Ephemeral Bits: Kinect®-based audio-visual interaction

Huynh Nguyen Hoang Minh
Creative Coding Group, EVL, UIC
Chicago, USA
hhuynh20@uic.edu

Angus Graeme Forbes
Creative Coding Group, EVL, UIC
Chicago, USA
aforbes@uic.edu

ABSTRACT
This paper presents Ephemeral Bits, an audio-visual interaction installation, using Microsoft Kinect camera. We present the piece in two versions: one with the Java voice recognition library Sphinx [3] from Carnegie Mellon University, that displays the participant's feeling when they speak to the artwork, and a purely visual performance version, that changes the graphic according to the sound of surrounding environment. Both versions use the depth data taken from the Kinect camera to visual the mesh in the version with voice recognition, and the point cloud in the latter version.

Author Keywords
Microsoft Kinect; audio-visual interaction; point cloud; Open Sound Control; voice recognition.

ACM Classification Keywords
J.5. Arts and Humanities: Arts, Performing arts. I.3.6: Computer Graphics: Methodology and Techniques: Interaction techniques. H.1.m Miscellaneous.

INTRODUCTION
Ephemeral Bits is developed as a semester-end creative project in the Computer Graphics II course at UIC. In this paper, we present the inspiration behind the installation, and discuss the interaction design, tools, as well as the implementation of Ephemeral Bits. We will also talk about the issues we faced when developing the installation, and improvements that we would like to achieve in future work.

RELATED WORK
Sound has always been a fascinating component in interaction artworks. And when combining it with visual element, it produces a dynamic and engaging work of art. In Messa di Voce [2], Golan Levin et al used sound from vocalists and visualized it with interesting graphic effects. Ana Rodrigues [1] visualized sound with a swarm of fireflies. Each firefly got its energy from sound and moved accordingly in their environment. A group of French artists also explored the connection between audio and video through the intricate sound art installation Murmur: Talking to the wall [9]. And Voice Array [10] is another pieces by Rafael Lozano-Hemmer that demonstrated the aesthetic aspect of the incorporating sound in interactive art installation.

On the other hand, with the advent of the Microsoft Kinect camera, an increasing number of visual, as well as recording artists have utilized the beauty of point cloud in their works. Among the most successful examples, we can count the work of Aaron Koblin in Radio Head's music video House of Cards [11], or the video Catalina from Moullinex [12]. Daniel Franke and Cedric Kiefer creatively used 3 Kinect cameras to create a mesmerizing experience in Unnamed Soundsculpture [13].

Taking inspiration from these amazing artworks, we created the Ephemeral Bits installation that combined the interactivity of sound and visual aspect of Kinect's point cloud. Furthermore, we also included the voice recognition engine that made the installation react according to the feeling the participant express in the sentence "I feel …". Ephemeral Bits is a unique artwork in a sense that it does not only help the participants to experience with sound and image, it also gives them the chance to express their feeling and immerse in their interactions with the installation, which has not been done before.

TOOLS
The main part of the project is written in the Processing [4] IDE. It handles the sound and depth data input from the Kinect camera. The voice recognition engine Sphinx is written in Java. Data transfer between the voice recognition engine and the Processing sketch is implemented with Open Sound Control [5, 6] libraries for Java and Processing. Processing libraries for handling Kinect data [7] and sound [8] are also used in this project.

IMPLEMENTATION
Inside the Processing sketch, there are thousands of particles whose targets are the location of points acquired from the Kinect camera. They also have fuel tanks that get refueled with the energy from the surrounding sound. With fuel, the particles move toward their respective targets to form the point cloud shape of the participant standing in front of the Kinect camera. When running out of fuel, they move away from their targets, and fall to the floor.

In the version with voice recognition, the Sphinx engine is run separately and its recognition result is sent to the Processing sketch using OSC. The text data is then visualized onto the mesh of the point cloud. The feeling is categorized into positive emotion and negative one. When positive feeling is spoken to the installation, the mesh turns green, and when negative feeling is detected, it turns to red.

In the live audio-visual performance, each particle gets its energy from surrounding sound, and moves toward its target. When they are in position, and get more energy from the sound, they grow bigger and change their colors to create a dynamic visual effect.

Figure 1. Voice recognition version: Point cloud mesh with participant's feeling.

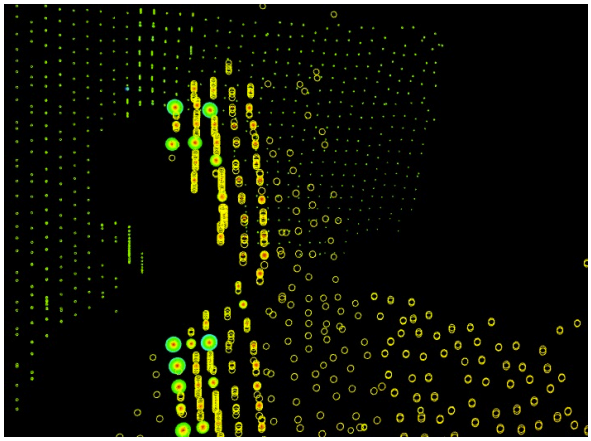

Figure 2. Audio-visual performance: Point cloud particles change their appearance according to surrounding sound.

DISCUSSION

For the semester-end exhibition, we have achieved our goal of presenting a workable version of our project. We have also succeeded in creating an original and engaging experience for our participants to explore both the audio-visual feature as well as the voice recognition one.

Our decision to use mesh instead of point cloud in the voice recognition version, with "feeling" word superimposed is based on aesthetic and design perspective. The mesh helps to create a minimalist composition and therefore do not interfere with the superimposed words.

On the other hand, we chose to use circle in the audio-visual performance based on the metaphor that we treat each point cloud particle as an agent that get energy from the environment and grow in size as well as change color depending on their current fuel levels.

We realize that in addition to audio-visual art installation, we can also use Ephemeral Bits in music video, dance performance video thanks to its dynamic sound interaction feature.

During the course of the project, we faced a number of issues that we had to make compromise:

1. Due to the lack of support for Kinect in Javascript, we had to turn to the Processing programming language, which has more support for the Kinect camera.
2. When we tried to increase the number of particles, or use sphere instead of circle, the program slowed down dramatically.
3. The voice recognition engine Sphinx has a low level of accuracy, which makes it unworkable in a noisy environment. And we had to limit the vocabulary the participant could talk to the installation to raise the accuracy.

In future work, we would try to bring the project into Javascript, which we can utilize the Google Web Speech Recognition to achieve higher accuracy in voice recognition. Also we realize that, with only one Kinect camera, the resulting visual looks flat and unattracting. In future development, we will try to use more Kinect cameras in order to get a better 3D effect.

CONCLUSION

In this project, we have successfully combined three interesting areas: audio-visual interaction, point cloud and voice recognition to create an immersive interaction installation.

However, there are still issues that we need to address in future version to make the work more engaging to visitors.

ACKNOWLEDGEMENT

I would like to thanks Professor Forbes for his indispensible advices during the course of the project. And many thanks to my classmates in the Computer Graphics 2 course for giving invaluable feedbacks about the artwork.

REFERENCES

[1] Rodrigues, Ana, et al. "Sound Visualization Through a Swarm of Fireflies."Progress in Artificial Intelligence. Springer International Publishing, 2015. 664-670.

[2] Levin, Golan, and Zachary Lieberman. "In-situ speech visualization in real-time interactive installation and performance." NPAR. Vol. 4. 2004.

[3] Carnegie Mellon University Sphinx Voice Recognition library: http://cmusphinx.sourceforge.net/

[4] Processing: https://processing.org/

[5] Java OSC library: https://wush.net/trac/lsu-idt/wiki/JavaOSC

[6] Processing OSC library: http://www.sojamo.de/libraries/oscp5/

[7] Processing Kinect library: http://shiffman.net/p5/kinect/

[8] Processing Sound library:
http://code.compartmental.net/tools/minim/

[9] Murmur installation:
http://www.everydaylistening.com/articles/2013/6/16/murmur.html

[10] Voice Array installation: http://www.lozano-hemmer.com/voice_array.php

[11] Radiohead's House of Cards music video:
https://www.youtube.com/watch?v=8nTFjVm9sTQ

[12] Moulinex's Catalina music video:
https://vimeo.com/19723907

[13] D. Franke and C. Kiefer's Unnamed Soundsculpture:
https://vimeo.com/38840688