objects with the power and flexibility that virtual overlays can provide.

An early example that demonstrated the effectiveness of tangible AR was the VOMAR system developed by Kato et al. with which a user with no previous experience of AR could easily create 3D virtual scenes using a real paddle to select and arrange virtual furniture. This was a result of the intuitive mapping of application commands to the physical affordances of the paddle [24]. A more recent demonstration of a tangible AR system is the Universal Media Book – a physical book that acted as a tangible interface to multimedia material. The "book" could be loaded with videos, images or even volumetric data such as MRI scans that could be explored by removing the page and using it as a 2D cross-sectional view onto the data [25].

While the concepts discussed here are applicable to handheld mobile AR, they have somewhat less relevance since one hand at least is required to hold the device, reducing the possibility of interaction with physical objects.

2.2.2 Mobile AR

Interaction techniques for handheld mobile AR typically focus on device-centric methods using the touchscreen as input. However, as highlighted in [22], [26] and [27], using touch input for control in handheld mobile AR scenarios can be problematic for the following reasons:

- Limited screen space – mobile devices typically have small screens so that they can be easily held in the hand which makes selecting smaller objects difficult especially because the finger can occlude large areas of the display.
- Handheld display – the fact that the display is held in the hand means that the view is shaky, again increasing the difficulty of selecting small objects.

Different attempts at tackling some of these problems have been performed and we will take a look at some of the most relevant subsequently.

Touch-based Interaction

Several approaches have been developed in recent years that attempt to combat some of the issues highlighted above for touch-based interaction. One such method first presented in [27] as "freeze-set-go" allows the user to freeze the live camera feed and yet continue to manipulate the AR scene while the viewpoint stands still, addressing the 'shaky view' problem mentioned above. The virtual objects in the scene can still be interacted with, and are registered to the frozen real world background. When the user can completed the desire manipulations, they could unfreeze the view and the scene would resume the live update using the new setup created whilst frozen. The results of the study showed that compared to without it, freeze-set-go doubled the accuracy of a line drawing task and that the users felt both less-tired and that the tasks were easier. A later study [28] further analysed this interaction method and discovered that users consistently found the freeze-view touch input to be easier to learn and use, more natural and intuitive, and less mentally and physically stressful than classical free-view touch input. However, the results also showed that users found freeze-view input to be less fun and engaging than free-view, showing that freeze-view input is perhaps better for serious tasks rather than for gaming or entertainment.

Another approach to the limited screen-space problem is view-zooming. When a user needs a closer view they will move the device closer to the scene, which can sometimes result in loss of tracking. In [29] Lee G. et al. proposed a solution to this using an automatic zooming method, where the zoom factor is updated smoothly based on the distance between the viewpoint and the target object. The study showed via user tests that the technique is considered natural and easy to use.

[22] and [26] describe similar device-based systems where instead of moving the target virtual object with the touchscreen the object is first 'picked-up' with either a screen tap or holding a cursor over the object for a period of time, and then manipulation by translating or rotating the device itself. User tests performed in [22] reported a 50% speed increase in object translation tasks, but a reduction in speed of rotation tasks. [26] reported similar findings that showed that the device-based approached showed significant improvement in the accuracy and speed of translation tasks over touch-based interaction.

One such successful application of a hybrid touch-based and device-based interaction method is MIT Media Lab's exTouch project [30], in which an actuated object could be manipulated by an embodied spatially-aware interface. The system seamlessly combined touch-based interaction allowing the user to position and rotate the tracked actuated cube by traditional

drag and two-finger swivel touch gestures, which a device-based interaction the allowed the user to grab the real object through the mobile device and “push” and “pull” it. This melding of the two approaches allows the best of both – the speed of the touch-based with the intuitiveness of the device-based.

Gesture-based Interaction

Gesture-based interaction, avoids the above problems by allowing the user to interact directly with the virtual objects by extending their finger or hand out into the space inhabited by the augmentations. Perhaps one of the most famous examples of this mode of interaction is SixthSense [31], a wearable gestural interface that augments the physical environment with digital information, published in 2009. The system comprises of a camera to perform the image processing and gesture capture, a miniature projector to display the augmentations on the real surfaces and a mobile computer. The user can use gestures that range from simple pointing to more complex gestures for taking photos, checking emails and other functions.

Another more recent example of a hand-based AR interaction system is presented in [32], in which owing to a marker-less finger-tracking algorithm the user can interact with virtual objects in the scene with gestures. However, more recent studies suggest that while users tend to find this type of interaction method more fun and interesting, gesture-based interaction has less usability than classic touchscreen interaction. In [28] this is reflected in longer times for performing object selection and manipulation tasks and lower reported intuitiveness, ease-of-use and ease-of-learning scores. The 2013 study by Hürst W. et al. [26] also report similar findings regarding the low usability of finger-based methods compared to touchscreen methods. Both qualitative and objective data collected for the report suggests a very low performance of gestural interaction when trying to perform canonical operations such as selection and translation of virtual objects, especially as the tasks became more difficult. Another major problem identified is that that the physical action of holding up the hand for longer periods of time can be tiresome, limiting this methods use for serious applications.

Body-centric Interaction

Body-centric interaction refers to using the space on and around the body as the interaction space, essentially extending the interaction out of the screen of the mobile device. The techniques described in this section aren’t technically AR since the real-world image isn’t displayed and the virtual objects aren’t registered to it, but instead positioned relative to the body. However, the concept is closely related and thus interactions are relevant to AR.

Fitzmaurice first proposed in 1993 the idea of spatially-aware devices as a potential solution to the limited display space and growing information density problem of mobile devices, where the idea is to track the device in space so that it can be physically moved around to reveal areas of a large workspace [33]. This work was further extended into the concept of “peep-hole” displays for use with pen-based input to ease the process of drawing and writing on documented larger than the display. Peep-hole displays allow the user to move the device around in space to show portions of a large canvas, the device having the role of a window onto this space [34].

In 2012, Chen X. et al. [35] combined all of the previous work in this field to develop the unified idea of *body-centric interaction*, a n interaction paradigm which placed the body at the centre of the interaction space. They demonstrated this concept by prototyping a body-

centric mobile web browser where the user could manipulate bookmarks, tools and tabs positioned on and around their body.

2.2.3 Collaborative AR

Single user AR applications had been explored for decades before the first collaborative AR applications were developed in the mid-nineties. Early projects such as Shared Space [36] and Studiersube [37] demonstrated that AR can facilitate both remote and co-located collaborative activities that would be impossible without it. This section focuses on the latest advances in collaborative AR systems, emphasising those that utilize mobile devices.

AR has been found to be useful interface for 3D computer-supported cooperative work (CSCW), where shared physical workspaces are augmented facilitating co-located collaboration. For example, in user tests conducted with the Shared Space application users found the interface very intuitive and conducive to real world collaboration because the groupware support can be kept simple and mostly left to social protocols, unlike other interfaces [38].

Five key features of collaborative AR environments have been suggested [37]:

- Virtuality: Objects that don't exist in the real world can be viewed and examined.
- Augmentation: Real objects can be augmented by virtual annotations.
- Cooperation: Multiple users can see each other and cooperate in a natural way.
- Independence: Each user controls his own independent viewpoint.
- Individuality: Displayed data can be different for each viewer.

Several user studies have compared collaborative AR interfaces to other technologies and demonstrated the value of these features. In one such study, Kiyokawa et. al. used their SeamlessDesign interface - in which users are seated across a table from one another and engaged in a simple collaborative pointing task - to compare gaze and gesture awareness between AR and VR conditions. The results showed that due to the improved perception of non-verbal cues supported by the collaborative AR interface, subjects performed significantly faster in the AR interface than in a fully immersive virtual reality setting, and also reported that it was easier to cooperate.

Collaborative AR interfaces have been shown to produce similar communication behaviours amongst the users to unmediated face-to-face collaboration, more so than screen-based collaboration [39]. When people collaborate at a table the task-space is shared with the communication space, whereas when collaborating in front of screen the task-space is part of the screen-space, which may be separate from the communication space. This may introduce a discontinuity that causes collaborators to exhibit different communication behaviours. The study performed in [39] involved asking the subjects to complete logic puzzle tasks in three conditions: face-to-face collaboration with real objects, co-located AR collaboration with virtual objects and co-located projection screen-based collaboration with virtual objects. The virtual objects used were clones of the real objects and to allow the use of tangible AR interaction methods, they were attached to real objects. Users exhibited deictic speech patterns (phrases that include the words "this" and "that" combined with pointing gestures) and pointing behaviours that were more similar in the unmediated and AR settings than in the projection setting, with the tangible interface being the most successful.

More recently, face-to-face collaborative AR has been implemented on mobile devices. Henrysson et al. describe in [40] how a custom port of the ARToolKit library to the Symbian mobile phone operating system was used to develop a sample collaborative AR game. Using this system they then performed a series of tests to determine the value of the AR. Pairs of subjects played the game in each of the following three conditions:

- Face-to-face AR – the virtual graphics are overlaid onto the live video view from the camera.
- Face-to-face non-AR – as above but with just the graphics.
- Non face-to-face gaming – as above but the users couldn't see each other.

The results showed that with face-to-face AR users found it easier to collaborate and had a stronger sense of awareness of each other's actions.

An example of a modern collaborative system is T(ether), developed at the MIT Media Lab by Ishii H. et al. which is described as "a novel spatially aware display that supports intuitive interaction with volumetric data" [41]. With this system, the display acts as a window onto a perspective view of 3D data made possible through head-tracking. The system creates a shared workspace in which co-located or remote users can collaborate in both the real and virtual worlds, with touch input via the display and gestural input via a motion-tracked glove.

2.3 Conclusion

From the exhaustive state-of-the-art review conducted for this thesis, it was apparent that while a lot of work has been conducted looking into advancing the enabling technologies behind augmented reality – which is of course extremely important – not as much has focused on how the technology can be used. Yes, there have been papers investigating the new interaction techniques made possible with this technology, but none have focused on musical interaction and the benefits of using an AR-based control mode in this area. This thesis will expand the research in this area, hopefully closing this gap somewhat.

3. Methodology

In order to properly evaluate the effect of an AR-based control mode on a real-time music system, there are two groups of properties that must be examined – the subjective and the objective. Subjective properties include concepts such as perceived ease-of-use, perceived performance, intuitiveness, immersion and level of engagement, and must be obtained by asking people who have used such a system. Since no prior AR-based music systems exist, one was first implemented and is discussed in full subsequently.

3.1 AR Reactable

An AR-based real-time music system was implemented based on the Reactable. The Reactable is a tangible table-top real-time modular music system developed in the Music Technology Group at Universitat de Pompeu Fabra [42]. The decision to base the AR music system on the Reactable was due to its accessibility and that fact that its physical nature lends itself suitably to augmented-reality.
The implementation itself was performed using a number of software packages namely:

- Unity 3D – a game engine used as the main development environment
- Qualcomm Vuforia Augmented reality library – used to perform all aspects of the AR, from the image tracking to the image registration
- Reactable synthesis engine – used to perform the synthesis for the music system

3.1.1 Unity 3D

Unity 3D is a game engine with a built-in IDE [43]. We chose to use a game engine as the main core of the implementation since they typically handle all of the rendering, the scene-graph, transformations and other important aspects that would have been time consuming to implement ourselves. Other useful features include the ability to code in C# thus making it easily extendable, and the ease of deploying the final product to numerous platforms such as Android and iOS.

Figure 3 - Unity 3D

3.1.2 Vuforia

The Qualcomm Vuforia library was chosen to handle the augmented reality parts of the AR Reactable because it was created by a company who design the CPUs in mobile devices and so is heavily optimized for them [44]. This makes it extremely efficient and so it is possible to achieve high-frame rates and as a result a greater user experience. It also integrates seamlessly with Unity 3D reducing development time and complexity. Vuforia uses image tracking techniques based on tracking natural features in the scene, so the tracking fiducial can be any image as long as it has sufficient detail.

Vuforia Tracking Algorithm

Vuforia tracks scale-robust corners extracted with a modified version of the FAST (features from accelerated segment test) corner detector [45]. Below is an image showing a typical output from the FAST corner detector. The corners that are chosen are done so because they are shown to be reliably detected at varying scales and rotations of the image.

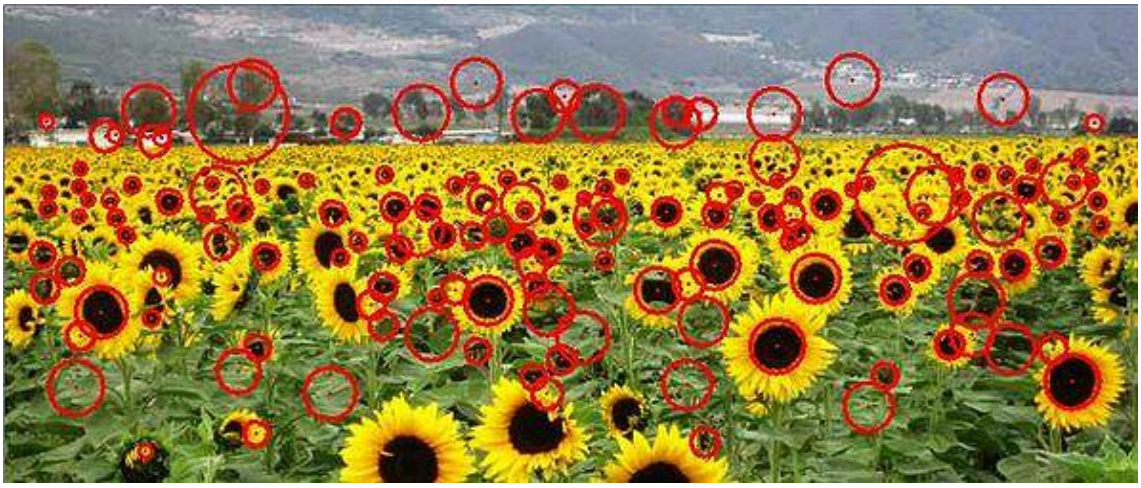

Figure 4 - Image with detected corners highlighted

Because the tracking is based on natural images rather than a specially designed fiducial such as with the ARToolkit [46], any image meeting certain criteria can in theory be tracked. However the image must have sufficient detail so that enough corners are detected to be tracked. Also the image must not be repetitive (e.g. a brick wall) as the tracking cannot distinguish between different parts of the image and the tracking can fail.

The tracking algorithm works as follows [47]:

- Firstly, the desired tracking fiducial has to be analysed with the corner detector and the detected points stored in a database. This is the reference data and is used for the fiducial detection and tracking. This must be done before the tracking begins.
- The video feed from the device is analysed in real-time with the corner detector and matching points between the detected points in the video feed and in the pre-calculated reference data are found based on comparing the pixels around each corner.

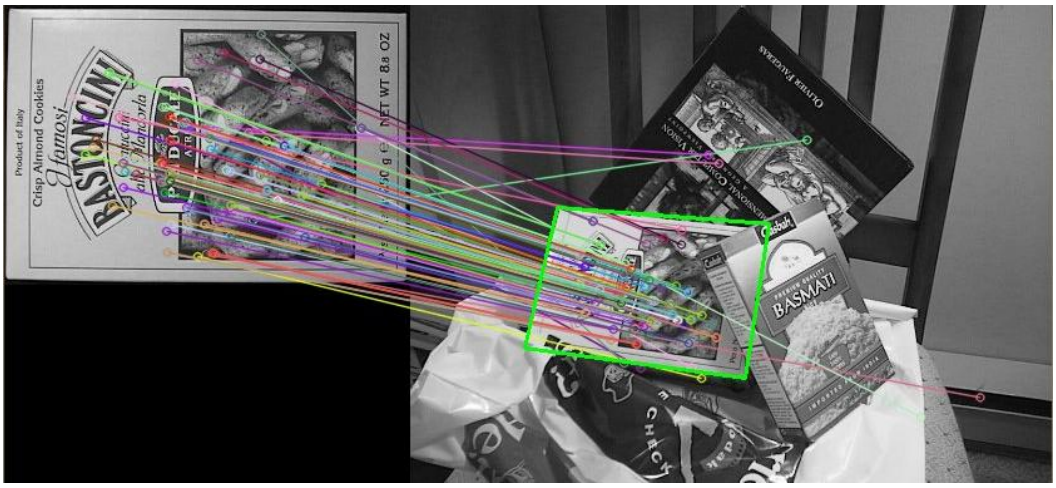

Figure 5 - Matching points detected in the reference image and the video feed

- Once a sufficient number of matching points have been located, the homography (a projective transformation matrix) that maps the set of points in the reference image to their pairs in the video feed is estimated using error minimization methods.

$$\begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

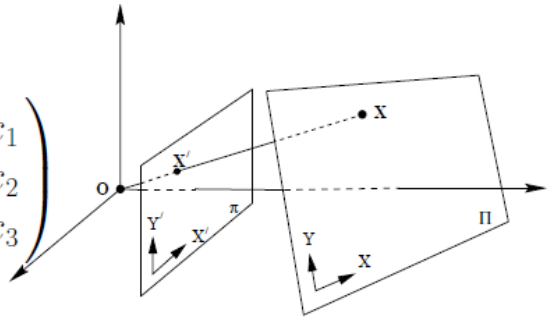

Figure 6 – The homography H that maps point x to x', or x' = Hx

- From the homography the camera pose is estimated using an appropriate camera model, and the virtual graphics to be overlayed over the video feed can be given the appropriate angle and perspective.

3.1.3 Synthesis Engine

Rather than implementing a real-time synthesis engine from scratch, it was decided to use the synthesis engine used by the Reactable as it is flexible, powerful and written in Puredata making it easy to modify. As this AR Reactable is just a proof of concept, it was decided to perform all synthesis off-device on an external computer. This was made easy by the fact that the Reactable synthesis engine is controlled via the TUIO protocol. Only three objects from the Reactable were implemented; a saw wave generator, a sine LFO and a bandpass filter. These were chosen because together they can produce an interesting array of sounds. Below is a diagram showing the full AR Reactable system.

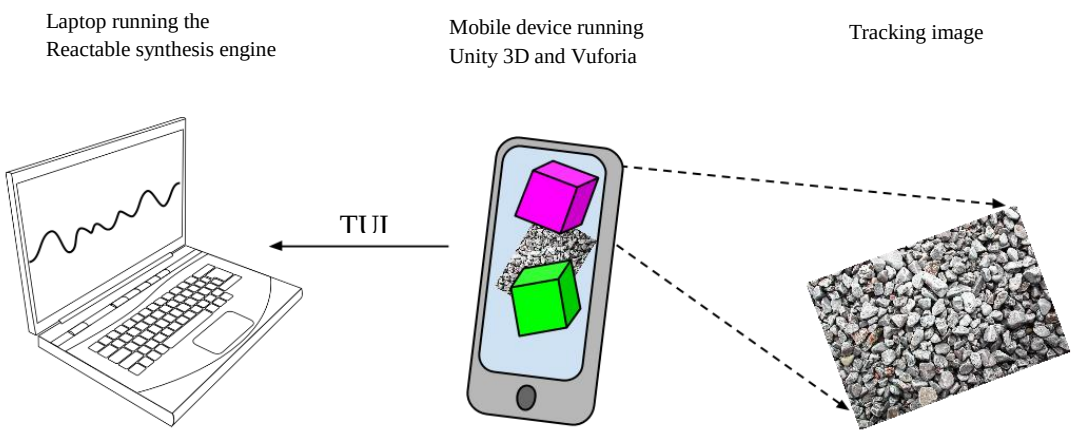

Figure 7 – AR Reactable system diagram

Figure 8 - AR Reactable, augmented reality mode

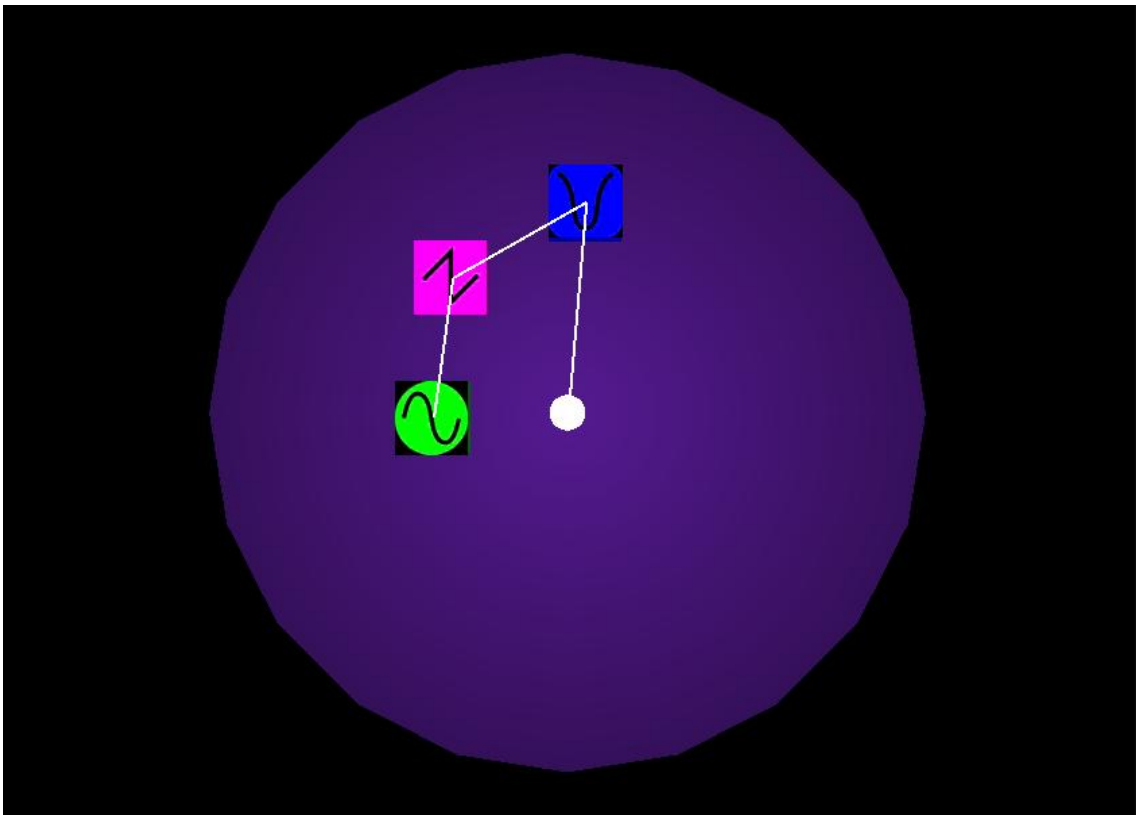

Figure 9 - AR Reactable, top-down view

3.1.4 Control Modes

Two other control modes other than the 3D augmented reality mode were implemented so that a comparison was possible. The three modes are detailed below.

Top-down Without Device-tracking

The first of the two modes is a top-down view similar to Google maps or Reactable Mobile. The objects are seen from above and the view is panned and zoomed with the traditional drag and pinch zoom gestures. The objects themselves are translated with a tap and drag, and rotated with a double tap hold and subsequent vertical drag. All interaction is screen-centric.

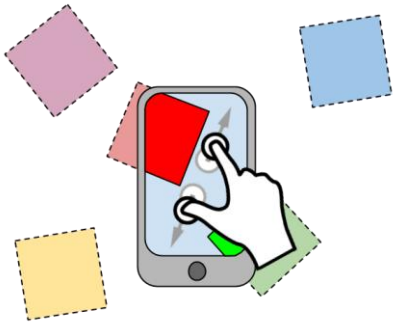

Figure 10 - Top-down without device-tracking

Top-down With Device-tracking

The second alternative control mode is similar to the top-down mode described above but with device-centric instead of screen-centric interaction. Specifically, moving the view is performed by physically moving the device – vertically and horizontally to pan and in and out to zoom. This view was inspired by the peep-hole view detailed in [34].

The tracking in this mode is performed using the Vuforia library to get the device position on a plane that is perpendicular to the camera direction.

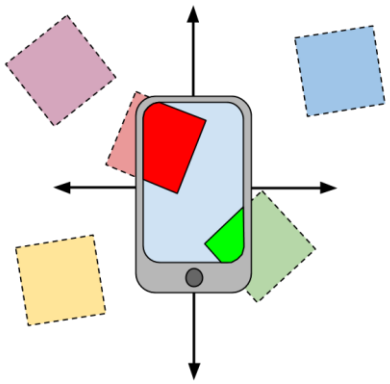

Figure 11 - Top-down with device-tracking

3D Augmented Reality

The final mode is the full augmented reality. The navigation has six degrees of freedom, and the scene is navigated by moving the device, with the screen of the device facilitating as both the view and the interaction. Control is device-centric as with the top-down with device-tracking mode, in which the objects can be picked by tapping on the screen, then moved by physically moving the device.

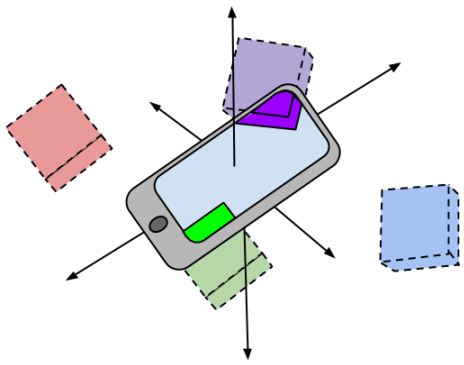

Figure 12 – 3D augmented reality

4. Evaluation

The evaluation of the AR Reactable took the form of subjective user tests in which 9 participants were asked to use the system in each of the three control modes – top-down, top-down with device-tracking, and 3D AR – and to complete various questionnaires regarding the modes. Before beginning the test, they were asked to complete a preliminary questionnaire to gauge their familiarity with music composition, real-time music systems and musical instruments in general.

After completing the preliminary questionnaire, the participants were asked to use each of the 3 control modes for 2 minutes, given the goal: “compose a short piece of music”. The modes were presented in a different order to each participant to reduce the effect of the order on their reactions to the control modes. After the 2 minute usage, a 7-point Likert-scale questionnaire was presented to them featuring statements that commented on the perceived engagement and fun, control, ease-of-use and learning, naturalness and responsiveness of interaction, and connection to the music being made. Once all 3 modes had been used, a final comparative questionnaire was presented that asked them to compare the 3 modes. This was to judge if they provided consistent answers.

All of the tests were performed in the same environment with the same lighting. The tracking image was printed at A1 size and placed flat on a table-top. The device used was an Android tablet with a 7” screen.

5. Results

5.1 AR Reactable Subjective Tests

The preliminary questionnaire results are shown in Table 1.

Question	Yes (%)
Have you ever used the Reactable before?	77
Have you ever used the Reactable Mobile application before?	67
Have you used pinch-zoom and drag gestures?	100
Are you familiar with augmented reality?	89
If yes, have you ever used an augmented reality system before?	56
Have you ever composed music before?	89
Can you play any instruments?	100

Table 1 - Results from preliminary questionnaire

These results show that generally the participants were musically-minded since all of them could play a musical instrument and most of them had composed music before. All of them were had used the pinch-zoom and drag gestures, but only just over half of them had used an AR system before.

As an introduction the post-test results, figure 8 shows the mean score for each of the statements in the questionnaire for each control mode, and figure 9 the standard deviations.

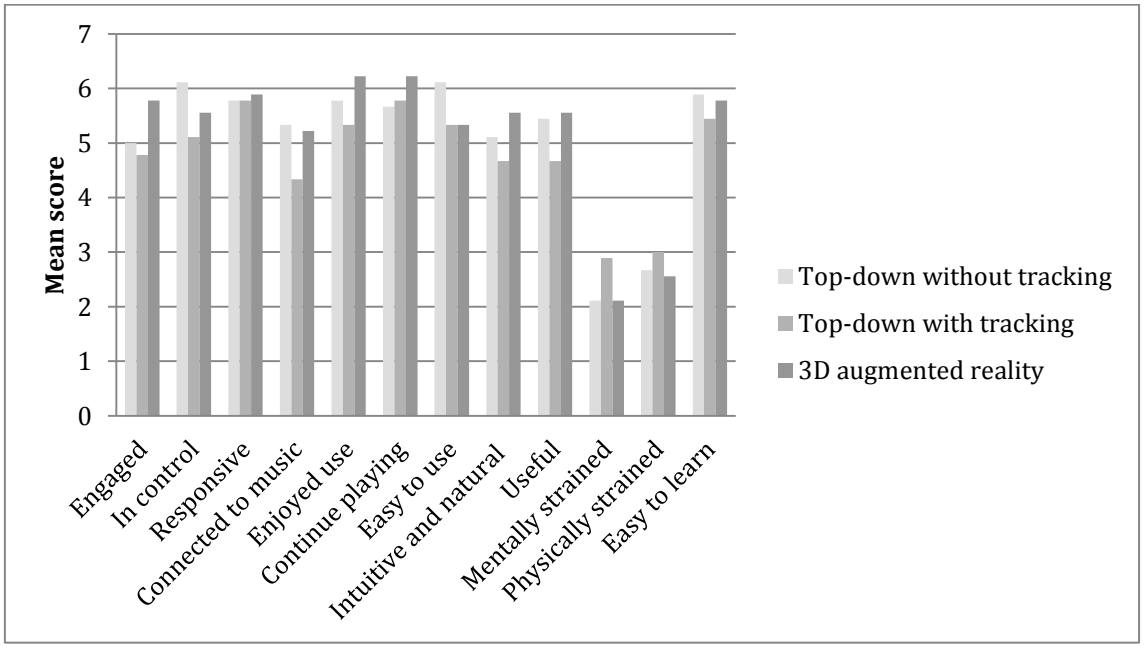

Figure 13 - Mean scores form the post-test questionnaires for each control mode

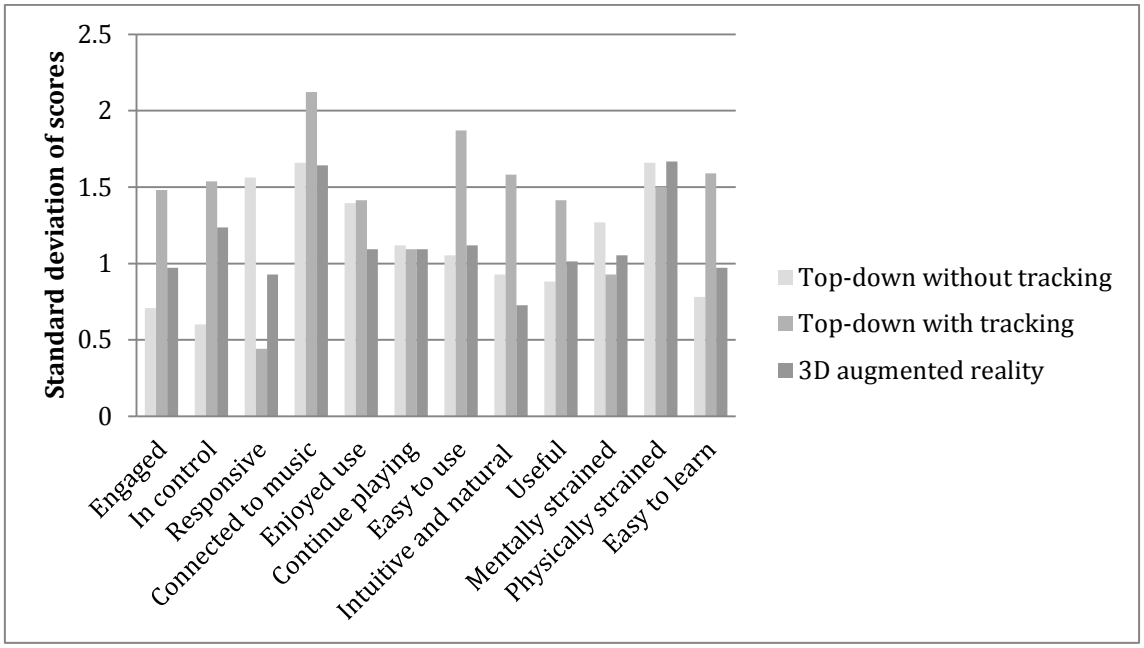

Figure 14 - Standard deviations of the scores from the questionnaires given after each control mode test

From the above figures, estimations of the differences between the control modes can be made. In terms of engagement, the 3D augmented reality mode comes out ahead, with the top-down modes rated a small amount lower. Explanations for this could be that the AR mode is perhaps more novel to the users as shown by the preliminary questionnaire, or that the AR mode is more engaging because of its less abstract nature.

The users appear to have felt that it was during the top-down without tracking mode they had most control. This is supported by the fact that there is a low standard deviation suggesting that this sentiment was shared. This is partly due to the familiarity of the pinch-zoom and drag gestures used to navigate this mode, shown in the preliminary questionnaire, and partly due to the unstable nature of the tracked modes. For example, the handshake that occurs when holding a device effects the selecting of objects, especially if they are small. This could lead to a lower feeling of control.

Interestingly, all 3 modes received similar mean responsiveness scores, but the top-down with tracking was most consistently scored shown by the lower standard deviation. Perhaps this is due to the novel sensation of the tracking (which was smooth owing to the responsiveness of the Vuforia library) and the familiarity of the top-down interaction leading to the perception of a more responsive experience.

Due to the high standard deviations of the perceived level of connection to the music, it is clear that the users had different interpretations of this statement. This statement is too vague (how do you define ‘connection’ in this context?) and will not be included in future questionnaires. As a result, the data from this statement will not be discussed.

The combination of the higher mean score and lower or same standard deviations for the ‘enjoyment’ and ‘desire to continue playing’ statements suggests that the users enjoyed the 3D augmented reality mode more than the others, possibly because of the novelty. Longer term studies would show if the enjoyment remained when the users were more familiar with the mode.

Generally users found the top-down without tracking mode the easiest to use and learn most likely due its familiarity. Users were contested about whether the top-down with tracking mode was easy to use and to learn, shown by the large standard deviations in each statement.

The users found the 3D AR mode to be the most intuitive and natural possibly as it's the least abstract and closest to interacting with real objects. The perceived utility of the top-down without tracking and the AR mode were approximately the same, however the top-down with tracking mode was lower and again with a high standard deviation. It appears that generally people are undecided about this mode. The mental and physical strain appears to be higher for the top-down with tracking mode than the other modes.

One thing to note is that the number of participants is small due to time constraints, and thus the results are not as solid as they could be with more data. To determine if the lack of data affected the validity of the conclusions, the Friedman test was performed on the data to determine if there is statistical significance in the differences between the control modes.

The Friedman test is a non-parametric statistical test used to detect differences in treatments across multiple test attempts, in our case the difference in questionnaire score across the user trials. The output of the Friedman test is the *p*-value, or the probability that the observed data would occur by chance in a given single null hypothesis. The null hypothesis in this scenario is that there is no difference between the 3 control modes. Table 2 shows the *p*-values for the post-test questionnaires.

Statement	*p*-value
I was engaged by the experience	0.0388
I was in control	0.2034
It was responsive to my actions	0.5682
I felt connected to the music I was making	0.1738
I enjoyed using it	0.0259
I want to continue playing	0.2592
It was easy to use	0.1353
The interaction felt intuitive and natural	0.209
The method of control was useful	0.0475
I felt mentally strained when using it	0.1426
I felt physically strained when using it	0.2801
It was easy to learn	0.8187

Table 2 - *p*-values calculated with the Friedman test for each statement in the post-test questionnaire

For example, the *p*-value for 'it was easy to learn' is 0.8187 which says there is an 81.87% chance that this data was appear even if there was no actual difference between the 3 control modes, meaning that any conclusions drawn from this result are weak. With a significance level of 0.05, the data collected for 'I was engaged by the experience' (0.0388) and 'I enjoyed using it' (0.0259) are significant, because it is unlikely that they could have happened by chance. However, in general the application of the Friedman test here highlights the fact that not-enough data was collected to provide strong evidence of the differences between the modes. This would need to be remedied in future studies.

To conclude, from the data obtained in the AR Reactable subjective user tests we can see that generally the users felt that the augmented reality control mode was the most enjoyable, natural and intuitive mode of the three, however the top-down mode without device-tracking

was the easiest to learn. This could be due to the familiarity of this mode. For real-time music systems, the level of engagement is essential to how connected you feel with what you are making, how ‘in the zone’ you feel. So for this reason, augmented reality could benefit this type of application. However, in terms of accuracy previous studies [28] have shown that with device-centric augmented reality interaction modes, object manipulation accuracy is harder to achieve. Sometimes in live-music situations, accuracy isn’t important as part of the time you may be experimenting and you don’t know exactly the sound you want. In this case, the lower accuracy wouldn’t be a problem. There are nonetheless times when the opposite is desired – when, for example you have an idea for exactly the sound you want and you know how to get it. In this situation, accuracy is paramount. Here is where the AR control mode could fail. This conclusion could be tested by performing the objective user tests with the game-with-a-purpose.

6. Conclusion

In order to determine the effects of an augmented reality control mode on real-time music systems, an AR version of the interactive table-top instrument, the Reactable has been created. To enable comparison two other control modes were also implemented, namely a top-down view with traditional pinch zoom and drag interactions, and a top-down view with device-centric interactions where the scene and objects can be navigated and manipulated by moving the device. These three modes were then compared in a series of subjective user tests involving 9 participants, in which the users were asking to play with each of the modes, and then complete a series of questionnaires on each of the modes, containing questions that focused on the intuitiveness, the ease-of-use, the enjoyment and other subjective aspects of the modes. The results of the tests show that users generally found the augmented reality mode to be the most intuitive and enjoyable of the three modes, and since these are essential for real-time musical instruments, imply that augmented has potential for these types of music systems. However due the low number of participants the validity of these conclusions is questionable, as shown by the high p-values in certain questions in the questionnaires. This would be remedied with repeat experiments with more participants.

A game-with-a-purpose was also developed in order to determine the objective effects of using AR in real-time music systems. This game - called 'Super Hyper Circle Mover' – is based on the classic Fits experiment, and challenges the player to move a circle to a given location and rotate it to a given angle. They are award points based on the speed and accuracy of the motions. During play, every gesture made is recorded to the phone for later analysis. This was to be released for two weeks and the data collected and analysed to make comparisons about the accuracy and speed of gestures in each control mode, however due to time restrictions this was never accomplished and is documented as part of this thesis purely for the aid of future researchers aiming to work in a similar area.

All-in-all, the work presented in this thesis shows that an augmented-reality based real-time music system in not only just possible, but also can have potential benefits in terms of enjoyment, intuitiveness and naturalness of interaction.

7. Future Work

7.1 Objective tests

The other side of analysing the potential of AR for real-time interactive music systems is the objective side – how quickly and accurately can the system be manipulated and navigated with AR as a control mode, in comparison with other control modes. In real-time music the speed and accuracy of the interaction is essential if the user is to feel in control of what they are making.

For this purpose a test environment was designed and implemented in the form of a game, however due to time constraints the tests were never performed, so the details of this game and how the data would be analysed are documented here solely for the assistance of researchers that are seeking to perform similar tests.

In order to obtain data on the precision and speed of interaction in an AR-based control mode compared to other modes, objective user tests needed to be performed, data collected and a test environment developed. Since the types of gestures that we wanted to analyse involved point and drag motions, a test based on Fitts' law would be suitable.

Fitts' law is a speed-accuracy model of human pointing gestures that predicts the time required for a rapid, aimed movement to a target area [48]. It was proposed by Paul Fitts in 1954 and since then has become a cornerstone in the HCI research community.

Equation 1 shows the Shannon form of Fitts' law.

$$T = a + b\log_2(1 + \frac{A}{W}) \tag{1}$$

T is the time taken, a and b are empirically determined constants, A is the amplitude (or distance), W is the target width, the $\log_2$ term is the index of difficulty which is task dependent, and $1/b$ is the index of performance. a and b are determined by plotting the recorded data from the experiment and performing line-fitting.

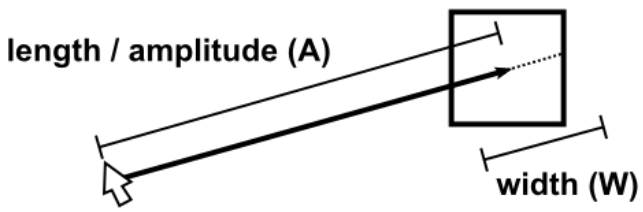

Figure 15 - Diagram showing the meaning of A and W in Fitts' law

The tests would be performed with each of the three control modes described previously – top-down, top-down with device-tracking and full 3D AR – so that comparisons can be made.

7.1.1 Game-with-a-purpose

Rather than create this test environment as a dry and clinical experience, it was decided to create it as a game-with-a-purpose, or a GWAP. A GWAP is a game that has been created to facilitate human computation in a more enjoyable way. One example of a GWAP is Phylo [49], a game in which players solve pattern-matching puzzles that represent nucleotide sequences of different phylogenetic taxa to optimize alignments over a computer algorithm. GWAPs have also be used to collect labelled data for experimental purposes, for example TagATune [50], an audio-based online game that aims to extract descriptions of sounds and music from human players.

Making our objective user tests as game has several advantages. Firstly, making it an enjoyable experience encourages users to play it more so that more data can be collected. Moreover, longer play times means that the users can improve their skill with the tasks, and analysing the rate of improvement can provide interesting information.

The game – called Super Hyper Circle Mover – is essentially a classic 2D Fitts' law test packaged up as a game, so the goal of the player is to move a circle to a given target zone, and then once positioned successfully, rotate it to a target angle. Since the user is given more points the more accurate and the faster the actions are performed, the user is encouraged to perform better. There are five levels, with each level the target zones for both the positioning and the rotation get smaller and smaller, thus the difficultly increases. The increasing difficultly can provide interesting data because users may perform better with certain control modes as the required accuracy increases.

During play, every gesture that the users makes is recorded including the accuracy, the time taken, the distance moved and the type of gesture (navigation or manipulation), into an XML file stored on the phone. The full path of the gesture is recorded at a sampling frequency of 20Hz so that anomalous gestures can be analysed more closely.

The game was implemented using Unity 3D and the Vuforia library as with the AR Reactable, and deployed on the Android platform.

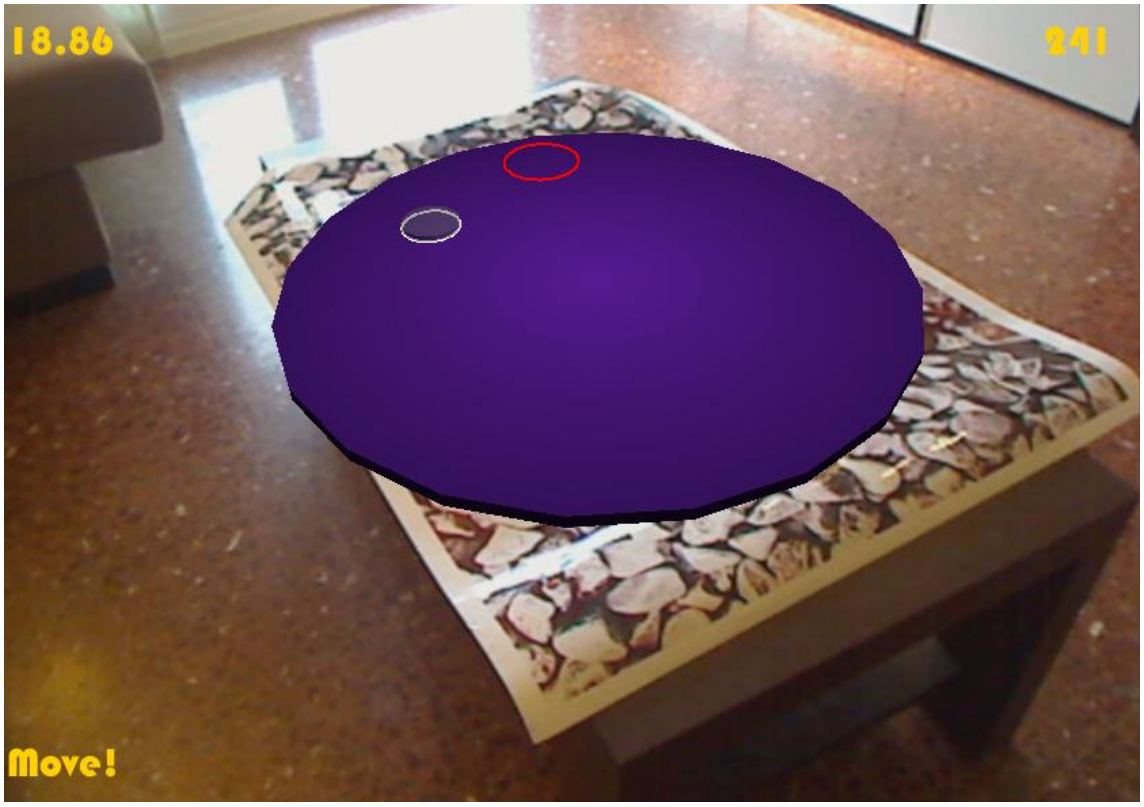

Figure 16 - Super Hyper Circle Mover move task

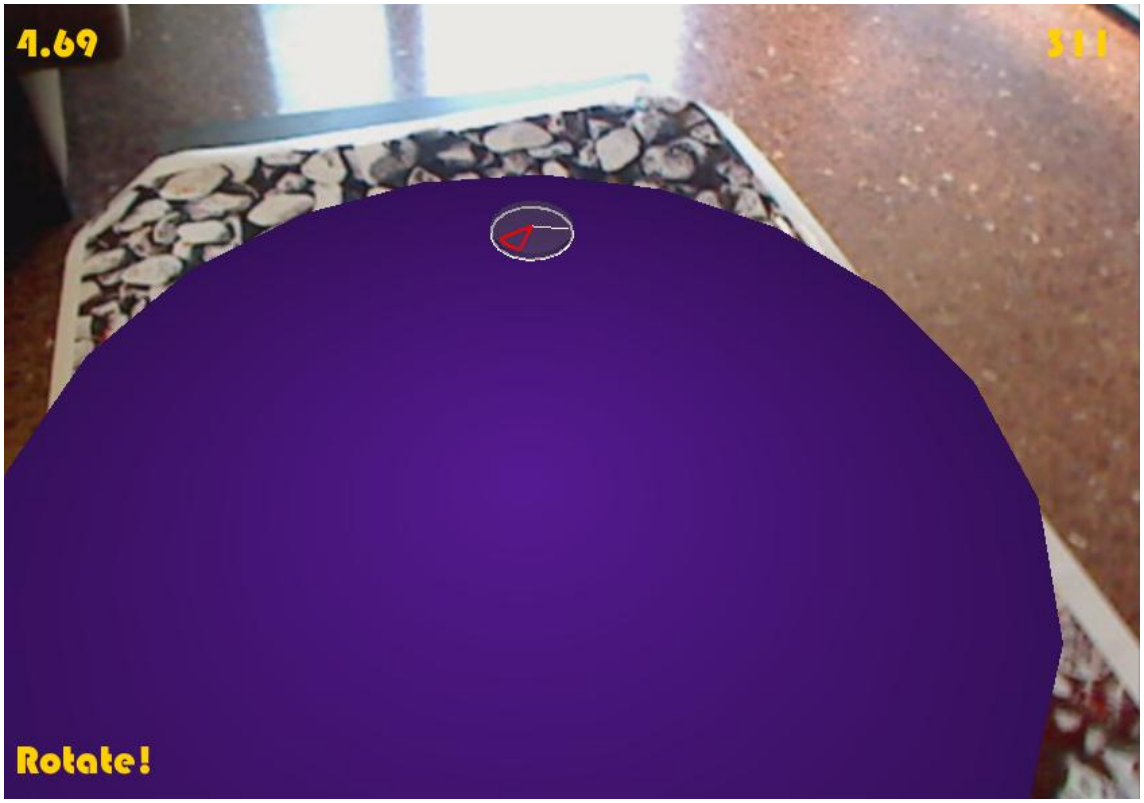

Figure 17 - Super Hyper Circle Mover rotate task

7.1.2 Evaluation

The game created to evaluate AR as a control mode objectively would be released for two weeks with a prize given to the highest scorers in each control mode at the end of that period. A separate prize for each mode would be awarded to discourage the users sticking to one mode. When the two week period was over, the users would be asked to fill out a short questionnaire and to return via email the XML file containing the recorded usage data retrieved from the SD card.

The questionnaire would contain questions regarding the perceived performance, the enjoyment, the intuitiveness and the levels of mental and physical fatigue of the users in the various control modes, information regarding the model of phone, and the prior experience the user had with each of the control modes.

The recorded usage data would be modelled with Fitts' Law, the values of a and b empirically determined with line-fitting techniques and then the index of performance ($1/b$) would be used to compare the three control modes.

7.2 Collaboration

One extremely interesting area which was not visited in the scope of this thesis is real-time collaboration. Studies have shown that AR interfaces have been shown to produce similar communication behaviours amongst the users to unmediated face-to-face collaboration, more so than screen-based collaboration [39]. Users of AR systems have been shown to exhibit more deictic speech patterns, pointing behaviours and other body-centric communications than with non-AR settings. Since collaborative music relies on shared communication, AR-based control modes could be extremely useful for this.

Since much work has been done studying the effects of AR on collaboration it wasn't chosen to be the main focus of the thesis, however it may be that the main benefits of AR may be in this area. As result of this, we implemented a real-time collaboration feature in the AR Reactable system described previously, however it was never tested and no experiments involving it were devised. In later work, this area would be examined in much more detail.

8. Bibliography

[1] K. Lee, “Augmented Reality in Education and Training,” *Techtrends: Linking Research & Practice To Improve Learning*, 2012.

[2] T. Technologies, “A TAXONOMY OF MIXED REALITY VISUAL DISPLAYS,” no. 12, pp. 1–15, 2003.

[3] R. T. Azuma, “A Survey of Augmented Reality,” vol. 4, no. August, pp. 355–385, 1997.

[4] Y. Vazquez-Alvarez, I. Oakley, and S. a. Brewster, “Auditory display design for exploration in mobile audio-augmented reality,” *Personal and Ubiquitous Computing*, vol. 16, no. 8, pp. 987–999, Sep. 2011.

[5] “Animoog Synth.” [Online]. Available: https://itunes.apple.com/us/app/animoog/id471638724?mt=8.

[6] “Lemur Controller.” [Online]. Available: https://itunes.apple.com/us/app/lemur/id481290621?mt=8.

[7] “Garageband App.” [Online]. Available: https://itunes.apple.com/en/app/garageband/id408709785?mt=8.

[8] O. Bimber, A. Grundhöfer, G. Wetzstein, and S. Knödel, “Consistent illumination within optical see-through augmented environments,” *Proceedings of the SIGGRAPH 2003 conference on Sketches & applications in conjunction with the 30th annual conference on Computer graphics and interactive techniques - GRAPH '03*, p. 1, 2003.

[9] S. Feiner, T. Terauchi, G. Rashid, and D. Hallaway, “Exploring MARS : developing indoor and outdoor user interfaces to a mobile augmented reality system Tobias Ho,” vol. 23, pp. 779–785, 1999.

[10] M. Möhring, C. Lessig, and O. Bimber, “Video See-Through AR on Consumer Cell-Phones,” pp. 3–4.

[11] Y. Gu, A. Lo, S. Member, and I. Niemegeers, “Wireless Personal Networks,” vol. 11, no. 1, pp. 13–32, 2009.

[12] H. B.-L. Duh and M. Billinghurst, “Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR,” *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp. 193–202, Sep. 2008.

[13] J. Newman, M. Wagner, M. Bauer, a. MacWilliams, T. Pintaric, D. Beyer, D. Pustka, F. Strasser, D. Schmalstieg, and G. Klinker, “Ubiquitous Tracking for

Augmented Reality," *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, no. Ismar, pp. 192–201, 2004.

[14] V. Radu, J. Li, L. Kriara, M. K. Marina, and R. Mortier, "Poster : A Hybrid Approach for Indoor Mobile Phone Localization," p. 2010, 2010.

[15] H. Wuest, F. Vial, and D. Stricker, "Adaptive Line Tracking with Multiple Hypotheses for Augmented Reality Department of Virtual and Augmented Reality," pp. 0–7, 2005.

[16] H. Kato and M. Billinghurst, "Marker tracking and HMD calibration for a video-based augmented reality conferencing system," *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)*, pp. 85–94.

[17] A. I. Comport, É. Marchand, F. Chaumette, and C. De Beaulieu, "A real-time tracker for markerless augmented reality," pp. 0–9, 2006.

[18] J. J. Leonard, I. Way, and H. F. Durrant-whyte, "Simultaneous Map Building and Localization for an Autonomous Mobile Robot," no. 91, 1991.

[19] N. Karlsson, E. di Bernardo, J. Ostrowski, L. Goncalves, P. Pirjanian, and M. E. Munich, "The vSLAM Algorithm for Robust Localization and Mapping," *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pp. 24–29, 2005.

[20] G. Klein and D. Murray, "Parallel Tracking and Mapping for Small AR Workspaces," *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 1–10, Nov. 2007.

[21] G. Klein and D. Murray, "Parallel Tracking and Mapping on a camera phone," *2009 8th IEEE International Symposium on Mixed and Augmented Reality*, pp. 83–86, Oct. 2009.

[22] M. Billinghurst, H. Kato, and S. Myojin, "Advanced Interaction Techniques for Augmented," pp. 13–22, 2009.

[23] H. Ishii and B. Ullmer, "Tangible Bits : Towards Seamless Interfaces between People , Bits and Atoms," 1997.

[24] H. Kato, M. Billinghurst, I. Poupyrev, K. Imamoto, and K. Tachibana, "Virtual object manipulation on a table-top AR environment," *Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)*, pp. 111–119.

[25] S. Gupta and C. Jaynes, "The universal media book: tracking and augmenting moving surfaces with projected information," *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp. 177–180, Oct. 2006.

[26] W. Hürst and C. Wezel, “Gesture-based interaction via finger tracking for mobile augmented reality,” *Multimedia Tools and Applications*, vol. 62, no. 1, pp. 233–258, Jan. 2012.

[27] G. A. Lee, U. Yang, Y. Kim, D. Jo, and K.-H. Kim, “Freeze-Set-Go Interaction Method for Handheld Mobile Augmented Reality Environments,” 2009.

[28] H. Bai, G. A. Lee, and M. Billinghurst, “Freeze view touch and finger gesture based interaction methods for handheld augmented reality interfaces,” in *Proceedings of the 27th Conference on Image and Vision Computing New Zealand*, 2012, pp. 126–131.

[29] G. A. Lee, H. Bai, and M. Billinghurst, “Automatic zooming interface for tangible augmented reality applications,” in *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry*, 2012, pp. 9–12.

[30] S. Kasahara, R. Niiyama, V. Heun, and H. Ishii, “exTouch,” 2012. [Online]. Available: http://tangible.media.mit.edu/project/extouch/.

[31] P. Mistry, “Sixth Sense.” [Online]. Available: http://www.pranavmistry.com/projects/sixthsense/.

[32] J. Choi, H. Park, J. P. Park, and Jong-Il, “Bare-hand-based augmented reality interface on mobile phone,” in *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium*, 2011, pp. 275–276.

[33] G. W. Fitzmaurice, “Situated information spaces and spatially aware palmtop computers,” vol. 36, no. 7, 1993.

[34] K. Yee, “Peephole Displays : Pen Interaction on Spatially Aware Handheld Computers,” 2003.

[35] X. “Anthony” Chen, N. Marquardt, A. Tang, S. Boring, and S. Greenberg, “Extending a mobile device’s interaction space through body-centric interaction,” *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services - MobileHCI ’12*, p. 151, 2012.

[36] M. Billinghurst, I. Poupyrev, H. Kato, and R. May, “Mixing realities in Shared Space: an augmented reality interface for collaborative computing,” *2000 IEEE International Conference on Multimedia and Expo. ICME2000. Proceedings. Latest Advances in the Fast Changing World of Multimedia (Cat. No.00TH8532)*, vol. 3, pp. 1641–1644.

[37] A. Fuhrmann and L. M. Encarnac, “The Studierstube Augmented,” pp. 33–54.

[38] M. Billinghurst and H. Kato, “Collaborative Augmented Reality,” 2001.

[39] M. Billinghurst, H. Kato, K. Kiyokawa, D. Belcher, and I. Poupyrev, “EXPERIMENTS WITH FACE-TO-FACE COLLABORATIVE AR INTERFACES .,” no. Inkpen 1997, pp. 1–21.

[40] a. Henrysson, M. Billinghurst, and M. Ollila, “Face to face collaborative AR on mobile phones,” *Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR’05)*, pp. 80–89, 2005.

[41] D. Lakatos and M. Blackshaw, “T(ether),” 2012. [Online]. Available: http://kiwi.media.mit.edu/tether/.

[42] “Reactable.” [Online]. Available: http://www.reactable.com/.

[43] “Unity 3D.” [Online]. Available: http://unity3d.com/.

[44] “Qualcomm Vuforia.” [Online]. Available: http://www.qualcomm.com/solutions/augmented-reality.

[45] “FAST Corner Detector Website.” [Online]. Available: http://www.edwardrosten.com/work/fast.html.

[46] “ARToolkit.” [Online]. Available: http://www.hitl.washington.edu/artoolkit/.

[47] T. Drummond, “High Speed Matching and Tracking.” [Online]. Available: http://www.qualcomm.com/media/documents/high-speed-matching-and-tracking-slides.

[48] P. M. Fitts, “The information capacity of the human motor system in controlling the amplitude of movement,” *Journal of Experimental Psychology*, vol. 47, no. 6, pp. 381–391, 1954.

[49] “Phylo.” [Online]. Available: http://phylo.cs.mcgill.ca/.

[50] E. L. M. Law, L. Von Ahn, R. B. Dannenberg, and M. Crawford, “TAGATUNE : A GAME FOR MUSIC AND SOUND ANNOTATION.”