Ricchiuti Mauro

Audio/Visual Generator

3 Types of Synesthesia

The focus of my generator is an attempt to visualize and recreate 3 different types of synesthesia: chromesthesia, pitch-size and pitch-location synesthesia.

Chromesthesia is the synesthesia which maps pitch or notes to colors. Multiple studies have been done on this topic, and the results of the notes-color mapping are slightly different. However, most of the results match among the different studies, meaning that most of the synesthetes do experience the same colors when exposed to the same notes. In my A/V generator, I followed Scriabin's model [1,2], which maps notes and colors as in figure 1.

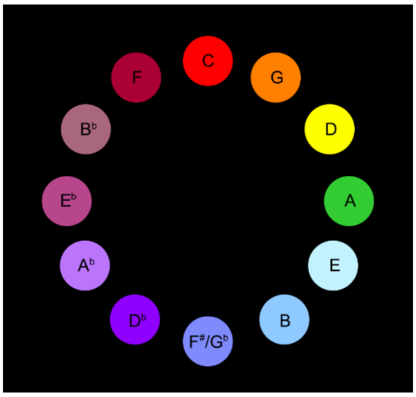

Figure 1. Scriabin's sound-to-color circle of fifths

Other 2 types of synesthesia which I have tried to visualize are the pitch-size and pitch-location (or pitch-altitude) [3]. Both phenomena are studied in Fernay's research [3]. In the said study has been found that these kind of synesthetes, experience a visual mapping of the type pitch-size and/or pitch-altitude, attributing high vertical location and/or smaller size to higher pitch sounds, and low altitude and/or bigger size to low-pitch sounds. In my visualization so, I have tried to reproduce this synesthesia scaling and/or translating the generated objects.

The Generator

For an explanation of the technical characteristics of the generator refer to the README file. Briefly, the audio for the model is generated through a Markov Chain probability model, which is learnt in VCVRack. The model is then triggered from a pseudo-random clock, and sends back notes according to the probability distribution. Since the inspiring file was made out of 3 MIDI channel, the model learns 3 different probability distributions. The system is thought to be used in real-time, but for simplicity, I am going to describe here the video-recording called Final.mp4 which I am handing in. The file I used as inspiring set is the song Everything you do is a Balloon from Boards of Canada. The MIDI file is composed of 3 channels and can be downloaded at the MIDM website (http://www.midm-database.co.uk/hiscores.html). The video looks like the screenshot in figure 2. And I will refer to the quadrants as top-left, top-right, bottom-left and bottom-right.

The visual generator, is receiving 3 MIDI channels containing the note which is currently played (it is always note-ON), the VCA output value and the corresponding audio channel. Additionally, the visual generator is receiving the Audio output from the mixer, but this is mainly needed for recording the video in sync with the audio. All the quadrants visualize 3 different shapes, and that is because I am visualizing 3 MIDI channels. The colors of the shapes are obtained from a table which maps notes-to-colors following the Scriabin's model [1] as shown in figure 1. Apart from the 4 models I am going to describe in the next lines, there is also a model which simulates chromesthesia with fluids, which is possible to be seen in the code, but which I am not reporting here.

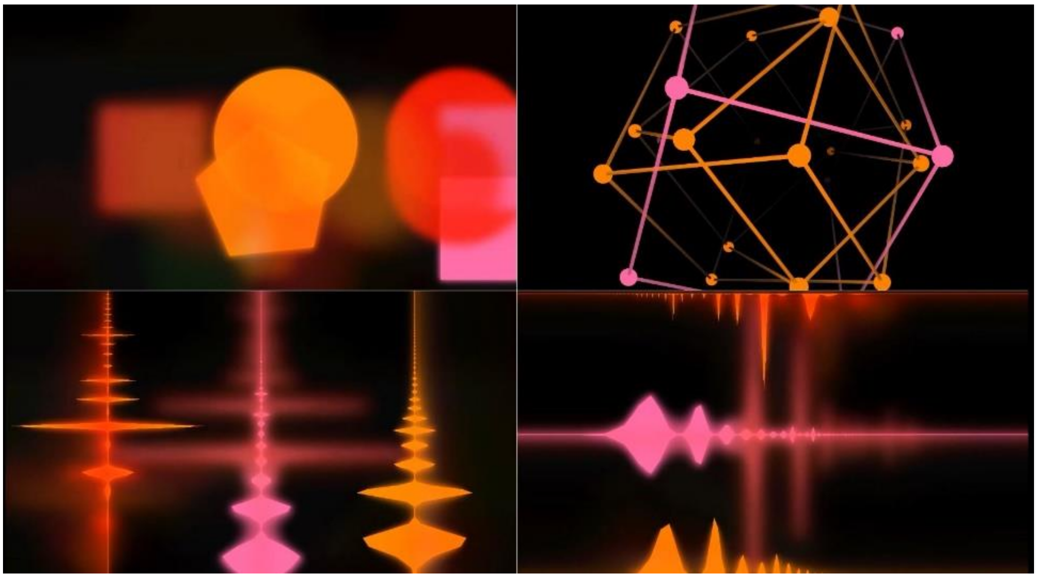

Figure 2. Screenshot of the recorded video

In the top-left quadrant, each channel is visualized with a different shape (circle, square and pentagon). The x location of the shapes is obtained as a random value, while the y location is obtained computing the FFT (Fast-Fourier Transform) on the Audio channel, taking the frequency of the highest peak in the FFT and remapping the value to a (0,1) range as for the v coordinate in uv coordinates. In this way, if the pitch is higher, the note would be drawn in a higher vertical position and so on. The scale of the shapes is indeed controlled from the output of the VCA of the channel that they are visualizing. This is done to try to better match the visual appearance with the audio. Feedback and blur are used to improve aesthetic. This visualization so, tries to visualize the pitch-to-color synesthesia (chromesthesia) and the pitch-to-altitude synesthesia.

In the top-right quadrant, sound is visualized as 3 rotating cubes, visualizing 1 MIDI channel each. The color of each cube changes according to the note they are visualizing, as in the other visualizations. The scale of the cubes however, is controlled by the pitch of the note they are visualizing. Using again an FFT analysis, I obtained the highest harmonic frequency in the spectrum, and I used this value, properly remapped, to scale the cubes. In this way, higher pitch frequencies, would scale down the cube, decreasing the size, while lowest pitch values would make the cubes bigger. This model so, is trying to visualize chromesthesia and pitch-to-size synesthesia.

In the bottom-right quadrant, the visualization is directly the output of the Audio spectrum for the 3 channels. Visualizing directly the FFT analysis, the sound-visual mapping is quite straightforward and does not need lots of explanations. I report this visual, mostly because I did like the visual appearance. Also, in the FFT analysis, lowest-pitch frequencies are already larger, so in a way this model is visualizing both pitch-size and pitch-to-color synesthesia.

Finally, the model which I think better visualizes the 3 types of synesthesia, even not being the best-looking one, is the one in the bottom-left quadrant. This model shows mirrored versions of the audio spectrums from the FFT analysis of the 3 audio channels. These graphs are colored with the color from the Scriabin's [1] maps note-color as for the other models. However, the graphs are oriented in vertical, because in this way it is natural that the low frequencies are located on the bottom part of the screen and the high frequencies on the top part of it, as described in pitch-to-altitude synesthesia. Also, the FFT naturally draws lower frequency with a larger shape and high-frequencies ones with very narrow shapes, resembling in this way what is described as pitch-to-size synesthesia [3]. This model so, is the most complete among the other ones, trying to visualize all the 3 types of synesthesia described.

The idea for this system is to run in real-time after being trained on a MIDI inspiring song. The user is invited to play with the Topograph's knobs and with the Basal oscillator's knob. Being the audio generated inside a modular synth, the possibilities of sound generation are infinite, and the system seems to have good potential for live audio/visual performances. However, as you will notice in the Final.mp4 file, it has some problems when it comes to recording. After the minute 3, you can hear some clicks and other noises. These noises are not present when running real-time, and the system runs smoothly at 60fps when not recording. VCVRack is very computationally intensive, since the appearance is quite heavy on the GPU while the audio is heavy on CPU. Same is valid for TD. Even if the visual algorithms are quite performative by themselves, it loses some performances when recording. I guess this is due to an high rate of IN/OUT data, as well as 3 different FFT computations and mainly the recording itself, which, is recording at a resolution of 2560x1440 and is also recording audio. The user so, is invited to play with it in real-time without recording, enjoying the best out of it. Additionally, if the system is run on a laptop, it is needed to run it with the power plug plugged in, or VCVRack would not render the modules.

References

[1]Galeyev, B. M., & Vanechkina, I. L. (2001). Was Scriabin a synesthete?. *Leonardo*, *34*(4), 357-361.

[2] https://en.wikipedia.org/wiki/Chromesthesia#cite_note-:6-26

[3]Fernay, L., Reby, D., & Ward, J. (2012). Visualized voices: a case study of audio-visual synesthesia. *Neurocase*, *18*(1), 50-56.