Team 9, Project Plan

Hearing colours in space

Harpo 't Hart & Philipe Louchtch
October 25, 2012

Concept summary

We want to combine the sensory capabilities of the Kinect with synthesising sound in real-time. And thus we've come up with the idea to research the translation of spatial and visual data (both of which are perceivable with the eyes) into sound in a natural way.

The concept is to create a system that allows people to perceive colours and localize them in space through sound. The project will focus on translating the space within the theoretical viewing space into sound with the finished product acting as the feedback device, essentially turning the room and its objects, a canvas or other (viewable) things into an interactive space. You can think of this system as a kind of augmented reality that makes use of sound instead of vision. We want to use a Kinect as an input source and translate this signal into sound for the user to perceive through a headphone. By attributing certain sounds to certain colours people should be able to differentiate between colours.

The product-to-be should be viewed as a kind of musical instrument that turns the space around the user into a computer interface that allows the user to create soundscapes. The interaction between user and world isn't limited by a single way of operation or use, just like a guitar isn't limited to plucking or a computer mouse to clicking on hyperlinks. Likewise, the abstract product-to-be is also not made to suit a single purpose function and leaves the interaction possibilities to the users. Although it should be noted that it is not an orthodox musical instrument; the user does not play it, he merely can use it to generate music or sound from his environment.

Few examples would be an artist drawing on a large canvas guided by the aural feedback, or placing the device facing the stage during a live performance.

We feel that this is a sufficiently unexplored field that gives us the possibility to do something fresh while leveraging interesting and new technologies such as the Kinect + .NET framework and the SuperCollider (sclang) real-time sound synthesiser.

Research question

The research question is mainly "How to sonify visual input data such as colour and a subset of its spatial position in a most natural way". By 'natural' we mean in such a way that it is easily understandable, logical or otherwise easy to get used to. This means that the final result should, ideally, be comfortably sounding to the human ear.

The results can hopefully be used for better sound-feedback, either by a continuation of this project or other systems dealing with sound-feedback from visual input, for example a Kinect-based game.

Description of interface design

Because we shall be dealing with translating visual data into sound, relying on a user interface for interaction seems very counter-productive to the project. Thus the intention is to rely on auditory in- and output. The Kinect has an excellent solution: speech recognition, which is also included in the "Kinect for Windows SDK". So the idea is to provide voice-commands to enable, disable or modify functionalities of the software, but a GUI for setting up the initial desired functionality is not completely excluded from the plan.

Functionality

The functionalities we are aiming at could be divided into two groups: ones that are related to the way in which the visual information is translated into sound and others that are related to the display of the sound itself.

These are some possible functionalities and or modes of operation that could be accessed through voice-commands:

Translation

The user can:

- Change the resolution of the colours the user wants to hear.
- Detect only one colour
- Detect only colour
- Detect colour + the distances
- Detect only distances (sonar-like)
- Zoom in

Sound

The user can change the sound to some extend and therefore he can adjust parameters such as:

- Volume (i.e.: “Volume Up”)
- Filters (i.e.: “Filters next”)

Note: Each voice command should start with a special “Voice command start” keyword.
(i.e.: “Xbox Bing <search query>”as used by Microsoft Xbox for voice searches via Bing on the Xbox, where both Xbox and Bing are keywords)

Most of the interaction with the sound will be done via the Kinect that detects the colours in space. The user can thus change the sound by looking at different coloured objects, or arranging coloured objects in such a way in a space that create interesting sounds or even by simply traversing the space.

We're not quite sure about the synthesis method to use, but at the moment we're thinking about granular synthesis. In this case the user might control parameters like grain density, number of grains and pitch or pitch deviation of the grains. This will depend on the user evaluations if people using the device feel the need to have more control over the sound than when they would only look at colours.

Low-profile function design

The final product should be able to map slices of its visual field into sound.
The slices were originally the image cut in 3, 5 or 7 vertical pieces, but on second though this seemed like a bad idea. So instead there will be 3 to 7 (the best amount is to be determined, too little vs. too much data) fixed dimension "target boxes" from which the average colour is to be calculated. These boxes should have such a size that would provide precise data. Another possibility that we are willing to try if time constraints allow for it, is the cutting the image into distance slices first. By that we mean that the image is cut spread into distance layers detected by the grayscale/Ir camera. Then the layer with the biggest presence in a "target box" is selected and has its average colour calculated. This way the average colour is truer to the observed surface.

The average colours are then mixed in a stereo soundscape according to the position of the "target box" in the image, so the first one would be 100~80% Left and the last 80~100% Right. The data extracted from the "target boxes" will differ according to the selected functionality such as, but not limited to: colours and/or distance. More functionality is still under consideration and might be implemented if the time constraints allow for it. Furthermore, the product should support speech-recognition to process commands by the user.

The Kinect-sensor is capable of 640x480 @30FPS RGB output, so the goal is to try and match that frame-rate or half of it in a worst case scenario. The idea is to calculate the average colours of the "target boxes", possibly round/clamp them to certain colours, send them to the SC-client and replace the "old" (previous frame) sounds with the new colour ones and repeat the whole process again for each frame. Hopefully, the latency will be tolerable.

The prototype shall consist of a Kinect-sensor taped to an over-ear headphone (AKG K240 MKII, most likely) and connected to a laptop, this way the positioning and orientation of the Kinect is the same as the head of the user.

We will leverage the .NET Framework 4.5 and the SCLang platforms during development on top of Microsoft Windows along with a Microsoft Kinect peripheral.

User group aimed at

The user group we like to address are creative people, especially musicians and visual artists. Basically, the people who are interested in new ways of mixing sounds with colours. They share our enthusiasm about the topic and they're willing to explore new technologies, it might give them new insights in their own respective fields and it can offer them a tool to make new sounds and use colours in an unconventional way. These people might not be the most experienced computer-users, but they're eager to learn and open to new experiences. We do have good access to test subjects from this user-group. Since Harpo is a musician himself he knows a lot of others who might be interested in such a project. Also we know quite some students at the KABK and the Rietveld Academy that we can contact.

Specification of the user-group

The users will be in the age group between 20 and 30 years old, most of them being mainly art and music students and people that just finished art-school. They could use the product for performances or maybe for themselves to listen to the sounds , experiment or otherwise just have fun with it. But the primary goal we have in mind is it to be a kind of musical device with which you can use to perform or to accompany their art. Just as with a musical instrument the users will not only use it at their performances but also at home or in a studio to rehearse.

Most art students speak English in school so they'll be able to speak it fairly well. The reasons for a user to use the system is an artistic reason they might be looking for new sounds, new ways to make sound, or maybe a painter might want to make a painting that sounds a certain way. These users must have some basic knowledge of computers, but more importantly they must be willing to learn something about sound synthesis so they can create some nice sounds with the system.

We choose this group as a target group for two reasons: The first one is quite obvious: if you make a musical instrument or device for performing music the users will be musicians. The second is more personal and comes from our interest in sound and new interactions and this group might share this interest. We are thinking of testing this on ten test-subjects.

Innovations aimed at

Our aim at innovations is in three ways. The first one is artistic, the second technical and the third is in the field of Augmented Reality. From an artistic point of view it would be very interesting if we find new ways of working with colours and sounds, giving musicians and visual artists a new tool to work with in the process. The technical innovation would be the spatial representation of the colours. In the field of augmented reality it would be quite new to use sounds instead of visual information. A somewhat similar device has already been made by Neil Harbisson and goes by the name of "Eyeborg". Through this device is only able to perceive one colour at a time. Our goal is to make a product that can convey information about multiple colours in an image and their position.

Time table

Week	Goal
26 Sep	Concept Summary *
5 Oct	Project Plan *
22 Oct	*Working OSC C# to SuperCollider message passing* *
26 Oct	Project Plan – Revised *
2 Nov	*C#: Averaging colours of from the RGB output of the Kinect + slicing logic* *SC: appointing sounds to colours.*
5 Nov	Paper Design
9 Nov	*C#: Distance tracking* *SC: Spatialization of sounds.*
16 Nov	*C#: Voice Recognition of commands + GUI window for initial configuration* *SC: Making sound parameters interactive*
Week 47	User Evaluation 1
26 Nov	15 min Presentation
Week 49	User Evaluation 2
21 Dec	Concept Project Paper
3 Jan	Final Project Paper
6 Jan	Final Product (Prototype)
9-11 Jan	Final Presentation to course Admin.

* - Indicating the completion of the item