Using Sound Streams as a Control Paradigm for Texture Synthesis

Cesar Renno Costa, Fernando Von Zuben, Jônatas Manzolli
Núcleo Interdisciplinar de Comunicação Sonora, NICS
Laboratório de Bioinformática e Computação Bio-inspirada, LBIC
Faculdade de Engenharia Elétrica e Computação, FEEC
Departamento de Música, DM/IA
UNICAMP
e-mails: {cesar,jonatas}@nics.unicamp.br, vonzuben@dca.fee.unicamp.br

Resumo:
Este trabalho apresenta um novo paradigma de controle dirigido a texturas de sintetizadores. O modelo ultrapassa a intuitividade dos métodos de desenho de som digital tradicionais permitindo ao usuário especificar as suas aspirações sônicas utilizando *streams* de som e não parâmetros numéricos. Apresenta-se também o como dita funcionalidade poderia ser atingida na síntese de textura e a sua implementação na síntese generativa baseada na população (*Population-Based Generative Synthesis* - PBGS).

Abstract:
A novel control paradigm for controlling texture synthesizers is presented. Such model increases the intuitiveness of traditional digital sound design methods by letting the user to specify its sonic aspirations using sound streams and not on terms of numerical parameters. How such functionality could be achieved on texture synthesis and its implementation on the Population-Based Generative Synthesis (PBGS) are shown.

Key-words: texture synthesis, granular synthesis, synthesis control, intuitive control

Introduction

Every design process objectives the materialization of designer's aspirations with the highest correlation level. Musical sound design's main goal is to obtain sonic material that matches some sort of aesthetical desire. Normally this cannot be quantitatively defined, but expressed as subjective sonic qualities with no possible mathematical description. Gray (1975) proposed a method to segregate different timbres with a limited set of linguistic and subjective qualities, on which he was successful. Nicol et al. (2004) worked on a synthesizer that mapped FM synthesis into a timbre space defined by fixed linguistic qualities. The context of online media repositories search is somehow analog to the sound synthesis practice. Traditional search paradigm consists of labeling media by name and genre, and guiding the quest on these parameters.. The content-based search has surged as a promising alternative by allowing the search procedure to find media that is related to some reference. In this context much effort has been applied on multimedia description formats, like MPEG-7 (Martinez, 2004) to permit media comparison by means of inner features. On the synthesis context, traditional control paradigms are still searching for names and genre, i.e., they use parameters that generalize some sound features to perform the generative process.

We propose a content-based synthesis that works as an alternative to the traditional parametrical approach. Instead of a model based only on the acoustic signal or the physical instrument we consider the auditory subjectivity too (Furlanete, 2000). We understand that using a search algorithm based on qualitative representation of sound streams could be more efficient than a simple parameter manipulation. By applying modern computational methods, such as bioinspired computing, it is possible to develop a synthesis process using sound samples as control set. We have studied this synthesis method with success in (Caetano et al. 2005a,b,c).

Texture Synthesis and Granular Synthesis

Sound texture consists of sounds with recognizable sonority with complex dynamical spectral variation that could not be defined as a punctual event or a gesture (Schaeffer, 1966). The sound of the rain, the sound of a waterfall, the sound of an unregulated air conditioner are samples of real-world sound textures. Beyond, there are textures in the context of l'écoute réduite (Schaeffer, 1966) that there cannot be associated to any physical source. Granular Synthesis is a great tool for Texture synthesis, and is normally used on the composition of soundscapes or musical objects.

On the Texture Synthesis model to be used, the inspiration was on Xenakis' Screens (Xenakis, 1971) and its variation Roads' grains (Roads, 1988). Based on Gabor's discoveries on the limitations of human's fast frequency variation perception (acoustic quanta theory) (Gabor, 1946), Xenakis wrote that complex sounds could be reproduced by playing a book of screens with a regular rate (just like a movie with frames, see Figure 2a. He defines a screen as a low-duration sound with well defined spectral distribution.

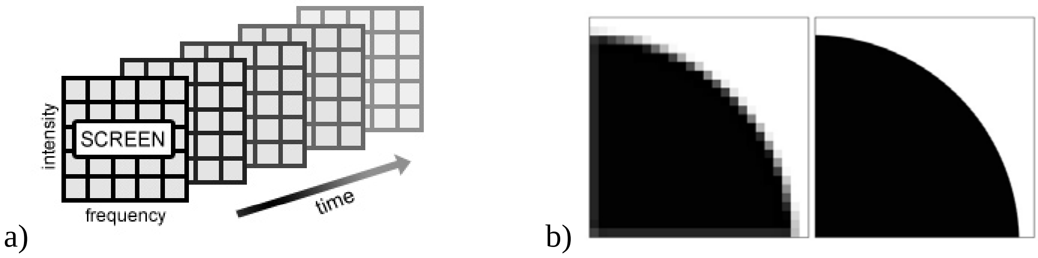

Figure 2: a) Book of Screens: sound seen as a movie b) Effect of resolution on perception

A direct analogy to the acoustic quanta theory is shown in Figure 2b. Human visual space resolution has equivalent limitations. On the left, a low-resolution quarter of circle is shown and quantization could be easily perceived. On the right, a high-resolution image is presented. Although quantized, it invokes a continuum perception. The way sound is perceived is equivalently limited, being in frequency or in time.

Derived from works of Gabor and Xenakis, GS synthesizes sounds using a rapid succession of tiny sounds, metaphorically referred to as grains or yet as microsound (Roads 1996, 2001). Truax (1988) proposed a different system and a real-time DSP implementation. As control paradigms for grain generation, mostly stochastic models were used, including Xenakis, Roads, Truax and Miranda. Non-stochastic models were used in (Miranda et al., 1995) with cellular automaton and also in Ecologically-based GS (Keller et al., 1988).

Sound Stream Control of Texture Synthesis

To provide the sound stream control function, it is necessary to find a methodology to automatically extract sonic features from a screen sequence and store them in a computer based structure. We define a screen as a low-duration sample frames extracted from a source stream and windowed by a Gaussian-like envelope. This extraction procedure is a hard task due to its high-dimensionality and to the fuzzy notion of what should be a relevant sonic feature for human perception. It is also necessary to develop a screen sequence generation technique guided by these sonic features. These demands are not fulfilled by exact mathematical procedures.

Bioinspired computation is a set of techniques based on natural processes such as evolution, self-organization and social behavior. The purpose is to bring, through computer simulation, attributes like self-adaptation. Our aim is to use transforming environments and self-regulation to develop new operational conditions (de Castro et al., 2004). Some common applications that have some relation to our needs are self-organization (in the self-organizing process of the Representative Structure) and pattern recognition (when automatically obtaining the relevant features). The main distinction between Xenakis, other reference works and our approach is the replacement of stochastic processes with sample frames generated by bioinspired algorithms to provide an interactive control methodology.

A population-based approach has been adopted. The idea is to obtain the most representative population of screens which could identify different details of the representative set. This way, the sonic features extracted from reference could be stored in the form of prototypes. The Representative Structure would be composed of a population of screens. In this task, self-organization has an important role on the process of identifying, organizing and separating screens with different features. These are well-known attributes of Self-Organizing Maps (SOM) (Kohonen, 2000). However, there are alternative population-based self-organizing algorithms, based on Artificial Immune Systems (AIS) (de Castro et al., 2002) and evolutionary computation (EC) (Goldberg, 1989). Under the existence of reference prototypes, the self-organizing process in denoted in the literature as Learning Vector Quantization (LVQ) (Kohonen, 1986). Figure 3 depicts the outcome of a two-dimensional LVQ process. The gray circles are the input samples that are represented by the black circles.

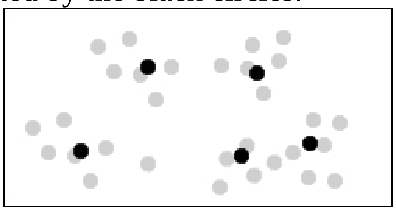

Figure 3: A pictorial view of Learning Vector Quantization. Gray: input samples. Black: representative prototypes.

The black circles correspond to the population of prototypes that will pass through a self-organizing process responsible for the final spatial configuration presented in Figure 3. Notice that the black circles are organized to capture the most relevant aspects of the input samples. They are called representative prototypes because they can be interpreted as concise representations of the input samples, generally expressing a consensual explanation of the local variability in the neighboring input samples.

An Implementation: PBGS

Population Based Generative Synthesis (PBGS) is a computational implementation of a texture synthesizer controlled by sound streams. It has two inputs: The Sonority Reference Stream (SRS) and the Dynamic Control Stream (DCS). The SRS works as the parametric set for the LVQ algorithm to produce the prototype population. Thus, it’s the SRS that defines the sonority of the output texture. The DCS lets the designer to control the global dynamic of the texture, being able to compose gestures or just maintain the texture as a stream. PBGS architecture is presented on figure 4.

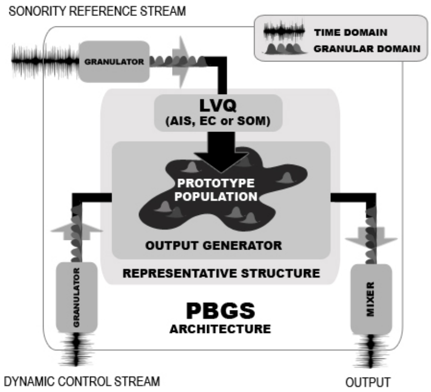

Figure 4: PBGS Architecture.

On prototype implementation of PBGS it was capable of generating complex textures with full dynamic control. Sound samples can be found online[1]. An extensive set of experiments has been applied over PBGS having two main targets: the understanding of its behavior over different sound entries, on which were included musical, voice, synthetic and natural sounds; verifying system performance boundaries for musical appliance.

On the first experiment the sample files presented on figure 5 have been used as sonority reference and as dynamic control streams. Figure 6 has some of the results. It has been possible to verify both spectral and dynamical traces of the reference on the output on all runs with strong correlation to the dynamic control dynamics. On some samples controlled by the speech signal, there was even possible to discern the spoken phrase. On other experiments, there where verified performance issues such as latency, effect of window size over spectral resolution and dynamic correlation and spectral tracking. A more comprehensive specification and a full set of experiments about PBGS implementation and real-time performance could be found at (Costa et al., 2006).

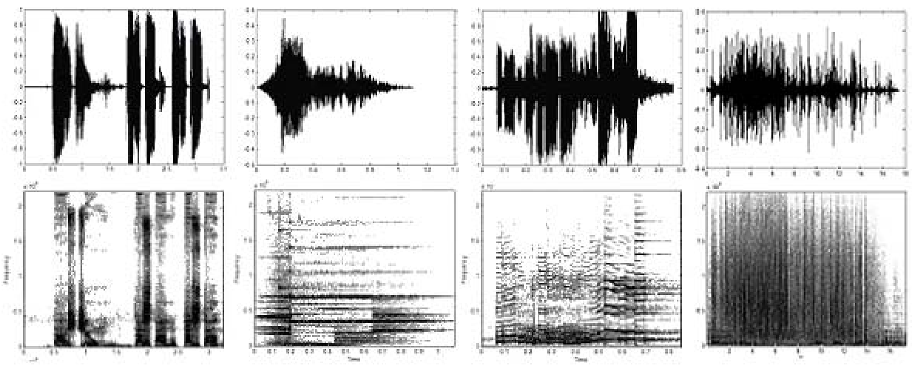

Figure 5: PBGS experimental set samples from left to right, waveform (first row) and sonogram (second row): speech, synthetic, guitar lick, water drops.

[1] http://www.nics.unicamp.br/~cesar/pbgs/

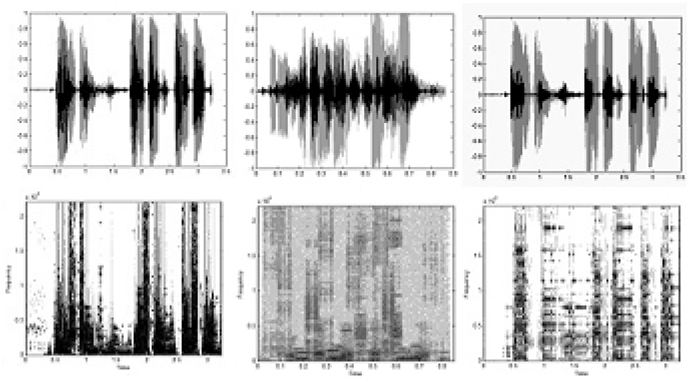

Figure 6: Some PBGS first experiment results waveform (first row) and sonogram (second row), from left to right: voice (Dynamical Control, DC) and guitar (Sonority Reference,SR); Guitar (DC) and voice (SR); Voice (DC) and synthetic sound (SR) .

Conclusion

The proposed control paradigm seems to be a good alternative to traditional synthesis control paradigm and is already implemented and functional on computational prototypes. However, there is still much room for improvement and development in order to become an established method. The main effort has to be in the sense of developing tools for efficient perceptually-driven sound manipulation and comparison. Some areas like Auditory Computing and Music Cognition might provide good material for such developments.

Notice that the model does not represent a unique neither definitive solution for sound synthesis control, but an alternative that may help designers to obtain their results with more efficiency and quality. It has its limitations: to be used with full experience it's needed a good sound archive and the quality of resultant sound is depended on the quality of the reference and on how close it is to the aspired sound. Thus, the designer must be efficient on reference definition and must have access to quality data to execute such task. However, with the ever-growing development of online repositories empowered with content-driven search engines, sound samples will be accessible for everyone. In this context our approach is a step-forward of traditional methods and represents a viable and promising sound synthesis control.

Acknowledgements

This work has been supported by grants from Fapesp and CNPq.

Reference

Caetano, M., Manzolli, J., Von Zuben, F. J. Application of an Artificial Immune System in a Compositional Timbre Design Technique, in *Proceedings of ICARIS 2005*, Alberta, Canada. In Press. 2005a.

Caetano, M., Manzolli, J., Von Zuben, F. J. Interactive Control of Evolution Applied to Sound Synthesis. In *Proceedings of the 18th International Florida Artificial Intelligence Research Society (FLAIRS)*, Clearwater Beach, EUA. 2005b.

Caetano, M., Costa, C. R., Manzolli, J., and Von Zuben, F. J. Self-Organizing Topological Timbral Design Methodology Using a Kohonen Neural Network. In *Proceedings of the 10th Brazilian Symposium on Computer Music (SBCM)*, Belo Horizonte, Brazil, 2005c, 94-105.

Costa, C. R., Manzolli, J. and Von Zuben, F. J. Population-Based Generative Synthesis: A Real-Time Texture Synthesizer based on Real-World Sound Streams. In *Proceeding of the 4th AES Brazil Conference*, São Paulo, 2006. 85-90.

Chowning, J., The Synthesis of Complex Audio Spectra by Means of Frequency Modulation, *Journal of the Audio Engineering Society* 21(7), 1973.

de Castro, L. N., Timmis, J. Artificial Immune Systems: A New Computational Intelligence Approach, Springer-Verlag, 2002.

de Castro, L.N., Von Zuben, F.J. (eds.) Recent Developments in Biologically Inspired Computing. Idea Group Inc., 2004.

Furlanete, F. Escritura Sismográfica. Master Thesis, PUC-SP, 2000.

Gabor, D. Acoustical Quanta and the Theory of Hearing. Nature, vol. 159, 1946, 591-594.

Goldberg, D.E. Genetic algorithms in search, optimization, and machine learning. Addison-Wesley. 1989.

Grey, J. M., "An Exploration of Musical Timbre." Ph.D. Dissertation, CCRMA/Dept. of Music, Stanford University, Stanford, CA., Dept. of Music, Report No. STAN-M-2, 1975.

Keller, D., Truax, B. Ecologically-based Granular Synthesis. ICMC 1998, Ann Arbor, Michigan. 1998.

Kohonen, T. Self-organizing maps. Springer-Verlag. 2000.

Kohonen, T. Learning Vector Quantization for Pattern Recognition. Technical Report TKK-F-A601. Helsinki University of Technology, 1986

Martínez, J. M., MPEG-7 Overview (Version 10) ISO/IEC JTC1/SC29/WG11 Palma de Mallorca, October 2004.

Miranda, E.R. Granular Synthesis of Sounds by Means of a Cellular Automaton. In Special Section: The Fourth International Symposium on Electronic Art (FISEA) Symposium Papers. Leonardo, Vol. 28, No. 4., pp. 297-300. 1995.

Nicol, C., Brewster, S., Gray, P., "Designing Sound: Towards A System For Designing Audio Interfaces Using Timbre Spaces" Proceedings of ICAD 04-Tenth Meeting of the International Conference on Auditory Display, Sydney, Australia, July 6-9, 2004.

Roads, C. Introduction to Granular Synthesis, Computer Music Journal vol.12 n.2, 1988.

Roads, C., Computer Music Tutorial, MIT Press,Cambridge, MA, 1996.

Roads, C. Microsound, MIT Press, Cambridge, MA, 2001.

Schaeffer, P. Traité des Objets Musicaux. Editions Du Seuil. 1966.

Truax, B. Real-time granular synthesis with a digital signal processor. Computer Music Journal, Vol. 12, no. 2, pp. 14-26. 1988.

Xenakis, I. Formalized Music. Indiana University Press. 1971.