

SCRIME STUDIO REPORT

*Myriam Desainte-Catherine, Pierre Hanna, Christophe Havel,
Gyorgy Kurtag, Mathieu Lagrange, Sylvain Marchand, Edgar Nicouleau,
Martin Raspaud, Matthias Robine, and Robert Strandh*

SCRIME – LaBRI
University of Bordeaux 1
F-33405 Talence cedex, France
www.scrime.u-bordeaux.fr

1. INTRODUCTION

The SCRIME¹ 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.

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 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. ADMINISTRATION OF THE SCRIME

Staff

Artistic director: Christian Eloy;
Scientific director: Myriam Desainte-Catherine;
Secretary: Annick Alexakine;
Sound Technician: François Dumeaux;
Software Engineer: Julien Villain.

Researchers

Researchers are professors and associate professors from the LaBRI² (Myriam Desainte-Catherine, Sylvain Marchand, and Robert Strandh), associate researchers (Pierre Hanna, Mathieu Lagrange), and PhD students (Joan Mouba, Martin Raspaud, and Matthias Robine). Furthermore, three composers are associate researchers (Christophe Havel, Gyorgy Kurtag, and Edgar Nicouleau).

¹ Studio de Création et de Recherche en Informatique et Musique Electroacoustique

² Laboratoire Bordelais de Recherche en Informatique, CNRS UMR 5800

3. FUNDAMENTAL RESEARCH

Concerning sound modeling, we are particularly interested in musical sounds. Polyphony (several sounds played together) is imperatively assumed. Therefore, the sound analysis methods need to be more complex. Moreover, the noisy component of musical sounds is assumed as quantitatively and qualitatively important. Models have to take into account all these constraints.

Furthermore, we choose to develop models that are closely connected to perception. Sounds have to be parameterized at the human receptor, not at the emitter (physical models). This choice often implies the spectral representation of sounds, and leads to collaborations with psychoacousticians.

The aim of our research is a complete analysis / synthesis scheme, at the contrary of other approaches classically chosen, like for example in the context of sound indexing. In these cases, the analysis method is the main point, and the modification of the parameters obtained for the resynthesis of the transformed sound is not desired.

Our efforts concern algorithms and data structures. Even if important knowledge in signal processing and statistics are necessary, we think that further problems may concern algorithmics and combinatorics.

In our research, priority goes to understanding (fundamental research), more than treatment (engineering). Once fundamental mechanisms are understood, results lead to several applications, often implemented as free softwares, developed under UNIX operating systems (Linux, MacOS X). These softwares can be used by both scientists and musicians.

3.1. 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.

3.1.1. Hierarchical Spectral Models

Spectral models based on additive synthesis provide general representations for sound well-suited for expressive musical transformations. We introduced the Structured Additive Synthesis (SAS) model which imposes constraints

on the additive parameters in order to make these transformations both computationally efficient and musically intuitive, in accordance to perception and musical terminology. This model breaks the arbitrary boundary between sound and music. Its applications in the fields of creation and education are numerous (see the Dolabip project below).

In the context of sinusoidal modeling, it is possible to hierarchically extract parameters of the sound. During the analysis of a sound, only a few parameters are extracted: mainly the frequency, amplitude, and phase of the partials that are composing the sound at a given time. These parameters stay at a very low level. Some application may involve the extraction of musical parameters such as the vibrato and the tremolo, or even higher level parameters.

A hierarchical analysis, involving at each analysis step the re-analysis of the sinusoidal model parameters obtained at the previous step, is of great interest for extracting this high-level parameters automatically.

3.1.2. *Statistical Modeling of Noisy Sounds*

Existing models only represent sounds composed of low noise levels. They consider natural sounds as mixes of sinusoids (deterministic part) and noise (stochastic part). In such approaches, the noisy part is assumed to be parts of the signal that cannot be represented with sinusoids whose amplitude and frequency slowly vary with time and is implicitly defined as whatever is left after the sinusoidal analysis-synthesis. These approximations lead to audible artifacts and explain why such models are limited to the analysis and the synthesis of purely noisy signals. Our research concerns improvements of the modeling of this noisy part.

An original noise model to analyze, transform and synthesize such noisy signals in real-time has been proposed. This spectral model represents noisy signals with short-time sinusoids whose frequency values are randomly chosen according to statistical parameters. Modifying these mathematical parameters extracted from sounds leads to original transformations which cannot be performed using the previously described models. Some of these transformations are related to the modification of the distribution of the frequency values of the sinusoids in the spectra.

3.2. **Sound analysis**

3.2.1. *Enhancing the Precision of Spectral Analysis*

A new analysis method was proposed in order to accurately extract the short-term spectral parameters in the context of sinusoidal modeling. This method extends the classic short-time Fourier analysis by also considering the derivatives of the sound signal, and it can work with very short analysis windows.

3.2.2. *Characterization of Peaks in Short-Time Spectra*

The analysis stage of spectral models based on phase vocoder techniques uses short-time spectra of the sound studied. An original approach has been introduced for determining which peaks of these spectra correspond to sinusoids in order to prevent bad connections between partials in successive frames. This approach differs from the existing

ones because it is based on the statistical analysis of the intensity fluctuations and not on the results of the Fourier transform. Experiments show that the methods proposed greatly improve the selection of peaks, especially in noisy conditions.

3.2.3. *Improving the Tracking of Partial*s

Most of musical applications that make use of the sinusoidal model are "short-term" based. The audio signal is divided in overlapping little frames where some sinusoidal components are identified. The parameters of these components are then modified to achieve meaningful musical transformations. This "short-term" representation of sound is very popular because of its robustness.

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 continuities between sinusoidal components of successive frames is needed. For example, the modification of the vibrato of a voice requires such a "long-term" knowledge of the characteristics of sinusoidal components that compose the voice sound.

The identification of these continuities can be achieved using partial tracking algorithms, which is a challenging part of the sinusoidal modeling of sounds that links successive peaks belonging to the same sinusoidal component (often called "partial"). Some efficient tracking algorithms are proposed in the literature to identify the musical structure of monophonic sounds.

Applied to polyphonic sounds, these algorithms often fail to identify reliably these continuities. Indeed, several constraints of the analysis of polyphonic signals lead to a spectral representation with artifacts, i.e. some short-term sinusoidal components (often called peaks) may be corrupted or missing. The tracking of partials across this corrupted spectral representation requires new tracking methods that use characteristics of the sinusoidal model. 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 in new algorithms useful for our purposes.

Thanks to these new partials tracking algorithms, the long-term sinusoidal components are extracted more reliably. As detailed below, this gain of quality can be exploited in many ways; the restoration of audio material can be enhanced and some meaningful informations concerning the musical structure of the analyzed sound can be extracted reliably.

3.2.4. *Statistical Analysis of Noisy Sounds*

Usual noise models represent sounds with filtered white noise and only extract the spectral envelope from sounds analyzed. Psychoacoustic experiments have shown the ability for humans to discriminate noise bands with different spectral density. This property may be taken into account for the mathematical representation of noises. An original method has been proposed to approximate this new parameter. The limitations of the discrete short-term transform lead us to consider a new approach. This technique is based on the statistical studies of the intensity

fluctuations which are theoretically related to the spectral density. Experiments show the quality and the limitations of this method. Although it may be improved, this technique of estimation is correct enough to be applied in a complete analysis / synthesis model for noisy sounds.

3.3. Sound Synthesis

3.3.1. Rapid Additive Synthesis

We are investigating a number of different methods that will allow for a very large number of simultaneous oscillators for additive synthesis. Currently, a reasonably fast machine can synthesize a few thousands oscillators in real time, and there are reasons to think that, because of psychoacoustic phenomena such as masking, no more than about 10000 are ever needed.

However, taking advantage of such phenomena to speed up synthesis is nontrivial. For that reason, we are also interested in increasing the number of oscillators that can be synthesized simultaneously. While we cannot improve on the current (optimal) method using one multiplication and one addition (plus another addition to sum up the individual oscillators) per sample and per oscillator, we can use techniques that allow us to avoid computing each sample for each oscillator each time, provided that we allow ourselves to introduce small inaudible errors in the computation. Preliminary results indicate that in the near future we will be able to increase the number of simultaneous oscillators by as much as a factor 10 or more. With tenths of thousands of simultaneous oscillators, we can hope to synthesize hundreds of individual sound sources simultaneously in real time.

3.3.2. Noise Synthesis

Existing models are specific to harmonic sounds and do not preserve the statistical properties of noisy sounds. We show that those transformations such as time stretching introduce artificial intensity variations. Another approach we propose suggests that the sinusoidal components are added using adapted overlap-add method to keep statistical moments constant. Time scaling operations using this approach have been developed. Experiments on artificial sounds (filtered white noises) as well as natural sounds such as consonants and whispered vowels, show impressive enhancement in quality. Infinite time stretching transformations of such noises can be perfectly performed.

3.3.3. Digital Audio Restoration

Known interpolations techniques of sinusoidal components are limited to small missing interval. For gap above 20ms, some musical modulations are not preserved during the interpolation. For example, the vibrato of a singing voice is interpolated as a steady voice using the state-of-the-art method.

We propose to exploit the predictability of the parameters of the partials to interpolate reliably musical sounds. Partial having simple modulations like vibrato or tremolo are interpolated with high quality to gap sizes up to one second. More complex modulations are harder to interpolate, but the proposed method shows a significant improvement over classical methods. Since these modula-

tions are important to perception, the sinusoidal interpolation of missing audio data is more realistic.

The listening tests showed that the proposed method provides fair interpolation quality for complex polyphonic signals for gap sizes up to 450ms and good interpolation for monophonic modulated tones for gap sizes up to 1600ms.

3.4. Musical Analysis

3.4.1. Analysis of Sound Sources

As stated above, the knowledge of the musical structure is important for relevant musical transformations. This is especially the case when polyphonic sounds are considered because each source should be considered separately.

We study the clustering of partials into sound sources using several psychoacoustic cues like their common onset, their harmonic relation, and the correlated evolutions of their parameters. We have shown that the identification of the onset of sources gains efficiency when the “long-term” evolutions of partials are considered. For example, the identification of smooth onsets are better identified.

Numerous musical analysis techniques considers the harmonicity as the most important cue for source identification and separation. However, this cue is not generic, because many natural sounds produced by popular musical instruments are inharmonic. The use of a “long-term” model allows us to study a more generic cue: the correlated evolutions of the partials of a source.

Several approaches are under considerations to propose algorithms to exploit this cue for a better analysis of the musical structure of sounds.

3.4.2. Realistic Interpretation

We have at our disposal a great number of musical scores stored in paper or digital form (MIDI files, etc.). This scores are aimed for human interpreter, and their performance by computers are not satisfactory. Indeed, even if the melody and the dynamics are perfectly followed, it ends as humanless, not taking the natural constraints of the instrument or of the instrumentalist into account. Thus, the note transitions, the sustain of these notes or the timbre of the produced sounds are not natural.

We are studying the possibility of automatically and realistically interpreting a digital score (e.g. a MIDI file). In a first step, we focus on piano and the position of the fingers on the keyboard. This study requires first the definition of instrument models, but also the definition of the instrumentalist and its own constraints. Then, we will focus on other instruments and their constraints (bow movement for the violin, air column for the voice, with natural vibrato and tremolo, etc.). In the meantime, we will program computer rules of interpretation from musicology books.

3.5. Computer-Based Composition

3.5.1. Score Editing

We are developing Gsharp, a score editor that is radically different from the ones that currently exist on the market. Gsharp is free software and uses the GPL license.

It is written in Common Lisp and makes heavy use of the CLIM (Common Lisp Interface Manager) graphics library. We are particularly interested in an editor that allows experienced users to be efficient in their daily work, as opposed to one that is easy to operate by inexperienced users.

In order for Gsharp to be able to recompute the layout of a large score (hundreds of pages if necessary) at interaction speed (0.1s), we have developed an incremental score layout technique based on an incremental version of the well-known dynamic programming algorithm.

Finally, we are paying great attention to respecting the rules (or at least one coherent set of rules) of music engraving. For that purpose, Gsharp contains sophisticated algorithms for computing spacing and page layout, beam slant, placement of accidentals, stem direction, and more.

Gsharp is still in development and not yet ready for end-users. However, we expect that by the end of July 2006, enough progress will have been made that a wider distribution could be possible.

3.5.2. *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, but also the difficulty of aiming to formalize and generalize these practices. Our objective is to try to formalize the notion of interaction and find out a model. We think that such a model would help to unify the different ways of interaction of musical and sound levels and bring a better understanding of musical interaction.

We start this study by focusing on temporal aspects of interaction, putting aside, for now, sound synthesis. We start with a piece which is completely composed and we want to provide the composer a way to define a set of possible interpretations of his piece by specifying simple temporal constraints between musical objects.

The interest of this study is twofold. Firstly, interpretation of a written piece can be formally defined by the composer as a set of eventual pieces resulting from the temporal constraints that bind musical events. In that way, the composer can describe a kind of degree of freedom given to the musician, while he gets the certainty that temporal relations he specified between musical events will always be satisfied. Secondly, the same piece can be interpreted in several ways, by varying interaction points and temporal relations by a simple edition of the interactive piece. Thus, the piece could be adapted to any situation in a very simple way, according to the material and musicians that are available for performance. Another area of application is music education. Whatever the studied instrument is, the learner has to begin playing very simple musical pieces to be able to play them. This is due to the fact that he has to control every parameter of the produced music: activation and releasing, as well as pitch, volume, modulation and timbral aspects of all the notes appearing in the score. The resulting system should provide a way to simplify the interaction of the learner without losing musical interest, so that the learner could enjoy playing a large set of pieces, even very difficult ones. Interaction could get more and more difficult and follow the learner's progress.

4. COLLABORATIVE PROJECTS

4.1. Dolabip

The Dolabip project is a multidisciplinary project for developing early-learning games for electro-acoustic music in schools. It is a project based on a new pedagogy, which is at the border between sound and music: we call it sound exploration. We introduce a musical instrument that provides children with a way to play in a virtual sound space by means of a joystick. The pedagogical goals of the exploration of sound are to develop the children's listening strategies as well as their capacity to pay attention to their sensations. The main musical ability resulting from these developments is musical composition.

4.2. Air Percussion Project

Since 2000 the composer Christophe Havel develops, with the metamorphoses project, a process of musical writing for electronic instruments through a succession of chamber music works performed in public. The setting is as following: to put the instrumentalists, musicians from contemporary chamber music and experienced virtuosos, in front of an electronic set up with the same ergonomics that their custom acoustical instrument but with sounds and control completely different. In this case, the musicians must play according to the reactions of the electronic instrument and the musical propositions of the other musicians inside a formal structure proposed by the composer.

In this context the composer wished to use percussion instruments and imagined to develop a system to collect the percussionist gestures. After a period of measurements and modeling an air-percussion device was implemented from the obtained results, including a system of sensors connected to a computer which analyzes and recognizes the gestures of the percussionist, a system of graphic edition of the instruments and visualization of the trajectories of the strikes, and modules of sound synthesis. The percussionist plays in the air with two electromagnetic sensors that are fixed at the end of drums sticks. The set of drums could be graphically defined before the performance and it is of course possible to change the setting during the performance.

This device is now ready for performance and was integrated in the metamorphoses project by the composer. Moreover it was used to perform the music of the movie "the greed" by Eric von Stroheim in a show premiered during the 2nd "Printemps des ciné-concerts" in Bordeaux.

4.3. Gesture-Controlled Diffusion

The main objective of this project consists in substituting the manipulation of the mixing table cursors with gestures inspired by an orchestra leader in front of instrumentalists (speakers). We use two *Flocks of Birds* sensors, one in each hand. These sensors transmit tridimensionnal coordinates and rotations every 10ms, and permit the control of the diffusion, for example the control of the sound level of each speaker.