H. Reiterer & O. Deussen (Hrsg.): Workshopband Mensch & Computer 2012
München: Oldenbourg Verlag, 2012, S. 319-322

Navigating audio-visual Grainspace

Max Neupert

Faculty of Media, Bauhaus-Universität Weimar

Abstract

Building upon the timbreID library for Pure Data, which is an audio feature analysis and classification tool with three-dimensional grading capabilities, I created an audio-visual instrument. Video frames can be cued to according to the auditory properties of the audio track by moving the hand inside a virtual cloud of snippets. This was achieved by combining three pre-existing things: 1. concatenative synthesis with a three-dimensional plot of the snippets 2. A Kinect sensor as an interface, and 3. Video playback along the audio.

1 Context

In a greater research project about video imagery used as sampled material in a musical context, I came across corpus-based concatenative synthesis. Concatenative synthesis is similar to granular synthesis, as it divides a longer sound into short snippets. In concatenative synthesis they are called *units,* in granular synthesis *grains*. Typically the length of the snippets in concatenative synthesis is longer (~ 10ms to 1sec) than in granular synthesis (~ 1 to 50ms). The actual difference of the two methods is, that in *concatenative synthesis* the units are graded according to their sonic properties and those vectors may be mapped on a two- or more dimensional space. Also the units may be of non-uniform size depending on the result of the analysis (Schwarz, 2006-3). *Granular synthesis* in contrast, only knows one grading dimension, which is the index of the grain (its temporal position in the whole sample). Granular synthesis offers rich sound experiences especially when a random jitter of grain size and position is applied. The downside is, that the sound is hard to control and reproduce, making it a difficult live instrument. This is where concatenative synthesis shines: having a cloud of units in a two- or three-dimensional space gives not only a visual impression of the sample's sonic characteristics, it also allows access to different sounds at specific coordinates. This allows for gestures, as we know them from real instruments (Schwarz, 2012).

2 My approach

I was dissatisfied with the ways to interact with the grains in a three-dimensional space and I wanted the video frames to be displayed along with the unit playback. When exploring the three-dimensional plots of units generated by the concatenative synthesis framework, I felt the need for a more suitable interface to the space than mouse, keyboard, track pad or multi-touch surfaces. I wanted to examine how different spatial categorisations of the sample would sound and feel when navigating through the 3D plots without the restriction of a two-dimensional input device. The novelty in my approach is the non-haptic interface, which utilizes the Kinect sensor to navigate in three axes and the synchronized video image display when the program is fed with audio-visual material.

3 Description

The clouds of grains are generated by an audio feature analysis tool called timbreID, conceived by William Brent. Detected sound features may be mapped along the X,Y and Z axes. Available features are cepstrum~, magSpec~, specBrightness~, specCentroid~, specFlatness~, specFlux~, specIrregularity~, specKurtosis~, specRolloff~, specSkewness~, specSpread~, MFCC~, BFCC~ and zeroCrossing~. A description of the analysis can be found in (Brent, 2009) Depending on which features are mapped on which axes, different plots are generated. In order to navigate through these clouds we are using a Kinect and a skeleton tracker (Synaptic, an application based on openNI) to define the right hand as the focal point of the playback. The head position is controlling the viewport towards the cloud, so that we can zoom-in and pan through our movement. This gives us a better sense for which units are in the front and which ones are in the back. The screen/monitor therefore acts like a window towards the space.

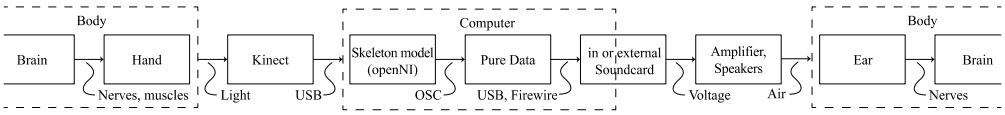

Figure 1: Interaction feedback – a flow diagram brain to brain.

3.1 Screenshot

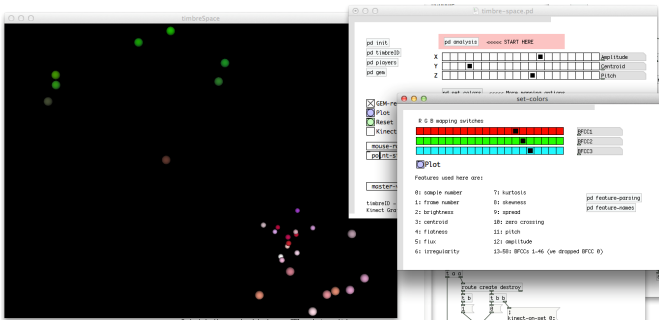

Figure 2: Units of a short sample in virtual 3D Space

4 Conclusion and prospect

I believe the sonic and visual result achieved is innovative and captivating. Drawback of the Kinect sensor is the additional latency added. This constrains the application scenarios as a musical Instrument. Future improvements may include the addition of the following features:

1. The possibility to grade the units in the plot according to video frame analysis.
2. Overall optimisation of the latency issue.
3. The use of a 3D-display or glasses for facilitated perception of the depth.

Acknowledgments

This work is building upon Pure Data by Miller Puckette, the timbreID external by William Brent with its extremely well structured example files, and the Synapse application by Ryan Challinor which is retrieving the skeleton data from the Kinect sensor.

References

Brent, William (2009). *Cepstral analysis tools for percussive timbre identification* – Proceedings of the 3rd International Pure Data Convention, São Paulo.

Schwarz, Diemo (2012). *The Sound Space as Musical Instrument: Playing Corpus-Based Concatenative Synthesis* – New Interfaces for Musical Expression (NIME)

Schwarz, Diemo (2006). *Real-Time Corpus-Based Concatenative Synthesis with CataRT* – Expanded version 1.1 of submission to the 9th Int. Conference on Digital Audio Effects (DAFx-06), Montreal

Schwarz, Diemo (2006-3). *Concatenative Synthesis: The Early Years* – Journal of New Music Research, 35(1):3–22, Special Issue on Audio Mosaicing.

Contact

Max Neupert · Marienstraße 5 · 99423 Weimar · http://www.maxneupert.de