INTERACTING WITH A CORPUS OF SOUNDS

Diemo Schwarz

Ircam–CNRS–UPMC, France

schwarz@ircam.fr

Abstract

Corpus-based concatenative synthesis (CBCS) is a recent sound synthesis method, based on descriptor analysis of any number of existing or live-recorded sounds, and synthesis by selection of sound segments from the database matching sound characteristics given by the user. While the previous phase of research focused mainly on the analysis and synthesis methods, and handling of the corpus, current research now turns more and more towards how expert musicians, designers, the general public, or children can interact efficiently and creatively with a corpus of sounds. We'll look at gestural control of the navigation through the sound space, how it can be controlled by audio input to transform sound in surprising ways, or to transcribe and re-orchestrate environmental sound, and at its use for live performance, especially in an improvisation setting between a instrumental and a CBCS performer, where recording the corpus live from the instrument creates a stronger and more direct coupling between them, and lastly sound installations that open up discovery and interaction with rich sound corpora to the general public and children.

Keywords: corpus-based synthesis, interaction, audio descriptors

1. Introduction

Corpus-based concatenative synthesis methods (CBCS) (Schwarz, 2007) are more and more often used in various contexts of music composition, live performance, audio–visual sound design, and installations. They take advantage of the rich and ever larger sound databases increasingly available today to assemble sounds by interactive real-time or off-line content-based selection and concatenation. Actual recordings or live-recorded audio are used to constitute the corpus, which makes the richness and fine details of the original sounds available for musical expression.

CBCS is based on segmentation and description of the timbral characteristics of the sounds in the corpus, and synthesis by selection of sound segments from the database matching sound characteristics given by the user. It allows to explore a corpus of sounds interactively or by composing paths in the descriptor space, and to recreate novel timbral evolutions. CBCS can also be seen as a content-based extension of granular synthesis, providing direct access to specific sound characteristics.

It has been implemented in various systems and environments (see Schwarz (2006) and http://imtr.ircam.fr/imtr/Corpus-Based_Sound_Synthesis_Survey) and notably by the author since 2005 in an interactive sound synthesis system named CataRT (Schwarz et al., 2006) at http://imtr.ircam.fr/imtr/CataRT running in Max/MSP with the FTM&Co. extensions.

While the research up to today focused mainly on the analysis and synthesis methods, and handling of the corpus, current research now turns more and more towards how expert musicians, designers, the general public, or children can interact efficiently and creatively with a corpus of sounds:

The use of CBCS as a new interface for musical expression (NIME) or digital musical instrument (DMI) for use in live performance by expert musicians introduces an important and novel concept that is the essence of the interface: the space of sound characteristics with which the player interacts by navigating through it, with the help of gestural controllers.

2. Principle and Motivation of CBCS

Corpus-based concatenative synthesis systems build up a database of prerecorded or live-recorded sound by segmenting it into units, usually of the size of a note, grain, phoneme, or beat, and analysing them for a number of sound descriptors, which describe their sonic characteristics. These descriptors are typically pitch, loudness, brilliance, noisiness, roughness, spectral shape, etc., or metadata, like instrument class, phoneme label, etc., that are attributed to the units, and also include the segmentation information of the units, like start time, duration, source sound file index. These sound units are then stored in a database (the corpus). For synthesis, units are selected from the database that are closest to given target values for some of the descriptors, usually in the sense of a weighted Euclidean distance.

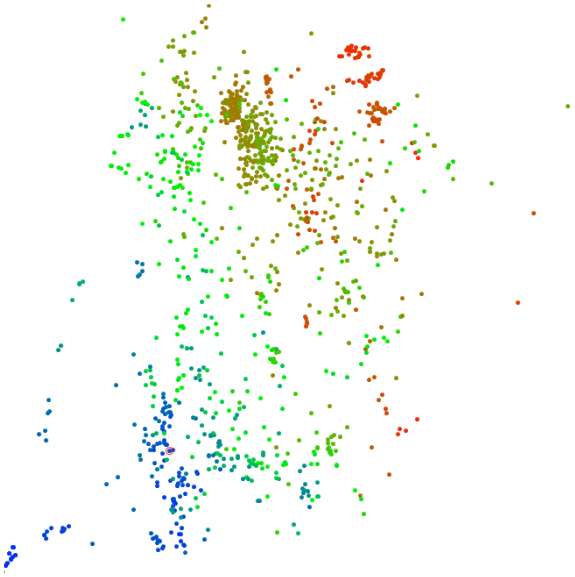

Figure 1. Example of the 2D visualisation of a corpus, plotted by Spectral Centroid (x), Periodicity (y), NoteNumber (colour).

Ever larger sound databases exist on all of our harddisks and are waiting to be exploited for synthesis, which is ever less feasible to do completely manually. Therefore, the help of automated sound description allows to access and exploit a mass of sounds efficiently and interactively, unlike traditional query-oriented sound databases (Schwarz and Schnell, 2009).

As with each new synthesis method, corpus-based concatenative synthesis gives rise to new sonorities and new methods to organise and access them, and thus expands the limits of sound art. Here, by selecting snippets of a large database of pre-recorded sound by navigating through a space where each snippet takes up a place according to its sonic character (see figure 1), it allows to explore a corpus of sounds interactively, or by composing this path, and to create novel sound structures by re-combining the sound events, proposing novel combinations and evolutions of the source material. The metaphor for composition is an explorative navigation through the sonic landscape of the corpus.

Last, using concatenative synthesis, as opposed to pure synthesis from a signal or physical model, allows a sound composer to exploit the richness of detail of recorded sound while retaining efficient control of the acoustic result by using perceptually and musically meaningful descriptors to specify the desired target sound features.

3. Interacting by Gestural Control

In a DMI context interacting with gesture sensors to control the navigation through the sound space, each combination of input device and synthesis mode redefines the affordances of the interaction and thus in fact a separate digital musical instrument (Schwarz, 2012).

The controllers fall into the two groups of 2D or 3D positional control, and control by the analysis of audio input (see next section). Many video and audio examples can be found at http://imtr.ircam.fr/imtr/CataRT_Instrument.

The most intuitive access to navigating the corpus is provided by XY controllers, such as MIDI control pads, joystick controllers, etc., for giving the target position in 2D. Better still are pressure-sensitive XY-controllers such as a graphics tablet, or some rare Tactex-based controllers, that allow to control also dynamics. Multi-touch controllers or touch-screens, especially when pressure sensitive, are the dream interface for navigation, providing an intuitive polyphonic access to a sound space.

Motion capture systems, either by cameras and markers, or depth-sensing cameras, offer a full-body access to a sound corpus mapped into physical 3D space. These interfaces have

not yet been frequently used for music performance, but are beginning to be used in installation settings (Caramiaux et al., 2011, Savary et al., 2013).

Accelerometer equipped devices such as game controllers, smartphones, or tablets can be used to navigate the 2D space by tilting and shaking.

How and when the units close to the target position are actually played is subject to the chosen trigger mode that, together with the control device, finally determines the gestural interaction. There are two groups of trigger modes that give rise to two different styles of musical interaction, that we will analyse according to Cadoz's framework (Cadoz, 1988, Cadoz and Wanderley, 2000).

With dynamic instrumental gestures, the specification of the target position can at the same time trigger the playback of the unit. Clearly, here the navigation in the sound space constitutes a selection gesture, and, in this group of trigger modes, it is at the same time an excitation gesture.

The other group of trigger modes separate selection from excitation, giving rise to continuous rhythms or textures. The navigational gestures are solely selection gestures, while no excitation gestures are needed, since the system plays continuously. However, the trigger rate and the granular playback parameters can be controlled by modification gestures on faders.

4. Interacting by Audio Input

When CBCS is controlled by descriptors analysed from audio input, it can be used to transform sound in surprising ways, to create augmented instruments, or to transcribe and reorchestrate environmental sound (Einbond et al., 2009). This special case of CBCS is generally called "audio mosaicing".

The possibility of choosing the mapping of input descriptors to target descriptors makes for a significant difference with the control of CBCS by audio spectrum analysis as in classical audio mosaicing, where the selection is by direct similarity between input and corpus spectra.

In a DMI context, we can make use of piezo pickups on various surfaces that allow to hit, scratch, and strum the corpus of sound, exploiting its variability according to the gestural interaction, the sound of which is analysed and mapped to the 2D navigation space. For instance, dull, soft hitting plays in the lower-left corner of the corpus descriptor space, while sharp, hard hitting plays more in the upper right corner.

Especially this latter mode of gestural control often creates a gestural analogy to playing an acoustic instrument, especially in a duo improvisation setting.

In a compositional context, corpus-based analysis and selection algorithms can be used as a tool for computer-assisted composition. In the work by Einbond (2009), a corpus of audio files was chosen corresponding to samples of a desired instrumentation. Units from this corpus were then matched to a given target. Instead of triggering audio synthesis, the descriptors corresponding to the selected units and the times at which they are selected were then imported into a compositional environment where they were converted symbolically into a notated score, that approximates the target, which could be an audio file, analyzed as above, or symbolic: an abstract gesture in descriptor space and time.

5. Live Recording the Corpus

For live performance, especially in an improvisation between an instrumental and a CBCS performer, recording the corpus live from the instrument creates a stronger and more direct coupling between the instrument and the laptop performer, compared to traditional acoustic improvisation. Whereas in the latter the coupling takes place in an abstract space of musical intentions and actions, live CBCS creates a situation where both performers share the same sound corpus as an instrument, thus the coupling takes place in a concrete space of sound, since the very timbral variation of the instrument player directly constitutes the instrument from which the laptop player creates music by navigation and recontextualisation. (Schwarz and Brunet, 2008, Johnson and Schwarz, 2011)

This setting could even be seen as an improvisation with two brains and four hands controlling one shared symbolic instrument, the sound space, built-up from nothing and nourished in unplanned ways by the sound of the instrument, explored and consumed with whatever the live instant filled it with. It creates a symbiotic relationship between the player of the instrument and the one playing the software.

6. Interaction for an Installation Audience

CBCS also found a very promising application in environmental sound texture synthesis (Schwarz and Schnell, 2010, Schwarz, 2011) for audio–visual production in cinema and games, and sound installations such as the Dirty Tangible Interfaces (DIRTI), that opens up discovery and interaction with rich sound corpora to the general public and children.

Dirty Tangible Interfaces (Savary et al., 2012, 2013) are a new concept in interface design that forgoes the dogma of repeatability in favor of a richer and more complex experience, constantly evolving, never reversible, and infinitely modifiable. In the first realisation of this concept, a granular or liquid interaction material placed in a glass dish (figure 2) is tracked for its relief and dynamic changes applied to it by the user(s).

Figure 2. Example of a liquid+granular interaction material (water and ink).

It allows the audience to interact with complex high-dimensional datasets using a natural gestural palette and with interactions stimulating the senses. The interaction is tangible and embodied using the full surface of the hands, giving rich tactile feedback through the complex physical properties of the interaction material.

Figure 3. Use of DIRTI with children at the event *Les petits chercheurs de sons*, in the cultural centre *104* in Paris, June 2013.

7. Optimisation of the Interaction Space

While a direct projection of the high-dimensional descriptor space to the low-dimensional navigation space has the advantage of conserving the musically meaningful descriptors as axes (e.g. linear note pitch to the right, rising spectral centroid upwards), we can see in figure 1 that sometimes the navigation space is not optimally exploited, since some regions of it stay empty, while other regions contain a high density of units, that are hard to access individually. Especially for the XY controller in a multi-touch setting, a lot of the (expensive and always too small) interaction surface can remain unexploited. Therefore, we apply a distribution algorithm (Lallemand and Schwarz, 2011) that spreads the points out using iterative Delaunay triangulation and a mass–spring model, the results of which are illustrated in figure 4.

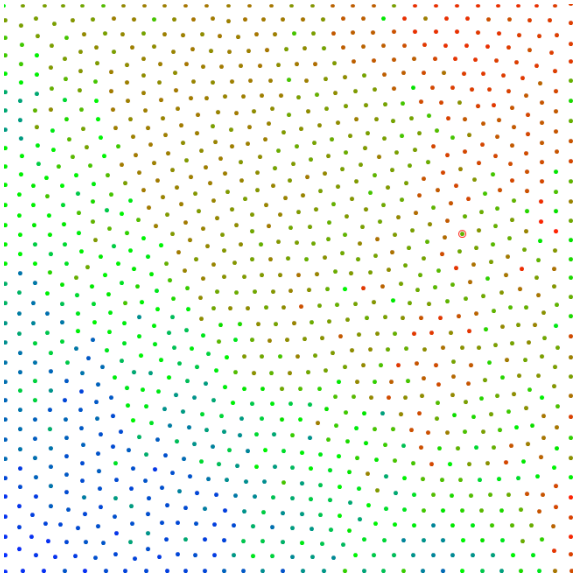

Figure 4. Distribution of the corpus in figure 1.

8. Discussion and Conclusion

From the variety of examples and usages, we can see that corpus-based methods allow composers and musicians to work with an enormous wealth of sounds, while still retaining precise control about its exploitation. From the author's ongoing experience and observation of various usages we can conclude that CBCS and CataRT can be sonically neutral and transparent, i.e. neither the method nor its software implementation come with a typical sound that is imposed on the musician, but instead, the sound depends mostly on the sonic base material in the corpus and the gestural control of selection, at least when the granular processing tools and transformations are used judiciously. As a DMI, it allows expressive musical play, and to be reactive to co-musicians, especially when using live CBCS.

A more general questioning of the concept of the sound space as interface is the antagonism of large variety vs. fine nuances, that need to be accommodated by the interface. Indeed, the small physical size of current controllers does sometimes not provide a sufficiently high resolution to precisely exploit fine nuances. Here, prepared sound sets and zooming could help, but finally, maybe less is more: smaller, more homogeneous corpora could invite to play with the minute details of the sound space.

One weakness in the typical interaction setup is that the interface relies on visual feedback to support the navigation in the sound space, and that this feedback is on a computer screen, separate from the gestural controller (except for the multi-touch screens), and thus breaking the colocation of information and action. For fixed corpora, this weakness can be circumvented by memorising the layout and practising with the corpora for a piece, as has been shown in the author's interpretation of the piece *Boucle #1* by composer Emmanuelle Gibello, where the computer screen is hidden, so the performer will base his interpretation solely on the sound, without being distracted by information on the screen, and can thus engage completely with the sound space he creates, and with the audience.

For future research and development in efficient interaction with a corpus of sound, application of machine learning methods seem to be most promising. These could assist in classification of the input gesture in order to make accessible corresponding classes in the corpus, or adaptive mappings between corpora (Stowell and Plumbley, 2010) to increase the usability of audio control. More advanced gesture analysis and recognition could lead to more expressivity and definition of different playing styles, and spatial interaction in an installation setting is largely left to explore.

9. References

Cadoz, C. (1988). Instrumental gesture and musical composition. In Proceedings of the International Computer Music Conference, pp. 1–12.

Cadoz, C. and M. Wanderley (2000). Gesture – Music. In M. Wanderley and M. Battier (Eds.), Trends in Gestural Control of Music, pp. 71–94. Paris: Ircam.

Caramiaux, B., S. Fdili Alaoui, T. Bouchara, G. Parseihian, and M. Rebillat (2011, September). Gestural auditory and visual interactive platform. In Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFx), Paris, France.

Einbond, A., D. Schwarz, and J. Bresson (2009). Corpus-based transcription as an approach to the compositional control of timbre. In Proceedings of the International Computer Music Conference (ICMC), Montreal, QC, Canada.

Johnson, V. and D. Schwarz (2011, November). Improvising with corpus-based concatenative synthesis. In (Re)thinking Improvisation: International Sessions on Artistic Research in Music, Malmö, Sweden.

Lallemand, I. and D. Schwarz (2011, September). Interaction-optimized sound database representation. In Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFx), Paris, France.

Savary, M., D. Schwarz, and D. Pellerin (2012, May). DIRTI Dirty Tangible Interfaces. In Proceedings of the Conference for New Interfaces for Musical Expression (NIME), Ann Arbor, MI, USA, pp. 347–350.

Savary, M., D. Schwarz, D. Pellerin, F. Massin, C. Jacquemin, and R. Cahen (2013). Dirty tangible interfaces: Expressive control of computers with true grit. In CHI '13 Extended Abstracts on Human Factors in Computing Systems, CHI'13, New York, NY, USA, pp. 2991–2994. ACM.

Schwarz, D. (2006, March). Concatenative sound synthesis: The early years. Journal of New Music Research 35(1), 3–22. Special Issue on Audio Mosaicing.

Schwarz, D. (2007, March). Corpus-based concatenative synthesis. IEEE Signal Processing Magazine 24 (2), 92–104. Special Section: Signal Processing for Sound Synthesis.

Schwarz, D. (2011, September). State of the art in sound texture synthesis. In Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFx), Paris, France.

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

Schwarz, D., G. Beller, B. Verbrugghe, and S. Britton (2006, September). Real-Time Corpus-Based Concatenative Synthesis with CataRT. In Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFx), Montreal, Canada, pp. 279–282.

Schwarz, D. and E. Brunet (2008). theconcatenator Placard XP edit. Leonardo Music Journal 18 CD track.

Schwarz, D. and N. Schnell (2009, July). Sound search by content-based navigation in large databases. In Proceedings of the International Conference on Sound and Music Computing (SMC), Porto, Portugal.

Schwarz, D. and N. Schnell (2010, July). Descriptor-based sound texture sampling. In Proceedings of the International Conference on Sound and Music Computing (SMC), Barcelona, Spain, pp. 510–515.

Stowell, D. and M. Plumbley (2010). Timbre remapping through a regression- tree technique. In Proceedings of the International Conference on Sound and Music Computing (SMC).