Freesound Radio: supporting music creation by exploration of a sound database

Gerard Roma
Music Technology Group
Universitat Pompeu Fabra
gerard.roma@upf.edu

Perfecto Herrera
Music Technology Group
Universitat Pompeu Fabra
perfecto.herrera@upf.edu

Xavier Serra
Music Technology Group
Universitat Pompeu Fabra
xavier.serra@upf.edu

ABSTRACT

The habit of sharing media online has created a platform with great potential for creative applications that are accessible to large numbers of users with very different backgrounds. As an example, a lively community has grown around *Freesound.org* to share sound files typically o be reused in music and multimedia content. However, in order to fully realize this potential, new interfaces are needed beyond concept searching to discover interesting multimedia content. We describe *Freesound Radio*, an experimental environment that allows users to collectively explore the content in *Freesound.org* by listening to combinations of sounds represented using a graph data structure. Users can create new combinations from scratch or from existing ones. A continuous supply of potential combinations is provided by a genetic algorithm for the radio to play.

INTRODUCTION

Media sharing has become one of the most prominent uses of the internet. Unlike traditional entertainment, media sharing is driven by users who enjoy uploading, commenting and rating all kinds of multimedia content. Started as an academic research project, *Freesound.org* has become one of the most widely used sites for sharing sound files licensed under a *Creative Commons* (CC) license. Sharing sounds, however has different implications than sharing other media such as video or images. Since the early experiments of *Musique concrète* in the 60s [Schaeffer 1966], sound recordings have been commonly used as building blocks for musical and multimedia compositions. As computers have gained a central position in music and sound production, the practice of reusing sound recordings has become standard. Hence, the activity of sound sharing is not only carried for the sake of sharing experiences and opinions, but also because of the possibilities for recombination of the sound materials. Because the interest of a given sound may not be directly related to its source (which may be not even recognizable in the final mix), text based search and retrieval may be limiting the use of large sound databases for creative applications.

A good deal of research has been done on the issue of content based retrieval to support creative uses of sound databases. Musical mosaicing [Zils and Pachet 2001] refers to a way of building music pieces by specifying a target using sound descriptors (typically extracted from an existing piece or performance). The target is then used as a set of constraints to automatically retrieve the samples that realize the piece. Concatenative sound synthesis improves on this concept by leveraging the tradition from speech synthesis to allow the modification of retrieved samples [Schwarz 2007]. The problem with this approach for retrieving sounds from a real world database is the requirement that the target is specified in advance, which often doesn't match popular practices in dealing with sound samples. For instance using descriptors to visually browse the database may require specialized knowledge that can't be assumed from internet visitors. On the other hand, the possibility to automatically describe units on the database typically relies heavily on a process that segments audio into small units, through which the database is artificially built. Hence, the problem of retrieving sounds from an audio database collaboratively built by internet visitors may benefit from a different approach.

The term *creativity* has traditionally been used to refer to unexpected events with regard to human actions. The idea that something is created out of nothing is usually a way to stress a disconnection between cause and effect, and the inability to offer a rational explanation. Nevertheless, attempts have been made to understand and describe creativity in computational terms. One of the most well known is due to Boden [Boden 2003], who explained personal creativity as an exploration of a conceptual space where unusual generative rules are used to find concepts that are novel and valuable. This perspective may be useful when looking at the problem of creative exploration of a sound database: as we have seen in many occasions sounds are sought to be used as building blocks. Often, the fitness of a given sound will only be known when listening to it in a given context. In this sense, it may be worth to explore the space of potential combinations of sounds. For creativity support tools, the decision on value and novelty may be let to the user. On the other hand, further research has tended to consider creativity more as a social phenomenon than a product of individual minds. Computer programs that support social creativity may then be thought as ecosystems that allow things to happen from the interaction between different agents. One possible framework was proposed by Kosorukoff [Kosorukoff 2001] with the definition of Human Based Genetic Algorithms (HBGA) as evolutionary algorithms where all genetic operations may be performed by users. Hybrid systems may balance work load between human and computational agents.

The *Freesound Radio*[1] project was developed in order to provide an alternative interface to the one of *Freesound.org* by creating a continuous sound stream. Instead of visual browsing, the radio allows to explore the database by listening to it. Instead of playing individual sounds, it generates potential combinations of them, defined by a graph data structure. Visitors can create and share such combinations which are seeded to a genetic algorithm that continuously generates new ones. All sounds used can be identified and downloaded. In the following sections we describe the main problems encountered in the development of this system and the solutions provided so far.

RELATED WORK

Several network music projects have investigated the the potential of the web for collective creativity. Because of the intrinsic delay in internet communications and the usage patterns of the web, involving users in different time zones, some research has focused on asynchronous sharing of musical creations. For example in [Jordà and Wüst 2001], the music for a theatre piece was composed by internet users using a custom synthesis engine. Users could depart from existing compositions to create their own. In [Miletto et al. 2005] the concept of *music prototyping* was proposed as a way to accommodate different levels of expertise in asynchronous creation of musical pieces. On the other hand, many projects have explored the potential of concurrent activity to drive or create a shared musical process, although this often needs to deal with relaxed synchronicity requirements in comparison to traditional music creation tools. Daisyphone [Bryan-Kinns 2004] allowed a group of users to concurrently modify a musical sequence displayed as a circle that is played in a loop. CC-Remix [Tanaka et al. 2005] explored online remixing of loops extracted from a single CD of songs released under the CC license. The advantage of our approach over using a synthesis engine or a MIDI player is that using samples allows visitors without any musical or specialized training to participate. On the other hand, the size and variety of the database is obviously very important in order to provide a greater vocabulary to people with different tastes. However, with large databases comes the need of appropriate description and organization. We now describe the approach followed in *Freesound Radio* for navigating sounds of many different kinds.

ORGANIZATION OF GENERAL AUDIO FOR CREATIVE APPLICATIONS

One important aspect of creative use of sounds is the possibility to cross established boundaries and use recordings of different types in unintended ways. Thus, in current popular music it is an established practice to use recordings from very diverse domains: voice recordings, environmental sounds or fragments of existing music are commonly reinterpreted by putting them in new auditory contexts. The *Freesound.org* database is actually a good example of this tendency towards eclectic use of sounds. The site contains as of january 2009 more than 60000 sounds of the most diverse nature, with the only common denominator that they are expected to be reused in some new context. Several different cultures can be identified on the site: many users upload sounds recorded from the environment with different levels of expertise and technical sophistication, other users upload sounds created with their synthesizers or computer programs, a number of users upload voice recordings, often upon request. Applications of these sounds range from media products (video, games) to mobile phone ring tones or music.

This diversity poses a problem on how to organize sounds. Our aim to support exploration of the database can greatly benefit from a measure of similarity between sounds, as well as some categorization. For an application based on collaborative media sharing it seems reasonable that any such categorization is not preset in advance, but emerges from community activity. Probably the most prominent feature used to describe and retrieve sounds is the tag folksonomy. Although rather noisy, tag folksonomies are an important sources of information because of the agreement they represent among users in using the same concepts to describe sounds [Heymann and Garcia-Molina 2006]. In this sense, Music Information Retrieval (MIR) methodologies used for automatic classification of music, typically based on machine learning algorithms that leverage automatic extraction of content descriptors, may be used also for creative organization of musical building blocks. An initial step in this direction was given in the development of *Freesound Radio* in order to identify a set of basic categories where different descriptors may be used. As an example, a monophonic pitch extractor may be used to describe a single piano note, a tempo induction mechanism may be more useful for a drum loop, and a detailed description of timbre may be needed for an environmental sound. For the task of telling apart a set of basic categories we use the decision tree learning algorithm so that the resulting models are understandable and may be developed and refined. We use some popular tags to train the algorithm. For example, 'note' and 'noise' are used on sounds with a single onset to tell apart pitched and unpitched ones, and 'drumloop' versus 'rain' to identify rhythmic sounds among those with multiple onsets. This technique can be extended to other tags and sound families in order to analyze and understand the relationships between concepts employed by users and automatic content descriptors. For simplicity, we use use the euclidean distance which is commonly used for sound segments (e.g. in [Jehan 2005]) for calculating the most similar sounds to a given one in its class, using an appropriate set of descriptors for that class.

[1] http://radio.freesound.org

PATCHING SAMPLES

A second problem that emerged in the development of *Freesound Radio* was to find a way to represent sound compositions as combinations of sound files. The required data structure had to be general so that the system would not be biased towards a specific musical genre, but simple enough to allow both human and computational agents to easily manipulate musical pieces. A *Sample Patch* is defined as a directed graph where vertices represent sounds from the database and edges represent transitions between them. Transitions are triggered when the sound of their source vertex finishes playing, and immediately start playback of the target vertex sound. Cycles are allowed, which results in patches of infinite duration. In particular, one vertex can be connected to itself so that it plays in a loop. Each vertex can have several incoming and outgoing edges, but only one edge in each direction is possible between two given vertices. Vertices are always followed and thus patches are deterministic. The idea is that this model is equivalent to the classic "brick wall" interface that is used by conventional audio sequencers, where building blocks are layered in several tracks, but here silence must be supplied by silent samples from the database. The main difference is that instead of focusing on the timeline, the interface is focused on the sounds and transitions among them, and durations and time structures are determined by the available samples. In order to provide the possibility of several entry points (i.e. not forcing all patches to start with one sound), a root node with no sound neither duration is defined. From the root, as well as from any other vertex, multiple outgoing edges will correspond to multiple audio tracks. [2]

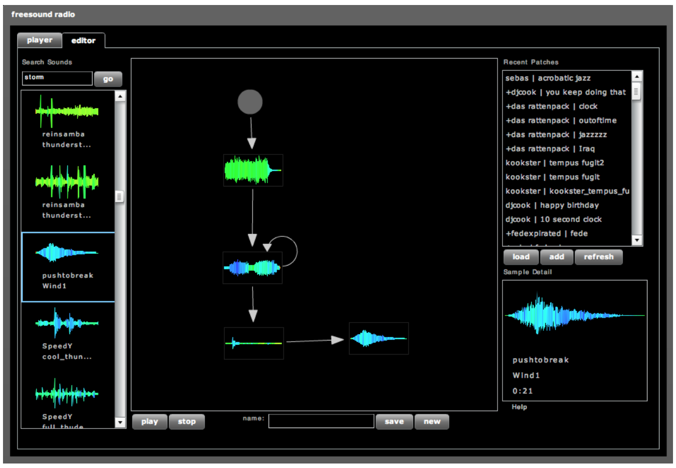

Figure 1. Editor interface

From the point of view of manual creation, this system provides a simple way to create sound compositions based on a large quantity of samples available in the database, requiring no specialized musical training from visitors. Instead of a full fledged musical sequencer, the radio provides the sample patch editor as a kind of sketchpad that allows ideas to be quickly and easily expressed, shared and reused, while heavily relying on the expressive potential of the database.

[2]Some examples of the patches created by users with this sysem can be heard in this web page: http://mtg.upf.edu/node/1202

EVOLUTION

Genetic algorithms have been used in many areas of computer music [Miranda and Biles 2007]. In some cases, when the fitness function can be specified computationally, the application of genetic algorithms to music is not essentially different from other domains. On the other hand, the specification of a fitness function may be also be used as the base for a music composition. Another possibility is offered by interactive genetic algorithms, where the fitness function is replaced by human evaluation. A difference of music with respect to other creative domains such as computer graphics is that all individuals must be heard, which may create a bottleneck. Also in some situations, the requirement of convergence to an optimal solution is removed. These differences have been thoroughly described by Biles [Biles 2003], since they pose important problems when using genetic algorithms for music performance. A similar approach to Biles' *GenJam* was developed in *Freesound Radio* so that a computational agent may provide an infinite supply of sample patches. One common problem with respect to understanding this kind of approach is that, as described in [Pearce et al. 2002], different motivations (e.g. using algorithms for as a means for artistic expression vs using them as tools for assisting the compositional process) tend to be confused with regard to algorithmic composition. In the context of creative exploration of a sound database we do not intend the outcome of the algorithm to be regarded as "easy listening" music. Each of the resulting patches may contain interesting sounds or interesting combinations of them. It is up to the users to participate in order to drive the algorithm to produce patches that match their taste. If an interesting patch is produced, the user may use it as a starting point for her own composition. In this sense, our goal is to support creativity. The main functions of the genetic algorithm may be described as follows:

- *Initialization.* Since the evolution process never ends, initialization is only needed once. Thus, better results are obtained by using manually created patches to seed the process. As new patches are created by users, they are fed into the system.
- *Selection.* Patches are selected according to users submitted ratings. Two buttons (thumb up or thumb down) allow each visitor to provide feedback about the patch that is being played at any moment.
- *Mutation.* The idea for mutation is inspired in Schaeffer's analysis of linguistics for application to *Musique concrète* [Schaeffer 1966]. A sample used in a given patch has been selected because of some characteristics to articulate a message. This means that it may

be replaced by a similar sound to play the same role. For example an instrument note may be replaced by the sound of a different instrument playing the same note, like a word may be replaced by some synonym in a given sentence. We implement mutation as substitution of one sample by one of its closest neighbors computed as we have described. Greater priority is given to samples that agree with the preferences expressed by users in the radio interface.

- *Crossover*. Creating a patch from two existing ones involves splitting both parents in two. The algorithm used was developed for crossover of graphs such as molecules and circuits [Al Globus and Wipke 2001]. In order to split a patch, two nodes are chosen, and all possible paths between them are broken. One half is then connected to a half of the other parent.

This algorithm is used by the radio to generate playlists of sample patches that are played continuously. When one playlist finishes, the patch with the best rating (or one of them) is repeated while the reproduction process generates a new playlist through mutation and crossover from the best patches of the previous one. A small sample of user generated patches is always included in each playlist.

CONCLUSIONS

Retrieving sounds from a database for applications such as music creation can sometimes be viewed as circular problem. In order to know if a given sound is good, one must listen to it in combination with the target piece. However, in order to build the piece, sounds must be retrieved first. With *Freesound Radio* we have devised an environment that allows to deal with this situation on one hand by placing creation in a social setting and on the other hand by providing an automated exploration of potential combinations of sounds. This is made possible by a representation that allows quickly sketching potential combinations and automatically evolving new ones. Visitors can create their own compositions from scratch or by departing from the ones created by other users or by the machine. This environment provides an alternative way of discovering many sounds in the *Freesound.org* database that would otherwise remain hidden.

REFERENCES

Al Globus, J. L., Sean Atsat and Wipke, T. 2001. Graph crossover. In *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO-2001* (2001).

Biles, J. 2003. Genjam in perspective: a tentative taxonomy for ga music and art systems. *Leonardo: Journal of the International Society for the Arts, Sciences, and Technology 36*, 1, 43–45.

Boden, M. 2003. *The Creative Mind; Myths and Mechanisms.* Routledge.

Bryan-Kinns, N. 2004. Daisyphone: the design and impact of a novel environment for remote group music improvisation. In *DIS '04: Proceedings of the 5th conference on Designing interactive systems* (New York, NY, USA, 2004), pp. 135–144. ACM.

Heymann, P. and Garcia-Molina, H. 2006. Collaborative creation of communal hierarchical taxonomies in social tagging systems. Technical Report 2006-10 (April), Stanford University.

Jehan, T. 2005. *Creating Music by Listening.* Ph. D. thesis, Massachussets Institute of Technology.

Jordà, S. and Wüst, O. 2001. Fmol: A system for collaborative music composition over the web. In *Proceedings of Web Based Collaboration DEXA 2001* (Munich, Germany, 2001).

Kosorukoff, A. 2001. Human based genetic algorithm. In *IEEE International Conference on Systems, Man, and Cybernetics*, Volume 5 (2001), pp. 3464–3469 vol.5.

Miletto, E. M., Pimenta, M. S., Vicari, R. M., and Flores, L. V. 2005. Codes: a web-based environment for cooperative music prototyping. *Organised Sound 10*, 3, 243–253.

Miranda, E. R. and Biles, J. A. 2007. *Evolutionary Computer Music.* Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Pearce, M., Meredith, D., and Wiggins, G. 2002. Motivations and methodologies for automation of the compositional process. *Musicae Scientiae 6*, 200–2.

Schaeffer, P. 1966. *Traité des objets Musicaux.* Seuil, Paris.

Schwarz, D. 2007. Corpus-based concatenative synthesis : Assembling sounds by content-based selection of units from large sound databases. *IEEE Signal Processing Magazine 24-2*, 92–104.

Tanaka, A., Tokui, N., and Momeni, A. 2005. Facilitating collective musical creativity. In *MULTIMEDIA 05: Proceedings of the 13th annual ACM international conference on Multimedia* (2005), pp. 191–198. ACM.

Zils, A. and Pachet, F. 2001. Musical mosaicing. In *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFx-01)* (Limmerick, Ireland, 2001).