MUSICAL APPLICATIONS OF REAL-TIME CORPUS-BASED CONCATENATIVE SYNTHESIS

Diemo Schwarz

Ircam–CNRS, STMS, Paris

Sam Britton

Composer, London

Roland Cahen

ENSCI Paris, France

Thomas Goepfer

Ircam, Paris

ABSTRACT

Corpus-based concatenative synthesis (CBCS) builds on a large database of segmented and descriptor-analysed sounds that are selected and played according to proximity to a target position in the descriptor space. This can be seen as a content-based extension to granular synthesis providing direct access to specific sound characteristics in real-time. The aim of the article is to show how CBCS supports—or actually inspires—new musical ideas by exploring the corpus interactively or via a written target score. We will show this with 4 musical examples of pieces that explore the concepts made possible by CBCS of live corpus recording, navigation in 3D in a metaphor of a score, interaction and corpus cross-synthesis, and harmonic selection tighly integrated with an orchestra score.

1. INTRODUCTION

The recent concept of corpus-based concatenative sound synthesis [1] is beginning to find its way in musical composition and performance. It permits to create music 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, such as pitch, loudness, brilliance. This allows to explore a corpus of sounds interactively, or by composing this path, and to create novel harmonic, melodic and timbral structures.

The database of source sounds is segmented into short *units*, and a *unit selection* algorithm finds the sequence of units that match best the sound or phrase to be synthesised, called the *target*. The selection is performed according to the *descriptors* of the units, which are characteristics extracted from the source sounds, or higher level descriptors attributed to them. The selected units are then concatenated and played, after possibly some transformations.

The CATART software system [2] described in section 3 realises CBCS in real-time. CATART inherits also from granular synthesis, adding the possibility to play grains having specific acoustic characteristics, thus surpassing its limited selection possibilities, where the only control is position in one single sound file.

CBCS allows new musical ideas to be experimented by a number of novel concepts it proposes, introduced below. These concepts will be expanded by giving concrete musical applications in four musical compositions and performances in section 4. The use of these concepts and conlusions that can be drawn will then be discussed in section 5.

Re-arranging is at the very base of CBCS: units from the corpus are re-arranged by other rules than the temporal order of their original recordings, such as given evolutions of sound characteristics, e.g. pitch and brilliance.

Composition by navigation through heterogeneous sound databases allows 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 a target in the multi-dimensional descriptor space.

Interaction with self-recorded sound: By constituting a corpus, live- or prerecorded sound of a musician is available for interaction with a musical meaning beyond simple repetition of notes or phrases in delays or loops.

Cross-selection and interpolation: The selection target can be applied from a different corpus, or from live input, thus allowing to extract and apply certain sound characteristics from one corpus to another, and morphing between distinct sound corpora.

Orchestration and re-orchestration: By descriptor organisation and grouping possibilities of the corpora, a mass of sounds can be exploited while still retaining precise control over the sonic result.

2. PREVIOUS AND RELATED WORK

Corpus-based concatenative sound synthesis methods are attracting more and more interest in the communities of researchers, composers, and musicians. The many recent approaches to CBCS, summarised in [3] and continuously updated,[1] fall into two large classes, depending on whether the match is descriptor- or spectrum-based: Descriptor-based real-time systems, such as CATART [2], the commercial corpus-based intelligent sampler *Synful*[2] [4], or the interactive concatenative drum synthesiser *Ringomatic* [5], use a distance between descriptor vectors to select the best matching unit.

Spectrum-based systems, on the other hand, perform lookup of single or short sequences of FFT-frames by a spectral match with an input sound stream. Although they have interesting musical applications, e.g. the *SoundSpotter* [6] system[3] with the *Frank* live algorithm, or the audio-visual performance system *Scrambled Hacks,*[4] descriptor-based systems, seem to be more readily usable

---

[1] http://imtr.ircam.fr
[2] http://www.synful.com
[3] http://www.soundspotter.org/
[4] http://www.popmodernism.org/scrambledhackz

for music because the descriptors make sense of the sound database, by pushing the representation higher than the signal level, and thus allowing a compositional approach by writing a target score in terms of sound descriptors.

3. CATART

This section describes the software system CATART [2] that realises real-time CBCS in a simple, intuitive, and interactive way. CATART's model is a multidimensional space of descriptors, populated by the sound units. To perform selection, the user controls a target point in a lower-dimensional projection of that space with a selection radius around it, and the selection algorithm picks the units closest to the target or a random unit within the radius. The actual triggering of the playback of the unit is independent of the selection and can happen at any rate, although for exploration of a sound database the most efficient way is to link both by playing a unit whenever the target point comes close.

To perform **analysis**, CATART first segments input sounds into units from 1/10 to several seconds, either by fixed-size, by pitch change, or by silence. It then calculates sound descriptors on short-time windows and represents each unit by the mean value of fundamental frequency, periodicity, loudness, and spectral descriptors: centroid, sharpness, flatness, tilt, high and mid frequency energy. Segmentation and analysis can also be imported from text or SDIF files. Complete corpora with their soundfiles and analysis data can be stored and reloaded.

Because all analysis takes place inside CATART, we can use real-time audio input that is segmented and analysed on the fly, to feed the corpus, as used in section 4.2.

CATART's **synthesis** can apply manipulations similar to a granular synthesis engine before playing: the length of the grain can be arbitrarily changed to achieve granulation effects or clouds of overlapping grains. Also, changes in pitch by resampling and loudness changes or application of an envelope are possible. Note that, because the actual pitch and loudness values of a unit are known in its descriptors, it is possible to specify precise pitch and loudness values that are to be met by the transformation.

In the standard explorative **interface** of CATART, the descriptor space is reduced to a 2-dimensional projection according to two selectable descriptors plus a 3rd descriptor being expressed on a colour scale (see figure 1). In these displays, the mouse serves to move the target point in the descriptor space and thus to control the playing of units. Additional control possibilities are MIDI input from hardware controllers, and a general message interface with which more than two descriptor target values and their weights can be set by automation envelopes or an external sequencer.

The CATART system is implemented as a collection of patches for *Max/MSP* using the FTM, *Gabor*, and MnM extensions,[5] whose optimised data stuctures and operators, statistical and matrix processing tools, and arbitrary-time grain processing allow efficient implementation of
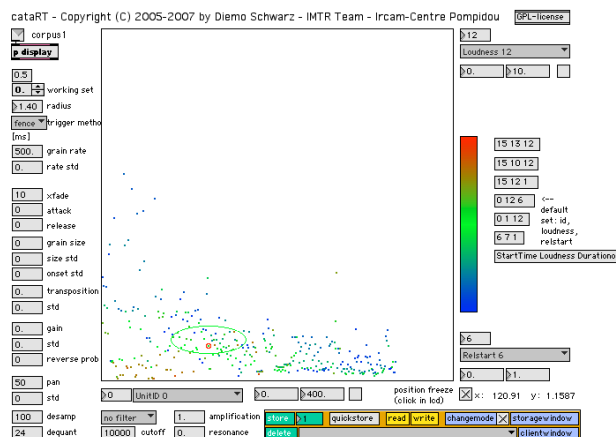

Figure 1. The graphic control interface of CATART.

the signal processing and data handling tasks necessary for such a complex application.

Its modular object-oriented architecture, where any number of selection and synthesis modules can work on one or several corpora, switchable at run-time, allows for the CATART components to serve as a toolbox to build individual applications, and to easily integrate only the necessary components in a concert production patch. CATART is released as free open source software under the GNU general public license (GPL).[1]

4. CONCEPTS AND APPLICATIONS

Of the various musical applications of CBCS that were made since its inception,[6] we have chosen four that, as we reckon, best illustrate the fundamental concepts that are made possible by it, and where the electronic part is exclusively produced by CATART. Other composers using CATART are: Hans Tutschku, Matthew Burtner, Sebastien Roux, Hector Parra, and Luca Francesconi.

4.1. Re-arranging

The concept of re-arranging the units of recorded sound is fundamental to CBCS so that it is at the base of each of the four applications. It can be seen as abolishing the temporal order—time is just another descriptor amongst many that can serve to make new sense of recorded sound.

4.2. Live Recording and Interaction

Two performances took place during the Live Algorithms for Music (LAM) conference 2006,[7] the first, to be released on CD, with the performers George Lewis on trombone and Evan Parker on saxophone, improvising with various computer systems. Here, CATART's live recording capabilities were put to use to re-combine events into harmonic, melodic and timbral structures, simultaneously proposing novel combinations and evolutions of the source material. The audio from the musician on stage,

---

[5] http://ftm.ircam.fr/

[6] See [3] for a corpus-based re-reading of electronic music since 1950.
[7] http://www.livealgorithms.org

was recorded, segmented and analysed, keeping the last several minutes in a corpus from which the system selected units, the target being controlled via a faderbox.

The second performance by Sam Britton and Diemo Schwarz, *Rien du tout*, draws on compositional models proposed by John Cage and Luc Ferrari. Through a process of re-composition it becomes possible to record environmental sounds and interpret and contextualise them into a musical framework. The performance starts with nothing at all (rien du tout) and by recording and re-composing environmental sound (here the sound of the concert hall and audience), evolves a musical structure by tracing a non-linear path through the increasing corpus of recorded sound and thereby orchestrating a counter-point to our own linear perception of time. The aim is to construct a compositional framework from any given source material that may be interpreted as being musical by virtue of the fact that its parts have been intelligently re-arranged according to specific sonic and temporal criteria.

4.3. The Navigable Score

While navigation is also at the base of all the examples, the *Plumage* project [7] exploits it to the fullest, making it its central metaphor. *Plumage* was developed within the ENIGMES project (Exprimentation de Nouvelles Interfaces Gestuelles Musicales Et Sonores) headed by Roland Cahen: a collaborative experimental educational project at the national superior school of industrial creation EN-SCI [8], bringing together design students with researchers from LIMSI and Ircam. Its subject was "navigable scores" or score-instruments, in which different kinds of users would play the sound or music, cruising through the score.

In *Plumage*, CATART was connected to and controlled by a 3D representation of the corpus, giving more expressive possibilities, more precision in the visualisation and interaction, and some new paradigms linked to 3D navigation. Yoan Ollivier and Benjamin Wulf imagined and designed this metaphor (*plumage* means the feathers of a bird) based on 3D modeled feather-like objects representing sound grains, allowing to place them in space, link them, apply surface colouring and texturing, rotation, etc., according to the sound descriptors of the grains they represent.

During first tests, the feathers happened to be quite spread out into the 3D world. Consequently, moving around a simple point trigger to play the sound grains one by one was not satisfactory. Several strategies where therefore tried: using variable-dimension triggers was not dynamic enough. We thus decided to build 3 controllable reading heads, each one composed of 3 triggers orbiting like electrons around a nucleus. Each orbit speed, size and orientation is controlable. Each head can follow a different cyclic trajectory. A little like a percussionnist who will play with three hands, each one holding three sticks. This enrichment shifted the sound result from elementary sound triggering to musical process.

How does Plumage realize the project of a navigable score? The navigation is not a simple movement of a cur-

Figure 2. *Plumage*'s 3D space.

sor because the score cannot be played by a single instrument. It is rather comparable to an orchestral score, like a collection of sounds, which can be played by a group of instruments according to certain rules.

The score is not fixed but in an open form, rather like Earl Browns approach: "It will never come out the same, but the content will be the same." "Is it more interesting to fill a form or to form a filling?" In *Plumage*, both the composition of the score and the navigation can be set very precisely or not, and the setting of the navigation can become a way to fix a composition, to study a sound corpus or to navigate freely as an improvisation. Concatenative synthesis can be compared to deconstructing a picture into small points and a classification of these points according to the chosen descriptors. Imagine separating all the brush spots of an impressionist painting and reordering them according to hue, luminance, saturation or other descriptors. The resulting work will be an abstract painting. Navigating through such a score will not easily rebuild the original figure and will need special processes such as the one we developed to make this reshaping significant.

4.4. Interpolation and Cross-selection

Stefano Gervasoni's piece *Whisper Not* for viola and electronics, created in April 2007 in Monaco, played by Genevieve Strosser, computer music realization by Thomas Goepfer, explores the interaction of the musician with her own sound, segmented into notes and short phrases. Here, CATART improvises a response to the musician as soon as she makes a pause, recombining her prerecorded sound according to a trajectory through the descriptor space, controlled via a faderbox.

Further on in the piece, the corpus of viola, with playing styles intended to create a resemblance to the human voice, is gradually interpolated with a second corpus of only pizzicato sounds, and then morphed into a third corpus of sounds of dripping water. Here, a new concept of corpus-based cross synthesis, or shorter *cross-selection* is applied: The descriptors of the selected response of CATART are taken as the target for the parallel third cor-

pus, such that the pizzicatos are gradually replaced by water drops, while retaining their timbral evolution.

4.5. Harmonic Selection and Corpus-Based Orchestration

Dai Fujikura's piece *swarming essence* for orchestra and electronics, created in June 2007 with the orchestra of Radio France in Paris, computer music realization by Manuel Poletti, uses 10 different corpora of pre-recorded phrases of 5 instruments (alto flute, bass clarinet, trumpet, violin, cello), segmented into notes. The phrases making up the sound base were composed to match the harmonic content of the orchestral part of the 10 sections of the piece, and to exhibit a large sonic variety by use of different dampers and playing styles.

The composer then explored each corpus graphically, recomposing and manipulating the sound material using CATART's granular processing capabilities. These trajectories were then transcribed into control envelopes for the concert patch (see figure 3). Each corpus was in-
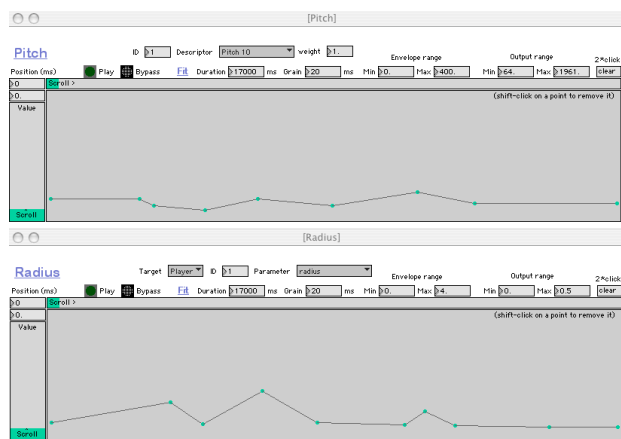

Figure 3. Graphic score for *swarming essence*.

ternally organised into sound sets by instrument, giving precise control of the orchestration of the electronic part by instrument-dependent routing, allowing their separate granularisation and spatialisation.

In this piece, the encounter of the composer with CATART also induced the inverse influence that the composition of the orchestral part was made to follow sonic effects that were to be obtained by CATART, in order to smoothly link both. For instance, the composer thought in terms of the "grain size" of the orchestra's playing.

5. DISCUSSION AND CONCLUSION

We can see that 3 of the 4 examples used the sound of the performers. For all three, the initial idea was to use the live sound to constitute a corpus from which CATART would then synthesise an electronic accompaniment. In the end, however, Fujikura and Gervasoni chose to prerecord the corpus instead, because of the better predictability of the sonic content of the corpus, in terms of both quality and variety. In Britton and Schwarz's use of

live corpus recording, the impredictability of the incoming material was either an integral part of the performance, as in *Rien du tout*, or inevitable, as with the LAM performance, because of the improvised nature of the music.

We see that precise knowledge of the corpus is a great advantage for its efficient exploitation. The 2D display of the descriptor space helps here, but can not convey the higher-dimensional shape and distribution of the space. We are currently exploring ways to represent the corpus by optimising its distribution, while still retaining access by musically and perceptually meaningful descriptors by dimensionality reduction.

CATART is a tool that allows composers to amass a wealth of sounds, while still retaining precise control about its exploitation. From the great variety of musical results we presented we can conclude that CATART is a sonically neutral and transparent tool, i.e. the software doesn't come with its typical sound that is imposed on the musician, but instead, the sound depends completely on the sonic base material and the control of selection, at least when the granular processing tools are used judiciously.

CATART's modular architecture proved its usefulness for its inclusion in concert patches, being able to adapt to the ever changing context of computer music production.

7. REFERENCES

[1] Diemo Schwarz. Corpus-based concatenative synthesis. *IEEE Sig. Proc. Mag.*, 24(2), March 2007.

[2] D. Schwarz, G. Beller, B. Verbrugghe, and S. Britton. Real-Time Corpus-Based Concatenative Synthesis with CataRT. In *DAFx*, Montreal, Canada, September 2006.

[3] Diemo Schwarz. Concatenative sound synthesis: The early years. *Journal of New Music Research*, 35(1), March 2006.

[4] Eric Lindemann. Music synthesis with reconstructive phrase modeling. *IEEE Sig. Proc. Magazine*, 24(1), March 2007.

[5] JJ. Aucouturier and F. Pachet. Ringomatic: A Real-Time Interactive Drummer Using Constraint-Satisfaction and Drum Sound Descriptors. In *ISMIR*, London, UK, 2005.

[6] Michael Casey. Acoustic lexemes for organizing internet audio. *Contemporary Music Review*, 24(6), 2005.

[7] Ch. Jacquemin, R. Ajaj, R. Cahen, Y. Ollivier, and D. Schwarz. Plumage: Design dune interface 3D pour le parcours déchantillons sonores granularisés. In *IHM*, Paris, France, November 2007.