SCRIME STUDIO REPORT

Myriam Desainte-Catherine, Sylvain Marchand, Pierre Hanna,
Mathieu Lagrange, Matthias Robine, Martin Raspaud, Robert Strandh,
Antoine Allombert, Guillaume Meurisse, Joan Mouba, Jean Louis Di Santo
and Gyorgy Kurtag
SCRIME – LaBRI
University of Bordeaux 1
F-33405 Talence cedex, France
www.scrime.u-bordeaux.fr

1. INTRODUCTION

The SCRIME project is the result of a cooperation convention between the Conservatoire National de Région of Bordeaux, ENSEIRB (school of electronic and computer scientist engineers) and the University of Sciences of Bordeaux. It is composed of electroacoustic music composers and scientific researchers. It is managed by the LaBRI (laboratory of computer science of Bordeaux). Its main missions are research and creation.

The SCRIME [1] project is funded by the DMDTS of the French Culture Ministry, the Aquitaine Regional Council, the General Council of the Gironde Department and ID-DAC of the Gironde Department.

Christian Eloy is the artistic director of the SCRIME. Myriam Desainte-Catherine is the scientific and administrative director. Researchers are professors and associate professors from the LaBRI [2] (Myriam Desainte-Catherine, Pierre Hanna, Sylvain Marchand, and Robert Strandh), associate researchers (Mathieu Lagrange, Matthias Robine and Martin Raspaud), and PhD students (Antoine Allombert, Joan Mouba and Guillaume Meurisse). Furthermore, three composers are associate researchers (Jean-Louis Di Santo, Christophe Havel, Gyorgy Kurtag, and Egdar Nicouleau).

The main objective of our research is to build sounds and scores modeling that are well-suited for various kinds of interactions. To reach this objective, several subjects of fundamental research are under study. In addition, several concrete artistic projects are under development in order to develop new ideas of interaction for creation.

2. SOUND MODELING

Sounds are physical phenomena belonging to the physical world. In order to manipulate digital sounds using a computer, we need a sound model, that is a formal representation for audio signals. Sound modeling draws the link between the real – analogical – and mathematical – digital – worlds.

Our research activities deal mainly with the modeling, analysis, transformation, and synthesis of auditory scenes. More precisely, we aim at allowing the identification of the several sound entities (sources) which are perceived in a binaural (stereophonic) mix. From the auditory scene composed by these entities, we would like to allow the musical manipulation of each individual entity and then resynthesize a different – transformed – scene, with transformations closely linked to the perception.

To reach our objectives, we propose to use spectral sound models since they rely on strong mathematical and physical bases, are closely linked to perception, are well-suited for musical transformations, and give rise to numerous interesting research topics in computer science.

2.1. Noise Analysis for Sinusoids+Noise Models

Nowadays, most of the sound models are hybrid models, where the noisy (stochastic) part is separated from the sinusoidal (deterministic) part. Practically, noise is estimated by subtraction of the sinusoidal component. Thus, the quality of the noise estimation relies on the quality of the estimation of the sinusoidal parameters. With high noise levels, errors in the estimation of the sinusoidal parameters cannot be avoided. This leads to errors in the estimation of the noise as well.

A new estimation methods has been proposed in [9]. This method relies on a long term analysis of the statistical variation of the amplitude spectrum. The noisy component is thus analyzed without any prior knowledge of the sinusoidal component.

2.2. Improving the Tracking of Partials

In the special case of deterministic sounds, the basic structure of these models is the partial, a pseudo-sinusoidal oscillator whose parameters (mainly frequency and amplitude) evolve slowly with time. Most of musical applications that make use of the sinusoidal model are frame based. However, the identification of the musical structure is generally a major prerequisite to the application of musical transformations and this identification requires knowledge that often cannot be extracted from frame-based representations of sounds.

The identification of these continuities can be achieved using a sinusoidal model and a partial tracking algorithm. In that case, this identification rely on the precise estimation of the frequency. We study in [6, 8], a class of estimators called phase-based. We demonstrate that they are

---

equivalent in theory, but diverge slightly in practise. Applied to polyphonic sounds, state-of-the-art algorithms often fail to yield reliable identification of these continuities. The predictability of the evolutions of the parameters of the partials as well as the theoretical lack of high frequencies in these evolutions are exploited [7]

2.3. Hierarchical Sound Modeling

During the sinusoidal analysis of a sound, only a few parameters are extracted by a partial-tracking algorithm, mainly the frequencies and the amplitudes of the sinusoids composing a sound at a given moment. However, there are also other parameters that could be extracted from a re-analysis of the estimated parameters estimated at the first step. Similarly, it must be possible to hierarchically extract parameters from sound.

A hierarchical sound model allows to model the sounds at several levels, i.e. at the sound level, but also at sound control level and at the variation level. This would allow to interact at all the levels of the sound, from microscopic to macroscopic. The research conducted by Martin Raspaud during his PhD thesis proposes such a model to perform time-scaling of high quality by acting at all the levels of the hierarchy. We have proposed a time-scaling technique preserving pitch, timbre and also micromodulations contained in the sound (vibrato and tremolo) [11].

These evolutions of parameters and there micromodulations have also been used to gather partials as "sound entities". Indeed, similar evolutions indicate a common sound source, and thus probably a single sound as origin.

2.4. Fast Additive Synthesis

In the synthesis stage, spectral modeling requires the computation of a large number of sinusoidal oscillations. The challenge is then to design an algorithm for generating the sequence of samples of each oscillator with as few instructions as possible.

We have proposed in [14] a new fast sound synthesis method using polynomials. This is an additive method where polynomials are used to approximate sine functions. Traditional additive synthesis requires each sample to be generated for each partial oscillator. Then all these partial samples are summed up to obtain the resulting sound sample, thus making the synthesis time proportional to the product of the number of oscillators and the sampling rate. By using polynomial approximations, we instead sum up only the oscillator coefficients and then generate directly the sound sample from these new coefficients.

Practical implementations show that Polynomial Additive Sound Synthesis (PASS) is particularly efficient for low-frequency signals. In the near future, we plan to combine the advantages of the PASS and Digital Resonator (DR) methods, since these methods both manipulate oscillators. For this hybrid method, the idea is, for a given partial, to use either PASS or DR depending on the frequency of the partial: for low frequencies, PASS will be preferred.

2.5. Sound Spatialization

Auditory scenes almost always consist of several sound sources located at different positions. For now, we consider that these sources are approximately in the same horizontal plane as the listener. This is the case in many musical situations, where both the listener and instrumentalists are standing on the (same) ground. Also, we focus on the direction of arrival, although we can also consider the distance between the source and the listener.

After Viste, we propose in [10] methods for the localization / spatialization of the several sources from / to a stereophonic (binaural) sound. These methods are based on general acoustic cues derived from specific Head-Related Transfer Functions (HRTFs) measurements. In addition to the direction of arrival, the enhanced model allows to accurately recover the energy of each sound source, so that it is now possible to hear the sources separately. The localization method is based on a Gaussian Mixture Model (GMM) on a power histogram, with the use of the Expectation Minimization (EM) algorithm and spatial filtering. In most cases we are able to isolate each sound source depending on its direction of arrival.

We aim at locating and isolating the sources of auditory scenes, in order to re-spatialize them a different positions. We want to extend the binaural spatialization to the case of multichannel diffusion. This will allow composers to ease the diffusion of electro-acoustic pieces, by the direct use of binaural recordings and with a more direct control, for example gestural. The computer is thus in charge of finding the best loudspeaker and output parameters for each spectral component of each source, to guarantee that the sound reaching the ears of the listener will be the one decided by the composer.

3. MUSIC INFORMATION RETRIEVAL

Allowing users to musically manipulate audio signals requires the development of systems that automatically analyze the musical properties of these signals. During the last few years, a field of research has emerged known as Music Information Retrieval (MIR), and this is an area in which research has been carried out at SCRIME."

3.1. Estimation of Symbolic Music Similarity

One of the key problems within MIR research is the estimation of the musical similarity between audio data. Measuring similarity between sequences is a well-known problem in computer science which has applications in many fields. However, musical sequences are characterized by specific properties. That is the reason why developing efficient and accurate algorithms entails taking into account areas such as sound perception and music theory.

Since 2005, a first collaboration with experts in string matching algorithms led to promising results. An algorithm for measuring monophonic symbolic music similarity has been developed [3]. It was entered in the symbolic music similarity contest during the second Music Information Retrieval Evaluation eXchange (MIREX 2006) and obtained the best results for the monophonic task. Furthermore, we propose an original algorithm that computes a similarity measure between polyphonic musical

sequences [4]. All the notes are taken into account, even if they sound at the same time. These methods are currently being used to evaluate the similarity between a monophonic query and a database of polyphonic musical pieces.

3.2. Evaluation of Musical Performance

Several parameters could be extracted from a musical performance. We focus on the technical and the physical part of the performance. In [12] we study the piano fingering. A lot of biomechanic studies have previously highlighted the influence of the physiology of a pianist in his performance, particularly due to his hands. This knowledge leads us to propose a new method of automatic fingering which uses dynamic fingering. We also introduce a method to recognize the fingering used by analysing the performance and we are now working on the physical and technical imprint of a musician in his performance.

We also propose to evaluate in [13] the technical level of a musical performer by analysing non expressive performances like scales. Our results are based on the analysis of alto saxophone performances, however the same approach can be used with other instruments. We did not want in this case to consider the physic behavior of the instrument or what is influenced by the physiology of the performer. Our aim was to highlight the technical part in the performance by considering the spectral parameters of the sound. However, the spectral envelop is strongly dependent to the physics of the couple instrument / instrumentalist and therefore often considered for instrument recognition. On contrary, the long-term evolution of the spectral parameters over time reflects the ability of the performer to control its sound production. Even if this ability is only one aspect of saxophone technique, it appears to be strongly correlated to the overall technical level in an academic context. Moreover, this evolution is relevant to evaluate performers over wide range from beginners to experts.

4. MUSICAL COMPOSITION

Nowadays, more and more musical pieces are created for heterogeneous musical ensembles including mechanical instruments, new electroacoustic virtual instruments and sound tracks. The projects we present in this section, address some problems arising from this situation, especially the assistance to the composer in the task of writing such pieces involving scores, sound processing and interaction between musicians and a computer.

4.1. Interactive Scores

Actual computers allow various kinds of real-time interactions with sound synthesis. The huge quantity of possible ways of interaction in contemporary music shows the great diversity of new possibilities provided by computers; our aim is to define a formalism for composition and performance of musical pieces involving temporal structures and discrete interactive events. In our model that we call interactive scores temporal structures and interactive events are bound with logical relations [1].

First we focus on temporal relations such as the Allen ones. Our model comprises two phases : a compositional one and a performance one. During the compositional phase, we allow the composer to design the score in adding temporal structures (musical notes), interactive events and in binding them with Allen relations [2]. During this phase, to maintain the constraints implying by the temporal relations, we use an incremental constraints propagation model based on the Gecode library. We can see the action of the composer during this phase as defining the degree of freedom given to the performer since he can increase this degree in adding many interactive events which will be triggered by the musician during the performance but on the other hand, he can limit the performer's freedom by adding temporal constraints through the Allen relations. For the performance phase, we develop a model based on Petri nets which allows us to manage parallelism between the temporal structures seen as autonomic processes that must synchronize at certain moments.

As far as the software development is concerned, thanks to a collaboration with Musical Representation team of the Ircam directed by Gerard Assayag, we implemented large parts of our model in OpenMusic (a graphic language for computer assisted composition developed by this team).

4.2. Structured Terminology and Symbolic Notation of the Electroacoustic Music

Linguistic and semiotic researches allowed to elaborate, for the language, various levels of unit, from the smallest to the biggest. This chain of units including one in the others can be applied to the music. It was thus necessary to define the minimal unit of the electroacoustic music as base of all these articulations . This definition bases itself on the various parameters defined by Pierre Schaeffer in the "Tableau Recapitulatif des Objets Musicaux" . The discovery of the minimal unit of the electroacoustic music has a double consequence. Firstly, the definition of all other levels of units which are strictly dependent on it. The minimal unit will be called *phase*, the combination of several units in the same matter *entity*, and the combination of several units in different matter *group*. Secondly, the possibility of elaborating a system of symbolic notation which represents each of the parameters of every sound, and which can be for the acousmatic music what a score is for the instrumental music, with all the advantages that it gets in an optics of abstraction - and thus of composition - and of musical analysis. Actually, the SCRIME develops a software with these elements. This system offers several conveniences: It can work sound by sound as a partition works instrument by instrument. Contrary to the instrumental partition, it does not work in a intermittent way, but describes processes, what allows to note the continuous. It allows a finer analysis of the electroacoustic music, so much from the point of view of the report and the function of the various elements of a composition between them, That from the point of view of a relation between the sound and the sense. It thus brings a supplementary tool to the *acoulogie* wished by P. Schaeffer.

4.3. GSHARP

Gsharp is a project the purpose of which is to create an interactive editor for music scores distributed as open-source

software. It is an ambitious project containing several innovative aspects in the form of incremental algorithms and data structures as well as novel interaction methods. Recently, we have made some important progress. In particular, we implemented a system for generating music symbols incrementally, with greatly improved visual appearance as a direct result. We also improved several visual aspects of Gsharp such as beam and clef placement. New functionalities have been implemented such as ties and the possibility of changing keys in the middle of a score. A new, very fast, incremental algorithm for page breaking has been implemented, allowing Gsharp to handle very large scores. An important aspect of the modularity of Gsharp has been published in [15] and we have had the opportunity to give a number of presentations of this work in several different countries during the past year and a half.

4.4. Pedagogical Projects

Several pedagogical projects are currently under development, especially in the domain of expressive interaction and in the domain of electroacoustic composition for very young children [5]. Those projects are based on the notion of musical gesture. The study of the gesture is the main part of the learning process for infants. By contrast, it is a first step before composition for young children. Thus, we use the computer as a musical instrument for infants, and as a tool for composition for young children.

5. REFERENCES

[1] Antoine ALLOMBERT, Gerard ASSAYAG, Myriam DESAINTE-CATHERINE, and Camilo RUEDA. Concurrent Constraints Models for Interactive Scores. In *Prodeedings of Sound and Music Computing Conferences (SMC)*, GMEM, Marseille, France, 2006.

[2] Myriam DESAINTE-CATHERINE and Antoine ALLOMBERT. Interactive Scores : A Model for Specifying Temporal Relations Between Interactive and Static Events. *Journal of New Music Research*, 34(4), December 2005.

[3] Pascal FERRARO and Pierre HANNA. Optimizations of Local Edition for Evaluating Similarity Between Monophonic Musical Sequences. In *RIAO 2007: Proceedings of the 8th International Conference on Information Retrieval - Large-Scale Semantic Access to Content (Text, Image, Video and Sound), Pittsburgh, PA, USA*, May 2007.

[4] Pierre HANNA and Pascal FERRARO. Polyphonic Music Retrieval by Local Edition of Quotiented Sequences. In *Proceedings of the 5th International Workshop on Content-Based Multimedia Indexing (CBMI'07)*, Bordeaux, France, June 2007.

[5] György KURTAG, Jean-Louis DI SANTO, Myriam DESAINTE-CATHERINE, and Philippe GUILLEM. Pédagogie de l'Electroacoustique, du Geste Musical à la Composition Assistée par Ordinateur. In *Proceedings of the Journées d'Informatique Musicale (JIM'07)*, Lyon, France, April 2007.

[6] Mathieu LAGRANGE and Sylvain MARCHAND. Estimating the Instantaneous Frequency of Sinusoidal Components Using Phase-Based Methods. *Journal of the Audio Engineering Society*, 2007. To appear.

[7] Mathieu LAGRANGE, Sylvain MARCHAND, and Jean-Bernard RAULT. Enhancing the Tracking of Partials for the Sinusoidal Modeling of Polyphonic Sounds. *IEEE Transactions on Audio, Speech, and Language Processing*, 2006. To appear.

[8] Sylvain MARCHAND and Mathieu LAGRANGE. On the Equivalence of Phase-Based Methods for the Estimation of Instantaneous Frequency. In *Proceedings of the 14th European Conference on Signal Processing (EUSIPCO'2006)*, Florence, Italy, September 2006.

[9] Guillaume MEURISSE, Pierre HANNA, and Sylvain MARCHAND. A New Analysis Method for Sinusoids+Noise Spectral Models. In *Proceedings of the Digital Audio Effects (DAFx'06) Conference*, pages 139–144, Montreal, Quebec, Canada, September 2006.

[10] Joan MOUBA and Sylvain MARCHAND. A Source Localization/Separation/Respatialization System Based on Unsupervised Classification of Interaural Cues. In *Proceedings of the Digital Audio Effects (DAFx'06) Conference*, pages 233–238, Montreal, Quebec, Canada, September 2006.

[11] Martin RASPAUD, Sylvain MARCHAND, and Laurent GIRIN. A Generalized Polynomial And Sinusoidal Model For Partial Tracking And Time Stretching. In *Proceedings of the Digital Audio Effects (DAFx05) Conference*, pages 24–29, Madrid, Spain, September 2005.

[12] Matthias ROBINE. Analyse automatique du doigté au piano. In *Proceedings of the Journées d'Informatique Musicale (JIM'07)*, pages 106–112, Lyon, France, April 2007.

[13] Matthias ROBINE and Mathieu LAGRANGE. Evaluation of the Technical Level of Saxophone Performers by Considering the Evolution of Spectral Parameters of the Sound. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Victoria, Canada, October 2006.

[14] Matthias ROBINE, Robert STRANDH, and Sylvain MARCHAND. Fast Additive Sound Synthesis Using Polynomials. In *Proceedings of the Digital Audio Effects (DAFx'06) Conference*, pages 181–186, Montreal, Quebec, Canada, September 2006.

[15] Robert STRANDH, Gilbert BAUMANN, and John HAMER. Using stealth mixins to improve modularity. In *Proceedings of the Australian Software Engineering Conference*, April 2007.