CORPUS-BASED CONCATENATIVE SYNTHESIS: PERCEPTUAL DESCRIPTORS AS AN APPROACH TO COMPOSING AND ANALYZING TIMBRE

Aaron Einbond

Computer Music Center (CMC)
Columbia University
New York
ae2341@columbia.edu

Diemo Schwarz

Ircam–Centre Pompidou
CNRS-STMS
Paris
schwarz@ircam.fr

ABSTRACT

How can sound descriptors arising from studies in music perception be applied to inform composing and hearing? Perceptual research by Wessel and Grey suggest timbre can be organized by listeners into a multi-dimensional spatial representation. Building on this work, we propose an approach to timbre that is based on computer analysis of perceptually-relevant descriptors. Using the recent concept of corpus-based concatenative synthesis (CBCS), a database of samples is plotted in a spatial representation corresponding to any two or three of these descriptors. A musical phrase may be generated by drawing a curve in the space or by closest matches to an external target sound file. While this technique can be used for more traditional sounds, it is especially effective for organizing non-pitched sounds based on their timbral characteristics, both in the contexts of electronic music and computer-assisted composition for acoustic instruments. We implement this technique with the *CataRT* package for computer programs *Max/MSP* and *OpenMusic (OM)*.

Recent works for instruments, interactive electronics, and sound installation illustrate this approach. Examples include *What the Blind See* for five instruments and live electronics, as well as a version of the same work for sound installation, in which the the listener is simultaneously the performer and part of the space itself. The title, taken from an article by neurologist Oliver Sacks, suggests perception as the focus of the musical experience. While these tools are already being effectively exploited by composers, they can be adapted for broader uses in improvisation, scholarship, and therapy.

Acknowledgements: This work is partially funded by the French National Agency of Research ANR within the project SampleOrchestrator. We thank Jean Bresson, Alexis Baskind, Eric Daubresse, John MacCallum, David Wessel, and Adrian Freed for their assistance implementing the software and their invaluable feedback on the manuscript.

INTRODUCTION

- There has been recent interest in the perception of *timbre*, or tone color, in musics as diverse as Tuvan throat singing [3], avant-garde jazz [4], environmental field recording, and ambient electronic music [5].
- However the manipulation of timbre as a structural musical element has long been a challenge for composers. In both instrumental and electronic music there is still a predominance of tools for organizing pitch and rhythm as compared to non-pitched materials.
- Wessel and Grey have suggested that timbre could be organized by listeners into a multi-dimensional spatial representation with axes corresponding to subjective timbral descriptors [2, 7]. Trevor Wishart has proposed this as a model for organizing electronic sounds [8].
- We present an approach to structuring timbre using corpus-based concatenative synthesis (CBCS), based on perceptually-relevant audio descriptors and controllable in real-time, applied in recent compositions for instruments and electronics.

CORPUS-BASED CONCATENATIVE SYNTHESIS

- The recent concept of corpus-based concatenative sound synthesis [6] makes it possible to create music by selecting snippets of a large database of pre-recorded sound by navigating through a space where each snippet is placed according to its sonic character in terms of sound descriptors, which are characteristics extracted from the source sounds such as pitch, loudness, and brilliance, or higher level meta- data attributed to them.
- CBCS can be advantageously applied interactively using an immediate selection of a target given in real-time as is implemented in the *CataRT* system [6] for *Max/MSP* with the extension libraries *FTM* and *Gabor*. See Figure 1 for an example of CataRT's sound browsing interface, where grains are played according to proximity to the mouse- or controller-driven target position in a user-selected 2-descriptor plane.

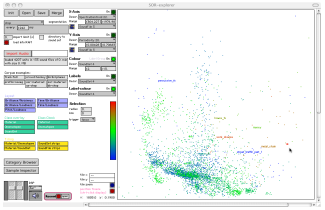

Figure 1. Screenshot of *CataRT 's* 2D navigation interface.
For documentation and downloads please visit: http://imtr.ircam.fr/index.php/CataRT and http://ftm.ircam.fr

REAL-TIME PERFORMANCE

- *CataRT 's* model is a multi-dimensional space of descriptors, populated by the sound units. They are selected by minimizing the target distance C^t , which is a weighted Euclidean distance function that expresses the match between the target x and a database unit u:

$$C^t(u_i, x) = \sum_{k=1}^{K} w_k' C_k^t(u_i, x) \quad (1) \qquad C_k^t(u_i, x) = \left(\frac{x(k) - u_i(k)}{\sigma_k}\right)^2 \quad (2)$$

- *Beside Oneself* for viola and live electronics and *What the Blind See* for five instruments and live electronics by Aaron Einbond are the first works to incorporate CBCS in real time, using a live analysis of 10 timbral descriptors to trigger *CataRT* synthesis from a corpus of close-miked instrumental samples (see Figure 2).
- *What the Blind See*, an interactive sound installation by Einbond uses the sounds of the public as well as the installation's outdoor setting as a target for *CataRT* synthesis from a corpus of filtered field recordings of insects and plants.

Figure 2. *Max/MSP* screenshot of live descriptor analysis using *FTM&Co*.

CORPUS-BASED TRANSCRIPTION

- *CataRT's* analysis and selection algorithms can be used as a tool for computer-assisted composition. Units are matched from a corpus of acoustic instrumental samples to a given target. The results are recorded to an SDIF (Sound Description Interchange Format) [1, 9] file using a specially-created recording module. This file can be read by other programs such as *OM*, or by *Max/MSP* using *FTM* data structures and externals.
- Once imported into *OpenMusic* the descriptors are converted symbolically into a notated score. The goal can be for the instrumentalist reading the score to approximate the target sound in live performance. The target could be an audio file, analyzed as above, or it could be symbolic: an abstract gesture in descriptor space and time, designed by hand with a controller such as a tablet or mouse. This process is summarized in Figure 3:

Figure 3. Flowchart for corpus-based transcription.

MUSICAL EXAMPLE: field recording of rain transcribed for ensemble.

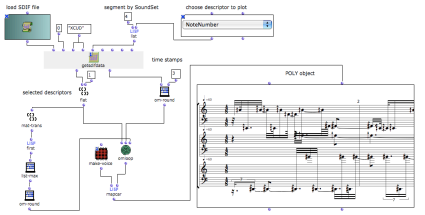

Figure 4. *OpenMusic* patch to generate an instrumental score based on perceptual criteria.

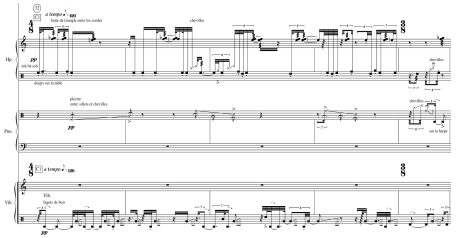

Figure 5. Corresponding score page from *What the Blind See* for chamber ensemble of harp, piano, and vibraphone.
For more information and audio examples, please visit: http://music.columbia.edu/people/bios/aeinbond

DISCUSSION

- *Descriptor Weighting:* In the applications presented, descriptors and weights are chosen subjectively. We hope to explore methods automatically to weight descriptors that best fit perceptual judgments.
- *Mapping Paradigms:* Instead of associating parameters of the target and synthesis directly, the target could be rescaled, "transposed," or "inverted" before mapping it to a corpus, through an linear transformation in descriptor space. A further step is to map one descriptor of the target analysis to a different descriptor in the corpus output to produce a kind of gestural "analogy."
- *Perceptual Tests:* While the sound descriptors used have been chosen based on music perception research, we would like to test the results of the technique to assess the extent to which subjects hear sounds synthesized using CBCS as perceptually similar to their targets.

REFERENCES

[1] J. J. Burred, C. E. Cella, G. Peeters, A. Röbel, and D. Schwarz, "Using the SDIF sound description interchange format for audio features," in *ISMIR*, 2008.
[2] J. M. Grey, "Multidimensional perceptual scaling of musical timbres," *J. Acoust. Soc. Am.*, vol. 61, 177.
[3] T. Levin and V. Süzükei, *Where Rivers and Mountains Sing*. Bloomington: Indian University Press, 2006.
[4] G. Lewis, "Purposive Patterning: Jeff Donaldson, Muhal Richard Abrams, and the Multidominance of Consciousness," *Lenox Avenue*, vol. 5, Chicago: 1999.
[5] F. López and B. Eno cited in C. Cox and D. Warner, eds., *Audio Culture*, London, 2004.
[6] D. Schwarz, "Corpus-based concatenative synthesis," IEEE Sig. Proc. Mag., vol. 24, no. 2, Mar. 2007.
[7] D. Wessel, "Timbre space as a musical control structure," *Computer Music Journal*, vol. 3, no. 2, 1979.
[8] T. Wishart, *On Sonic Art*. London: Harwood Academic Publishers, 1996.
[9] M. Wright et al., "Audio applications of the sound description interchange format standard," in *Audio Engineering Society 107th Convention*, New York, 1999.