

AUDIOGARDEN: TOWARDS A USABLE TOOL FOR COMPOSITE AUDIO CREATION

Christian Frisson¹, Cécile Picard², Damien Tardieu³

¹ Laboratoire de Télé-détection et Télécommunications (TELE), Université catholique de Louvain, Louvain-la-Neuve, Belgium

² Freelance researcher / [pl-area](#)

³ Laboratoire de Théorie des Circuits et Traitement du Signal (TCTS), Université de Mons, Belgique

ABSTRACT

This project presents a new approach to sound composition for soundtrack composers and sound designers. We propose a tool for usable sound manipulation and composition that targets sound variety and expressive rendering of the composition. We first automatically segment audio recordings into atomic grains which are displayed on our navigation tool according to their timbre. To perform the synthesis, the user selects one recording as model for rhythmic pattern and timbre evolution, and a set of audio grains. Our synthesis system processes then the chosen sound material to create new sound events based on onset detection of the recording model and similarity measurements between the model and the selected grains. A large variety of sound events such as those encountered in virtual environments or other training simulations.

KEYWORDS

MediaCycle, Multimedia Databases, Content-based Navigation, Interfaces, Sound Design, AudioGarden

1. INTRODUCTION

Soundtrack composers and sound designers aim at creating auditory experiences [2]. In order to produce soundtracks for movies or video games, Foley artists mainly rely on prerecorded sound material, or record it themselves. While the use of prerecordings is easy to implement, the number of samples in a database is often limited due to memory constraints. Another possibility to generate such sounds is sound synthesis. A large variety of synthesis methods exist, but each of them is usually more suited for a reduced range of sounds. A very common technique for texture synthesis is the data driven concatenative synthesis, also referred to as mosaicing [8]. Concatenative synthesis approaches aim at generating a meaningful macroscopic waveform structure from a large number of shorter waveforms. They typically use databases of sound snippets, or grains, to create a given target phrase. Unlike granular synthesis where no analysis is performed on the audio units and where the unit size is defined arbitrarily [7], concatenative synthesis selects the audio units according to a set of audio descriptors. Physical modeling can be introduced to further refine granular synthesis [3, 1]. A very important issue for applications of granular synthesis to sound design is the control of the synthesis process. Vocem, introduced by Lopez et al. [5], is one of the first graphical interfaces for real-time granular synthesis, with high-quality audio output and very short latencies. Parameters allow the user to easily control the creation and the distribution of the grains. With MoSevius, Lazier et al. [4] first attempt to apply unit selection to real-time performance-oriented synthesis with direct and intuitive controls based on descriptor values such as energy, spectral flux or spectral centroid, as well as voicing and instrument name. For a more musical context, Misra et al. [6] focus on a single framework

that starts with recordings and proposes a flexible environment for sonic sculpting in general. Another class of control methods relies on a wise visualisation of the grains database in order to adequately select them. In Catart, Schwarz proposes to display the grains in a two-dimensional space according to descriptor values or output of dimension reduction techniques such as multidimensional scaling analysis or principal component analysis [8]. Following these ideas, we propose an approach that combines hypermedia navigation and a synthesis process into an adequate multimodal user interface for sound composition and design.

Our specific contributions are:

- a method for automatic analysis of audio recordings, extraction and classification of meaningful audio grains as new database.
- a technique for automatic synthesis of coherent soundtracks based on the arrangement of audio grains in time.
- a usable interface for database manipulation and sound composition.

2. CREATING SOUNDS

2.1. Synthesis Method Overview

The synthesis method is based on content driven concatenative synthesis. The main idea is to segment a target sound, that will be used has a time structure model and timbre evolution model, and to replace each segment by a grain contained in a database. This method involve four steps: segment the target, extract feature from the target segments and the grains, choose the grains that will replace each target segment and finally concatenate the chosen grains to create the final sound.

2.2. Sound Segmentation

2.2.1. Method

Sounds are segmented by finding the local minima of the spectral flux. This simple method allows to find onsets quite reliably. One problem is that onsets do not always coincides with energy minima resulting in a segmentation after the actual beginning of the sound and to small clicks in the synthesis. An alternative method would be to segment by finding the local minima of the energy of the sounds. This method results in less clicks in the synthesis but the rhythm feeling can be lost in the synthesis because energy minima do not correspond to perceived onsets. So the best solution would probably be to segment using energy minima and keep onset information (obtained using spectral flux) as a feature to be used during the synthesis process. The synthesis method thus needs to be adapted.

2.2.2. Segmentation In Mediacycle

Segmentation utilities have been added to Mediacycle (see Fig. 1). First, each media can now have children represented by a vector of pointers to media. Second, the support for segmentation plugins has been added, allowing various kind of segmentation depending on the type of media and on the application. The role of the plugins is to make all the computation allowing to find the segment boundaries and then to add these segments to the children vector. We developed one such plugin that implement the segmentation method previously described.

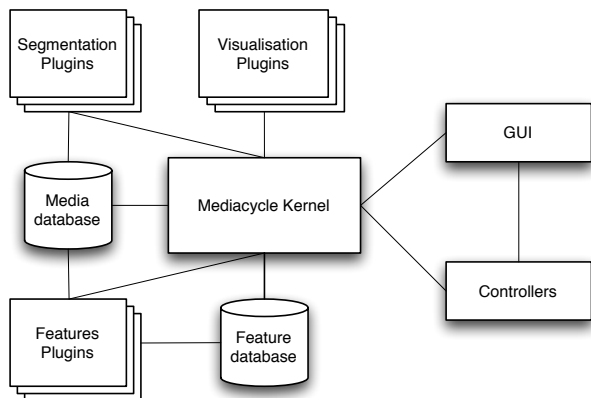


Figure 1: Mediacycle Architecture

2.3. Audio Features

To describe the grains and the target segments, the following features are computed:

- MFCC : to describe the spectral envelope of the sounds,
- SFM : to describe the noisiness of the sounds,
- Duration,
- 10-point temporal envelope interpolation.

2.4. Choosing Grains

The choice of the grain that will replace a segment is a very important issue in the synthesis process. It raises the problem of similarity measurement, and further the problem of similarity measurement in context. That is, if one chooses a grain to replace the first segment this will impact the subsequent similarity measurements. An other example of the kind of problem that can arise is given in figure 2. The grains and the target segments are displayed in a 2 dimensional feature space. In this example, if we choose the grain by similarity, the same one will always be chosen whatever is the segment. So we have to provide either transformation of the feature space or mapping function that allows consistent choice of the grains.

We propose three different methods:

- subtract the mean of both feature sets,
- subtract the mean of both feature sets and normalize the standard deviation,
- do nothing.

Those are very simple and further research still has to be done, but the second one, for instance, give very good results when the target and the grains are drum sounds from different drum kits. By normalizing the mean and standard deviation, the sounds of similar items of the different drum kits (snare drum, kick drum . . .) end in the same region of the space, so the grain choice is very consistent (example).

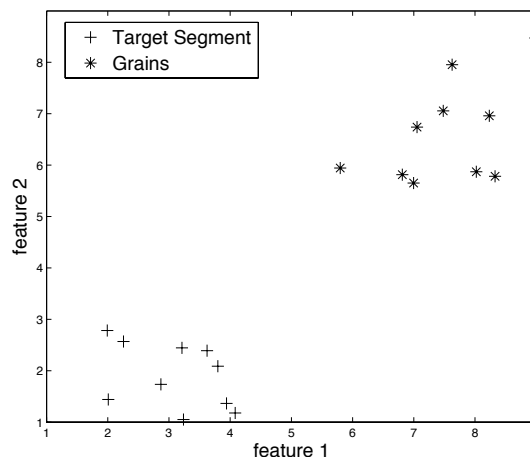


Figure 2: Example of similarity problem

2.5. Adding Grains

After selecting the grains, we need to add them in order to create the final synthesized sound. First of all we apply a tapered window on the grain in order to guarantee the smoothness of the composition. Then we propose three different addition methods:

- Simple: Grains are positioned at the same position as the original segment,
- Squeezed: Grains are concatenated, so rhythmic properties of the target is not preserved,
- Padded: for each segment, the closest grain is added, then if this first grain is shorter than the segment, the second closest is concatenated.

In the first and third cases, grains iteratively added to the current sound, starting from an empty sound, to allow superposition between grains.

3. USER INTERFACE DESIGN

3.1. Prototyping with mockups and storyboards

As there are very few computerized systems or analog practices that propose a workflow similar to the method we described here, we had to design a user interface fed by our own creativity. To achieve a certain level of mutual understanding of what we believe to be a suitable design, we produced throughout several brainstormings many mockups of the visual user interface and a storyboard of the expected scenario of usage, as illustrated in Figure 3, using paper [9] or whiteboards (shooting backups with cameras).

Drawing mockups prevented us from diving directly into the implementation of software prototypes, particularly a two-browser solution (one for selecting rhythmic patterns from sound events, the second for timbres from audio grains) that would have been harder and slower to implement and less straightforward in terms of interaction than the solution we opted for, a single browser revealing temporal and timbral features.

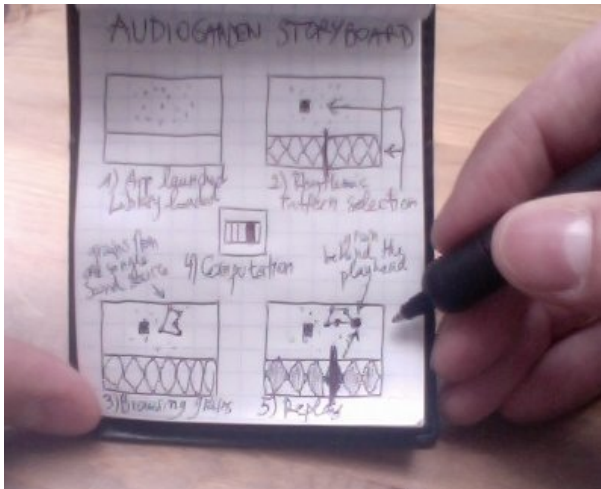


Figure 3: Storyboard of an expected scenario of usage of the desired workflow quickly drawn on a small notebook.

3.2. Proposed scenario of interaction

The scenario consists in:

1. browsing, listening to and selecting:
 - (a) one sound event for its rhythmic pattern,
 - (b) several audio grains for their timbral character,
2. easily constructing a new sound event that “updates” the chosen sound event with different timbral features;
3. listening to the new sound event, optionally saving it (thus making it appear on the browser);
4. renewing the aforementioned cycle (steps 1. and 2.), by either choosing another sound event or different grains, or starting again with no audio content set.

3.3. Proposed Visualisations

3.3.1. Disc

The first proposed visualisation is shown on Fig. 4. The position of the points is computed as follows:

- The radius is proportional to the inverse of the logarithm of the duration of the sounds. Thus long sounds, that can be used as targets are positioned in the center of the display and short sounds that can be used as grains are on the periphery of the circle.
- The angle depends on the timbre features (for now only MFCC). It is proportional to the coordinates of the sound on the first principal component of the MFCC.

This visualisation is useful to explore the grain database and to experiment with the synthesis method.

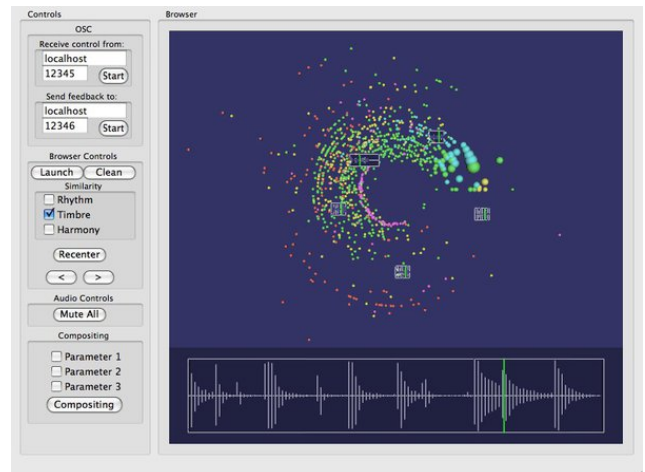


Figure 4: Screenshot of an early prototype of the user interface, featuring a two-pane view: audio database browser by similarity on top, waveform of the sound being “composed” on the bottom.

3.3.2. Flower

The second visualisation is shown on Fig. 5. Each long sound (a sound that has been segmented) is represented by a circle. The long sound itself is in the center of the circle, while the segments of this sound form a circle around it. The grains in the circle are in chronological order. The circles are placed in the 2D space depending on the average timbre, that is, the coordinates of the center of the circles are proportional to the two first principal components of the MFCC and SFM features.

In this visualisation it is possible to select the grains one by one, or if one clicks on the center of a circle while hitting a special key, the entire circle is selected at once. This visualisation is very useful to mix sounds, i.e. using one sound as a target and all the segments of one or several other sounds as grains.

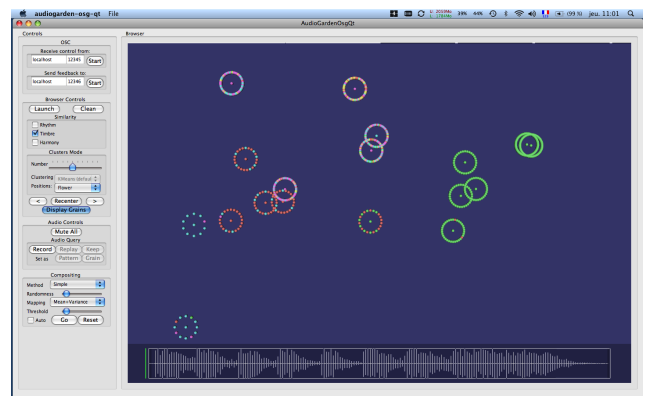


Figure 5: Screenshot of an early prototype of the user interface, featuring a two-pane view: audio database browser by similarity on top, waveform of the sound being “composed” on the bottom.

4. SUMMARY OF MEDIACYCLE IMPROVEMENTS

- Waveform display at the bottom of the window to show one sound of particular interest, such as the target or the synthesis in audiogarden,
- Possibility to record a sound from the computer input, play it back and add it to the library,
- Internal support for media segments,
- Support for segmentation plugins.

5. CONCLUSION AND PERSPECTIVES

In this project, we designed and developed a tool for sound creation. First of all, it has been the occasion to test the flexibility of the Mediacycle framework. In the course of the development, Mediacycle has proved to be a very good tool to quickly design new audio application and test various kind of sound analysis and data display on screen. By allowing a very fast prototyping, it allows to test various configurations and then select the best one in a very short period of time. In addition with paper mockups, such frameworks can be very useful tools for research and development. Some improvements have also been done, such as the support for segments that will be useful for many other applications. But the main achievement of the project is the AudioGarden software. Standing on content driven concatenative synthesis, this software proposes a new way to create sounds. Informal tests showed that a large variety of sounds can be created, but still more formal tests needs to be performed. Some aspect needs improvements. The synthesis method could be improved in many ways. First the segmentation and grain addition could be changed such as described in section 2.2 to allow both a segmentation on low energy part and a synthesis that preserve the rhythm of the target. New mappings between the target and the grains could also be explored. We proposed three simple ones, that give very good results for some kind of sounds, but sometimes unexpected results on other sounds. This may involve sound similarity perception researches and experiment. Finally different visualisation could be proposed for different usages, this would need to give the tool to different kind of users and look at the way they use it and listen to their suggestions.

6. ACKNOWLEDGMENTS

numediart is a long-term research program centered on Digital Media Arts, funded by Région Wallonne, Belgium (grant N°716631).

C. Picard obtained a Short-Term Scientific Mission (STSM) funding from the COST Action Sonic Interaction Design (SID) ¹.

We want to thank the One Laptop Per Child Project (OLPC) for providing the Free Sound Samples Library under a Creative Commons license ².

7. REFERENCES

- [1] P. R. Cook. "Toward Physically-Informed Parametric Synthesis of Sound Effects." In: *In Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA-99)*. 1999. Pp. 1–5. P.: 33.

- [2] Christoph Cox and Daniel Warner, eds. *Audio Culture: Readings in Modern Music*. Continuum International Publishing Group, 2004. ISBN: 9780826416148. P.: 33.
- [3] D. Keller and B. Truax. "Ecologically-based Granular Synthesis". In: *Proceedings of the International Computer Music Conference (ICMC)*. Ann Arbor, USA 1998. P.: 33.
- [4] Ari Lazier and Perry Cook. "MOSIEVIUS: Feature driven interactive audio mosaicing". In: *Proceedings of the International Conference on Digital Audio Effects*. London, UK 2003. P.: 33.
- [5] Daniel Lopez, Francesc Marti, and Eduard Resina. "Vocem: An Application for Real-Time Granular Synthesis". In: *Proceedings of the Digital Audio Effects (DAFx)*. 1998. P.: 33.
- [6] A. Misra, P. R. Cook, and G. Wang. "Musical Tapestries: Re-composing Natural Sounds". In: *Proceedings of International Computer Music Conference (ICMC '06)*. New Orleans, USA: International Computer Music Association, 2006. P.: 33.
- [7] Curtis Roads. "Introduction to Granular Synthesis". In: *Computer Music J.* 12.2 (1988). P.: 33.
- [8] Diemo Schwarz. "Concatenative Sound Synthesis: The Early Years". Ed. by Adam T. Lindsay. In: *Journal of New Music Research* 35.1 (Mar. 2006). Pp. 3–22. P.: 33.
- [9] Carolyn Snyder. *Paper Prototyping: The Fast and Easy Way to Define and Refine User Interfaces*. Morgan Kaufmann, 2003. ISBN: 1558608702. P.: 34.

¹COST SID: <http://www.cost-sid.org>

²OLPC Sound Sample Library: http://wiki.laptop.org/go/Sound_samples