Sound Cube: Interact With Virtual Object In Real World

Yu-Jen Lin
School of Information
University of Michigan
YJLINTW@umich.edu

ABSTRACT

The Microsoft Kinect is an easy and powerful tool to tracking and locating human body's movement and position. This paper describes the prototype using Kinect to track user's movement in the space, to interact with invisible virtual objects with audio feedback in physical environment. . In this project, it uses either ego- and allocentric spatial reference to manipulate the virtual sound object either directly or using gestured-based control. The basic idea is to let the user uses their body movement to produces sound in real-time, so a mime actor or a dancer can have improvised movement with correspondent sound effect

Keywords

NIME, Movement-based interaction, Kinect

1. INTRODUCTION

Motion capture is a technique to recognize the movement of objects or people. Unlike traditional motion capture used in filmmaking, which requires user places markers on their bodies and calibrate bunch of cameras before using it, Microsoft Kinect provides a cheap and effective solution for markerless 3D motion capture. Currently, Kinect is widely used in interactive installation and stage performance.

In this project, I used human body's position in 3D space as well as their limbs' movement as input, to interact and manipulate the virtual object in the physical environment. I would like to simulate an environment where the mime actors live in, so the users cannot actually see and feel the objects in the area. However, the providing audio feedback will give the users information about their interaction between their body and the virtual objects.

In the rest of the paper, I first present a review of related project, and state the difference between the existing ones and this project. Next, I present the design scenario and design decision of this project. Then, I go over the system, and describe the interaction used in this project. Finally, I discuss issues I discovered and encountered while developing and testing the system and future works.

2. RELATED WORKS

As Kinect was launched onto the market, there are a lot of interactive projects using motion tracking to let the users get involved into the interaction without using tangible controllers. "Dancing with swarming particles" by Carvalho [1] captures the motion of the dancer and morph the particles based on the dancer's body movement. "Mirage Theater" by Huang[2] integrated the motion tracking with 3D stereoscopic projection and projection mapping, which provides vivid visual feedback to the audience.

To cooperate motion tracking with sound, Bevilacqua et al. [3] use the body position to trigger different sound clips, and uses body movement to modify the timber of the sound. The LoopJam (Frisson et al., 2012)[4] turns the dance floor into a two-dimensional collaborative instrumental map, where the users' position will trigger different sound loop, and their movement speed will alter the tempo of the sound.

Reactor for Awareness in Motion (RAM) by YCAM InterLab [5] is a framework that can be used to create an environment, where dancers can create and interact with virtual objects in. In order to detect the actor's motion accurately, it requires the actor to wear the MOTIONER, which is an inertial motion capture system. "The Third Room" by Honigman et al. [6] introduces a new music interface where the users can perform the music instrument based on the body motion with some physical instruments.

My project is different from the above works, since I focus more on exploring the invisible virtual world based on audio feedback. Instead of asking users to produces precise and accurate pitch and tempo from their movement, the idea of the project is to let the user interacts with virtual object in the environment with different sound feedback. And different from most Kinect application/project which is using the body movement as an input to control the interface on a 2D screen, this project let the users can freely move around the 3

3. SCENARIO AND DESIGN DECISION

The project is first inspired by the movie, "The Dancer" (2000), where a young scientist designed a system for a mute and deaf dancer to compose the music by using her body movement. Therefore, I was thinking if we can get into the mime actors environment, and hear the sound that produced while the actors interact with those imaginary object, it would make the performance become more vivid to the audience.

At the beginning of the design, I was thinking about whether providing some object in the physical environment to give the users some tactile feedback while the user interact inside the area, since moving around in the mid air is hard for general people to accurate locate object while the object is invisible and non-touchable. As it might contradict with the nature of mime, I decided not to provide physical object in this stage, and will seeking some other solution to provide some other invisible tactile feedback in the future.

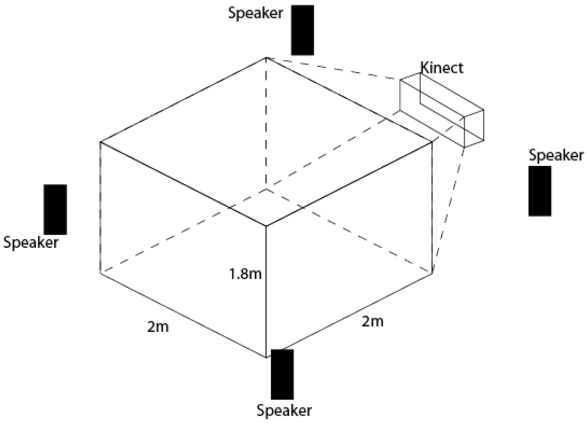

Figure 1. Physical Environment

4. SYSTEM OVERVIEW

4.1 Hardware and Physical Environment

The system consists of a laptop running OS X, a Microsoft Kinect, and a set of quadraphonic speakers. The user does not require wearing anything on to interact with the system. The physical space is roughly a 2m x 2m square. Due to the limitation of Kinect's viewing angle, the vertical range of the detectable area is around 1.8m, figure 1 shows the physical environment. The 4 quadraphonic speakers are placed on the four corner of the interactive area, and Kinect is placed in the front side, about 2m away from the center of interactive area, and about 1.5m high from the floor.

Since the idea of this project is to create a mime-like environment, so the feedback provided from the system should be acoustic only. However, in the demo version of the project, the laptop monitor still provide some visual feedback to indicate the relative position between user and those invisible objects.

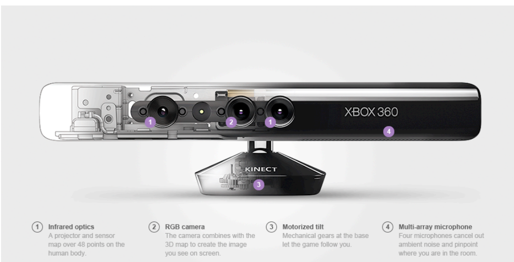

Figure 2. Microsoft Kinect contains infrared optics, RGB camera, multi-array microphones and motors

4.2 Software System Schema

As shown in figure 3, I used Kinect's depth camera to collect the user's movement input. The OpenNI SDK provides support for body motion tracking, which will generates user's skeleton position based on the raw data from Kinect, and than Zigfu wrapper parse those data to the game editor, Unity 3D. In the application, I track the limbs and head, and each tracked body part has its own id. To recognition different gestures, I use both the position of the body part in real world as well as the moving speed of each part. After recognition certain gestures or

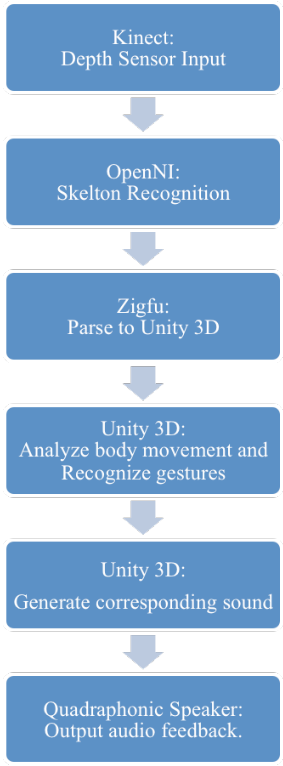

Figure 3. System Flow

movement, the system will generate corresponding sound and output it to the quadraphonic sound based on the location of the sound source in the virtual space.

4.3 Sound Generation

There are two type of sound generate methods in this project: one is generate from sample, another is generate by using different waveform and frequency modulation. The sampling method directly plays music files from the computer. The second method is generating waves directly from code. By using different formula, we can generate different waveforms, such as sine, triangle and white noise.

5. INTERACTION

5.1 Motion Tracking in 3D Environment

The system created a 1:1 virtual space directly mapping to the tracking area. Therefore, the system provides the tracked users to interact with the real object in the physical environment, as well as the virtual object in the virtual space at the same time. Before the system start, the users can add different Sound Objects, which will be discussed in next section, into the spaces they would like to interact with. In both the physical and virtual environment, the user has 3 degree of freedom to move their body inside the space: the X-axis is the direction perpendicular to Kinect's camera facing direction, the Y-axis is parallel to the normal vector of ground, and the Z-axis is the direction is same to the Kinect's camera facing direction.

Currently, limited by the viewing range of Kinect and the space of the area, the interaction in this project is best and limited to one person at a time. If there are more than one person in the tracking area, only the first person will be tracked. If two or more people appear in the scene at the same time, the front one will be tracked.

5.2 Sound Object

In this project, there are three different types of sound objects: Sequence, Continuous, and Percussion.

The Sequence sound object is like a music sequencer which plays loop of beats. All of the Sequence sound objects on the scene are synchronized with same timer. The length of the loops and the pitch of the notes can be edit in editor mode. The users use their body touching the sound object to toggle it on and off. While the user toggle the Sequence sound object on and stay their hand inside the sound object, it will enter modify mode. A tick-tock sound feedback will tell the user the object is currently in modify mode, and the user can use their second hand to modify the tone of the sound. The X-axis direction affects the low-pass filter's cutoff frequency; The Y-axis direction affects the low-pass filter's Q-factor; and The Z-axis affects the reverb mix ratio between dry and wet sound. User can define the length of the loops, and the pitch of the notes before starting the interaction.

The Continuous sound object will generate sustained sounds once the user touch the object, and the sound will stop immediately stop once the user stop touching it. Moreover, based on the body parts (i.e. left hand, right hand, left foot, right foot, etc.) that touches the Continuous sound object, it will produces different sound. User is allowed to use more than one body parts to interact with the Continuous sound object, and multiple sound will be mixed and play simultaneously. While the body parts are inside the Continuous sound object, user can move them along the Y-axis to alter the low-pass filter's cutoff frequency.

The Percussion sound object, as its name, will generate one and only one sound every time the user hits it. If user wants to generate the second sound, they have to move their body outside the Percussion sound object and hit it again. The Percussion sound object also detects the user's incoming speed. If the user hit it hard and fast enough, it will turn on the echo effect. The speed that is over the threshold will determine the delay time of the echo effect: faster speed will have shorter delay time, and vice versa.

Table 1. Sound objects' interactive methods and gestures

	Sequence	Continuous	Percussion
Trigger Method	Touch it to toggle on/off	Touch it to play the sound continuously, leave it to stop	Touch it to play the sound once
Enter Modify Mode	Stay the hand that toggles the object on inside the object.	While the sound is playing.	Move the body part that hits the object faster than the threshold.
Modify Sound Method	Move the second hand freely	Move the body part along the y-axis	Adjust the body part hitting speed
Modified Features	X: Cutoff Frequency Y: Q Factor Z: Reverb Dry/Wet Mix	Cutoff Frequency	Delay Time of Echo Effects

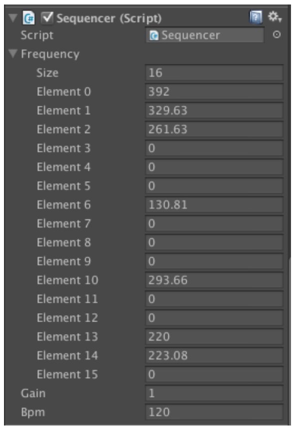

Figure 4. Sequence sound object setup interface. The user can define the length of the loop, the pitch of the note and the tempo of the loop.

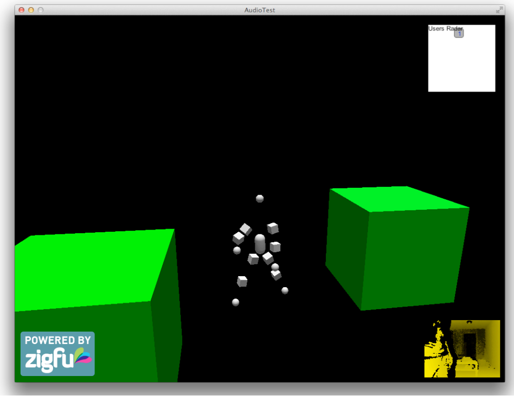

Figure 5. Visual Information and Sensor data.

6. DISCUSSION

In the process of developing and testing the system, I discovered several interesting issues that should be improve and need further exploration. First, the interactive method of the Sequence sound object is not intuitive enough. The current feedback provides to the users is not clear enough for them to accurately locate the position of the sound object. Most users are frustrated in toggle the sound object on/off accidently while moving around the space. In the first version, there is more space between different Sequence sound objects, so the problem is not obvious. In the second version, while all of the Sequence sound objects are align together, most users accidently trigger more than one sound object at the same time. The unclear boundary of the sound object also make it harder to entering the modify mode, since it requires the users to stay their hand inside the object.

Second, while the user move his/her body too fast, the skeleton tracking sometimes will lost the user which will make the speed detection have strange behavior. Another problem is while the user not facing Kinect's camera, the skeleton's orientation will sometimes be wrong and will incorrectly maps the user's right-hand side to left and left-hand side to right.

Another issue is, without having any physical constraint in the environment, it is hard to limit the users from the not desired, not expected, but sensed movement [8]. However, putting too many physical objects in the environment might also limit the actors and dancers movement on the stage. Therefore, we need to find a solution that balanced between free space and physical objects.

7. CONCLUSION AND FUTURE WORK

By using Kinect, we can easily track and locate the user's body movement in the environment. The Sound Cube project provides only acoustic feedback to let the user reconstruct and interact with virtual objects without visual clue. Different from most Kinect-based interaction, which is mapping the users movement from 3D space to 2D/3D screen interaction, this project construct a 3D space in real world let the users actually play inside it. In the future, I would like to make the sound cube in a larger scale and allowed multiple users interact inside the area together at the same time.

Another direction I would like to working on is to provide tangible or haptic feedback to the user without losing the freedom to moving around inside the area. A potential solution is combined with the research "AIREAL" from Disney Research [9], which provided tactile experiences by using a vortex of air.

Moreover, this system could be used into sonification, where take the users' movement and position as input, and generate different kind of sound based on their movement. For example in the surveillance system, if there is some people approaching the restrict area, it can produces different sound to alert the security guars. Overall, the Sound Cube project provides a new potential to use body movement to produces corresponding sound effect on stage in real-time.

8. ACKNOWLEDGMENTS

Thanks Erick Sheffield for setting up the quadraphonic speakers.

9. REFERENCES

[1] Carvalho,R. *Dancing with swarming particles,* 2011 http://visiophone-lab.com/wp/?portfolio=dancing-swarming-particles .

[2] Huang,H.C.*MirageTheater*, 2012 http://www.storynest.com/pix/_4proj/per_mirageTheater/p0.php?lang=en .

[3] Bevilacqua, Frédéric, Lisa Naugle, and Isabel Valverde. "Virtual dance and music environment using motion capture." In *Proc. of the IEEE-Multimedia Technology And Applications Conference, Irvine CA*. 2001.

[4] Frisson, Christian, Stéphane Dupont, Julien Leroy, Alexis Moinet, Thierry Ravet, Xavier Siebert, and Thierry Dutoit. "LoopJam: turning the dance floor into a collaborative instrumental map." In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, vol. 4. 2012.

[5] YCAMInterLab.*ReactorforAwarenessinMotion*, 2012. http://interlab.ycam.jp/en/projects/ram

[6] Honigman, Colin, Andrew Walton, and Ajay Kapur. "The Third Room: A 3D Virtual Music Paradigm." Presentation at the *New Interfaces for Musical Expression (NIME), 2013*

[7] Besson, Luc. *The Dancer*. DVD. Directed by Frédéric Garson. Los Angeles: 20th Century Fox, 2000.

[8] Benford, Steve, Holger Schnädelbach, Boriana Koleva, Rob Anastasi, Chris Greenhalgh, Tom Rodden, Jonathan Green et al. "Expected, sensed, and desired: A framework for designing sensing-based interaction." *ACM Transactions on Computer-Human Interaction (TOCHI)* 12, no. 1 (2005): 3-30.

[9] Sodhi, Rajinder, Ivan Poupyrev, Matthew Glisson, and Ali Israr. "AIREAL: interactive tactile experiences in free air." *ACM Transactions on Graphics (TOG)* 32, no. 4 (2013): 134.

10. Appendices may follow the references

Video links: http://youtu.be/XzPS0-9cPNw