Colours in Space

Harpo 't Hart[1], Philipe Louchtch[1]

[1]LIACS, Leiden University, Niels Bohrweg 1, Leiden, the Netherlands
harpo_t_hart@hotmail.com, coolicerz@gmail.com

Supervised by Hanna Schraffenberger, HCI Course Administration.
hanna@schraffenberger.de & http://hci.liacs.nl

Abstract. How would visual information such as colour in a given 2D (screen) or 3D (world) space be translated into sound which would enhance and argument the visual perception of a person? A program for the Microsoft Kinect has been written as means to test different translation schemes and their granularity. [*Preliminary*] results show that this varies from person to person. Both the resolution of the visual to aural mapping and its tonality seem to be the most important factors.

1 Introduction

The mapping, or translating, visual information such as colour in real-time doesn't seem as an area where a lot of work is being done. We feel that some interesting research can be done in this area. Our work aims to help other applications such as games or natural user interfaces (NUI) augment their visual component with sound, allowing them de-emphasize the visual elements by providing easy to understand aural feedback. Another possible area of direct application would be art or music, where art can be made with a aural component in mind or allowing the world around the user to generate new sounds.

This paper will fist provide some further background on the problem before explaining our approach followed by the results and conclusion.

2 Mapping visual input to aural output

Problem scope

How can colour from a 2D image be mapped to an aural output in such a way that is easy to understand or otherwise the most logical to a user? There seems to be fairly little work done on this, safe for devices or software to assist colour-blind people.

Just like the addition of natural speech recognition or otherwise just normal voice commands can greatly enhance the usability of an application, by complementing the visual output with aural feedback we feel that a user interface experience can be greatly improved.

Typically, an interface with a continuous stream of visual information, such as a continuous video feed or a game, requires the user to pay great attention to the screen. If such an interface is augmented with aural feedback, a user would be able to pay less attention and still be able to notice when something important can be happening.

A great and already applied example of this would be a sudden change of BGM of a video game just before a boss battle that prompts the user to be alert.

Focus and User Analysis

Our intention is not to immediately provide a solution to this. Instead, we choose to focus on creativity initially. This gives us room to experiment with the approaches and possible other applications of this and future research.

By focusing on creativity allows our user group to be creative people. Their experience with visual, audio and even audio-visual areas should provide us with helpful feedback and suggestions.

We define our "creative" user group to be mainly art students, this group is further subdivided into visual artists and aural artists (performers). It is interesting to note the somewhat diverging feedback from both groups.

3 Colours in Space

To be able to experiment we first have written a piece of software that interfaces with a Microsoft Kinect and synthesizes sound from its input in real-time with SuperCollider.

A similar device already exists, it is used by Neil "Eyeborg" Harbisson (todo: add reference to some page about him). However this device is limited to a very basic tone and a single colour at a time. In our approach we choose the Kinect due to its easy API, development tools and other features like depth tracking and good microphone array for speech recognition.

By building our solution with a Kinect we have a lot of freedom concerning extra functionality. Because our user group is creative people, we don't expect them to be very skilled with technology, but we also realize that in order to make the device usable to them, they should be able to easily express themselves with the device so good configuration should be made possible.

Our user interface for the device is twofold, the first one is a Graphical User Interface that allows for overview of settings and a quick and easy way of setting the initial parameters. The other one is a voice controlled interface, we feel that during the operation of the device, the user should be required to walk back to the computer to change something. He or she should be able to change some single parameter easily wherever he is.

We shall use the “sonochromatic” scale (todo: reference to the page about it) devised by Neil Harbisson to map the colour to a sound. It is important to evaluate this scale with other tonalities.

4 Results and Evaluation

The first user evaluation was conducted with a basic functioning version of the device. The basic functionality implies simple mapping of parts of the current frame to a sound. This is accomplished by taking a piece of the frame, calculating the (literally) average colour and using the Harbisson’s sonochromatic scale to generate the corresponding sound with the sound placed at the relative position of the image (ie. Leftmost part will be heard left and rightmost right).

The evalutation focused on both the configuration graphical user interface prototype and the mapping of the visual data to sound, the tonality and granularity. Our first evaluation group consisted of a visual artist and a musician. Whilst our device is able to generate three to seven (in steps of two) mappings per frame, the visual artist felt that anything more than three mappings sounded too chaotic to distinguish. This contrasted with the feedback of the musician, who felt that more mappings didn’t make it more chaotic and more musically rich.

The feedback was somewhat within our projected estimations, if a user is mentally mapping the aural feedback to what he sees or expects to see, a large amount of sounds will be confusing and chaotic. While a large number is more harmonically rich when used to as way to produce sounds from the environment for musical purposes.

The feedback concerning the configuration graphical user interface was mostly positive due to its intuitiveness and the amount of parameters presented to the user. However, they felt that the some parameters should be less implicitly defined by adding more labels. Due to not being implemented yet, the voice commands option has been mentioned as a future feature which was met with a lot of positivity.

Second evalutation yet to be done.

Our single largest limitation is the processing power available to us. The laptop that is used to run the software on doesn’t have a lot of computing power compared to a desktop, especially due to its weak graphics card due to it missing GPGPU features.

Because we are doing computations on images, it would have been nice to have some of the most expensive computations run on the graphics processing hardware since we have to process up to 30 images per second in the .NET language.

5 Conclusion and Discussion

As of this writing, we feel that we failed to provide a natural sounding mapping of colours from a video stream into sound. There are three possible causes for this, first of all due to the fairly mediocre quality of the RBG camera video stream from the Kinect to the computer and basic color averaging algorithm, the computed colours feel a bit off and furthermore, not very stable.

The second one being the granularity of the possible colours that we can detect, twelve in total, one colour per semitone. By decreasing this amount and tweaking the scales used we could achieve more comfortable sounding mappings.

Third reason would be the monotonous sound generated, by having different types of synths, samples or instruments mapped to a colour as well as pitch, we could achieve better separation of sounds.

So these would be the most rewarding things to fix in a future version of the software.

6 Future work

Nothing yet.

References

<To be added in final version>