Responsive audio feedback for hand gesture interaction

to enhance immersion in audio-only games

A Thesis

Submitted to the Faculty

of

Drexel University

by

Wenjie Wu

in partial fulfillment of the

requirements for the degree

of

Master of Science in Digital Media

May 2015

Dedications

I dedicate this project to my dearest dad and mum for their unconditional support and care.

My Dad often took me outdoors to embrace the beauty of nature. We went climbing for sunrise, we went fishing for sunset. My childhood was full of outdoor adventures until I met the piano. I like being immersed in the imaginative world of music. However, constant practice always kept me indoors. And I gradually cultivated my habit of sensing the outside world through sounds.

Though sitting indoors, I always wish that music can lead me out of the window. Through imagination, sometimes I can row the boat in Tchaikovsky's June, and sometimes I can appreciate the Luna under Debussy's sky.

"There was no obvious evidence to show how much you wanted to be in that different world until one night you sleepwalked to the balcony, hummed that song and grabbed the clothing fork as your paddle to start rowing in the air," my dad remembered.

"That awkward scene did shock us in the middle of the night, I dare say. But once you told me you want to row a boat again in your game, I knew that you can definitely do it. Since you have already been there," my mum said.

Thanks for believing in my awkwardness and for your inspiring impact on my life.

And now I hope you would like to row my boat.

Acknowledgments

I would like to thank my committee for their support: Dr. Stefan Rank, Professor Jervis Thompson, Professor John Avarese and Austin Seraphin.

Thank you to Dr. Stefan Rank for constantly encouraging me to achieve what I wanted to do but seemed almost impossible for me to achieve at the very beginning. Without his faith and visionary advice, I may have never realized that I can reach this far in this field in this two-year limited time period.

Thank you to Professor Jervis Thompson for introducing me to the exploration of new technology. Without his timely recommendation and suggestive feedback, I may not have the courage to keep on practicing my ideas on new devices.

Thank you to Professor John Avarese for helping me in developing the magical soundscape for the game. Without his professional suggestions and patience, my game would never be as meaningful or unforgettable.

Thank you to Austin Seraphin for his great passion to keep improving the project. Without his unique suggestions based on his life-long audio-only experience, the immersive experience of this project would never be so realistic.

I would like to thank all Drexel Faculty who have helped and guided me in one way or another: Dr. Glen Muschio, Professor Jeremy Fernsler, Justin Patterson, Professor Robert Lloyd, Professor John Berton, Dr. Paul Diefenbach, Dr. Michael Wagner and Dr. Jichen Zhu.

I also want to thank my fellow graduate students. I am truly thankful to have been able to cross paths with all of you. It was really great to be part of such a vibrant and creative community: Ariel Evans, Ben Goldberg, Caroline Guevara, Cathy Lu, Natalie Lyon, Jun Ma, Shannon Sepelak, Ning Shao, and Junyu Zhu.

And finally, special thanks to my dearest 16-year-old cousin, Annie, for contributing her adorable voice to my game. Without her great help, I may have never been able to achieve my dream goal in this project.

Table of Contents

List of Tables

List of Figures

Abstract

Responsive audio feedback for hand gesture interaction
to enhance immersion in audio-only games
Wenjie Wu
Stefan Rank, Ph.D., Jervis Thompson, John Avarese, and Austin Seraphin

This thesis investigates different audio feedback designs for touch-less hand gesture interaction to enhance immersion and encourage intuitive user participation in audio-only environments. In particular, we aim to assess whether replacing explicit audio instructions for hand positions and movements with responsive audio feedback using environmental story-related audio cues that are suggestive of interaction methods leads to measurably higher immersion. The potential of using audio cues as feedback to improve the responsiveness of touch-less gesture interaction has been an active topic of study. We focus on methods for effectively informing players about gesture interaction methods through audio cues without breaking game immersion. As part of our research, we developed a framework for hand gesture detection based on the Leap Motion hardware and software. We also present a design approach for audio feedback that considers three stages, before, during and after a gesture sequence, and proposes corresponding design principles. In order to test different audio designs and support our hypothesis, two game versions using non-diegetic or diegetic audio feedback were created as testbeds for a between-subjects study. The study found that the version using environmental feedback led to higher immersion in terms of story and was consistently rated higher. In particular, compared to explicit vocal instructions, environmental story-related audio cues were found to be more helpful in guiding gesture interaction and in enhancing how real the game feels.

Chapter 1: INTRODUCTION

Although immersion in games has been studied for many years, our understanding of immersion and of the factors that influence the degree of immersion can still be refined. At the same time, as motion tracking technologies developed and have been applied towards increasing immersion in games through natural motion control, audio-only game designers can now consider studying gesture interaction in audio-only environments to enhance immersion. Apart from controlling game navigation, gesture interaction has by now also been used as main mechanics for some games, in particular after the release of the Kinect and the Leap Motion. In this context, the following question arises: How should audio feedback be used to react to natural gesture interaction? This question also evokes new ideas for interaction in audio-only game design.

Due to the lack of visual cues, it is hard for designers to structure the interaction so as to suggest when and how to interact in a gesture-based audio-only environment. How to convey gesture interaction messages to the player through auditive cues without interrupting the immersive experience in audio-only environments has become an interesting problem that should be tackled. In addition, the potential inaccuracies involved in gesture control also lead to a higher risk of negatively affecting immersion.

The study reported on in this thesis was designed to work towards a solution to the above problem. Our approach is to test out useful designs of audio feedback that is responsive to hand gestures to both encourage user participation and maintain immersion in gesture-controlled audio-only environments. Our research hypothesis is replacing explicit audio instructions about hand positions and movements with responsive but less intrusive audio feedback, by suggesting interaction methods using a combination of environmental story related audio cues, leads to higher immersion.

In order to test this hypothesis, we created a game scenario for players to do hand gestures according to an audio story. Two versions of the game were built to implement different audio cues designs for testing effects on immersion in a two-group between subject study. Results show that, compared to explicit vocal instructions for hand gestures, environmental story-related audio feedback responsive to hand gestures are more helpful in suggesting interaction methods and enhancing immersion in the way of better achieving a realistic game context.

The following two chapters cover relevant theoretical aspects of gesture interaction, audio-only environments and the notion of immersion (chapter 2), followed by related work that captures the state of the art in the field (chapter 3). The rest of the thesis lays out the addressed problem in more detail (chapter 4) and then shows the approach to design, development, and experimental study (chapter 5). Finally, the results of analysing the data are presented (chapter 6), followed by concluding remarks about the significance and implication of the work (chapter 7).

Chapter 2: ON GESTURES, AUDIO-ONLY AND IMMERSION

This chapter covers several areas relevant to the study in order to frame a solid background to introduce the targeted problem and research question. We will discuss the following topics: Gesture Interaction, Audio-only Environments and Immersion.

2.1 Gesture Interaction

Definition of Gesture Interaction Motion-based interaction has been described by many terms. Norman [2010] sums everything under the plain term: gestural interaction. Hartson and Pardha [2012] define one of the terms, embodied interaction, which means a possibility to use a physical body for interaction with technology. O'Hara et al. [2013] speak about touchless interaction referring to a controller-free way to interact with applications. Gestures are a good example of basic intrinsic needs to manipulate objects, either digital or physical, with hands [Hartson and Pardha 2012].

Types of Gesture O'Hara et al. [2013] lists gestures for different purposes: gestures for pointing, manipulative gestures for controlling an object, and different gestures for communication or language use. The first two categories are commonly used in current applications and games, possibly because they are easier to implement due to their similarities to traditional point and click user interfaces.

Natural Gesture Interaction Interface As technology develops, more "natural" gestural interfaces become possibly achievable. Designers keep looking for new ways of implementing motion or gesture control to enhance control intuitiveness. Cranor et al. [2010] used a self-built detection hardware, a pair of gloves with several microcontrollers sewed in the locations of wrist and finger joints, to access data of the two hands. Though the gesture recognition system needed to work under a wire-connected condition, the glove installation still helped in capturing natural gesture input. Hantrakul and Kaczmarek [2014] used Leap Motion, a hand skeletal tracking system based on monochromatic IR cameras, to help in building intuitive gestural interface for music composition and live music performance. Results showed that natural gesture control did provide musicians' with higher working efficiency and convenience.

Natural Gesture-Based Game Play The application of sensor devices has also helped to enhance game control intuitiveness, which suggests players' expectation on natural gesture-based game play. As designers, we should address the player beyond mechanical presentation of the movement. Body movements impact the player cognitively, affectively and physiologically [Bianchi-Berthouze 2013]. So-called "universal" gestures are simple enough to remember, which is a good choice for achieving natural gesture-based game play. The context of gestures is also an important factor, as making a move or grabbing something needs to feel logical in that environment and response immediately. Further, gesture control often lacks in accuracy, making the use of effective feedback important to maintain positive participation, for example sound effects that are closely responsive to the speed or angle of the movement. Furthermore, in order to help players get engaged quickly and effortlessly, it is advantageous to increase the intelligence of the feedback system. The price to pay for too demanding user interface is losing the user's interest, which leads to a loss in natural game play.

2.2 Audio-Only Environments

Computer games often put their main focus on graphical contents. Sound and music are sometimes only used to keep up the atmosphere and strengthen the graphics. However, sound is a very capable element of game design. People have an enormous library of associations and memories related to sounds. This is so strong that a simple rectangle on the screen can suddenly acquire weight, texture and resistance based on sound cues. Papworth [2010] indicated that audio provides information that triggers viewers' emotion and degrees of perceived realism that pure visuals cannot. For instance, we all determine values such as surface, friction, weight and impact from the audio elements in films rather than the visuals we are presented with. Audio's ability to manipulate and affect the emotional state of the user presents us with some new and unique opportunities in both the gameplay design and its execution in the appliance [Papworth 2010].

Audio-only games are a special case in that they rely purely on audio output to present the player with a game world. They help players to create an inner picture of the environment. According to [Liljedahl et al. 2007], user investment plays an important role: by making the users invest both time and imagination in the game experience, they are given space and opportunity to share in the creative process. Potentially this creates a much more compelling game experience for the player, as personal imagination is far more dynamic and compelling than anything that a game designer can create. They showed that audio-focused games can convey a richer and more satisfying experience when the player's own imagination is given free range. For example, in audio-only environments, objects and drama in the game world become more abstract for players to perceive. As there is no dominant visuals to convey message, players are provided with more space to fill the image of the story with their own investment.

2.3 Immersion

Definition of Immersion Though studies and discussions on immersion have increased, there is no commonly accepted definition of the term among researchers. Janet Murray gave her definition of immersion in [Murray 1997]. She referred to the experience of being transported to an elaborately simulated place that takes over our whole perceptual apparatus as immersion. Virtual reality research has an interesting perspective on immersive experience. The concept of presence arises and is defined as the "extent to which a person's cognitive and perceptual systems are tricked into believing they are somewhere other than their physical location" [Patrick et al. 2000].

Types of immersion in games Immersion taking place in a virtual game world can be divided into two main types: Environmental immersion and social immersion. Environmental immersion concerns technical aspects of the game including aesthetic detail, sensory stimulation, and narratives. Social immersion evolves as an opposition of cooperation and intimidation. The immersion included in my study mainly relates to environmental immersion.

Factors relating to Immersion Radford [2000] relates Immersion to the ability to enter the game through its controls. In [Winograd and Flores 1986], the authors discussed the invisibility of the tool. They claimed that essentially there needs to be invisibility of control for total immersion to take place. However, it is not clear enough to show to what extent this invisibility is an essential part of immersion or merely a contributing factor. In Murray [1997], Murray indicated that achieving immersion involves a combination of explicit time and space. In participatory media such as video games, this combination would be heavily affected by participatory activities. According to [Brown and Cairns 2004], immersion is usually used to describe the level of involvement with a game. This involvement moves along time and is controlled by barriers that are generated by either elements of game construction or gamer investment.

Levels of Immersion In [Brown and Cairns 2004], there are three game involvements: engagement, engrossment and total immersion. Each level is only possible if the barriers to the level are removed. According to the study's interview, gamers' feelings and behaviors would change when reaching different levels. In the engagement level, gamers would get interested in the game and want to keep playing. When entering to the engrossment level, gamers would get less aware of the surroundings. They would feel emotionally drained whey being stopped from playing. Some of them even purposefully constructed a distraction free environment for gaming. And once arriving in the total immersion level, gamers would feel being cut off from the reality, nothing except the game can impact their thoughts and feelings.

Immersion and Participation One of the ways of sustaining immersion in traditional media is to prohibit audience participation. Former Scholar Langer attributes her stress of inviting audience participation into theatre performance. She believes that art is dependent on establishing distance, the improvisation caused by inviting participation would violate the art form, which may increase the risk of disrupting immersion. However, in [Murray 1997], Murray argued that immersion can be empowered by game interactivity. She indicated that digital media are providing us with a new stage for creation of participatory activities. She demonstrated how to preserve immersion in participatory narrative by discussing several important factors that should be included in the formation of boundary conventions: finding the borders, structuring participation as a visit or with a mask, the active creation of belief, structuring collective participation with roles and regulating participant's arousal. She believed that if we properly define boundary conventions in participatory narrative and follow them to structure participation, we can sustain and even deepen the enchantment of immersion in the digital space.

Chapter 3: RELATED WORK

This chapter provides an overview of the state of the art regarding sound design and approaches to control in audio-only games, as well as gesture detection, and pointing out how the previous approaches relate to the one presented in this thesis.

3.1 Sound Design Conventions in Audio-Only Games

Following sound design conventions is the most common way that game designers achieve immersion in audio-only environment. For example, the method of designing immersive sound for iSpooks, which is a mystery audio game for iPhone, was discussed in [Papworth 2010]. In [Liljedahl et al. 2007], researchers present their sound designing conventions for Beowulf, another adventure audio game. In order to achieve a more immersive game experience, the individual sounds that build the total Beowulf game soundscape were chosen for their imagistic value, the potential to trigger associations in the imagination of the player.

In[Roden et al. 2007], researchers showed their approach of building the structure of an audio only virtual world. They analyzed the types of sounds and the functionalities of each type. For example, there should be some sounds that convey information and knowledge, ambient sounds played in loops that trigger emotion and responsive sounds that are reactive to the players' behaviors. This inspired me that as a design of audio only game, it should present players with an imaginable game world with various pieces of information that are provided by rich sound sources.

3.2 Control in Audio-Only Games

Keyboard, hand-held controllers or touch input on touch screens are common game controls in many existing games, and there is no surprise when designers apply those in controlling audio-only game worlds. However, as technology develops, players' expectation on natural control grows. Therefore, new approaches of fulfilling this growing desire appeared. Jumper is an iPhone audio-only game that uses voice input as game control. In the game, players will play as one LAPD officer and will be asked to prevent a desperate man from committing suicide by voice persuasion [1].

Apart from voice recognition technology, there were also technologies having been applied to track and recognize body motion input in audio games. In some motion-controlled game works, players can only use simple motion to navigate the game world. In other works, motion input is applied as the main game mechanics. For example, the pressure control in games that focus on dance experience, players can dance by stepping on a pressure-measured board. In [Baldan et al. 2013], designers built a tennis game based on cellphone accelerometer. As the system provides 3D tennis sound feedback which depend on the detected phone accelerations, players can use the phone as tennis racket to physically play air-tennis with other players in realistic space. Most of these game play variations were achieved by players' touching or holding the motion tracking devices. However, hand-held devices or touch-based tracking may be in some way limiting the possibilities of addressing more natural gestures. Considering the views on invisibility of tool in [Winograd and Flores 1986], the unnatural touch of devices may always remind players of the boundary between game world and reality, which may cause the problem of keeping players stuck in the lower level of immersion.

In recent years, more and more motion tracking technologies that are capable of capturing natural motion input, such as kinetic and leap motion, have been applied in game design. This fact indicates game designers' intention of enhancing game immersion by approaching enhanced naturalness of user input. And there should be no exception for the audio-only game future development. Manuel et al. [2012] showed an approach of enhancing immersion in audio-only environments by replacing the hand-held control with natural gesture control captured by Kinect. The results of the study showed that soundfield interaction through motion tracking is capable of producing a more enjoyable and immersive experience when compared to a traditional handheld analogue controller.

As part of my work, I investigate if a higher level of immersion can be achieved by improving the sound responsiveness to motion.

[1] See https://www.youtube.com/watch?v=TQ9F1v8B6MI for a gameplay video.

3.3 Detecting Gestures in Interactive Projects

In [Clifton et al. 2013], ”Don’t open that door”, a gesture-based interactive narrative project, was shown, see Fig. 3.1. The research mainly focuses on exploring the methods for instructing the interactor when and how to interact in interactive narratives by using both visual and audio cues. Four ”scripting” techniques were discussed in the study, which are the following:

- Verbal - A character verbally directs the interactor to perform an action.
- Audio-visual - The setting and soundscape indicate that some action can be taken.
- Reactive - A sudden dramatic change triggers an action of the interactor without thinking about it.
- Mimetic - The interactor performs an action after first seeing it performed by a character.

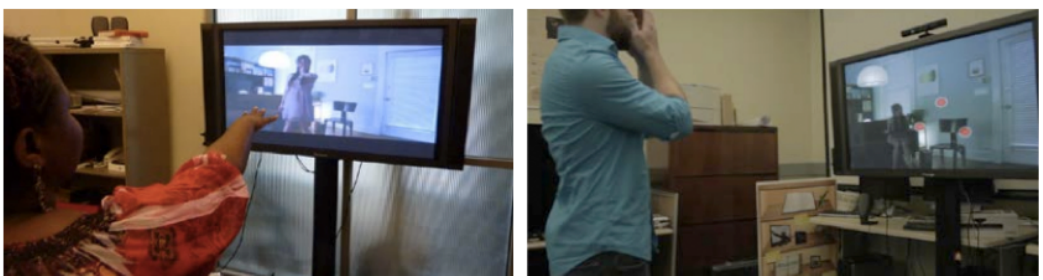

Figure 3.1: Gameplay from [Clifton et al. 2013]. Both Take hand interaction and Peek-a-boo interaction use mimetic scripting method to inform the interactor of the gestures.

Figure 3.2: Screenshots of Kinetic Story, showing the visual motion sequence at the top of the screen that conveys gesture instructions to the player.

Kinetic Story, see Fig. 3.2, an award winning interactive storybook series for 3- to 6-year-olds was presented in [Huhtanen 2014]. Children are put in the shoes of the main character, performing the character’s gesture to interact with the story. Instructions of gestures are shown by visual motion sequence located on the top of the screen. Gestures are captured by Kinect, and story branches are unfolded according to the child's gestures. Though players can not control the main story line of the whole game, this design still enables them to participate in parts of story telling and explore self-driven game play that affected by their own movements.

These gesture-based interactive projects mainly used visual cues to show the gesture instructions to interactors, while due to the audio-only feature of my project, I have to present the gestures only using audio cues.

As part of my work, I investigate how to effecively instruct the interactor on when and how to interact in gesture-based audio-only environments. I also consider whether the techniques presented in above studies can be applied in my research.

Chapter 4: PROBLEM STATEMENT

This chapter covers the research questions of the study, in part derived from the problems raised by related work. We distinguish the main research question from subsidiary questions as well as our design objectives for development and the hypothesis that guides the design of the experimental study.

4.1 Research Question

The following research questions were used as the key guidance that led me through the process of finding design approaches, preforming experiments and setting evaluation criteria. They were answered by the study results from the experiments, which are covered in detail in the later chapters of the thesis, in order to find out a useful solution to the problem of breaking immersion in audio-only environment using intrusive audio feedback for hand gesture interaction.

Main Question What kind of design for an interactive audio experience responsive to hand gestures can be used to both encourage user participation and maintain immersion in a gesture-controlled audio-only environment?

In order to find an immersive method to convey hand gesture interaction message to the players in audio-only environments, I evaluated different approaches for designing the responsive audio feedback. In the study, two responsive audio feedback designs were separately implemented in two game versions. Both game versions were based on the same audio story told by a main character, and they were played by two different subjects groups in order to test which design can lead players to a higher immersion level. Here are the descriptions of two different designs on audio cues:

1. Explicit vocal instructions (non-diegetic), responsive to hand gestures, related by a character in the story.

2. Environmental audio cues, responsive to hand gestures caused by objects collisions and scenarios in the game environment and story (diegetic).

Subsidiary Questions Here are some subsidiary questions that I considered as well in order to better answer different aspects related to the main research question.

1. According to the immersion levels defined in [Brown and Cairns 2004] and immersion measures in both [Jennett et al. 2008] and [Brockmyer et al. 2009], what is the most significant difference on the immersion degree achieved by each of the two designs mentioned above?
2. Can the study results provide new knowledge about enhancing game immersion in audio-only games through gesture interaction for both the audience and game designers?
3. Will the designs tested in this study be useful as reference for further experimenting immersion creation in motion-controlled audio-only environments?
4. Can the audio cues design that leads to higher immersion be applied to enhance immersion in gesture-based virtual reality?
5. Can the game version that leads to higher immersion be applied in enlarging the blind community's access to motion-controlled audio games?

4.2 Design Objectives

The objective of our project was to create a responsive audio feedback system for hand gesture interaction in audio-only games. We adopted the following quality criteria for our design:

- Helpful in guiding hand gesture interaction
- Helpful in enhancing immersion in audio-only environment
- Helpful in encouraging intuitive participation

Details on the approach to designing this project are presented in chapter 5 of this thesis, which includes both game construction and implementation. In particular, details on design principles for responsive audio feedback and the technique implementation for the hand gesture detection framework are covered by the implementation section 5.1.2.

4.3 Thesis Statement

In this section, two complementary hypotheses on audio cue design are presented that guide the implementation of versions of our game. The hypotheses capture the problem we are addressing and lead to the claim of our thesis.

Hypotheses

1. Explicit vocal instructions for hand gesture interaction will weaken immersion in audio-only games.
2. Environmental audio feedback responsive to hand gestures and reasonable for the story setting can both help in conveying hand gesture interaction messages and in enhancing immersion in audio-only games.

Thesis claim

> Compared to explicit audio instructions on hand positions and movements, a combination of responsive audio feedback with narrative and environmental audio leads to higher immersion in gesture-controlled audio-only environments.

In the rest of the thesis, the design approach for a two-group between subject study is presented in chapter 5 explaining the structure of our comparative experiment on two audio cue designs. The results of the experimental evaluation relating to both the hypotheses and the thesis claim will be discussed in chapter 6, showing support for our claim, as well as in the concluding chapter 7, section 7.1.

Chapter 5: APPROACH

This project focuses on the interaction design between hand/finger-based gestures and audio, in order to create an immersive audio experience responsive to players' hand gestures. Two game versions are used to compare two approaches for audio feedback. Several iterations of game concepts preceded these final versions, and Appendix A details relevant aspects of that development.

In this chapter, we present the design and development of the two game versions including the framework we developed for gesture detection and the design approach for audio feedback. Further, we will cover the approach to the experimental evaluation of the game versions, including the type of data that was collected in questionnaires and by video taping and automatic in-game metrics.

5.1 Project Design and Development

This section will focus on unfolding the methodology on building the whole project. The content is divided into two main parts: Game Construction and Technique Implementation.

5.1.1 Game Construction

Game Story As a concrete example of the feedback design and as a testbed for different approaches, we use the following game scenario that prompts players to perform certain hand gestures to move the story forward:

A little boy called Dean wakes up on his boat. To his surprise, he finds out that his magic watch starts talking to him. He follows the watch's suggestions and embarks on a journey of time travel using his hands. He can throw stones, pinch petals, catch birds, row a boat, shake a tree, tie a rope, ride on a bike and more. By performing gestures differently, he can go faster or slower than his normal pace in time. At the of the game, little Dean eventually travels back in time, to a time before he was born.

In order to maintain a fluent game experience, we intended to implement an interactive way of storytelling into the game. The narrator would tell the audio story, and she would give responsive feedback to the player's gestures as well. The story-related audio cues were written in the form of talking dialogue, and the narrator's voice-acting was performed as a child's voice.

In terms of the story context, the main character is situated in natural environments, and interacting with nature. For example, players can get interactive reactive sound feedback from the trees or animals around them when they perform the hand gestures. This approach was designed to create an organic game context for players to perform natural gesture in a logical and immersive experience.

The story structure was refined by adopting the Zelig Interaction Metaphor from [Llobera et al. 2013]. According to this metaphor, participants would be treated as the protagonist and easily get immersed with the role and the situation in the virtual game environment. The development of the story basically depends on the behavior of the participant, which was thought appropriate for the gesture-driven feature of our game story. In our game, new story elements would not be unfolded until the player performs the gestures correctly. Llobera also pointed out that as designers we should be cautious about player's passive attitude. And we considered using the following approaches to avoid it:

1. Arrange for an appropriate stopping time period between gestures.

   Here stopping time means, in this period of time, players can enjoy the audio story with no gestures involved.

2. Refine the storyline depending on the difficulty curves of gestures

   Consider some modification for the order of gestures, making sure to keep the game fun by letting players experience simple one hand gestures and complicated two hand gestures in turn.

3. Create a separate branch of story hidden in the game which would include some new rewarding gestures to play with.

   Players who finish one main gesture sequence without making a mistake can unfold parts of the secret story branch, and when they finish each new part, they will return to the main story line. Such a hidden story branch may be applied as a reward system for keeping the participants active in the whole game play.

Gesture Design With the intention of creating an immersive experience compelling enough for testing the audio cues design within a short play test section, we tried looking for universal hand gestures that can easily trigger people's imagination and resonate with their real life experience. In addition, considering Leap Motion's limited detection range, we tweaked the story line and experimented with different combinations of gesture sequences to find a balance between a logical unfold of the story and a smooth learning curve for gesture interaction. Fig. 5.1 and Fig. 5.2 are illustrations of the final gesture interaction sequences that players can experience in our game:

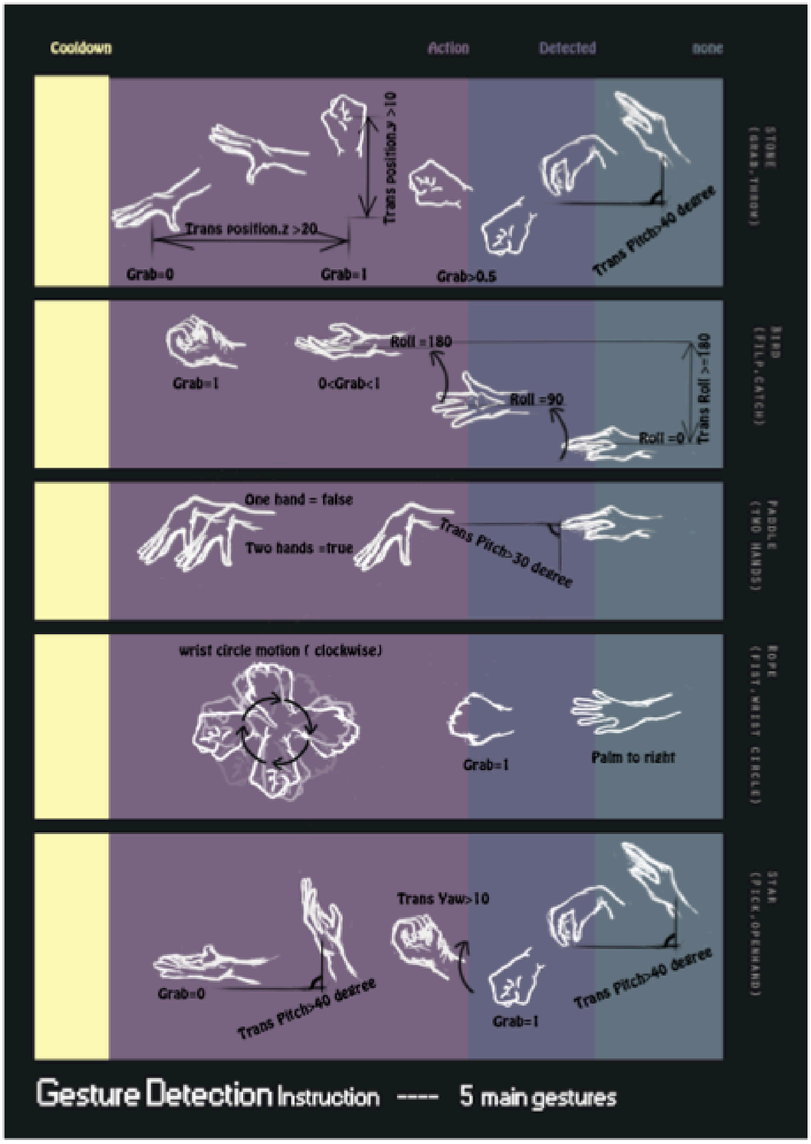

Figure 5.1: Five main gesture interaction sequences in the game: Stone Throwing, Bird Catching, Paddle Rowing, Rope Binding and Star Picking.

Soundscape Creation In [Roden et al. 2007], the authors analyzed the layers of music and the functionalities of different kinds of auditory elements. For example, there should be narrative speech played

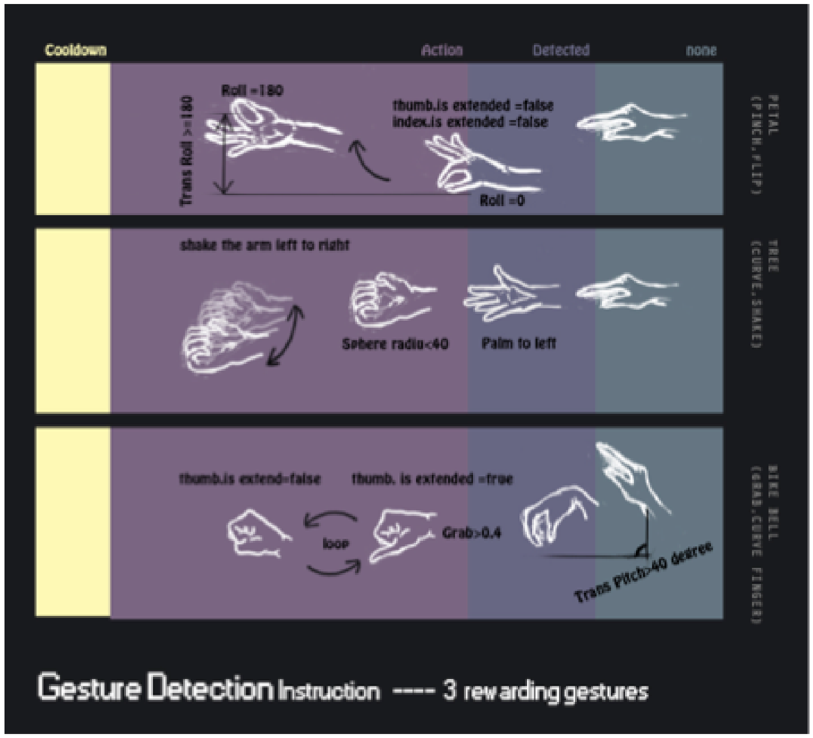

Figure 5.2: Three rewarding gesture interaction sequences in the game: Petal Pinching, Tree Shaking and Bike Bell Ringing.

reactively to convey information and knowledge, background music plays in loop to trigger emotion and sound effects that depict process or objects. We applied Roden's method to structure the soundscape of our project, which in order to present players an imaginable game world with various information that provided by rich sound sources. In addition to hearing the object-related sounds that they are interacting with, in our game, players can also hear ambience which informs them of the surrounding game environment. They can even hear some accompanying motions the character is doing besides those related to hands. For instance, when the players get to land from the boat, although they can only control characters' hand motions, they can still hear the character jumping and walking on the grass according to the story. We set special triggers for the sounds of these accompany motions, and have them triggered by players' activities at some special moments of the story to achieve deeper engagement.

Game Versions As preparation for an experimental study, designs for responsive audio feedback were implemented in two game versions. Both game versions were based on the above audio story, as narrated by the magic watch. See Fig. 5.3 for an illustration showing the different implementations of audio feedback responsive to the stone-throwing gesture in the two game versions.

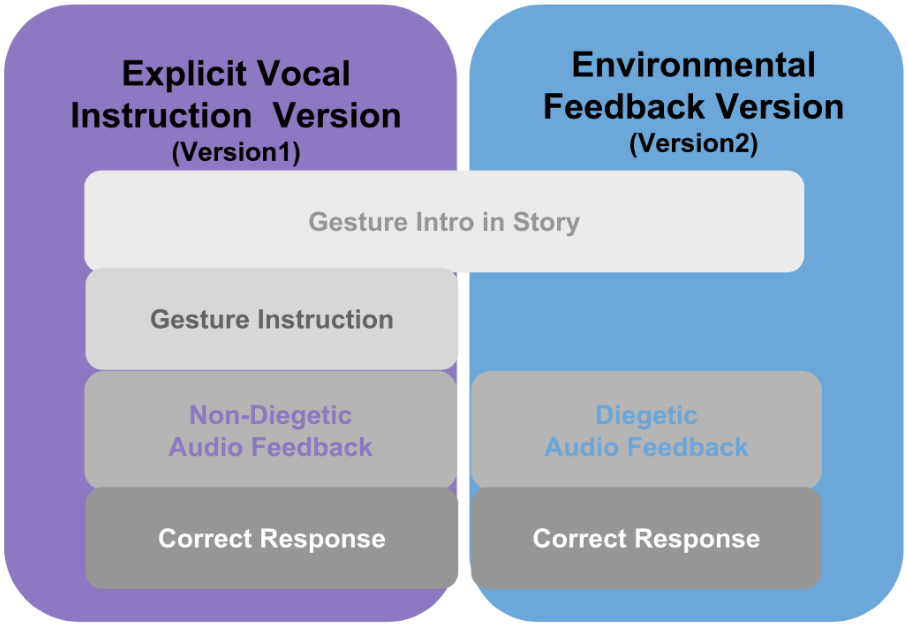

Figure 5.3: Explicit Vocal Instruction Version vs. Environmental Feedback Version

1. Explicit Vocal Instruction Version

   When playing this version, apart from the audio story content, the player can hear explicit vocal instructions that are relevant to hand positions and movements from the magic watch.

2. Environmental Feedback Version

   Various audio cues are used that are responsive to hand gestures caused by objects collisions and scenarios in the game environment and story.

Story-related audio cues and gesture instruction The following tables present all the vocal scripts in the game, which include both the diegetic and non-diegetic audio cues in both game versions: the game story, gesture instructions, story-related cues during gestures told by the narrator and slightly different correct responses for finishing the gesture in both versions. The tables cover the gestures related to stone throwing, Table 5.1, flower picking, Table 5.2, bird catching, Table 5.3, paddle rowing, Table 5.4, rope binding, Table 5.6, bike riding, Table 5.7, and star catching, Table 5.8.

Table 5.1: Audio cues and gesture instructions: Stone Throwing Gesture

Stone

Ver.1	Ver.2
(kid snoring sound) *(frog cry fade in)* *(Once the player put hands on the camera, kid snoring sound stops)* ***Watch: (whispering)*** **Wake up Dean, wake up!** **See buddy, the moon has come out…You slept over in the boat.** *(Fish water splash rock the boat)* ***Watch:*** **Can you hear that? A big fish is rocking our boat.** *(boat rock by the fish again)* ***Watch :*** **Let's throw stones to scare it away.**	
Stretch your finger in a 45 degree angle and move your hand downward to grab the stone in your fist. Raise your arm higher to throw the stone forcefully towards the front	
Instead of having your hands to the side, direct it so it is in front of you. Continue the motion of grabbing the stones like before.	*No, dont play around the water. There are stones on the head of the boat.*
In order to have a strong throw, you need to angle your arm in a 90 degree angle, so that it is close to your body, but below your head.	*Wow, the leaves were just hitten down by your stone. You threw it too far away. Can you hear it? The fish is close to the boat.*
Remember to keep your wrist aligned with your arm by preventing it from rotating to the sides, Also, lift your arm above your head before throwing.	*Stone skipping is fun, but we need a stronger hit to scare the fish. I think probably lifting your arm higher would help!*
Good job on raising your arm above your head, it will help exert more strength into your throw towards the front of the boat.	*Good! I can hear your arm lifting. Now try to throw it forcefully towards the head of the boat.*
Um... I think you're forgetting to elevate your arms before throwing..	*Heh, seems like the fish didn't get scared. Try with more strength.*
Oh no, you miss the fish, try again!	Yea, you almost hit it! Try this one more time!
It is almost there, one more time!	So smart! Keep doing this!
Yea, you hit it! Well done Dean!	Well done! the fish just jumped away.

Table 5.2: Audio cues and gesture instructions: Flower Petal Pinching Gesture

Flower	
Ver.1	**Ver.2**
(fish jump away , rocking fade out) ***Watch :*** **Ah~ there are ducks playing with lovely yellow petals in the water, let's pinch some, they look nice!** *(duck sounds starts)*	
Pinch the petal between your thumb and index fingers only. Examine it by flipping your palms over. Do this 4 times.	
Okay, since you are holding the petal. Inspect it more closely by flipping your hands so your palm is facing upwards.	*Eh? I see something blinking on the other side of it. Why not flip it over and take a look!*
Humm~~~~~~ they smell so good! Don't let the ducks eat them all, pinch some for Cinna, you know he is the most adorable dragon that always waiting you home.	Wow, it glows once you turn it back! And it smells so good! I guess we should pinch more before the ducks actually finish them.
...	Nice! Never see scent can sparkle like this! Pinch two more!
Well done Dean!	Humm~I just can't stop smelling them.

Table 5.3: Audio cues and gesture instructions: Bird Catching Gesture

Bird	
Ver.1	**Ver.2**
(duck sounds ends) ***Watch:*** **Oh man, there are birds pecking our boat!** *(pecking sounds starts)* ***Watch:*** **Wait, you don't need to drive these fellows away~just catch two for Cinna, you know he would be lonely when we are out.**	
Stretch both of your arms straight in front of you with your palm facing down.Flip your hands over so your fingers and palms are curled together cupping the bird.	
Great, you are doing an awesome job! Just remember to use two hands while cupping this chubby bird.	*Oh, he is walking away! Seems like you can't grab this chubby guy with one hand. Try to wrap it with both hands.*
Oh no, you are using too much force. Try to be more gentle by flipping your hands over to cup it.	*No no no, you have to be more friendly. Let them sit on your palms first so that you can wrap it gently.*
Woah there! Be gentle with that chubby thing. You don't want to squeeze it that tightly. Release it and try to gently cup it again.	*You're scaring the bird! Don't hold it that tightly*
Be careful, the bird cannot breath! Relax your palms to give it air.	*The bird can't breathe, try to hold it more gently*
Haha, you did it! Catch one more! Good job Dean!	Hehe, you did it, catch it one more time. Good job Dean!

Table 5.4: Audio cues and gesture instructions: Paddle Rowing Gesture

Paddle	
Ver.1	**Ver.2**
(pecking sounds ends) *(thundering)* ***Watch:* *Gosh, it's about to rain. We should roll back to the riverside now!***	
Stretch your arms straight in front of you with your palm facing down, rotate your wrist 90 degrees, downward towards yourself and move it back and forth	
...	*It seems like we need two paddles to roll the boat.*
Oh you dropped your paddles. Reach forward to grab it back then continue the rowing motion.	*Seems like you just drop your paddles.Try grabbing them inside the boat first before rowing.*
Your arms are too far apart. Keep them in front of you while rowing.	*Now you are just rowing with your hands, we need paddles to speed up. And they are lying in the head of the boat.*
We are getting close, just roll three more times.	Haha, good job! I can see the moon get reflected on the water. Just keep rowing!
...	Just love these waves so much! You know what Dean, I would like to stay on this peaceful lake forever. Oh, don't stop, I am just kidding. Please just keep moving!
Good! we have arrived the riverside. Now we can get off the boat now.	When I am thinking of your mum's face, I just can't bear you being late. Hurry up man, Two hands, back and forth. You can do it!

Table 5.5: Audio cues and gesture instructions: Tree Shaking Gesture

Tree	
Ver.1	**Ver.2**
(frog cry getting weaker) ***Watch:*** **Hmmm~The scent is getting strong here! Oh!it's the tree of frangipani! But it's dangerous to climb at night, I think we can just rock some from the tree.**	
Grab the tree trunk with both hands and begin to sway it left to right. Stop until you hear the wind.	
It seems like shaking the tree isn't helping the petals fall towards the ground. Why don't you try to sway it from side to side?	*Oh, the flowers grow on the inner parts of the branches, so sway the tree more vigorously.*
...	Yeah, you made it! The flowers are falling! Keep shaking it, we definitely need more!
Cool Cool! I love the scent!	Cool Cool! I love the scent!

Table 5.6: Audio cues and gesture instructions: Rope Binding Gesture

Rope

Ver.1	Ver.2
(leaves sound ends) *(Wind Blowing fade in)* ***Watch:*** **Hey, it is getting windy, don't let them blow away the boat. Now pick up the rope and bind it to the tree.**	
Have your palm facing downwards and fist your hand into a ball. Then circulate your arm in a churn.	
You forgot to pick up the rope! Tie it around the trunk by moving your arm in a circle from side to side.	*No no no! The branches are too weak for the rope to be wrapped around. Instead coil it on the rims of the trunk.*
Yea, you got it! Do this two more rounds and then it would be fine.	Nice, try some more rounds left to right to make it tighter.
Well done Dean!	Good, you got it! But it doesn't seem can hold the boat in this strong wind. Just keep binding!
Now come on, we need to hurry home!	Right, tight enough! Now come on, we need to hurry home.

Table 5.7: Audio cues and gesture instructions: Bike Bell Ringing Gesture

Bike

Ver.1	Ver.2
Watch: **Get on the bike now, it is under the tree.**	***Watch:*** **Oh, you left your bike under the tree, remember? Just grab your handles to get on it, then we can get home sooner.**
Pick up the bike handles by curling both of your hands in to a fist. leaving only the right thumb out to ring the bell. Then jab it towards the index knuckle 6 times.	
I see you found the bell. But safety first! Always keep both hands on the bicycle.	*Cool! Seems like the bell woke up the fireflies. Their dance is so pretty! It will help light up your way home. So keep ringing the bell while riding your bike.*
Awesome you just hit the brakes. But right now we do not want to do that. So relax your grip slightly on the handles.	*Oh, you are just hitting the brake right now. Try holding both your handles to get on the bike first, we need to to home soon.*
...	*Ouch~it really hurts to fall down in the darkness! Try to get on the bike again.*
...	*Oops~ you are getting out of balance! It is dangerous to just hold one handle!*
	Humm, I just like the feeling of riding! Oh god, I can feel I am flying with the fireflies. Just ring the bell to attract more!

Table 5.8: Audio cues and gesture instructions: Star Picking Gesture

Star

Ver.1	Ver.2
(blowing Reed Getting stronger) *(space atmosphere sound fade in)* *(thundering starts)* *(wheels running ends)* ***Watch:*** **Oh my god! It is not the rain! Stars are falling from the sky! Wow, so beautiful! they are making the grass glow. let's pick some before we go home!** *(steps walking on the grass/reed starts)*	*(blowing Reed Getting stronger)* *(space atmosphere sound fade in)* *(thundering starts)* *(wheels running ends)* ***Watch:*** **Oh my god! It is not the rain! Stars are falling from the sky! Wow, so beautiful! they are making the grass glow. let's pick some to take a closer look!** *(steps walking on the grass/reed starts)*
Now start with palm facing downwards in a fist. Rotate it 360 degrees, resulting in open palm facing upwards.	
Great, you have just picked it up. Now try to flip your fist over and open your hand.	*Wow, I can't believe that you can catch it so easily! They are just so fast falling down! Wait, it has moving streaks, just flip it over and see how it goes!*
...	*The wings are starting to grow inside of it, just open your hand and let it fly!*
Ok, one!	Great, it start to fly up high again! But wait, some annoying fireflies are dragging it down. Oh no, it is struggling on the grass. Let's pick it up again!
Two!	Yea, flip it over and let go, so that it can fly again
Three! AH~~~~~~it is the time crack!	Oh my god! its streak is getting long, and the sky is cracking!! just let it go! Otherwise you will be dragged into the time crack! Ah~~~~~~~~~~~~

Game Mechanics The game mechanics exhibit the following three characteristics.

1. Audio-Only

   Once the game starts, players would follow the narrative voice and some informative sounds to finish the game by doing the hand gestures step by step. There is no visual representation of the game environment.

2. Gesture Interaction

   Players hear the story progress if they perform the gesture correctly. If they perform it wrong, either during or after the whole gesture sequence, they would hear audio cues, from both the narrative voice and the game environment responsive to their current hand gestures, which suggest them interaction method.

3. Game Progression

   Players need to finish a series of hand gestures so that they can get to the end of the story. Players have unlimited time to explore the gestures in the game. However, depending on the duration and correctness of every gesture sequence, the game story will be slightly different as secret story branches will be unfolded as reward for players with better performance. They will be sent back to the main story line once they finish experiencing the rewarding gestures.

Actual Gameplay Audio plays a very important part in such a gesture-controlled audio-only project, as it is the only interface for gesture guidance in the game. The core of this design is to let people get both their body and mind relaxed, and the gameplay should be described like this: Follow the narrative voice and let the sound and motion create a picture in your mind[1].

In order to trigger players imagination with an immersive and reasonable context, this project integrates environmental sound effects with an interactive audio story. The spatial recognition of physical hand gestures presented by the responsive audio feedback was designed to help enhance immersion and entertainment of the interactive experience.

[1]See https://www.youtube.com/watch?v=Qmj7Z38VAh4 for a gameplay video.

5.1.2 Implementation

Gesture Detection - Hardware Setup Fig. 5.4 shows the hardware setup for the game which includes using a Leap Motion Controller[2] for hand tracking and data input, a MacBook Air for data recognition and processing of the game, and noise-cancelling stereo headphones for audio output.

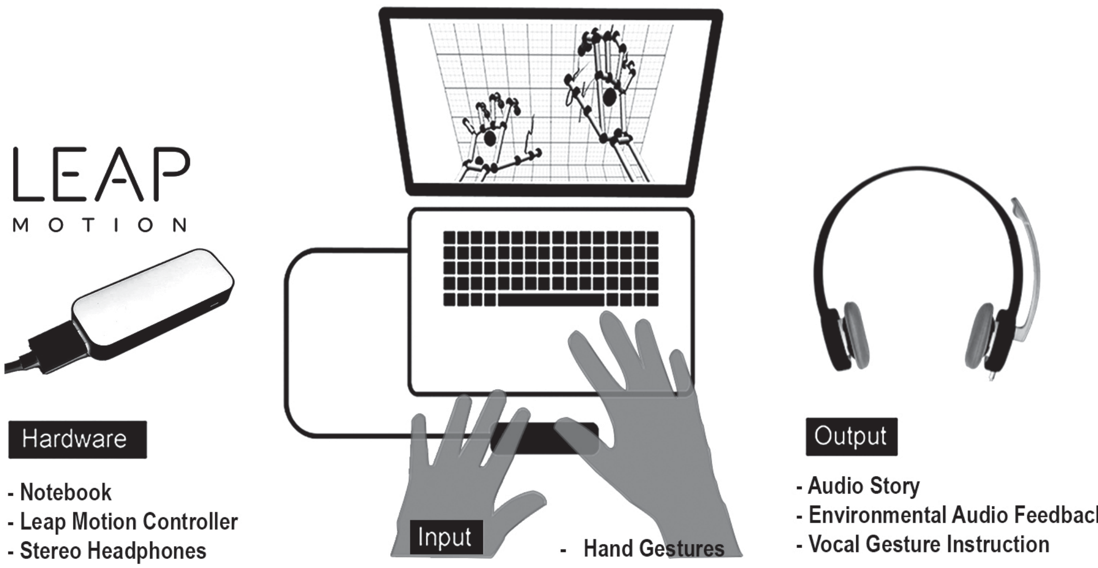

Figure 5.4: Game development set up: hand motion input, processing, and audio output.

Gesture Detection Framework As our project has the focus of investigating audio feedback for natural hand gesture interaction and has the objective to encourage intuitive participation, we considered the importance of building a detection system supportive for providing responsive audio cues for various reactive motion from different players.

Fig. 5.5 demonstrates three different possible approaches that players might use to go through a single gesture sequence in the game. If the player has a similar memory association with the correct gesture in their experience, he might perform his gesture smoothly and go straight to progress the story and hear the correct response from the game. This is the basic path (the blue arrows in the first diagram of Fig. 5.5) that we expect the players to walk through and how the game should progress. However, players' possible reactions might vary depending on their own experience and interpretation of the gesture. Our approach

[2] See `https://www.leapmotion.com/product` for product information.

for keeping the game progressing is to guide players back to the normal track by providing audio cues responsive to their current hand gestures. Responsive audio cues are presented as the purple arrows in both the second and third diagrams in Fig. 5.5.

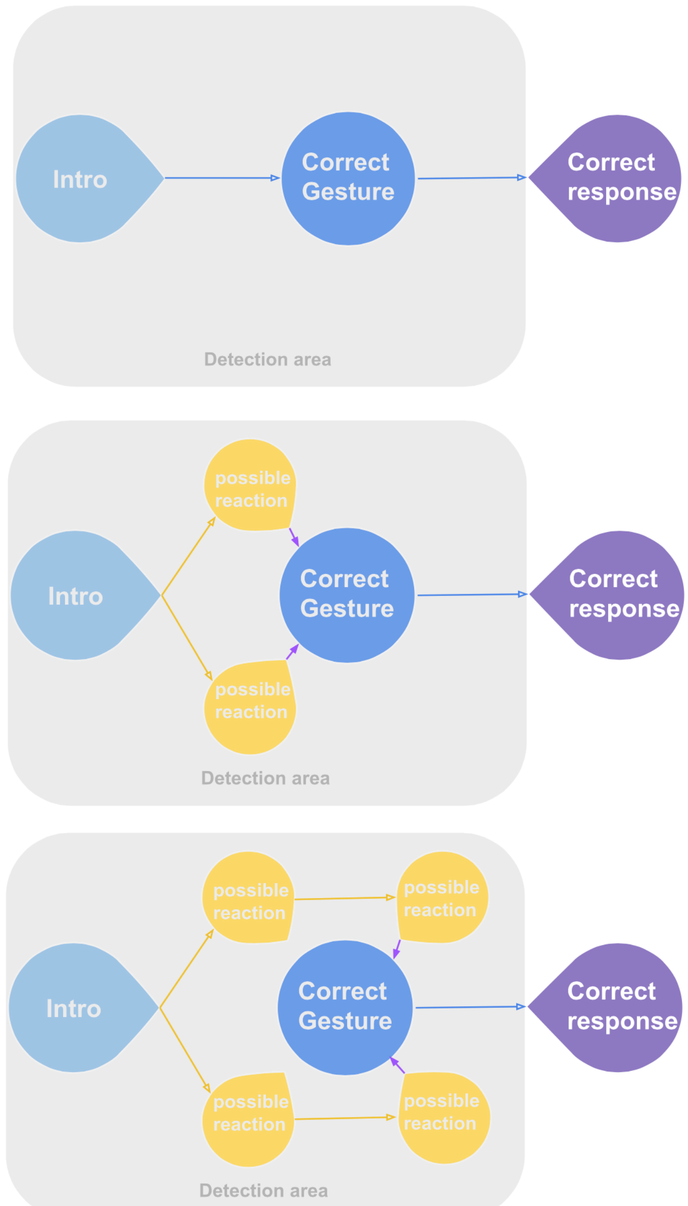

Figure 5.5: An overview on gesture detection structure and game progression.

In order to execute our approach, we built a real-time hand gesture detection system based on Leap Motion API. The Leap Motion API enables developers to get access to the raw data of both hand-skeleton, which includes grab strength, pinch strength, palm position, finger position, finger recognition, pitch yaw and roll data as measures for joints rotation, and palm radius. All of the data are captured by Leap Motion's frame-based tracking system.

Frame-based tracking has a high sensitivity to noise with small movements and it does not provide data on longer-lasting movements over time. In order to improve the stability of the detection, we created a new movement calculation unit, gesture state, and used several of them to build a gesture-state-based detection system based on Leap Motion's frame-based structure.

Figure 5.6: Gesture states, correct and incorrect sequence loops in detection system.

Our system supports 3-5 layers of detection by checking transitions between different sub-states of one gesture sequence, see Fig. 5.6, labelled: None, Detected, Action, and Cooldown as correct gesture sub-states; Ready, Other and "Ing" as incorrrect gesture sub-states. In the figures, blue arrows indicate correct loops, while yellow and red arrows connect the wrong loops.

The system checks the gesture sequences one by one according to the current story section. In every gesture state, we expect certain small movements preparing for various possible hand gestures from different players. Events triggered by those movements include audio feedback for specific movements, story content triggers for game progression, keeping track of gesture progress and the number of correct and incorrect attempts and the most likely next gesture state predicted by current movement.

In the first layer, before getting any gesture input, the system sets the current gesture sequence to the None state. When the player starts moving their hands and triggers a certain set event in the correct loop, he can reach the second layer and go into Detected state. If the player continues to trigger the following expected events in the correct loop, they can be sent to third layer: Action state. This layer is usually designed as a safety layer which can save all of the player's previous steps coming to this stage. If the player continues to follow the correct loop, they can be smoothly sent to the Cooldown state, the fourth layer that confirms a correct gesture input by playing the correct audio response. Players will stay in the Cooldown state for a waiting period until being sent to a new story section, whose length may vary in different gesture sequences. If the game requires the player to repeat a gesture for another time, they will be sent back to the None state to go through the detection process again.

If the player follows wrong loops and triggers various possible reactions, which are the preset events in the wrong gesture states, they will hear audio-cues responsive to their current movement in specific states either from the narrator or the game environment. These audio-cues, which are presented as Purple arrows in Fig. 5.6, were designed to suggest an interaction method and guide the player's performance to get close to the correct gesture. Players will have 1-2 failing chances before being sent back to None state, so they can try 1-2 possible reactions that are different from the expected gesture before failing the whole sequence. Once the player reaches the None state, the detection will start all over again. However, the system will not always send the player back to None state, as it is reasonable for the player to start re-performing gestures in the middle of some sequences. For example, when the player start to hold two fists side by side above the Leap Motion Controller, they have already gone into the Detected state of the Bike Bell Ringing Gesture. Sometimes, the player would fail to ring the bell because of an incorrect finger rotation. If player's two hands remain fists, the system would recognize the gesture as partly wrong, which will just send the player back to Detected state when the two-hand Bike holding gesture is being checked, rather than sending them back to None state and detect the two-hand states from open hand to fists again.

Gesture Library In addition to tracking raw data from both hands, Leap Motion software can recognize four basic patterns as default gestures: circle, swipe, key tap and screen tap[3]. However, this limited gesture library could not support the intended natural gesture design in our project. We therefore built our own gesture library by creating a frame-based movement calculation method and defining the story-related gestures based on the raw data as mentioned above.

The basic method of arriving at stable measures for detecing gestures from frame-based data consists of relying on the differences of values between a variable number of previous data frames to detect relevant movement. Based on this movement calculation method, we can get single-value changes within certain frame periods based on the Leap Motion API. After that we can start to define parts of the gesture sequence by putting several different data value changes together as a detection condition. And finally we can define a complete gesture sequence in our gesture library by combining all of the gesture parts together, see Fig. 5.7 for an illustration.

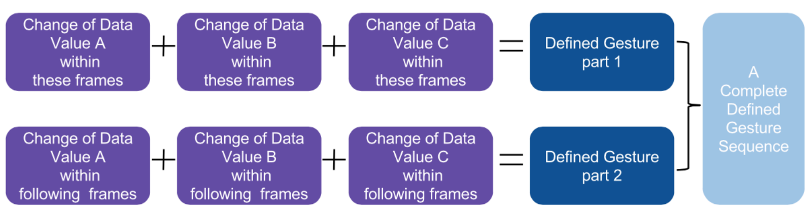

Figure 5.7: Gesture definition for the our gesture library.

Audio feedback Dimensions and Categories In order to realize our design goals, we include different kinds of responsive audio feedback. For a principled approach towards the design of audio feedback for hand gesture interaction, we distinguish potential feedback along two dimensions:

1. vocal instructions vs. environmental audio cues

2. non-diegetic vs. diegetic (i.e. story-related) audio cues

[3]See https://developer.leapmotion.com/documentation/csharp/devguide/Leap_Overview.html for more details.

We claim that the second variant of each dimension, respectively, can lead to a more immersive experience for the players. In the following, we present our design approach for an experimental setup that distinguishes between different types of audio feedback, by using the example of the stone-throw gesture in our story testbed.

In order to improve the responsiveness of audio feedback for hand gesture interaction regarding immersion, we consider audio feedback not just as a result of gesture interaction, but as an interactive signal that guides the player while going through the whole process of a gesture sequence: before, during, and after the gesture. All of these stages need an explicit auditory feedback. The following illustrates the design principles applicable to the three categories for environmental diegetic cues, using the stone-throw as an example.

Example Design for Environmental Diegetic Audio Cues To illustrate how we applied principals specifically to different stages of a complete gesture sequence for designing audio cues, we will focus on the Stone Throwing gesture sequence that represents grabbing and throwing of stones. In Fig. 5.8, the Stone Throwing gesture sequence is shown as seen from the side and, reading left-to-right, split into three main sections, pre-gesture, during-gesture and post-gesture.

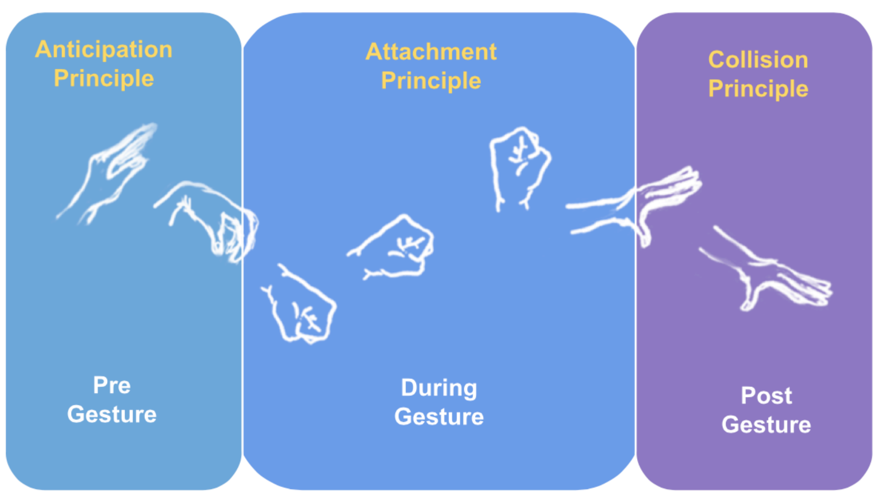

Figure 5.8: Design Principles for stages of the stone throwing gesture: pre-, during- and post-gesture.

In the story context, the player is prompted to grab a stone from inside their boat and to throw it towards a fish. The design of audio cues for this whole sequence will be discussed according to three principles:

Anticipation, Attachment and Collision.

Anticipation - pre-gesture Humans constantly use environmental signals or social cues given in advance by other humans or the surrounding environment to anticipate forthcoming events in part in order to know how to behave during the event. Ideally, effective audio design for games can leverage this understanding of anticipatory signaling.

1. Object-related audio for introducing the forthcoming gesture

   Sound derived directly from or related to the main object of the forthcoming gesture prepares the player: the player can hear sounds of fish moving and swimming in water, combined with an ambient sound of a boat creaking before doing the grabbing gesture. Through the combination of these introductory sounds with the audio story told by the character, the player can get a better idea of the setting for performing the forthcoming gesture.

2. Movement of object-related audio positions for suggesting the direction and scale of the forthcoming gesture

   The movement of the sound source is proportional to the forthcoming gesture: Fish sounds are concentrated in front of the boat, the intended direction of the stone throw. If the player keeps listening to the moving effect, they can get an understanding of the scale of the forthcoming movement as well.

3. Volume changes for informing the player of the availability of the forthcoming gesture

   Relative volume changes corresponding to the distance of the gesture's main object to the player can indicate if a particular gesture is currently available or not. For example, when the player is instructed to do a stone-throwing gesture several times in a row in order to scare away the fish, the volume of its swimming sound will be reduced after each successful gesture, as the fish is getting more and more scared. So after throwing the stone for the third time, the volume of the swimming sound will gradually fade out, suggesting that the fish has been scared away and there is no need or opportunity to successfully perform the stone throwing gesture again.

Attachment - during-gesture In real-world situations, hand gestures usually provide only minimal auditory feedback. Responsive audio feedback during a gesture sequence can only be expected when performed by someone or something that produces constant movement sounds, such as someone with noisy clothing or a metal robot. In order to enable instant feedback from small hand movements of the player, we need to include suitable attachments on arms or wrists that generate sounds in the story-setting. For example, a bracelet with a ringing bell would be helpful for telling the movement of the wrist. We employ this attachment idea for designing the during-gesture feedback and it is particularly interesting to our investigation as it provides the largest potential for improvements in making a gesture more easily understandable. In the game, during-gesture feedback in particular plays the role of guiding players towards going through a complete gesture sequence. So in addition to conveying the message of the current ongoing movement, the design of during-gesture feedback should also be able to reflect positive or negative reinforcement that affects players' next movement.

1. Real-time mapping of hand movement to audio parameters for enhanced responsiveness

   In the stone-throwing gesture sequence, we assume people will perform either an over-head throwing gesture or attempt to do a gesture suitable for skipping stones after grabbing a stone. In order to suggest that the over-head throwing gesture is suited to scare away the fish and, thus, to move the game's story forward, we use a sound of the player's coat rubbing and creaking as feedback for the hand lifting movement. Further, as the hand is moving upward, the rubbing audio volume increases. Without the help of visual information, humans can still infer the structure of and distances to their surrounding environment. In some cases, blind people are able to perfect this as an echolocation method to astounding levels of accuracy. As another example, in a paddling gesture sequence, players need to row the boat from a cave to the riverside. In order to suggest to the player that they are successfully performing the paddling gesture as audible movement of the boat and a change of spatial surroundings, we map the echo of the water wave and the paddle creaking to the dynamics of the wrist movement. Thus, players can hear the echo effect change according to their successful gesture performance.

2. Noticeable positive or negative feedback for guiding interaction

Apart from enhancing immersion using story-related audio cues, we also provide positive or negative audio feedback during a gesture sequence to better affect the players' next step. In our game, players are told that the main character is wearing a magic watch. In this case, we apply the attachment concept into designing the sounds of the magic watch to be responsive to the players' hand movement. We designed two types of sounds for the watch to convey movement feedback that are currently subjected to play testing: ticking and ringing tone. If the player performs a gesture sequence in the anticipated way, the ticking of time will remain at normal speed. If a player hears the time ticking getting faster, this suggests that they are wasting more time than they are supposed to use, or that they may be going in a wrong direction. The ringing tone uses a similar notion of suggesting correct vs. incorrect performance by a higher or lower pitch.

Collision- post-gesture Blind people usually get a sense of surrounding environments by touch. Apart from the sense of touch, this can also provide sound feedback, e.g. when a stick collides with the environment. Based on the fact that sound feedback triggered by any kind of object collisions can provide blind people with information of further interaction in the environment, we assume that providing the player with audio feedback of collisions between their hands or extensions thereof and other objects in the game environment setting can be a way of informing them of hand movement messages needed for performing a complete gesture sequence.

1. Collision audio feedback for informing hand movement direction

   When the player is prompted to throw a stone to scare away a fish which is blocking his boat from going forward, they need to locate stones. In order to tell the player that stones are located in the front of the boat, we provide different audio feedback based on three location sections: left-front, middle-front and right-front. Only when the player performs a grabbing gesture in the middle-front section, they can hear an audio related to stone movement, suggesting the grabbing of a stone. In any other section, they can hear water splashing, which suggests the collision between their fingers and the water around the boat.

2. Setting of spatial surroundings to enable border collisions that indicate available moving space

In addition, the surrounding spatial setting is designed to help reminding players of the height and width of the available moving space. For example, if players attempt over-head throwing after grabbing, they might lift their arm before letting go of the stone. In order to indicate that there is limited space for lifting arms, we create a cave setting for this gesture sequence, and provide hand-cave collision feedback when a players' hand position reaches that boundary.

5.2 Experimental Approach

In this section, we focus on the methodology, the design and the execution of our experimental study comparing the different designs.

5.2.1 Study Design

After game development, we set up a between-subjects experiment to test the immersion degree of two game versions.

Subject Requirements and Recruitment

- Age 18 or over
- Literate in English
- Near perfect hearing (corrected or uncorrected)
- Normal full range of motion of the hand and arm

All of our potential subjects are recruitment via email, announcements at the university and flyers. Consequently, most of our subjects are expected to be students of Drexel University.

Play Test Setup

1. Experimental Condition

   Experiments were held in a quiet non-glass-wall room in order to limit outside interference. Brightness of the room was adjusted by the experimenter to prevent subjects' from being distracted by lights. As subjects' physical gameplay needed to be recorded as part of data for project evaluation, we kept the light bright enough to see the subject's movements. Subjects wear noise-cancelling stereo headphones to experience the game.

2. Before the Test

   The game was set up in a PC plus Leap Motion installation before subjects get in the room. After giving informed consent, the experimenter would introduce the whole testing section to the subjects, including the study objective and experimental devices. In addition to this, a pre-test questionnaire on subjects' experience of computer gaming, gesture interaction and audio-only environments needed to be filled before playing the game.

3. During the Test

   No other people were expected to be with the subject except the experimenter in the test room. Subjects were given as long as they wanted to complete the game, as our study's objective is to investigate audio feedback's impact on immersion, a time limit would at some point become another variable affecting the study results. Once the subjects were prepared and put on the headphones, they were randomly assigned to finish one of the two game versions by the system. Directions to start the game were given by the game itself. No conversations between the subjects and the experimenter were involved during the gameplay except subjects' requesting help. The experimenter was in charge of recording each game play section by using a digital camera.

4. After the Test

   A post-test questionnaire would be presented by the experimenter and needed to be filled by the subjects after game play to record their evaluation of the game version they played and the impact of different audio designs on game immersion.

5.2.2 Data Collection

In order to evaluate our study as comprehensively as possible, we set up several approaches for collecting data that can measure immersion and user participation.

Questionnaire Design In our study, two questionnaires were designed as part of the data collection approach: A Pre-test questionnaire and a Post-test questionnaire. Before the game, each subject was asked to fill in the Pre-Test questionnaire. Questions in this questionnaire can be divided into three parts: 4 questions on physical and statistical characteristics of the subject, 7 questions on experiences of general computing, computer gaming, audio-only and gesture interaction, and finally 2 questions on subject's immersive tendencies. The immersive tendencies questions were designed based on Immersive Tendencies Questionnaire from [Witmer and Singer 1998], intended to see if there are differences between people that have a tendency towards immersion vs. those that do not.

After playing the game, each subject was asked to fill in the Post-Test questionnaire. This questionnaire was designed to interpret subjects' feelings of immersion. 24 questions in total were used to evaluate the following qualities of two game versions: Basic Usability (BU), Basic Enjoyment (BE), User Participation (UP), and Immersion (general IM, control IM-C, story IM-S, flow IM-F, environment IM-E). The evaluation of User Participation was mainly based on automatic measures (metrics) done by the game system and video analysis, but can be related to parts of the immersion measure (engagement) as well.

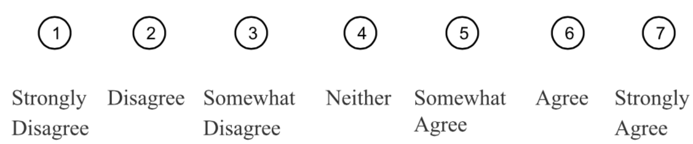

Figure 5.9: 7 possible answers to the Likert-style questions.

We used Likert-style questions aggregated into Likert-items throughout: Answers from "Strongly Disagree" to "Strongly Agree" (7 possible answers, See Fig. 5.9).

Basic Usability and Enjoyment questions were developed by us, specific to the game, but relate to previously used game evaluation questionnaires. Engagement and Immersion questions were directly adapted from existing questionnaires: the Game Engagement Questionnaire [Brockmyer et al. 2009] and work on an immersion questionnaire [Jennett et al. 2008]. We used two measures: one single-question measure that directly asks about immersion, and a combination measure based on questions used in those studies, adapted to this game. Below are the 24 questions, including a two- or three-letter indicator of which questions contributed to which aggregated Likert score (BU, BE, IM, IM-C/S/F/E), as well as references to the source of the question and additional notes where applicable.

- Q1. I could easily hear everything the character said. BU
- Q2. I would need outside help to be able to play this game. BU
- Q3. It was hard to follow the story. BU
- Q4. I would need to learn a lot before I would be able to enjoy this game. BU
- Q5. The gesture interaction in the game felt intuitive. BU
- Q6. Listening to sound hints for the gestures was helpful. BU
- Q7. Playing the game was tiring. BU
- Q8. I enjoyed playing the game. (from [Jennett et al. 2008]) BE, IM
- Q9. I enjoyed the gestures and would have wanted to repeat gestures more often. BE
- Q10. I really got into the game (Immersion GEQ from [Brockmyer et al. 2009]) IM
- Q11. The game felt real (from [Brockmyer et al. 2009]) IM-F, IM-E, IM-S
- Q12. I played without thinking about how to play (Flow GEQ from [Brockmyer et al. 2009]) IM-F
- Q13. To me it felt like only a very short amount of time had passed. (from [Jennett et al. 2008]) IM-F

  Better suited than: I lost track of time (Presence GEQ from [Brockmyer et al. 2009]) or: Time seemed to kind of stand still or stop (Absorption GEQ from [Brockmyer et al. 2009])
- Q14. It was as if I could interact with the world of the game as if I was in the real world. (from [Jennett et al. 2008]) IM-C
- Q15. I enjoyed the narration and sounds in the game. (adapted from [Jennett et al. 2008] to audio-only) IM-E
- Q16. I felt that I really empathised with/felt for the character. (adapted from [Jennett et al. 2008] to "character" instead of "game") IM-S

- Q17. I became unaware that I was even using any controls. (from [Jennett et al. 2008]) IM-C
- Q18. I felt detached from the outside world. (from [Jennett et al. 2008]) IM-E
- Q19. I did not feel any emotional attachment to the game. (from [Jennett et al. 2008]) IM-S
- Q20. It felt like the world reacted to my hand movements. IM-C
- Q21. I did not feel like I was in the real world but the game world. (from [Jennett et al. 2008]) IM-E
- Q22. I was interested in seeing how the game's events would progress. (from [Jennett et al. 2008]) IM-S
- Q23. I felt very immersed in the game. (single question immersion measure, adapted from [Jennett et al. 2008]) IM (single-question measure)
- Q24. I would like to play this game again. BE

Further Data: Video and Metrics Apart from questionnaires as our current immersion evaluation approach, video was used to collect data for evaluating user participation of two game versions in the future. During the test, a digital video camera was set up in the corner of the test room to record the subjects' physical movements, see Fig. 5.10.

Automatic metrics (participation indicators) of player behavior were also used to complement manual video evaluation. Several metrics that reflect game progression were automatically recorded by the system for each subject's game play. Metrics include duration of each single gesture sequence, the count of wrong attempts of gesture performance and the frequency of bare hand gestures, i.e. how often hands entered and left the detection area.

Figure 5.10: A digital camera was set up for video data collection in the test room.

Chapter 6: RESULTS

In this section, we will present the evaluation of experimental data in the order of the above collection approaches, and analyse our study results in two main aspects, immersion and user participation, in order to show how our project helps to support our study hypothesis.

6.1 Pre-Test Questionnaire Results

Study Subjects 26 subjects joined our experiments. 13 subjects were randomly assigned for testing each game versions and evaluating the game immersion and audio design in that version. The age range of the whole group is from 19 to 34, and the median age is 22. Among the subjects, 6 are females and 20 are males. Only 1 of the subjects is left-handed.

Experience levels Most subjects have high experience levels with computers and games. However, almost all of them have low or even no experience with gesture interaction and audio-only.

Immersive Tendencies The collected data on immersive tendencies was not evaluated as part of the thesis work, but is part of future work on the dataset.

6.2 Post-Test Questionnaire Results

Significant differences in single questions Significant differences between two game versions were found in question 6 (Listening to sound hints for the gestures was helpful) and question 11 (The game felt real). Even though directly comparing single Likert-style questions should not be used as conclusive evidence, we report the findings here to show the clear difference despite the small number of subjects.

In Fig. 6.1, the numbers on the y-axis refer to 7 Likert scale answers, from strongly disagree to strongly agree, with 4 being the neutral choice. For Q6, both the mean and median in version 2 (the Environmental Feedback Version) are higher than those in version 1 (Explicit Vocal Instruction Version), which means diegetic audio feedback in version two was more helpful than the non-diegetic ones in version one in sug-

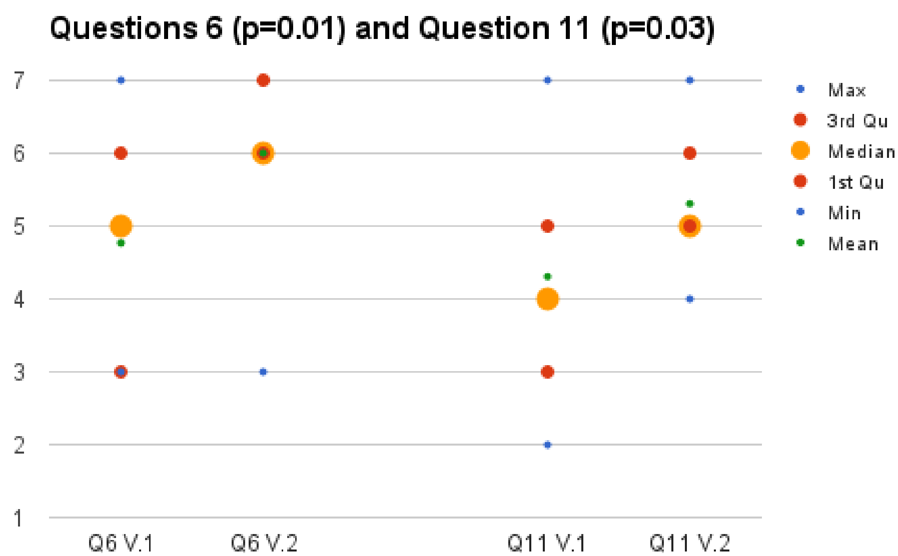

Figure 6.1: Significant differences between game versions for questions 6 and 11.

gesting gesture interaction methods. For Q11, the mean and median in version2 are higher than those in version1 as well, which indicates that version 2 also felt more real than version 1.

Signifiance testing using a t-test results in p-values of 0.01 for Q6, and 0.03 for Q11. Both are lower than the traditional threshold of 0.05, which indicates that the differences between the two game versions reflected in these two single questions were significant, or at least notable.

Regarding direct feedback for the design of the game, the distribution of the answers is particularly encouraging. The answers for version 2 are strongly positive as can be seen in Fig. 6.1 by the quartiles. Almost all answers for both questions were in the positive range.

Aggregate Measures In Fig. 6.2, we aggregated the answers to questions that evaluate the same qualities, showing the averages. Negatively formulated questions were adequately flipped. The first two items from the left are two basic measures: Basic Usability (BU) and Basic Enjoyment (BE). The rest are all measures for immersion, from left to right: Immersion Single Question measure, Immersion overall averaging all other immersion-related questions, Immersion regarding Control (IM-C), Immersion regarding Envi-

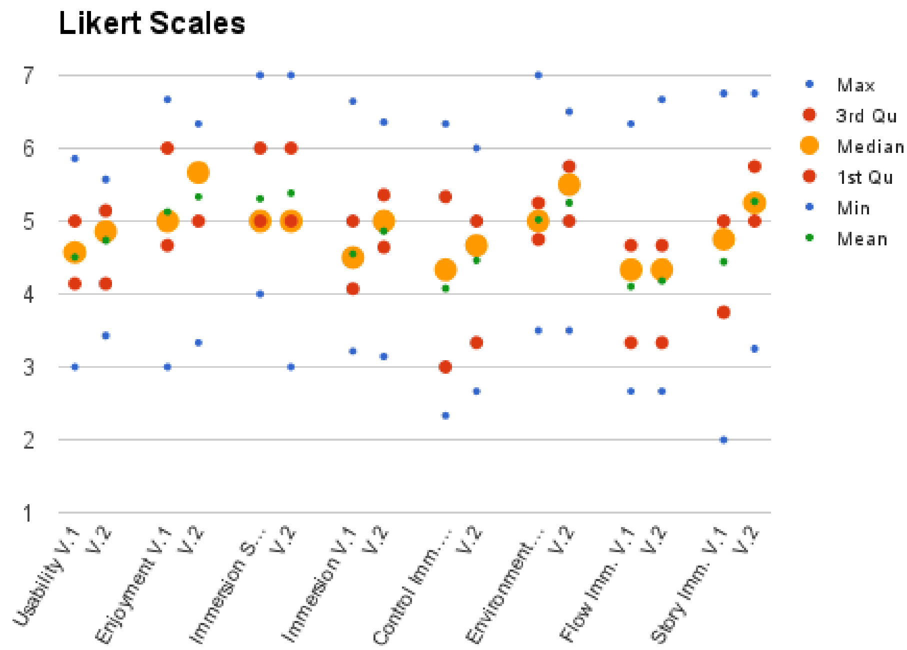

Figure 6.2: Likert items from all 24 questions comparing both versions. See the text for detailed explanation of the x-axis categories.

ronment (IM-E), Immersion regarding Flow (IM-F) and Immersion regarding Story (IM-S). Responses to all the 24 questions indicate that version 2 is consistently rated higher on average in all aspects compared to version 1.

In addition, the mean and median values to all the 24 questions in both versions are above 4 (the neutral answer), which shows that the feedback for both game versions is overall mostly positive, especially in some measures, such as Immersion Environment and Immersion Story, which were close to 5.

Story Immersion Item Fig. 6.3 shows the aggregated answers to questions that evaluate immersion regarding story, an aggregate of answers to questions Q10, Q11, Q16, Q19, and Q22.

Both the mean and median values of answers to these above questions from subjects who played version 2 are above 5 (somewhat agree) while those from subjects who played version 1 are below 5. All these results reflect that the story element in version 2 get subjects more immersed than the one in version 1. Among all the immersion measures, the story immersion item, in particular, is actually p-significant (p=0.03). This result indicates that the impact of the design for audio cues had a significant role in enhancing story im-

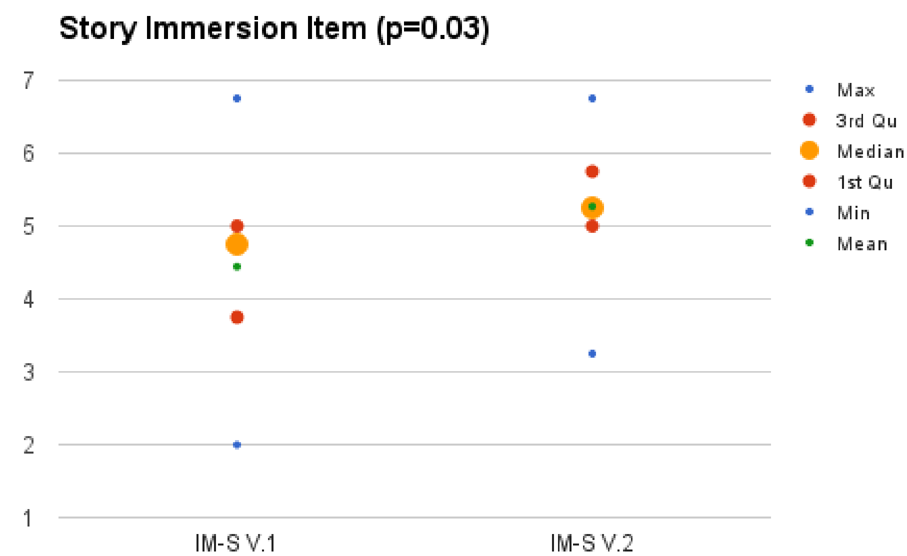

Figure 6.3: Story immersion is p-significant.

mersion.

6.3 Future analysis

Automatic Metrics Preliminary Results During each subject's game play, the game system automatically generated a txt file containing several specific metrics that reflect the subject's game progression and participation. We integrated all the metrics data and performed a preliminary analysis on the differences between two game versions. We found that durations of some gestures (paddle, rope and bike) are significantly shorter in the version 2 (the Environmental Feedback Version). This may suggest that environmental audio feedback for these gestures are better at making the game context feel real, better at triggering their imagination to do the gesture intuitively. We also found that players are better at performing gestures that they are familiar with in real life. However, the analysis was not conclusive as part of the thesis work. Future work will be to consider rest of the collected data in combination with the questionnaire results as well as the collected video data.

Video Future Analysis Due to time constraints, video footage has not been used in the current evaluation of our study. For future work, they can be used to observe subjects' intuitive reaction towards different gesture sequences in order to improve the current gesture detection system by enriching the variety of gestures in the gesture library. Comparing suggestive results from game metrics with qualitative video analysis would be another interesting way to further evaluate this study.

Chapter 7: CONCLUSION

This thesis presented different audio feedback designs for touch-less hand gesture interaction. The main goal was to enhance immersion and encourage intuitive user participation in audio-only environments. An experimental study found that the version using environmental feedback led to higher immersion in terms of story and was consistently rated higher. In this section, we provide some further thoughts about the significance and the implications of this research and summarize the contributions on both the theoretical and practical aspects. Finally we list some possible directions for extending this work in the future.

7.1 Research Significance

Significant Evaluation Results The two-group between-subject study has successfully supported our research hypothesis that replacing explicit vocal instruction for hand gesture interaction with story-related and environmental audio cues suggestive of interaction methods leads to higher immersion. In particular, we got significant evaluation results for story-immersion between two design approaches, indicating that a story context which consists of diegetic audio cues is more effective than an informative context that consists of non-diegetic audio cues in getting players engaged in intuitive hand gesture interaction. We hope our study results can strengthen our understanding on game immersion, at least in terms of story immersion.

Techniques Applicable beyond audio-only environments

Audio Feedback Design Principles for Enhancing Immersion in VR Experience Our study used an audio-only game as testing platform in order to focus on the investigation of different designs of audio feedback. Results showed the important role that audio elements play and their impact on game immersion, especially the environmental aspects of immersion. As environmental immersion is one main feature of virtual reality, and touchless gestures, especially hand gestures, is becoming and important interface as part of the immersive experience, we believe that our design principles for audio feedback, focusing on before-, during- and after-gesture elements, can be applied for enhancing immersion in VR experiences. We also hope our approach of achieving immersion can make a small contribution to attracting more designers' attention to audio elements as one of the important factors for immersion qualities.

Detection Framework for complex gesture sequence interactive training purpose In addition to the study results, our research also set up a framework for detecting complex hand gestures and for providing responsive feedback. This framework can be used as base in designing more complex feedback systems for gesture sequences with long durations and various reactions. For instance, it can be applied into detecting sign languages or martial arts hand forms, which offers designers new possibilities of making gesture-based educational games or interactive training application for physical movements.

7.2 Research Implications

Potential Application for Blind Community We created an audio-only game using a hand-gesture control interface. Both the game environment and responsive feedback for hand gestures are all constructed purely using audio elements which leads to accessibility for both people with normal and impaired vision. For the future, this work can be extended to become a specific study on motion-control games for the blind community. New design requirements for the audio feedback will have to be integrated to better serve this specific community and enlarge their access to motion-control gaming.

Comparative Experiments in VR Touchless hand gesture interaction is becoming an important part of most VR experiences. However, feedback for the hand gestures are mostly visuals, which is in some way distracting the player's concentration on observing the game environment or achieving the main goal. For future work, in order to test whether audio cues for hand gestures can help to convey interaction methods without affecting player's attention, we suggest comparative experiments between two game versions: gesture-based VR without audio feedback vs. gesture-based VR with audio feedback.

Detection Hardware alternatives Although the Leap Motion hardware can support hand and finger motions as input and requires no hand contact or touching, it is still a computer hardware sensor device, which needs to work under the cable-PC connection condition. For future work, devices such as depth cameras in Microsoft phone [Fanello et al.] and the leap motion mobile version, which might be released in the near future can be applied to improve game mobility.

In addition, the Leap Motion is a controller based on camera tracking technology, its detection area is not flexible enough for supporting total immersion. For future work, researchers can apply devices such as Myo[1], a newly released gesture control armband, using a wireless connection, to help improve the degree of movement freedom for players to enjoy immersive gesture-based experiences.

[1]See https://www.thalmic.com/en/myo/ for product information.

Bibliography

S. Baldan, A. de Götzen, and S. Serafin. Mobile rhythmic interaction in a sonic tennis game. In *Proceedings of the CHI '13 Extended Abstracts on Human Factors in Computing Systems, New York, NY, USA*, pages 2903–2906, 2013.

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

J. Brockmyer, C. Fox, K. Curtiss, E. McBroom, K. M. Burkhart, and J. N.Pidruzny. The development of the game engagement questionnaire: A measure of engagement in video game-playing. *Journal of Experimental Social Psychology*, 45:624–634, 2009.

E. Brown and P. Cairns. A grounded investigation of game immersion. In *Proceedings of CHI'04 Extended Abstracts on Human Factors in Computing Systems*, pages 1297–1300, New York, NY, USA, 2004. ACM. doi: 10.1145/985921.986048.

B. Caramiaux, F. Bevilacqua, T. Bianco, N. Schnell, O. Houix, and P. Susini. The role of sound source perception in gestural sound description. *ACM Transactions on Applied Perception*, 11:1, 2014.

P. Clifton, J. Caldwell, I. Kulka, R. Fassone, J. Cutrell, K. Terraciano, J. Murray, and A. Mazalek. Don't open that door: designing gestural interactions for interactive narratives. In *Proceedings of the 7th International Conference on tangible, embedded and embodied interaction., New York, NY, USA*, pages 259–266, 2013.

D. Cranor, A. Peyton, A. Persaud, R. Bhatia, S. Kim, and V. M. Bove. Shakeonit: an exploration into leveraging social rituals for information access. In *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction. New York, NY, USA*, pages 277–278, 2010.

S. R. Fanello, C. Keskin, S. Izadi, P. Kohli, D. Kim, D. Sweeney, A. Criminisi, J. Shotton, S. B. Kang, and T. Paek. Learning to be a depth camera for close-range human capture and interaction. *ACM Transactions on Graphics*, 33.

L. Hantrakul and K. Kaczmarek. Implementations of the leap motion device in sound synthesis and interactive live performance. In *Proceeding of the 2014 International Workshop on Movement and Computing., New York, NY, USA*, page 142, 2014.

R. Hartson and P. Pardha. *The UX Book: Process and Guidelines for Ensuring a Quality User Experience*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc, 2012.

M. Huhtanen. Kinetic kid – guiding child interaction within a motion game. https://aaltodoc.aalto.fi/handle/123456789/13992, 2014.

C. Jennett, A. L. Cox, P. Cairns, S. Dhoparee, A. Epps, T. Tijs, and A. Walton. Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies*, 66:641–661, 2008.

M. Liljedahl, N. Papworth, and S. Lindberg. Beowulf – an audio mostly game. In *Proceedings of the international conference on advances in computer entertainment technology., New York, NY, USA*, pages 200–203, 2007.

J. Llobera, K. J. Blom, and M. Slater. Telling Stories within Immersive Virtual Environments. *Leonardo*, 46: 471–476, 2013.

D. Manuel, D. Moore, and V. Charissis. An investigation into immersion in games through motion control and stereo audio reproduction. In *Proceedings of the 7th Audio Mostly Conference: A Conference on Interaction with Sound., New York, NY, USA*, pages 124–129, 2012.

F. F. Mueller and K. Isbister. Movement-based game guidelines. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems., New York, NY, USA*, pages 2191–2200, 2014.

J. H. Murray. *Hamlet on the Holodeck.* New York: Free press, 3rd edition, 1997.

D. Norman. Natural user interfaces are not natural. *Interactions*, 17:6–10, 2010.

K. O'Hara, R. Harper, H. Mentis, A. Sellen, and A. Taylor. On the naturalness of touchless: Putting the "interaction" back into nui. *ACM Trans. Comput.-Hum. Interact*, 20:no.5, March 2013.

N. Papworth. ispooks: an audio focused game design. In *Proceedings of the 5th Audio Mostly Conference., Piteå, Sweden*, pages 1–8, 2010.

E. Patrick, D. Cosgrove, A. Slavkovic, J. A. Rode, T. Verratti, and G. Chiselko. Using a Large Projection Screen as an Alternative to Head-Mounted Displays for Virtual Environments. *CHI Letters*, 2:478–485, 2000.

A. Radford. Games and learning about form in architecture. *Automation in Construction*, 9:379–385, 2000.

N. Rober and M. Masuch. Interacting with sound: An interaction paradigm for virtual auditory worlds. In *Proceedings of ICAD 2004, Sydney,Australia*, pages 277–278, 2004.

T. E. Roden, I. Parberry, and D. Ducrest. Toward mobile entertainment: A paradigm for narrative-based audio only games. *Science of Computer Programming*, 67:76–79, 2007.

T. Winograd and F. Flores. *Understanding Computers and Cognition.* Norwood NJ: Ablex, 1986.

B. Witmer and M. Singer. Measuring presence in virtual environments: A presence questionnaire. *Teleoperators and Virtual Environments*, 7:225–240, 1998.

Appendix A: Game Development Process

Several iterations of the game concept were designed and implemented before the final version used for the reported study.

A.1 Logo Design

Though I was building an audio-mostly game, I had designed a simple set of UI elements for early versions of my game as well. Here is the logo I designed for the game using Photoshop CS5 (see Fig. A.1) and some screenshots of the start screen (see Fig. A.2), game screen, and the final screen of a prototype intended for the iPhone. As for the game screen, I used different color schemes to create a basic atmosphere for the environment of each part of the story.

Figure A.1: Logo design: 5 music notes compose a five fingers hand form, colorful circling curves symbolize the waves of both motions and music.

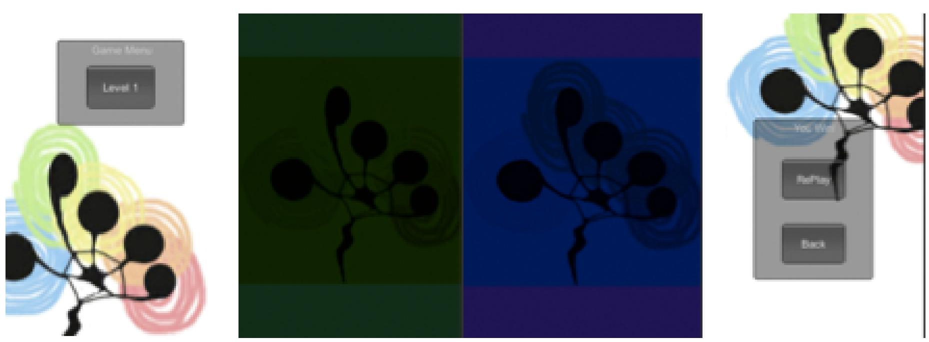

Figure A.2: Screenshots of start screen, game screens and the win screen.

A.2 Hand Gesture Design

I visited the Wingchun martial arts training club in Philadelphia to look for a gestures design reference when I built the first game version on iPhone. Since iPhone accelerometers can only provide single-point detection, I designed simple gestures that just feature the wrist rotation in Siu Lim Tao hand forms to build the entry level of my game. See Fig. A.3 for a decomposition of one Siu Lim Tao hand form that I use as core features for the gestures design.

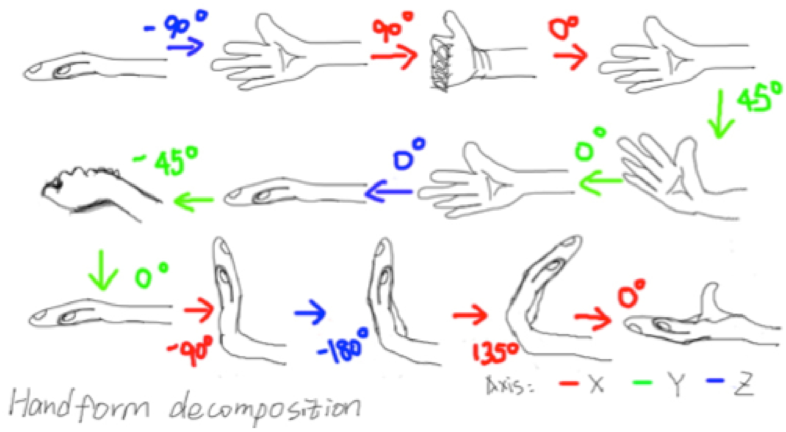

Figure A.3: Draft of hand form decomposition, based on the original form of Siu lim Tao.

A.3 Game Development

Software Architecture I chose the Unity3D game engine as the building software for my game design. As Unity is an established platform for mobile game design, it provides me with references of effective basic scripts as well as a plugin for Leap Motion speeding up development. As the audio elements would play an essential role in the design, I chose both CoolEdit Pro and Adobe Audition CC to compose the background music and edit all the sound effects. CoolEdit Pro provided me with a multiple sound tracks composition setting and a variety of sound filters, which satisfies most of the possibilities of editing sound sources. For the narrative voice and the motion guiding audios, I used Adobe Audition CC to adjust the pitch, stretch the duration and also change the reverb of recordings from the voice actor in order to fit them in the story and the style of the game.

Wrist Motion Recognition (for iphone prototype) In my iPhone prototype, the basic concept of motion recognition would be presented in the following steps: 1. Move the wrist. 2. Hear the sound. Based on this idea, I named my game Swing and Ring.

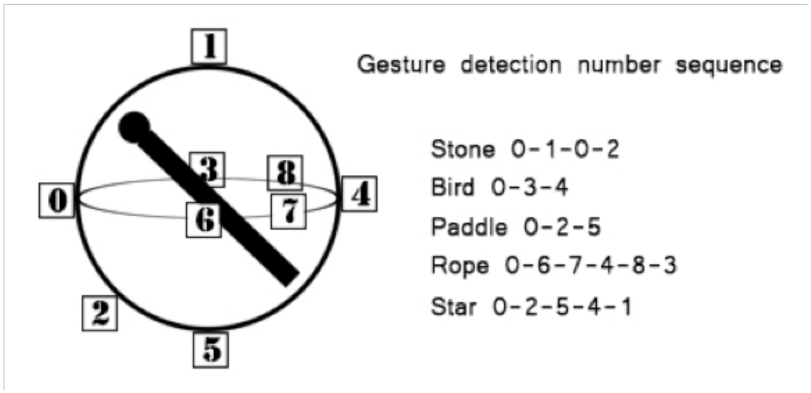

Figure A.4: The series of colliders around the users hand that correspond to the motions needed for successful gesture detection in the prototype.

In the prototype level of my game, there are stone throwing, bird catching, paddling, rope throwing and star picking gestures, see Fig. A.4. When players perform each motion, they can get either a sound response of the item involved in the motion for showing correctness or a guiding voice which shows them clues for correct motion when they perform it wrong and let them continue the story by following the guidelines. When players finish doing every motion for several times according to guiding voice, the story continues. After interacting with all the sounds by performing all the required hand motions, they can hear the whole story section of the level and win the game.

In terms of the novelty of game design, Roden et al. [2007] argued that comparing to realistic user interface of visual games, a simple interface enhances both immersion and entertainment value, making audio-only games more practical for mobile computing. So following this principle, I created a really simple scene in Unity for my game: a cube, a sphere and several colliders.

The cube represents the player's hand. As for showing the orientation of the hand, I attached a small sphere to the cube to simulate the finger tips, which is also set as the touch point to enter the sound triggers. Then I used C# Script to code for its angle rotation, which intends to simulate the rotation of the wrist. Here I coded the acceleration data captured by the accelerometer in iPhone as input to control the rotation of the cube. And I set up several colliders as sound triggers around the cube, for the different gestures, see Fig. A.5

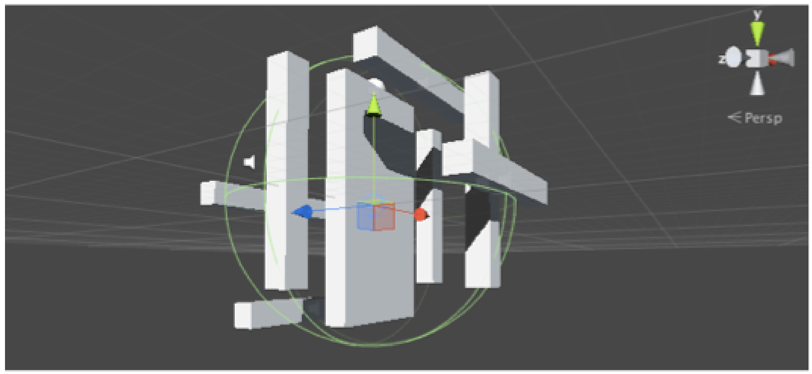

Figure A.5: The game scene for the motions of stone throwing, bird catching and paddling.

and Fig. A.6. Their positions are arranged within the 360 degree reach range of the sphere.

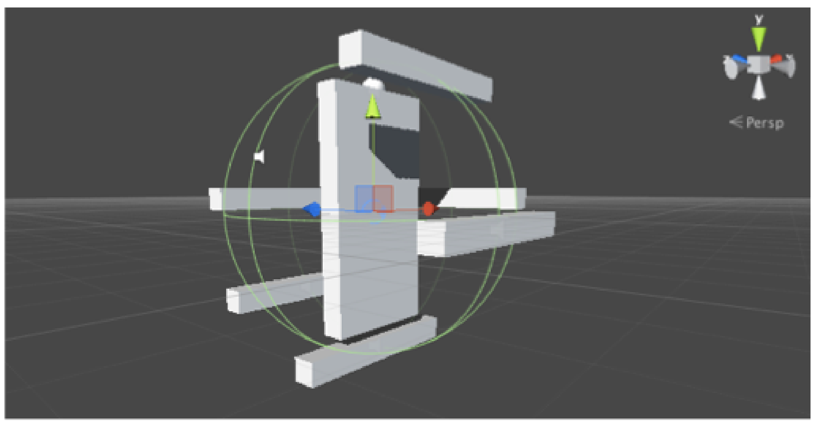

Figure A.6: The game scene for the motions of rope throwing and star picking.

I portrayed each motions with several colliders. When players use their hand to hold the phone and start rotating their wrists according to the narrative voice, the sphere would start enter the sound triggers constantly. The way I used to judge the correctness of the motion is to check the sequence of the conveyed trigger numbers in the sphere. If the sphere enters the triggers in a determined sequence I set in advance, then the motion is performed right, otherwise it is performed wrong.

In addition to relevant sound effects of the items and the responsive feedback from the narrative voice, feedback for wrong gestures may include the phone vibration as well to take advantage of haptic feedback and a richer somatic experience on a mobile platform.

Hand/finger based gesture recognition (in the Leap Motion version) Thanks to the Leap's skeletal tracking system, the previously described method of using collisions for motion detection was replaced. Using the API of Leap Motion, I have access to almost all the position, rotation or even velocity data of all the elements from two hands. In addition to the basic yaw, roll data of the palm, we can also get these data of every fingers. Leap also provides some other detecting variables such as the sphere radius, the grab strength and the pinch strength. Apart from accessing hand-finger data from every frame using leap API, I also implemented a new method for gestures detection, as described in the main part of the thesis in the section on Implementation. Fig. 5.1 and Fig. 5.2, show the gestures that were transformed from the early prototype to the final version using the new features, including: Grabbing stones before throwing, closing the whole hand into fist to catch bird, two hands paddling instead of one hand, wrist circle motion and picking before opening hand.

A.4 Design Influences on Gestures and Sound

Gesture Interaction

1. Detection method

   From the preliminary results during development, I learnt the problem of delay and overlapping audio feedback which was fixed for the final game version. Despite possible technical solutions, this problem suggested a focus on avoiding frustration by promoting immediacy of feedback and enriching the variation of feedback according to players' intuitive reaction. Based on movement-based game guidelines in [Mueller and Isbister 2014] and views on natural gesture design in [Bianchi-Berthouze 2013], I considered to focus more on the aspect of maintaining the immersive experience rather than achieving the realism of gestural performance when revising the detection method, as I don't want to build up a demanding detection system which will cause loss of participants' interest.

2. Instructing method

   The "scripting" techniques for interaction design in "Don't open that door" [Clifton et al. 2013], served as a reference to follow when revising the detection method of my game. The research mainly focuses on exploring the method of instructing the interactor when and how to interact in interactive narratives, and there are four scripting techniques (verbal, audio-visual, reactive, mimetic) in the interaction design of this project. The difference to this project is that researchers can script the interactor by using a combination of visual and audio feedbacks, while in my design, I should consider the question of scripting the player to perform the determined gestures in my game just using audio cues. So far, verbal technique has been used in scripting my game's gestures in the current game versions. In terms of reactive technique, I can address it as making sudden moves of the audio sources to inform players of spatial changes of determined gestures. As for mimetic technique, I can address it as having player previously listen to a sequence of sound effects, then reproduce the sequence themselves by performing the right gesture. Caramiaux et al. [2014] investigated people's gestural description of different sound stimuli. According to their results, when people listen to sound sources with identifiable causal actions, they will find it easy to mimic the sound by performing the causal actions that cause the sound source. If exposed to sounds with no association of causal actions, people will mimic the sound by acting with abstract gestures. Therefore, when applying the mimetic scripting technique in audio-only environment, I should consider the identifiability of causal actions that cause the sound source. Here the causal actions will be referred to the gesture that should be scripted to the player to perform.

Sound and Game Environment From the preliminary results I learnt the problem of informing players of the gesture movement and position through explicit instructions will in some way prevent players from reaching a higher level of immersion. I intended to replace some of the instruction (ideally all of them) with environmental audio feedback responsive to hand gestures in order to maintain the immersive experience.

1. Using environmental audio feedback to suggest gesture message

   In the current game versions, gesture instructions told by the narrator include explicit degree of rotations and demanding sentences, which may overemphasize the participation boundary and keep players from reaching a higher immersion level. Our new approach is to use environmental audio feedback to suggest factors, such as position, force, distance, that may influence players to perform the determined gesture that moves forward the story. For example, the game's story has players throw a stone to hit a fish in the river, when players perform a gesture that triggers water splashing and fish swimming sound effects, the story will continue. However, players would perform “stone-throwing” differently depending on their intuition and interpretation. So based on the game environment, supplementary environmental audio feedback, such as a stone hitting the boat or hitting the tree which leads to leaves falling, or hitting a duck etc, is responsive to different players’ intuitive movements, to suggest to the player that they have not fulfilled the goal of hitting the fish in the river. After interacting with these supplementary audio elements, players should get a better sense of the gesture context and the location and distance of the goal object, which helps them in performing the determined gesture correctly.

2. Convey space and distance through sounds

   Since there are no visual cues to show the game, players can only perceive the game environment by listening to sounds’ echoes or ambient sounds. Depending on different spatial characteristics of the places, I use different filters to adjust the sounds’ echoes according to their locations in the game world: the lake, the land under the tree and the opening grass field, and so on. The idea of using echo to convey space and distance message is inspired by echolocation, a method of perceiving environment features used by blind community.

3. Convey time message through sounds

   In [Roden et al. 2007], time-of-day is mentioned as a sound attribute. It was used to vary the way sounds are played at locations depending on the time of day in the game world. Each sound has an associated volume, ranging from inaudible (0) to full volume (100), which would change according to three time states: day, night, and all. This assumes that time condition can be conveyed to the players through sound volume. According to [Rober and Masuch 2004], music played as background sound is primarily used for emotional impact but can also be used to suggest a change in light intensity. For example, slow low-pitched music can evoke a darker environment whereas faster higher-pitched music can signal a brighter atmosphere. This principle may also help in designing background music for describing daytime and nighttime.

In addition to the above methodologies to present time messages, season conditions can be used to convey the time concept. For example, the sounds of crickets and frogs to convey the message of summer time, or replacing the footsteps on dry grass with footsteps on snow to represent winter. As there is a time-travel scenario embedded in the story, changing the pitch of character's voice between different time period would be a way of showing the growth of age.