SPATIALIZING TIMBRE WITH CORPUS-BASED CONCATENATIVE SYNTHESIS

Aaron Einbond

Computer Music Center (CMC)
Department of Music, Columbia University, New York

Diemo Schwarz

Ircam–Centre Pompidou CNRS-STMS
1 Place Igor Stravinsky, 75004 Paris

ABSTRACT

Corpus-based concatenative synthesis presents unique possibilities for the visualization of audio descriptor data. These visualization tools can be applied to sound diffusion in the physical space of the concert hall using current spatialization technologies. Using CataRT and the FTM&Co library for Max/MSP we develop a technique for the organization of a navigation space for synthesis based on user-defined spatial zones and informed by the perceptual concept of *timbre space*. Spatialization responds automatically to descriptor data and is controllable in real-time or can be recorded for later playback. This work has been realized in recent compositions for instruments, electronics, and sound installation using *Wave Field Synthesis* and *Vector-Based Amplitude Panning*. The goal is to place the listener in the midst of a virtual space of sounds organized by their descriptor data, simulating an immersion in timbre space.

1. INTRODUCTION

Composers from Varèse to Chowning have proposed strategies for recording and reproducing spatial trajectories adapted to the spatialization technologies available to them [6, 2]. Recent research involving sound descriptors and the availability of tools for their rapid calculation suggest broad applications for real-time electronic music. It is therefore tempting to look for a model for spatialization taking advantage of this rich source of data.

1.1. Timbre Space

One promising model can be drawn from studies of music perception: *timbre space*. According to research by Wessel and Grey, listeners' grouping of sounds of disparate timbres is consistent with a low-dimensional spatial model [14, 4]. This research prompted Wessel to propose a system that would allow the user metaphorically to "take a walk in timbre space." [1] Pursuing the implications of this idea, we propose timbre-space as the point of departure for the parametrization of spatial trajectories according to sonic descriptor data.

Taking advantage of the recent technique of corpus-based concatenative synthesis, large databases, or *corpora*, of recorded sound can be analyzed, stored, and retrieved based on the descriptors of each sound sample. Descriptor data is mapped to a low-dimensional user interface to facilitate the selection and re-synthesis of units. Through a suitable choice of axes, this interface can be mapped to similar descriptors to those identified by Wessel and Grey for a two- or three-dimensional timbre space.

Beyond this literal application, we have developed tools for more elaborate spatial navigation: multiple sub-corpora can be placed in the navigation space and superposed, allowing a more creative mapping. This organization can be used for both the user interface from which units are synthesized and the spatialization during synthesis. Each sub-corpus can still itself be organized by timbre.

1.2. Corpus-Based Concatenative Synthesis

The concept of corpus-based concatenative sound synthesis (CBCS) [9] 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 metadata attributed to them. This allows one to explore a corpus of sounds interactively or by composing paths in the space, and to create novel harmonic, melodic, and timbral structures while always keeping the richness and nuances of the original sound.

The database of source sounds is segmented into short *units*, and a *unit selection* algorithm finds the sequence of units that best match the sound or phrase to be synthesised, called the *target*. The selected units are then concatenated and played, possibly after some transformations.

CBCS can be advantageously applied interactively using an immediate selection of a target given in real-time as is implemented in the CataRT system [10] for Max/MSP with the extension libraries FTM&Co and Gabor,[2] making it possible to navigate through a two- or more-dimensional

[1] David Wessel, personal communication.

[2] http://imtr.ircam.fr/index.php/CataRT, http://ftm.ircam.fr

projection of the descriptor space of a sound corpus in real-time, effectively extending granular synthesis by content-based direct access to specific sound characteristics.

2. ORGANIZING DESCRIPTOR SPACE

To better control spatialization, zones may be defined corresponding to sub-corpora that are spatialized independently. The play position is tracked relative to these zones and this information is used to control sound synthesis.

2.1. Defining Zones by Points and Radius

A simple way of defining zones is to define a number of points p_i that give the zone centers, and the zone radius or "influence" r_i, as in the famous GRM parameter interpolator in the SYTER system [13], or several other current implementations for MAX/MSP [12, 7], and *pMix* by Oliver Larkin.[3]

The squared proximity d_i^2 of point x to each zone i, weighted by influence r_i is then calculated as

$$d_i^2 = \frac{(x-p_i)^2}{r_i^2} \tag{1}$$

This formula is efficiently implemented in FTM&CO by the mnm.mahalanobis external.

2.2. Defining Arbitrary Zones

In order to organize a navigation space into several arbitrarily-shaped non-overlapping zones, we present a method that allows one to draw the outline of the zones and then performs an efficient lookup of the zones closest to the navigation position by k-nearest neighbor search in logarithmic time using a kD-tree index [11]. This simplified representation avoids more complex geometric algorithms such as ray testing or hull calculations. The resulting distances to the closest zones can then be used as an interpolation control (see below).

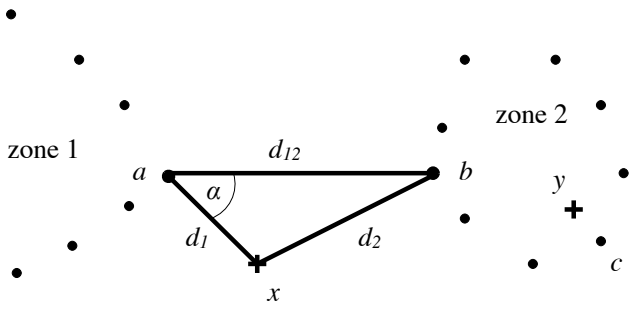

Figure 1. Distances of point x to closest points of two zones.

[3] http://www.olilarkin.co.uk/index.php?p=pmix

Figure 1 shows two zones, defined by their outline of points, the current position x, and the distances d_1, d_2 that are calculated by the nearest neighbor search. However, in point y, d_2 would be wrongly calculated to point c at the opposite border of zone 2. To avoid this, we find for each closest zone point p_i the *induced closest zone points* q_{ij} to all other zones j. In the figure, $p_1 = q_{21} = a$, $p_2 = c$, but $q_{12} = b$. These points are then used to calculate the distances, which are set to zero if the following test finds that we are inside a zone:

Using the law of cosines on the angle α in point a

$$\cos\alpha = \frac{d_{ij}^2 + d_j^2 - d_i^2}{2 d_{ij} d_j}, \tag{2}$$

and dropping the denominator that doesn't change the sign, we can test if $|\alpha| < 90°$, which we take as the approximate criterion that tells us if the current point is beyond zone i's border seen from zone j:

$$\text{beyond}_{ij} \iff d_{ij}^2 + d_j^2 - d_i^2 < 0 \tag{3}$$

From this flag, we can also deduce whether the point is inside a zone, namely if it is seen beyond a zone from all other zones. That is, if $beyond_{ij}$ is true for all j, the point is within zone i.

The final answer of proximity to a zone i can be either taken directly from the clipped distance d_i, or we can again use the law of cosines to calculate a normalized interpolation factor f_{ij} between each pair of zones i and j by (see figure 2 for an example):

$$f_{ij} = \frac{d_i^2 - d_j^2}{2 d_{ij}^2} + \frac{1}{2} \quad \text{clipped between 0 and 1} \tag{4}$$

The used distance values are all known from the calculation of the nearest neighbor and induced closest zone points, and conveniently, the mnm.knn and mnm.mahalanobis externals both output square distances, such that above formulas can be calculated directly by efficient matrix operations using the FTM&CO framework.

The simplicity of our representation of zones as a point contour leads to a low computational complexity of $O(K^2 \log N)$, with K the (small) number of zones, and N the number of points in each zone, or $O(K \log N)$ when the induced points are not computed. Conventional ray-tracing point-in-polygon algorithms have a complexity of $O(KN)$.

One limitation of the algorithm is due to the discrete point representation. It is up to the user to draw the contours dense enough to avoid jump effects when the closest point changes. Another limitation is that concave areas of a zone are considered inside the zone, unless another zone follows the cavity (i.e. winding corridors between two zones are indeed possible). This can be avoided by placing a point of an additional "dummy" zone in the cavity, such that the

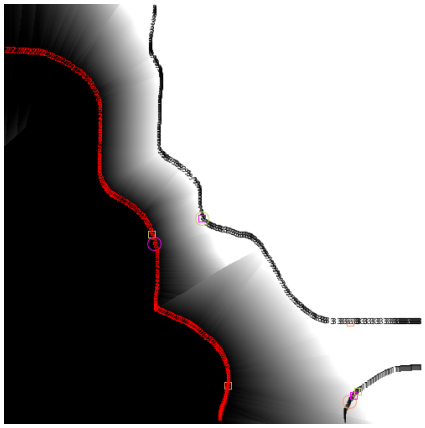

Figure 2. Map of the interpolation factor f_{23} as greyscale values (black is one).

point is never hidden beyond the zone borders from all other zones.

2.3. Interpolation Between Zones

We now turn to the use of the zones defined above for the organization of the sound space by assigning a sub-corpus to each zone. Given a position between zones (separate as shown above or possibly overlapping as when defining layers at different heights of the space), there are several ways to interpolate between the corpora assigned to each zone: First, the selection can take place independently in each corpus, and the resulting sounds are mixed with levels anti-proportional to the distance. Second, when mixing sounds is not desired, we can perform a selection in each corpus, and then choose randomly between the selected units with probabilities anti-proportional to the distances. Generalizing this to random selection interpolation between several zones is possible using a distance-mapping function. This function maps the distances to zones to a likelihood, starting at 1 for 0 distance, and descending to 0 at a cutoff distance. The mapped distance values need to be normalized to sum 1 to become a PDF (probability density function), from which we calculate the cumulative sum to obtain the CDF (cumulative density function). We then draw random bin indices accordingly by accessing the CDF by a uniformly distributed random value. This method is implemented in the mnm.pdf abstraction.

3. DESCRIPTOR-BASED SPATIALIZATION

CataRT has powerful capabilities for real-time synthesis in live performance [10, 3]. A live audio signal can be analyzed for its descriptor content and used to power the synthesis from a pre-recorded corpus, or itself to define a new corpus. A mouse, drawing tablet, or other controller can be used to navigate through a corpus based on its two- or three-dimensional spatial layout. Both a single corpus organized by timbral descriptors, and multiple sub-corpora arranged in a more complex navigation space, can be mapped to a corresponding live spatialization system.

Synthesis can be spatialized according to a single-source or multi-source metaphor. In the former, the output of CataRT is conceived of as an "instrument" with a moveable spatial location unique at any moment. In the latter, a CataRT corpus is envisioned as collection of instruments, each with a fixed location, excited by a metaphorical body moving through the space.

In addition to real-time capabilities, the catart.select.record module can be used to record the timing of grain selection and descriptor data to an SDIF file which can then be played back in deferred time [3]. Using the spatial mapping described, this presents a technique for recording and reproducing spatial trajectories.

3.1. Single-Source Model

A single-source model is well-suited to a spatial implementation using *Wave Field Synthesis* (WFS). The mono output of CataRT is treated as a *point source* by the WFS system. With each grain selection, descriptor coordinates are sent to the WFS system normalized to a range chosen by the user. By default the ranges correspond to the current values of CataRT's 2D interface.

3.1.1. Break

This model was used in Aaron Einbond's *Break* for baritone saxophone and live electronics, written for saxophonist David Wegehaupt and premiered at the Technische Universität, Berlin, at the *Lange Nacht der Wissenschaften*, 2008. Using the WFS system at the TU powered by Wonder software [1], corpora of pre-recorded saxophone samples were selected for re-synthesis and moved about the two-dimensional plane of the WFS system according to the real-time descriptor analysis of the live saxophone. Short samples containing rapid, dry attacks, such as close-miked key-clicks, were especially suitable for a convincing impression of motion of the single WFS source. The effect is that of a virtual instrument moving through the concert hall in tandem with changes in its timbral content, realizing Wessel's initial proposal.

3.2. Multi-Source Model

For a multi-source spatial model, the CataRT synthesis module catart.synthesis is replaced with catart.synthesis.multi. This module receives a vector of channel numbers and amplitudes, such as *Vector-Based Amplitude Panning* (VBAP) coefficients generated using the Max/MSP object vbap [8]. Taking advantage of the polyphonic capabilities

of FTM&CO's overlap-add module gbr.ola, each CATART grain is re-synthesized in a multi-channel system with its amplitude scaled according to VBAP coefficients. In this way an arbitrary number of grains can be simultaneously reproduced with distinct virtual positions (see Figure 3).

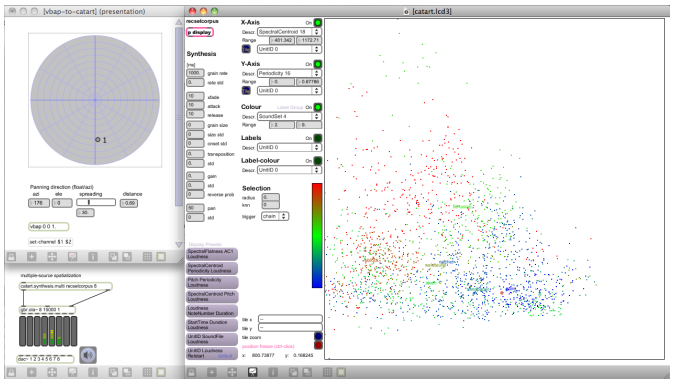

Figure 3. Screenshot of CATART connected to vbap and ambimonitor to control spatialization in eight channels.

3.2.1. What the Blind See

This multi-source model was realized in Einbond's *What the Blind See* for bass clarinet, viola, harp, piano, percussion, and live electronics commissioned by IRCAM/Centre-Pompidou for Ensemble l'Instant Donné at the Agora Festival 2009. An eight-channel system was arranged in a near-circle around the audience. As in *Break*, two-dimensional VBAP coefficients were calculated in real time corresponding to the descriptor values of CATART grains. Each grain was synthesized to occupy a distinct virtual position for its duration, in general overlapping with other grains in other positions. This model was well-suited to corpora of percussion samples, each grain with a sharp attack and long decay that continued playing as new grains were added.

4. DISCUSSION

Moving beyond a literal timbre space model, the tools described could be easily adapted to navigation spaces that rotate, deform, and respond live in performance, creating a dynamic approach to spatial mapping. The organization of the navigation space into zones is to be used in the interactive sound installation *Grainstick* [5] by composer Pierre Jodlowski, utilizing IRCAM's WFS system, and suggests rich possibilities for future work beyond the concert hall.

5. ACKNOWLEDGEMENTS

We thank Eric Daubresse, Sylvain Cadars, Jean Bresson, Emmanuel Jourdan, Pierre-Edouard Dumora, Florian Göltz, Eckehard Güther, Wilm Thoben, Volkmar Hein, Stefan Wienzierl, John MacCallum and David Wessel. The work presented here is partially funded by the *Agence Nationale de la Recherche* within the project *Topophonie*, ANR-09-CORD-022.

6. REFERENCES

[1] M. Baalman and D. Plewe, "Wonder - a software interface for the application of wave field synthesis in electronic music and interactive sound installations," in *Proc. ICMC*, Miami, USA, 2004.

[2] J. Chowning, "The simulation of moving sound sources," *JAES*, vol. 19, no. 1, 1971.

[3] A. Einbond, D. Schwarz, and J. Bresson, "Corpus-based transcription as an approach to the compositional control of timbre," in *ICMC*, Montréal, 2009.

[4] J. M. Grey, "Multidimensional perceptual scaling of musical timbres," *J. Acoust. Soc. Am.*, vol. 61, 1977.

[5] G. Leslie *et al.*, "Grainstick: A collaborative, interactive sound installation," in *Proc. ICMC*, NYC, 2010.

[6] V. Lombardo *et al.*, "A virtual-reality reconstruction of *Pome lectronique* based on philological research," *Computer Music Journal*, vol. 33, no. 2, 2009.

[7] A. Momeni and D. Wessel, "Characterizing and controlling musical material intuitively with geometric models," in *Proc. NIME*, Montreal, Canada, 2003.

[8] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *JAES*, vol. 46, no. 6, 1997.

[9] D. Schwarz, "Corpus-based concatenative synthesis," *IEEE Sig. Proc. Mag.*, vol. 24, no. 2, Mar. 2007.

[10] D. Schwarz, R. Cahen, and S. Britton, "Principles and applications of interactive corpus-based concatenative synthesis," in *JIM*, GMEA, Albi, France, Mar. 2008.

[11] D. Schwarz, N. Schnell, and S. Gulluni, "Scalability in content-based navigation of sound databases," in *Proc. ICMC*, Montréal, Canada, 2009.

[12] M. Spain and R. Polfreman, "Interpolator: a two-dimensional graphical interpolation system for the simultaneous control of digital signal processing parameters," *Organised Sound*, vol. 6, no. 02, 2001.

[13] D. Teruggi, *Vorträge und Berichte (eds.) Neue Musiktechnologie II*. Schott Musik International, 1996, ch. The technical developments in INA-GRM and their influences on musical composition.

[14] D. Wessel, "Timbre space as a musical control structure," *Computer Music Journal*, vol. 3, no. 2, 1979.