Audio browsing strategies for soundscape real-time composition

email@panayiotiskokoras.com
Aristotle University of Thessaloniki

ABSTRACT

This paper discusses audio browsing strategies used in the soundscape composition with the title *City Fables: Corfu*. The author used *Soundtorch* to demonstrate the strategies. This is a software application that uses a sophisticated suite of algorithms to analyse and intelligently classify an audio sound database. Examples from the work *City Fables* provide a platform for the analysis and explain the proposed formulations. The strategies used in the piece comprised: 1) the preparation of a sound database, 2) the preparation of a software session, and 3) performance strategies such as Texture / Polyphony, Sound in Space, Similarity and Control / Gesture. Audio browsing strategies could become integrated into specific methods of composition and performance, and this could aid the composer towards sophisticated structures, control, and coherence in a work, and the same time provoke intuition, serendipity and creative thinking.

1. Introduction

Developing tools that use automated acoustic content analysis to control and structure large sound collections during real-time composition is of interest to composers, performers and researchers (Eigenfeldt, Pasquier, Birchfield, Janer, Schwarz). However, the challenges are many, as Barry Truax has written "… arbitrary juxtaposition of the sounds prevents any coherent sense of a real or imagined environment from occurring" (Truax, 2002). It is a great challenge for the composer to maintain coherence throughout a work and keep its structure clear and musically meaningful. It is possible to end up with a mess of sound particles looking for a reason for their existence. How can the composer and the performer handle a large collection of environmental sounds in real time? How can they organize the material in a meaningful, intuitive and efficient way for real-time composition performance?

The aim of this article is to show how audio browsing strategies were used in the piece *City Fables: Corfu* (Kokoras, 2011). It will describe what steps were taken before the actual performance and what techniques were used during the performance. To exemplify the strategies, I used *Soundtorch* (Heise *et al.*, 2011). This is a software application that uses a sophisticated suite of algorithms to analyse and intelligently classify the audio collection. It displays the sounds in 'bacteria'-like form and the user can play them by hovering over them with a torch-like pointer.

2. The Tools

There are several recent approaches that use a database of source sounds, segmented or not into short units. They are arranged according to their timbre proximity into a multidimensional space of descriptors and played by the user hovering over them using the mouse. As well as the variety of software applications in development that can handle audio browsing, there is also in development the techniques and strategies which a composer/performer can use in his/her compositions using this software.

CataRT[1] Real-Time Corpus-Based Concatenative Synthesis (Schwarz, 2007) is probably the only programme of its kind made specifically for real-time composition, among others taking advantage of audio content-based analysis methods. The concatenative real-time sound synthesis system CataRT plays grains from a large corpus of segmented and descriptor-analysed sounds according to proximity to a target position in the descriptor space.

[1] http://imtr.ircam.fr/imtr/CataRT

The TimbreID[2] (Brent, 2007) is a collection of timbre feature analysis externals for [Pd]. It is able to find the best match between an input feature and previously-stored instances of training data. It is also designed to facilitate real-time concatenative synthesis and timbre-based orderings of sound sets.

Soundscrotch[3] and some media players such as Mufin Player[4] are totally different in scope from real-time composition. In Mufin Player (Schönfuß, 2011), the user can navigate through the audio in a 3D space. Each track is displayed as a circle. Tracks are arranged according to their audio properties: from sad to happy (left to right), synthetic to acoustic (bottom to top), and from calm to aggressive (front to back) (*see* Figure1).

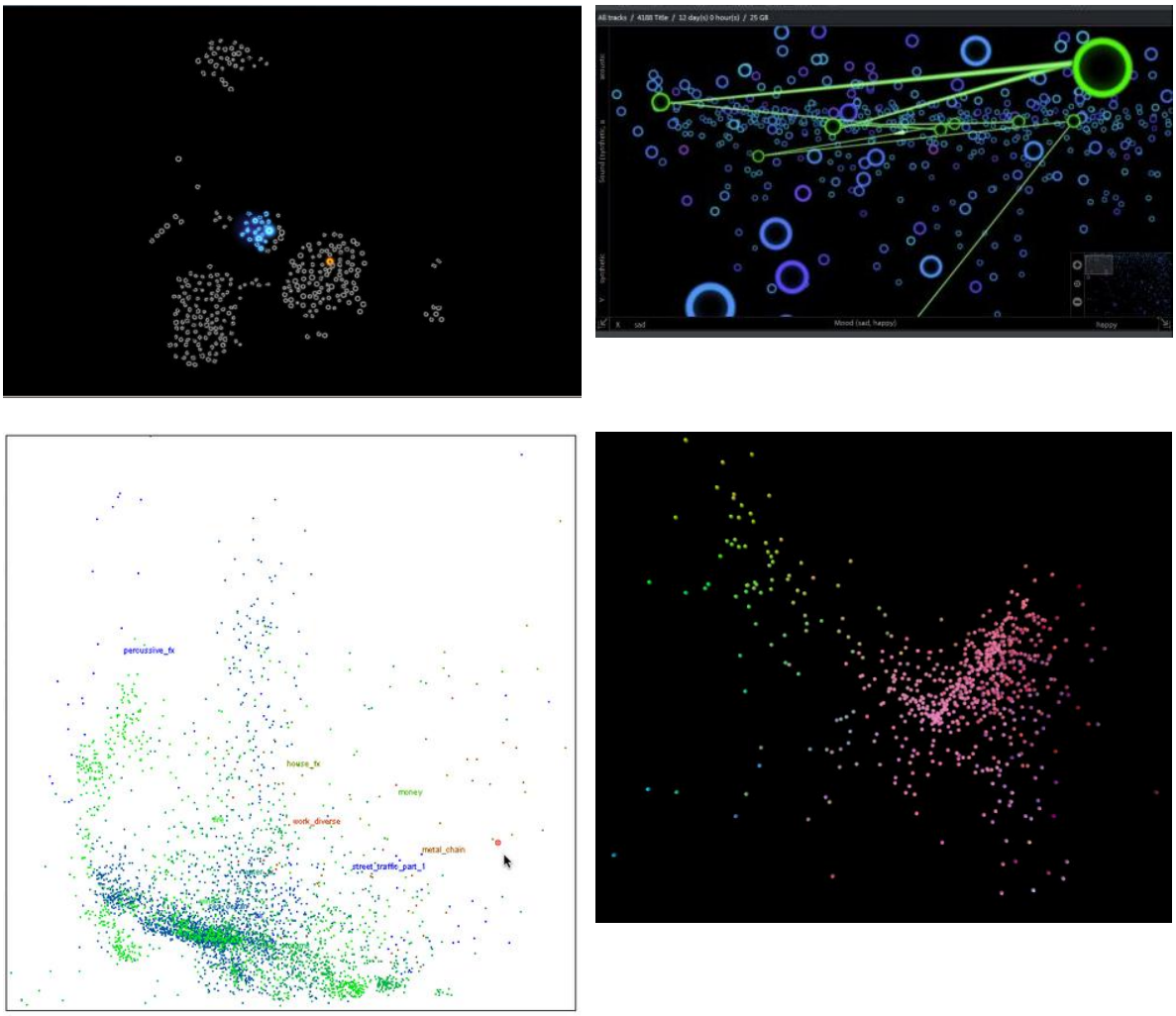

FIGURE 1. From top left clockwise: Soundtorch, Mufin Player, CataRT and TimbreID

Finally, the Soundtorch (Heise *et al.*, 2011) is basically made for SoundFX audio browsing to help sound designers to find a sound for any particular project more quickly and more easily. It is clear that this software is not made for real-time composition/performance; on the contrary, it is dedicated for studio use. However it was a pleasure to use it as a real-time composition performance tool. Soundtorch is unique of its kind, simple and extremely efficient. Its immediacy allowed me to think right away about the composition and the performance.

[2] http://williambrent.conflations.com/pages/research.html
[3] http://www.soundtorch.com
[4] http://www.mufin.com/gb/software/mufinplayer

3. Audio browsing strategies

One of the main difficulties a composer would face during the performance of a real-time soundscape composition is to find the appropriate sounds within his/her collection and control them in a meaningful way. Most of the time, the conventional browsing techniques of naming or tagging might be imprecise, general and very slow. Software applications that provide sound content analysis, fast auditioning on 2D or 3D plots, and mouse navigation for playback and gesturing are of great help.

In this section, I shall explain some of the strategies which I applied to my composition *City Fables: Corfu* which I prepared for the World Forum of Acoustic Ecology Conference in October 2011, held on the island of Corfu, in Greece.

The methodology of completing a real-time composition has three main stages:

- Preparation of a sound database;
- Preparation of software session; and
- Performance strategies.

The final stage will be further expanded by analysing several strategies for creative audio browsing. I am going to provide a platform for the analysis and explain the proposed formulations. Strategies of this kind used in the piece included the Texture / Polyphony Sound in Space, Similarity and Control / Gesture.

3.1. Preparation of a sound database

The first step is to prepare the sound material in an appropriate way that will be useful during the performance. The sound database is prepared carefully in advance and includes from short simple sounds to long articulated sound gestures. There is no *a priori* segmentation on the sound files. The duration for which a sound is looping depends on the time a sound is illuminated by the light of the torch. A fast hovering across the sound 'bacteria' is played only at the beginning of the sound files. To make gestures of this kind more effective, the sound files should be prepared with suitable attacks so as to obtain the expected results.

Although it might sound inconvenient, this phase requires a lot of time. The sound database might consist of several hundreds of sounds properly edited and processed. A fixed electroacoustic piece of the same duration (about 10 minutes) could require about 700 sounds (Kokoras, 2010). A similar number of samples could give a good analogy. The main steps of the preparation of a sound database are:

- Feed the system with sounds that will loop;
- Edit simple sounds so the combination of many of them creates a rich texture without them being overwhelmed; and
- Prepare similar sounds by hovering around them and creating gestures.

3.2. Preparation of the software session

The second step is to set up the sessions of the software to be used for the piece in question. This phase first involves the loading of the sound files prepared earlier into the system for analysis and placement into the multidimensional plot. After that has been done, the following actions should be taken:

- Save the initial map for immediate use on subsequent occasions;.
- Create sub-maps with different clustering of the sound database to use it in parallel as a second or third instance of the application;
- Choose particular sounds and set them to function as magnets to attract the sounds which have similar audio content;
- Create various sound clusters within the plot, as structural pillars for the composition, ready for playback.

The capability to apply colour and/or labelling on areas that have a structural significance could be a helpful feature.

3.3. Performance strategies

In the final step, the user should become familiar with the sounds in relation to their placement in the multidimensional space. In the following paragraphs, four main elements will be explained that can contribute to the performance of the piece: Textures / Polyphony, Sound in Space, Similarity and Control.

3.3.1. Textures / Polyphony

The vertical organization, known as harmony, of the sound material is one of the most important elements. The simultaneous playback of a great number of samples clustered according to similar sound content allows the creation of different textures from organic soundscapes to schizophonic sonic layers, or just focuses on only one sound at a time. When the Soundtorch plays back simultaneously a large number of samples with similar sound content, it is easy to create a natural soundscape. Moreover, with small changes to the torch pointer the sound loops are always different.

Another way of layering sound material is the use of multiple instances. A constraint in this case was that Soundtorch could only playback a single sound stream from the instance that was in the foreground. The advantage was that I could prepare different arrangements of sounds and easily play them. In the work *City Fables: Corfu*, the output of the Soundtorch was routed to another application so to be able to record and play back specific passages.

3.3.2. Sound in Space

Another parameter is the use of space. Soundtorch positions each playing sound on a surround system. The right front area of the illuminating light corresponds to the right channel, the left to the left channel and so on. As a result, each sound is dynamically diffused according to the position of the virtual torch. Moreover, zooming in and out of the torch adds more sounds in the middle of the torch and therefore in the middle of the surround, and *vice versa*. With one cluster of percussive sounds from the right and another cluster with continuous sounds from the left, it is easy to obtain a spatial polyphony.

An extension of this idea would be to assign the physical space of the hall to the 2D timbral space with the four corners representing the LR front and rear speakers in a quadraphonic setup. This arrangement will give the option to set the spatial polyphony right from the beginning and instead move the sounds or the clusters in the 2D plot to achieve additional motion.

3.3.3. Similarity

Further homogeneity or contrast of the sonic material can be achieved by applying suggestions of the most similar sounds within the whole sound collection in relation to a particular sound. In other words, specific sounds are pointed out to function as magnets which attract the sounds that have similar audio content.

This is something which enables the user to know and control everything and nothing: in the use of Soundtorch the user needs to know or care nothing about the similarity algorithms and procedures. Soundtorch is driven by a C.A.S.E. (Computer Aided Sound Exploration) engine, a sophisticated suite of algorithms that analyse and intelligently classify an audio collection. All sounds are immediately arranged in a meaningful way and are ready to be played back. All the user has to do is to try it for a while and understand how it functions. The logic behind it will be easily understood and taken into account.

Further homogeneity and at the same time variety of the sonic material can be achieved by applying suggestions of the most similar sounds within the whole sound collection in

relation to a particular sound. If a sound is of particular importance in the piece, its functionality can be enhanced by asking Soundtorch to suggest the most similar sounds from the whole collection in the hard disk, and then import them to the timbre space and cluster. Furthermore, a sound can be magnetized and Soundtorch will find and attract all the similar sounds from within the plot. As the magnetized sound changes position, all similar sounds follow and create new clusters.

3.3.4. Control of the Gesture

The mouse wheel is used to zoom in or focus the light beam to a single sound: the closer the focus of the light, the louder the sound. This effect provides a powerful tool to mute or unmute a number of tracks with a single movement. A single water drop can be slowly turned into heavy rain with wind and thunder. The free navigation of the torch around the plot at various speeds and in various directions can produce gestures from bright to dark, noisy to harmonic or low to high. Moreover, fast small movement of the torch can create a granular effect with an adjustable thickness.

The simultaneous use of more than one torch in one space by selection could provide richer sonic results. In addition, the possibility of creating trajectories within the space and triggering them at will would generate rather elaborate articulations and gestures. An extra feature of midi mapping capabilities would give more options to the control parameters. Another feature for real-time composition which uses audio browsing technology would be the addition of spot-lights which add effects and at the same time instantly rearrange the sounds to a new position in the timbre space as a result of the effect.

4. Conclusion

The strategies described above were helpful and suitable for a particular soundscape composition and the particularities it required, but they could easily be adapted to any kind of real-time composition that uses audio browsing technology.

It should be mentioned here that Soundtorch is not made for real-time performance. On the contrary, it is a studio tool for content-based audio search intended for sound designers who deal with large numbers of sound effects. The limitations and intentions described above by no means cover all the possibilities of this excellent piece of software. It is been used as a simple paradigm to show how the author used audio browsing strategies to structure his composition.

Audio browsing strategies are essential for the development and refinement of musical ideas and might help with the development of relevant software tools. They could become integrated into specific methods of composition and performance. They can support the composer in creating sophisticated structures, in refining the control of parameters, and in maintaining coherence in a work. At the same time, they can provoke intuition, serendipity and creative thinking in the composer.

5. References

William Brent (2009). *Cepstral analysis tools for percussive timbre identification*, in Proc. 3rd International Pure Data Convention, Sao Paulo, Brazil.

William Brent, *A Timbre Analysis and Classification Toolkit for Pure Data* <http://williambrent.conflations.com/papers/timbreID.pdf> (accessed on 19 September 2011).

Arne Eigenfeldt (2008). *Intelligent Real-time Composition*. eContact 10.4 Live-electronics-Improvisation-Interactivity in Electroacoustics <http://cec.concordia.ca/econtact/10_4>. (Accessed on 1 September 2011).

Mikael Fernström (2005). ‘Reflections on Sonic Browsing: comments on Fernström and McNamara’, ICAD 1998. *ACM Trans. Appl. Perception* 2, 4, 500-504.

Barry Truax (2002). ‘Genres and techniques of soundscape composition as developed at Simon Fraser University’. *Organised Sound* 7(2): 5-14.

Sebastian Heise, Michael Hlatky & Jörn Loviscach (2009). *Aurally and Visually Enhanced Audio Search with SoundTorch*. CHI 2009, April 4–9, Boston, Massachusetts, USA.

Sebastian Heise, Michael Hlatky & Jörn Loviscach (2008). *Sound-Torch: quick browsing in large audio collections*, in Proc. 125th Convention of the Audio Engineering Society, Paper 7544.

Panayiotis Kokoras (2010). Program notes of *Magic for tape*. <http://www.panayiotiskokoras.com/en/works.html> (Accessed on 15 September 2011).

Thor Magnusson (2006). *Screen-based musical interfaces as semiotic machines*, in Proc. NIME, 162-166.

Dirk Schönfuß (2011). ‘Content-Based Music Discovery’, in Ystad, Sølvi *et al.*, *Exploring Music Contents*. Berlin / Heidelberg: Springer, 356-360.

Diemo Schwarz, Sam Britton, Roland Cahen & Thomas Goepfer (2007). *Musical applications of real-time corpus-based concatenative synthesis*, in Proc. International Computer Music Conference, 47-50.

Klaus Wassermann, Kynan Eng, Paul Verschure & Janatas Manzolli (2003). ‘Live Soundscape Composition Based on Synthetic Emotions’. *IEEE Multimedia* 4, 82–90.