# STATISTICAL APPROACH FOR SOUND MODELING

*Myriam DESAINTE-CATHERINE*

SCRIME - LaBRI
Université de Bordeaux 1
F-33405 Talence Cedex, France
myriam@labri.u-bordeaux.fr

*Pierre HANNA*

SCRIME - LaBRI
Université de Bordeaux 1
F-33405 Talence Cedex, France
hanna@labri.u-bordeaux.fr

**ABSTRACT**

This article introduces a mathematical approach to extracting some statistical parameters from noise-like sounds. This approach could define a new spectral model or extend existing ones to analyze and synthesize such complex sounds. In the future, this method could also permit electro-acoustic composers to make musical transformations. We have performed several synthesis tests with synthetic and natural sounds using these parameters. The main assumptions is that analyzed sounds must not contain any transient or slow-varying deterministic component. The natural sounds resynthesized during our experiments sounds like the original ones even if some defects are perceptible because of the analysis limitations. We propose some initial solutions for future works.

## 1. INTRODUCTION

Electro-acoustic composers use any type of sounds : natural ones from instruments or from other natural sources, or synthesized ones. These sounds from natural sources are very difficult to define mathematically. Although other types of models exist, this paper focuses on spectral models. The limits of existing spectral models are related to the assumption that non-sinusoid components are always filtered white noises. Moreover this assumption does not take account in spectral density which seems to be relevant to human sound perception. Indeed experiments show the ability of human listeners to discriminate between sounds with different number of spectral components in a same band [1]. The statistical representations can extend existing synthesis techniques to be applied to wider variety of sounds, which may be neither deterministic nor filtered white noise.

Some recent works [2], based on time-domain manipulations, explore such methods with some effective results. In this paper, we consider a sound as a random process and use some statistical tools to represent noisy sounds in the spectral domain with mathematical parameters.

This article presents a statistical approach to analyzing and synthesizing sounds from non-harmonic music instruments like percussions for example. The model deducted from this approach defines high-level parameters, which have to be validated by composers but which could be modified to transform sounds in a natural and musically expressive way.

After having explained some limits of existing models in section 2, we introduce statistical parameters in section 3. Analysis and synthesis method are then explained in sections 4 and 5. We expose some experiments in section 6 and the limitations. Some initial solutions are proposed in section 7.

## 2. BACKGROUND

Recent modeling techniques have been proposed to extend the additive model.

The *Spectral Modeling Synthesis* (SMS) [3] separates the analyzed sound into deterministic and stochastic parts.

$$x(t) = d(t) + s(t) \quad (1)$$

The stochastic part is obtained by subtracting the magnitude spectra of the deterministic signal from that of the original. The main assumption about this component is that it is *stochastic* so it can be fully described by its spectral envelope. In fact stochastic processes also have probability density functions that are independent of their spectra as well as other higher moments than just second moments (that power spectra are related to). Once the deterministic part of the analyzed sound has been detected, a frequency-domain subtraction from the magnitude spectra of deterministic and original signal. The synthesis of the stochastic part is then obtained by inverse-Fourier transform from spectra defined by the original stochastic part (see figure 1).

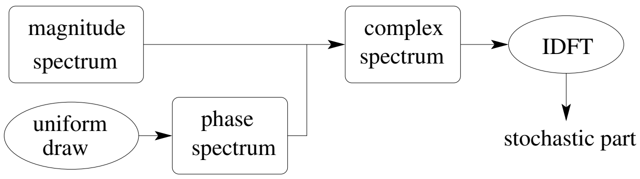


Figure 1: *The SMS synthesis method for the stochastic part.*

The *S+T+N model* [4] is an extension of SMS which includes an explicit model for transients. The assumptions and limitations about noise-like sound are the same except for transients.

$$x(t) = d(t) + t(t) + s(t) \quad (2)$$

The *Structured Additive Synthesis* (SAS) model [5] is a spectral sound model based on additive synthesis. It can be seen as an extension of the SMS model concerning the deterministic part. It is based on first-order Fourier transform [6]. Four parameters are extracted from the analyzed sound. These parameters can modify the sound in an intuitive and musical way. The SAS model can not produce every sound such as noises or transients but it could be extended to include noises according to the assumptions from the SMS model.

All these models can be used with non-deterministic sounds with the assumption that everything which can not be represented by

sinusoids whose amplitudes and frequencies evolve slowly in time, is filtered (or colored) white noise. Limitations of these approaches can be clearly heard with applications on natural sounds or dense musical instruments (cymbals).
Furthermore these models represent sounds in two or three different parts and it is obvious that separately editing these parts can lead to an inadequate fusion of re-synthesized noise and sinusoidal components [7]. It can particularly be heard when some transformations such as time scaling or pitch shifting are performed. Indeed some spectral components from the deterministic part seems to be related to components from the stochastic part, for example turbulence around a harmonic frequency in a wind instrument would make some noise band appear in its spectrum. That's why every musical transformation must be applied not only to both deterministic and stochastic parts but also must take into account the relation between the two parts [8].

## 3. STATISTICAL AND SPECTRAL PARAMETERS

Considering every non-deterministic sound as filtered white noise implies that every frequency has the same probability of being a component of such sounds; a questionable assumption. For example, this assumption is invalid for most musical instrument [7]. Moreover it implies that spectral density is constant and depends on the short-time Fourier transform properties.
For that reason we propose a statistical approach to considering the possibility for each frequency to have different probabilities and for a sound spectrum to be composed of many bands with different width or different density.
Analyzed sounds (sample rate $R$) are here considered in the spectral model as random processes $X$. Each frequency component $F_i$ is a random variable with fixed amplitude $a_i$ and random phase $\Phi_i$ has its own probability of being a component of the analyzed sound.

$$X_k = \sum_{i=0}^{N} a_i \sin(2\pi F_i \frac{k}{R} + \Phi_i) \quad (3)$$

In this paper we focus our attention on the frequencies. In order to simplify this approach, the amplitude is not considered as a random variable.

### 3.1. Main assumptions

The high-level parameters presented here could be part of sound analysis and synthesis models. The only assumptions here are related to statistical mathematic tools and to the fact that a sound is considered as a random process. This process must be ergodic for practical reason, which implies that statistical properties are the same when extracted from many realizations as temporally from one realization. It is obvious that this assumption is invalid in most of natural sounds. It could be interesting in the future to work with many samples from the same source under the same conditions.

### 3.2. Model parameters

Using statistics naturally leads to defining and extracting the four following parameters, some of which can be directly related to perception:

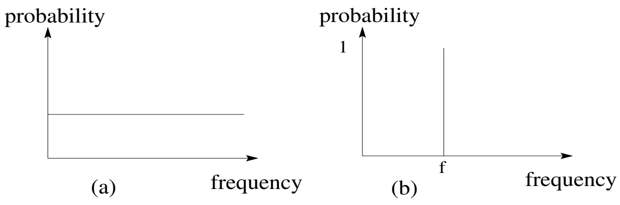


Figure 2: *Theoretical PDF from (a) white noise and (b) sinusoid.*

#### *3.2.1. Probability Density Function*

***PDF* : *frequency* $\times$ *time* $\rightarrow$ *real***
For each spectral analysis window, prominent components are extracted. The first parameter is the probability for each frequency values to be detected in several short-time transform windows. In what follows, this sequence of STFT windows will be called a *PDF window*. This probability density function can be related to musical parameters by describing sound harmonicity. For example, concerning a sound from a musical instrument, a dirac function as probability density function will model a perfectly harmonic instrument whereas a band around an harmonic frequency will represent possibilities for this frequency not to be exactly harmonic. This inharmonicity can also be controlled by the SAS model (*warping* [9]) but it seems interesting to re-synthesize a new sound with a random inharmonicity which could make it more realistic.

#### *3.2.2. Frequency count*

***FC* : *time* $\rightarrow$ *integer***
For each spectral analysis window, the number of prominent components is also a useful parameter, linked to the probability density function. For example, a wind instrument is represented by a spectrum whose some harmonic components are narrow bands. These bands can be synthesized by increasing the frequency count and by broadening the probability density function around the harmonic value.

#### *3.2.3. Continuation probability*

***CP* : *frequency* $\times$ *time* $\rightarrow$ *real***
For each spectral component, if its frequency value is selected in two successive windows (length $N$, sample rate $R$), the associated phases $\phi_i$ are tested according to equation 4 in order to determine whether they suggest two different partials or just only a multiple-window long one. These results determine the continuation probability for each frequency.

$$\phi_n = \phi_{n-1} + 2\pi f \frac{N}{R} \quad (4)$$

#### *3.2.4. Color*

***C* : *frequency* $\times$ *time* $\rightarrow$ *amplitude***
Color is one of the parameters of the SAS model [5]. It coincides with an interpolated version of the spectral envelope. We call it color by analogy between audible and visible spectra. This analogy is already well-known for noises (white, pink, ... ). It is directly related to the complex notion of *timbre*.

## 4. ANALYSIS

The analysis part starts with computing a set of magnitude spectra using the short-time Fourier transform. An interpolation of these spectra is obtained with *zero padding*, i.e. filling out of the precision wanted with zeros. The choice of the type if the analysis window is beyond the scope of this article. We will, however, discuss the length parameter. On the one hand, as many sound spectra such as natural noise spectra vary rapidly, it seems relevant to choose the shortest analysis window. On the other hand, this size must be large enough so that a minimum of frequencies must lay in different Fourier transform bins in the case of dense spectra so that every perceptually relevant component can be detected.

### 4.1. Peak detection

Once the set of spectra of a sound is computed, the prominent peaks of each spectrum are extracted. A peak is simply defined as a local maximum in the magnitude spectrum. This method is based on the assumption that each peak corresponds to a sine component of the analyzed sound. It is clear from studying complex sound spectra that they are certainly dense enough so that many frequencies other than the local maximum are perceptually important. That is why another selection criterion is introduced here.

### 4.2. Amplitude thresholding

As the magnitude spectrum is only an estimation using Fourier transform, some others frequencies must be detected to try to correct errors introduced by too low a frequency resolution. Another selection is made with an amplitude threshold arbitrarily defined. Each spectral component whose amplitude is higher than this value is saved.

### 4.3. Methods

Using peak detection or/and amplitude thresholding depends on the nature of the sound studied. It is obvious that different amplitude threshold will result in completely different extracted parameters.

For example, analyzing perfectly harmonic sounds requires peak detection **and** amplitude thresholding in order to detect all sine components but no peaks introduced by the estimation window from the Fourier transform.

On the other hand, dense or noise-like sounds have to be modeled taking care of many components, even if they are in the same analyzing bin. That's why peaks **or** component whose amplitude seems to be high enough are relevant as shown in figure 3.

Once the spectrum components detected, the statistical parameters have to be extracted. Each one is calculated from an arbitrarily defined number of STFT window. It is obvious that statistical parameters like probability density function is all the more relevant when the number of analysis window is larger. But it depends on the nature of the studied sound too.

### 4.4. Limitations

Analysis is the most difficult and uncertain part of this investigation. Some more advanced works will be needed to be sure about the quality of the relevance of the detected frequencies. Indeed the analysis limitations imply the main assumptions about sounds: they must not contain any transient or slow-varying deterministic component. Indeed many clicks can be heard because of the variations of the amplitude component.

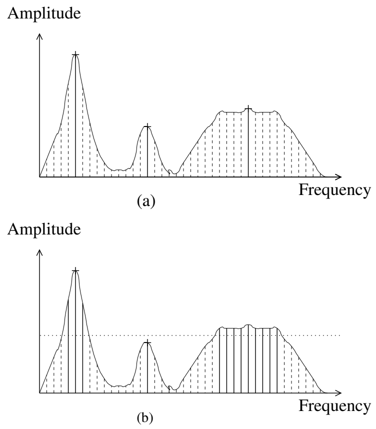


Figure 3: *Illustration of the analysis method for noise-like sounds by (a) peaks detection and (b) peak detection added to amplitude thresholding. Detected frequencies are represented with solid lines and undetected ones with dash lines.*

## 5. SYNTHESIS

With the four parameters defined above, it is possible to synthesize the analyzed sound or, by modifying these parameters, to obtain a new sound with different statistical properties. Figure 4 represents the synthesis method based on additive synthesis.

**Algorithm 1** Synthesis algorithm

**Require:** Probability Density Function **PDF**
**Require:** Frequency Count list **FC**
**Require:** Continuation Probabilities **CP**
**Require:** Color **C**
**for** each PDF window **do**
  **for** each STFT window **do**
    **for** frequency count FC **do**
      random frequency following PDF : $f_i$
    **end for**
    **for** each determined frequency $f_i$ **do**
      random real $a$ between 0 and 1
      **if** $a \leq$Continuation Probability CP$[f_i]$ **then**
        random phase between 0 and $2\pi$
      **else**
        Calculation of the phase from the precedent STFT window
      **end if**
      amplitude = Color C$[f_i]$
    **end for**
  **end for**
**end for**
**Ensure:** partials (frequency $f_i$, amplitude C$[f_i]$, phase $\Phi_i$)

First, the number of spectral components is set to the parameter *frequency count*. Each one is drawn according to the *probability density* function and the cumulative function. Once a frequency value is determined, another draw (*continuation probability*) is done to decide whether this component will be part of the next frame of the spectrum. In this case, the phase is calculated, otherwise it takes a random value between 0 and $2\pi$. Then an amplitude is set according to the spectral envelope (*color*). An additive synthesis is then performed.

$$x(n) = A \sin(2\pi f \frac{n}{R} + \phi)$$

As the frame size is constant, some undesired periodicities could appear. That is why we have chosen to randomize the frame size $N_0$,

$$N = N_0 + \epsilon \qquad \epsilon \in [-0.1N_0; 0.1N_0]$$

where $\epsilon$ is a random value.

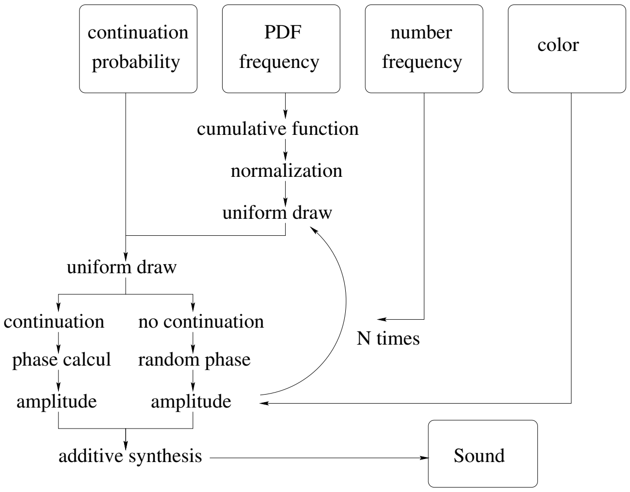


Figure 4: *Synthesis method.*

## 6. IMPLEMENTATIONS AND EXPERIMENTS

The method of extracting the four parameters presented have been implemented and tested [10] (figure 5). First of all, analyzing and synthesizing non-varying sinusoids and filtered white noises have been successful. Of course, these examples on synthetic sounds can also be performed with existing models like SMS and its extensions. A more interesting example come from synthetic sounds created with mixed white noise and sinusoids. Improvement of our technique appear in the cases of musical or sound transformations, when, for example, the deterministic fundamental frequency is modified. SMS-like synthesis would create a sound with the old fundamental always perceptible mixed with the new one because of the errors of the spectrum subtraction [11]. The importance of considering such a sound in its integrity is confirmed.

Moreover the parameters introduced here improve the result on some others synthetic sounds which can not be successfully analyzed with existing models. The main example is the case of a noise which is not a filtered white noise because its probability density function is not uniform (figure 6). Tests were executed in the case of a noise whose probability density function is constant except for a frequency band which is characterized by a most important probability. This type of sound can be perfectly analyzed and synthesized with the proposed statistical approach whereas existing techniques fail (figure 7).

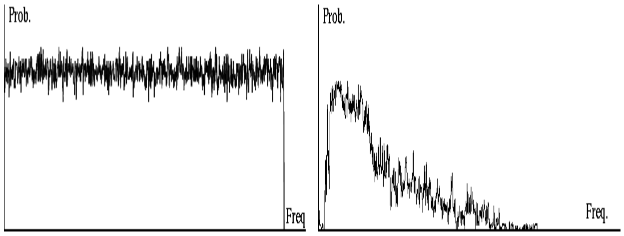


Figure 5: *Analyzed PDF of white noise (left) and percussion sound (right) (analysis parameters: STFT window size =* $2048$ *samples, PDF window size =* $65536$ *samples).*

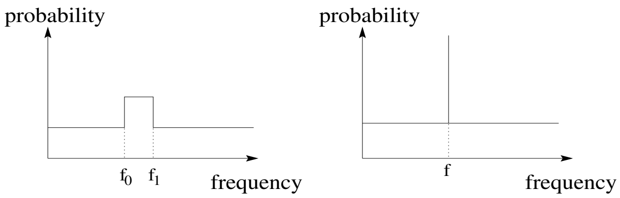


Figure 6: *Frequency PDF of synthesized sounds : white noise with non-uniform PDF (left) and mix of one sinusoid with white noise (right).*

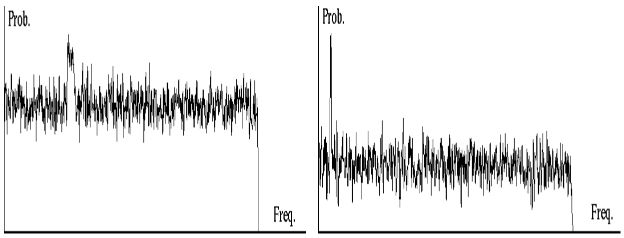


Figure 7: *Analyzed PDF of white noise with non-uniform PDF (*$500$*Hz band) (left) and mix of one sinusoid and white noise (right).*

Several more experiments with a wide variety of natural noise should be done to find sounds which could have the same properties as the synthetic examples presented above.

Some tests on natural sounds like rain, crash cymbal hits or insect noises were promising but make some defects appear because of the limitations of the assumptions described. Indeed many natural sounds contain transients, have deterministic components , or otherwise can not be presented as ergodic processes.

Here are some details about our experiments, especially a few representations of analyzed parameters and synthesized sounds:

| | |
|---|---|
| PDF window size | 65536 samples |
| STFT window size | 2048 samples |
| STFT window type | Hann |
| STFT zero padding | 44100 |
| Analysis method | peak or amplitude threshold |

Table 1: *Analysis parameters used for sound from crash cymbal hit.*

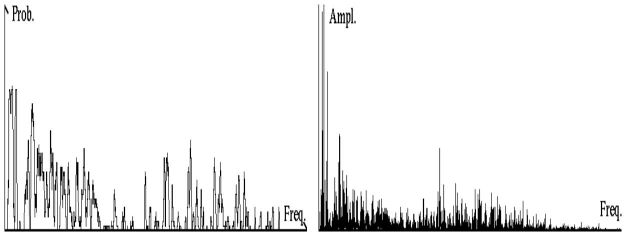


Figure 8: *Extracted parameters from cymbal crash hit at time $t = 0s$ from 65536 samples ($1.49s$) : frequency probability density function (right) and amplitude spectrum or color (right).*

### 6.1. Crash cymbal hit

We have performed several experiments with a sound from a crash cymbal hit, which is often considered as a noise [7]. Although its spectrum is dense, some sine components can clearly be heard. The synthesized sound is promising and has many similarities with the original. However some defects are revealed certainly because of the limitations of the analysis, particularly concerning the transients.

The analysis parameters are listed in table 1. The continuation probability is zero except for a few frequency values corresponding to the deterministic part of the sound. Moreover the number of frequencies is around 1000 and decreases with time.

The figure 8 shows the PDF parameter extracted from the crash cymbal sound and the spectrum associated. We can notice some similarities between the two representations. More studies would explain whether these similarities exist because of the analysis limitations or just because of the sound composition.

### 6.2. Seashore

We have also performed other experiments with seashore noises. As crash cymbal sounds, seashore noises have dense spectrum. The synthesized sound is very similar to the original.

The analysis parameters are listed in table 2. The continuation probability is equal to zero for nearly every frequency value. Moreover the number of frequencies is around 1500 and fluctuates between 500 and 2500.

The figure 8 shows the PDF parameter extracted from the seashore sound and the spectrum associated. We can also notice some similarities between the two representations.

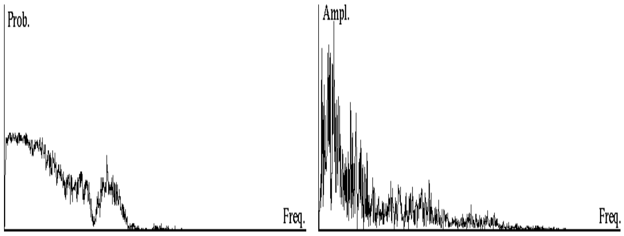


Figure 9: *Extracted parameters from seashore noise: frequency probability density function (left) and amplitude spectrum or color (right)*

| | |
|---|---|
| Analysis window size | 65536 samples |
| STFT window size | 2048 samples |
| STFT window type | Hann |
| zero padding | 44100 |
| Analysis method | peak or amplitude threshold |

Table 2: *Analysis parameters used for seashore sound.*

### 6.3. Limitations

One of the main assumption of this statistical approach is the process ergodicity. This approximation exists mainly for practical reasons. A better experiment would be to extract statistical parameters from many sounds from the same source.

It is obvious that our experiments on natural sounds were limited by fact that assumptions for the parameter extraction method are that sounds mustn't contain any transient or slow-varying deterministic component. But these constraints could be relaxed in the future by extending the analysis method.

Furthermore analysis limits exist when some spectral components are inside the same bin. For example, two components separated by 50 Hz are estimated as only one component by a short-time Fourier transform with window size of 512 samples (sampling rate 44100 kHz). The frequency resolution implied by size of the analysis window introduces a few spectral density errors. Some others methods should perhaps be investigated to try to correct these problems.

## 7. FUTURE WORK

More experiments have to be done with a wide variety of sounds in order to demonstrate the hypothetic link between probability density function and perception. That is why some other psychoacoustic tests should indicate the validity of this statistical parameter. Furthermore amplitude has not be considered as a random variable in this paper. Some works about this parameter will be done to try to treat it as a statistical parameter.

A more advanced statistic model using the parameters introduced in this paper could constitute a solid base for investigating scientific and musical research on the concept of noise. The parameters can already be useful to classify natural noises in different categories defined by statistical properties.

Moreover the synthesis method presented using statistical parameters could lead to an experimental synthesis tools which could permit composers to create and edit sounds. Once the analysis method improved with extended techniques based on short time Fourier transform like *n-order* Fourier transform or techniques for adapting window length, a certain number of musical manipulations would be possible. Transformations of natural sounds will be experimented with electro-acoustic composers in order to validate the utility of statistical parameters and to relate it to musical manipulations and definitions.

## 8. CONCLUSIONS

Our results suggest that a statistical approach is possible for modeling sounds. This can lead to a sound model that does not require a separate treatment of periodic and stochastic components. However, these works must be improved to create a useful tool for electro-acoustic composers by relating the modifications of statistical parameters to musical manipulations.

## 9. ACKNOWLEDGEMENTS

This research was done at the SCRIME (*Studio de Création et de Recherche en Informatique et Musique Electroacoustique*) and was supported by the *Conseil Régional d'Aquitaine, the Ministère de la Culture*, the *Direction Régionale des Actions Culturelles d'Aquitaine*, and the *Conseil Général de la Gironde*.

## 10. REFERENCES

[1] W.M. Hartmann, S. McAdams, A. Gerzso, and P. Boulez, "Discrimination of spectral density," *Journal of Acoustical Society of America*, vol. 79, no. 6, pp. 1915–1925, 1986.

[2] Sh. Dubnov, R. El-Yaniv, Z. Bar-Joseph, D. Lischinski, and M. Werman, "Granular synthesis of sound textures using staistical learning," *Proceedings of International Computer Music Conference (ICMC'99, Beijing)*, pp. 178–181, 1999.

[3] X. Serra and J. Smith, "Spectral modeling synthesis : a sound analysis/synthesis system based on a deterministic plus stochastic decomposition," *Computer Music Journal*, vol. 14, no. 4, pp. 12–24, 1990.

[4] A.S. Verma and T.H.Y. Meng, "Time scale modification using a sines+transient+noise signal model," *Proceedings of the Digital Audio Effects Workshop (DAFX'98, Barcelona)*, pp. 49–52, 1998.

[5] M. Desainte-Catherine and S. Marchand, "Stuctured additive synthesis: Towards a model of sound timbre and electroacoustic music forms," *Proceedings of International Computer Music Conference (ICMC'99, Beijing)*, pp. 260–263, 1999.

[6] M. Desainte-Catherine and S. Marchand, "High precision fourier analysis of sounds using signal derivatives," LaBRI, to be published in Journal of Audio Enginering System 2000, 1998.

[7] A. Freed, "Spectral line broadening with transform domain additive synthesis," *Proceedings of International Computer Music Conference (ICMC'99, Beijing)*, pp. 78–81, 1999.

[8] X. Serra and J. Bonada, "Sound transformations based on the sms high level attributes," *Proceedings of the Digital Audio Effects Workshop (DAFX'98, Barcelona)*, 1998.

[9] S. Marchand, "Musical sound effects in the sas model," *Proceeding of the Digital Audio Effects Workshop (DAFX'99, Trondheim)*, 1999.

[10] P. Hanna, "Modélisation de bruits," Mémoire de DEA, LaBRI, 2000.

[11] X.. Serra, *Spectral Modeling Synthesis Tools*, 1997.