Hearing Colours

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, Fons Verbeek.
hanna@schraffenberger.de & fverbeek@liacs.nl
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. The results show that the way a person interprets sound varies but still have a lot things in common. Most importantly, different applications of sonification have varying constraints and requirements and thus need different approaches. The resolution of the visual to aural mapping, its tonality and the relations between the tonalities of colours seem to be the most important factors.

1 Introduction

We now live in a world where we are bombarded with visual information, mostly by our gadgets and computers. Large advances have been made in displaying information as well as the means to display it with. Larger, faster, more brilliant and even capable of showing 3D content displays are being mass produced and viewed by us daily. While still within the limits of our visual perception, processing that viewed information is becoming harder and harder, not to mention the burden of designers to layout more information in more efficient means. The question that comes to mind is, “Can we offload some of that visual information to another sense?”

What if we could translate some of that visual information into sound? Enter the world of sonification. The field of sonification is “the use of non-speech audio to convey information or perceptualize data” [1], and while it has been around for some time, surprisingly little developments have been made. Neither is there a tool or framework in place for the sonification of data [2]. To us, this opens up a lot of possibilities for experimenting (and failing), being creative and creating something new or unheard of (pun intended).

Being fans of music in its whole and involved with it (one more than the other), a field that allows for large amount of creativity and the usage of cutting-edge technology certainly is appealing, more so if it deals with audio. So we stepped up to the plate for the challenge.

While there certainly has been work done on sonification, 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 as opposed to specialist uses for sonification. Specialist uses being instruments whose output is a single aural mapping of its input(s).

An example that has inspired us and thus closest to our approach in this paper, is Neil Harbisson's Cyborg Eye [3], a device that allows him to hear colours, one at a time and given that Neil is fully colour blind, allows him to recognize colours.

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 (visual) art can be made with an aural component in mind or allowing the world around the user to generate new, inspiring 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 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 background music 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 for real-time mapping of various visual components to sound. Instead, we choose to focus on creativity, by building an initial platform to see how people react to sonification. 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 Hearing Colours

To allow ourselves to familiarize with the field of sonification, realize what our possibilities are as well as to simply experiment, we have chosen to build a device that maps the visual field of the user into sound.

This is achieved as follows, the device consists of a Kinect mounted on top of an over-ear headphone facing forward to capture the field of vision of its wearer and thus user. The visual data, frame, is mapped to sound by taking selections of the data, referred to as “Boxes”, “Targets” and “Target Boxes”, in a horizontal progression. Per box, the pixels contained within are averaged and clamped to one of the twelve colours described by the “sonochromatic scale” [4]. The relative position of the box from left to right in the frame to other boxes is mapped to a position in the stereo field, with the left-most box being 100% Left, the middle 0% L/R and right-most 100% right.

As already mentioned in the introduction, a similar device already exists, it is used by Neil “Eyeborg” Harbisson. However this device is limited to a very basic tone and a single colour at a time. In our approach we choose the Kinect as a camera 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.

The Kinect detects twelve different colours. These colours are all being mapped to a musical pitch in a chromatic scale with the range of one octave. The mappings are done in two ways. The first one is a mapping according to the 'sonochromatic scale' by

Neil Harbisson [4]. This is a scale that maps colours in spectral order of lightness to pitch in chromatic order so the colours with the shortest wavelength (red) will sound like the lowest note. The tonal scaling goes chromatically, so a colour that is close in wavelength to another will sound like a minor second.

The above mentioned mapping can give some very dissonant sounds when looking at objects that have the same kind of colours. Therefore another mapping was used. This is the mapping that Alexander Scriabin used for his "*Clavier a lumiere*" [5]. This mapping uses a harmonic relationship between colours that are close to each other. This mapping gives more harmonic sounding effects.

4 Materials & Methods

As mentioned earlier, we make use of the Microsoft Kinect (for Xbox) as a camera for its rich API and feature set. Using the Microsoft Kinect for Windows SDK implies the use of Windows as operating system. In our case the version of Windows used is, Windows 8 Professional due to it being the most recent release of Windows.

The device is connected to a Dell Studio 1558 laptop, from 2010, with a modest CPU (Intel i5 M520, dual-core with four logical threads) and a very low-end embedded GPU, Radeon 5450.

The CPU allows us to spread out some computationally intensive parts of the software of the device over multiple threads to ensure smooth operation. However due to the lackluster feature set of the budget GPU, we are unable to offload (and implement efficiently) image processing to it.

Furthermore, the software stack consist of two platforms, the .NET Framework 4.5 (in our case, the C# language) for the UI and processing, and SuperCollider for real-time synthesis of audio.

For the development of the C# backend and UI components, Visual Studio 2012 Ultimate Edition was used.

5 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 evaluation focused on both the configuration graphical user interface prototype and the mapping of the visual data to sound, the tonality and granularity of the

soundscape. Only three and five “Target Boxes” were tested. 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 five mappings sounded too chaotic to distinguish from each other. This contrasted with the feedback of the musician, who felt that more mappings didn’t make the soundscape sound more chaotic, merely containing “more tones”.

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 and was met with a lot of positivity.

It is worth mentioning that an interesting feature request has been made. In order for the device to be usable in a musical setting, e.g. a musical instrument; there must a mechanism in place to toggle the sound of each “target box”. Whilst it is easy to implement this in a non-obtrusive way by implementing this as voice commands, it is not logical to do so due to usability constraints, namely the switching should very fast and straightforward. The best way to implement this functionality would be by using an off-the-shelf USB Numpad and mapping the keys to toggle a “target box”.

The second evaluation was more about usability and ease of use of the interface. It became clear that the graphical user interface was for the most part quite easy to work with. The part that caused more concern were the voice commands. The microphone mounted on the Kinect didn't pick up the voices from the users so well, because the microphones weren't positioned right for this purpose. In this evaluation another colour to sound mapping was added. This was done according to the technique of Scriabin’s “*Clavier a lumiere”*. This was a far more pleasant sound for people to listen to than the sound in the previous test session. This provided a soundscape that made it for the test persons more relaxed to walk around and explore the room with the device.

Surprisingly the Sonochromatic scale with its less pleasant sound was perceived as a more clear representation of the colours in the room. The test persons found it pleasant to have the option to switch between the different sounds. On the one hand to give their ears some rest from the more dissonant sonochromatic sound and on the other hand to be able to listen to the room in different ways as to look at an object from different angles. So after all this possibilities to change could provide some clarity in translating sound to colour.

6 Conclusion and Discussion

As of this writing we feel that we have gone in to some very interesting subjects and gained insight in ways of mapping sound to colour, but we strongly feel that we didn't come up with a satisfying colour to sound mapping. 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 colour 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 and probably most importantly would be the sound. The different types of mappings we have tried showed clearly that there are better and worse mappings for this, so there should be a best option we haven't found. The 'sonochromatic scale' provided quite harsh sounds, but was relatively clear in comparison with the Scriabin mapping.

This research opened the way to a lot more questions. As said in the introduction, there hasn't been much previous work in the field of mapping colour to sound in real-time. We have touched on this field and set some first steps in it, but there is a lot more work to be done.

7 Outlook

A lot of valuable insight has been gained from this project. Many of the issues can be attempted to be solved with better technology such as a good quality (web) cam or even a next generation version of the Kinect, once it is out. As well as choosing a different target machine to increase the available computing power, especially on the GPGPU (the use of a graphics processing unit for general-purpose computations) [6] side.

Having more computing power would allow us to utilize better ways to average colours and be able to do some post processing on each frame before feeding it to the averaging algorithm(s).

The inclusion of depth data was planned for this device and still is, but we have decided to put more work into the colour mapping first. This is due to the fact that depth data is useful when used in large environments, then with the depth data, only the centermost object of each selection will have its colours averaged thus resulting in less contamination of the colours of an object and its (distant or close) surroundings.

Alongside the technical issues, a lot of further research can be done on the amount, selection and tonality of colours. Meaning that we will have to address questions such as:

How many and which colours specifically do we recognize?

Are there colours that we can ignore?
How do we clamp/round off a colour to a supported value?
Which tonality does each colour get?
What musical scales can we use to describe all the defined colours?

And possibly, most of these questions cannot be answered in the global case. So the questions must be answered in a case-by-case basis. An artist might expect a different tonality of colours than a musician.

The USB Numpad “target box” toggling is fairly easy to implement and could be added in a future version as well.

References

[1] "Wikipedia, Sonification," [Online].
Available: http://en.wikipedia.org/wiki/Sonification.

[2] J. Flowers, "Thirteen years of reflection on auditory graphing: Promises, pitfalls, and potential new directions.," Eoin, Brazil, 2005.

[3] *TED Talks, Neil Harbisson: I Listen to color.* [Film].

[4] "Wikipedia, Neil Harbisson: Sonochromatism," [Online].
Available: http://en.wikipedia.org/wiki/Neil_Harbisson#Sonochromatism.
[Accessed 6 January 2013].

[5] "Wikipedia,
Alexander Skrjabin (Александр Скрябин): Infuence of colour," [Online].
Available: http://en.wikipedia.org/wiki/Skrjabin#Influence_of_colour.
[Accessed 6 January 2013].

[6] "Wikipedia, GPGPU," [Online].
Available: http://en.wikipedia.org/wiki/GPGPU.
[Accessed 6 January 2013].