EURASIP Journal on Applied Signal Processing 2005:12, 1794–1806

A Statistical and Spectral Model for Representing Noisy Sounds with Short-Time Sinusoids

Pierre Hanna

SCRIME, Laboratoire Bordelais de Recherche en Informatique (LaBRI), Universite Bordeaux 1, 33405 Talence Cedex, France
Email: hanna@labri.fr

Myriam Desainte-Catherine

SCRIME, Laboratoire Bordelais de Recherche en Informatique (LaBRI), Universite Bordeaux 1, 33405 Talence Cedex, France
Email: myriam@labri.fr

Received 12 March 2004; Revised 10 March 2005; Recommended for Publication by Mark Kahrs

We propose an original model for noise analysis, transformation, and synthesis: the CNSS model. Noisy sounds are represented with short-time sinusoids whose frequencies and phases are random variables. This spectral and statistical model represents information about the spectral density of frequencies. This perceptually relevant property is modeled by three mathematical parameters that define the distribution of the frequencies. This model also represents the spectral envelope. The mathematical parameters are defined and the analysis algorithms to extract these parameters from sounds are introduced. Then algorithms for generating sounds from the parameters of the model are presented. Applications of this model include tools for composers, psychoacoustic experiments, and pedagogy.

Keywords and phrases: stochastic part of sounds, analysis and real-time synthesis of noisy sounds, spectral models, spectral density, musical transformations of sounds.

1. INTRODUCTION

Computers offer new possibilities for sound processing. Applications are numerous in the musical field. Digital sound models are developed to represent signals with mathematical parameters in order to allow composers to transform the original sound in a musical way.

Noises are used more and more frequently in contemporary music, especially in electroacoustic music. A new vocabulary describing noisy sound properties has been proposed during the twentieth century [1]. We consider as noisy sounds the natural sounds such as rubbing or scratching, but also some parts of instrumental sounds like the breath of a saxophone, and speech phonemes such as consonants or whispered voices.

Existing models only represent sounds composed of low noise levels. They consider natural sounds as mixes of sinusoids (deterministic part) and noise (stochastic part). They first extract sinusoids and model the residual using a low noise model such as LPC [2] or piecewise-linear spectral envelopes [3, 4]. 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. The goals is the extraction of the structure of the pseudoperiodic components and to propose a reasonable approximation of the residual relying on psychoacoustics. We focus on robust stand-alone noise modeling.

In this article, we present an original noise model to analyze, transform, and synthesize such noisy signals in real time. 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. Psychoacoustic experiments show that this distribution is perceptually relevant and mainly depends on the number of sinusoidal components. For this reason, we focus on the spectral density and the mathematical parameters related to it.

After presenting existing models and their limitations in Section 2, we present theory behind the representation of noise with short-time sinusoids in Section 3. The new model and its mathematical parameters are defined in Section 4. Then, in Section 5, we propose an original method to extract these parameters from analyzed sounds. In Section 6, the synthesis algorithms are detailed before presenting the limitations of this model and two applications in Section 7.

2. BACKGROUND

Many model types have been considered for music synthesis. In this section, we summarize previous approaches to analyzing, transforming, and synthesizing noise-like signals and indicate their limitations.

2.1. Temporal models

The existing models for analyzing, transforming, and synthesizing noisy sounds are temporal or spectral models. Temporal models generate noises by randomly drawing samples using a standard distribution (uniform, normal, etc.). Then, they may be filtered (subtractive synthesis). The main temporal models use linear predictive coding (LPC) to color a white noise source. These approaches are common in speech research but are less closely linked to perception and are less flexible [5, 6].

2.2. Spectral models

We are particularly interested in spectral models because they are useful for (mostly) harmonic sounds [7]. These sinusoidal models are very accurate for sounds with low noise levels and are intuitively controlled by users [8]. Therefore, it seems interesting to adapt them to the representation of more complex sounds.

Several advances have been proposed in the area of sinusoid-plus-noise models [9]. Macon has extended the ABS/OLA model [10] to enable time-scale and pitch modifications to unvoiced and noise-like signals by randomizing phases [11]. Another extension proposes to modulate the frequencies and/or amplitudes with a lowpass-filtered noise [12].

In 1989, research led to *hybrid* models, which decompose the original sounds into two independent parts: the sinusoidal part and the stochastic part [13]. Extensions have been proposed to consider transients separately [14, 15]. The stochastic part corresponds to the noisy part of the original sound. It is entirely defined by the time-varying spectral envelopes [6]. Other methods use piecewise-linear spectral envelopes [3], LPC [2], or DCT modeling of the spectrum [16]. Another residual model related to the properties of the auditory system is proposed in [4]. The noisy part of any sound is represented by the time-varying energy in each equivalent rectangular band (ERB). However, because of such approximations, artifacts may result if this model is applied to sounds with high-level stochastic components.

Hybrid models considerably improve the quality of synthesized sounds, but it is desirable to present more parameters for the user to control musical sounds. The only musical parameters presented to the composers are amplitude (related to the volume) and spectral envelope (related to the color). We propose to develop a robust noise model that allows the largest possible number of high-fidelity transformations on the analysis data before synthesis.

Experiments have demonstrated the importance of the spectral density of sinusoidal components [17, 18]. A spectral model for noisy sounds is adequate to control these parameters. Furthermore, the color of the noise, related to the spectral envelope, is intuitively represented on the frequency scale. For these reasons, the modeling of noisy sounds in the spectral domain is justified.

However, the mathematical justification of the representation of any random signal by a sum of sinusoids with time-constant amplitude, frequency, and phase, is not straightforward. Similar models have been developed with theory in physics [19].

3. JUSTIFICATION OF A SPECTRAL AND STATISTICAL MODEL FOR NOISES

In this section, we present the justifications from the fields of statistics, physics, perception, and music, for the proposed spectral model of noisy sounds.

3.1. Thermal noise model

Thermal noises can be described in terms of a Fourier series [19]:

$$X(t) = \sum_{n=1}^{N} C_n \sin\left(\omega_n t + \Phi_n\right), \tag{1}$$

where N is the number of frequencies, n is an index, ω_n are equally spaced component frequencies, C_n are random variables distributed according to a Rayleigh distribution, and Φ_n are random variables uniformly distributed between 0 and 2π. The samples X defining the signal are distributed according to a normal law. This definition is the starting point of our work.

This definition represents a noise by a finite sum of sinusoids. It justifies a spectral model for the noisy sounds and is the central point of the model presented in this article. Nevertheless, the thermal noise model does not specify the number of sinusoids and the difference between frequencies. It is obvious that choosing $N = 2$ sinusoids in a frequency band whose width is 20 kHz is not sufficient to synthesize a white noise that is perceptually equivalent to a white noise synthesized by randomly distributing samples according to a Gaussian law.

So long as the number of sinusoids is not small, the synthesized samples are normally distributed because of the law of large numbers [19]. Nevertheless, this normal distribution is not sufficient to define colored noise. The perception is sensitive to the number N of sinusoids (for a given bandwidth). It is detailed in the next section.

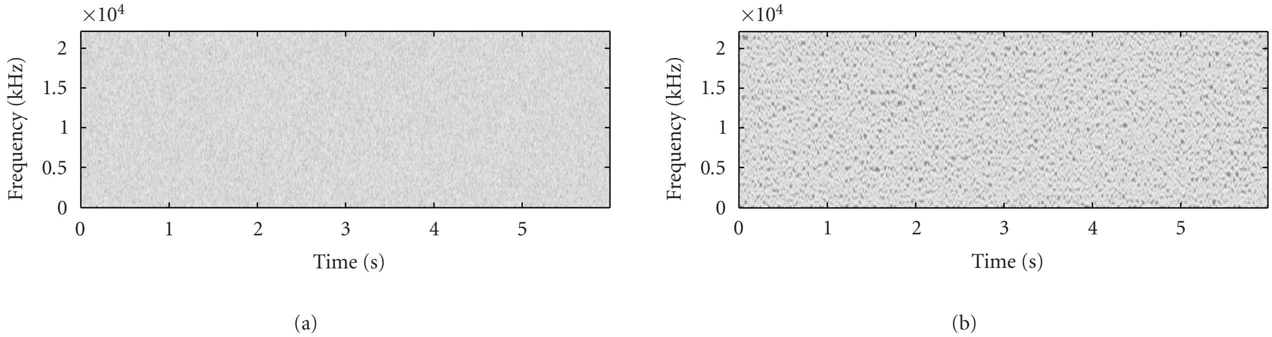

Figure 1: Illustration of the spectral density: spectral differences between (a) Gaussian white noise and (b) white noise whose spectral density is low (synthesized with the CNSS model). The black parts indicate gaps of energy. These gaps are perceived and allow human ears to differentiate these two sounds.

It is important to note that the representation of the stochastic part for the hybrid models (e.g., the SMS model) is implicitly based on the thermal noise model. The only difference comes from the deterministic definition of the amplitudes of the sinusoids. Indeed, the resynthesis part [3] generates short-time spectra from the spectral envelopes by randomly distributing phase values according to a uniform law. Then an inverse Fourier transform is computed. This mathematical operation consists of summing a fixed number of sinusoids whose frequencies are equally spaced, and whose amplitudes are fixed. We denote by F_s the sample rate and by W_s the size of the synthesis window. The difference between successive frequencies is F_s/W_s, and the number of sinusoids is $W_s/2$ (number of sinusoids in the audible frequency range). This synthesis model can be described by the equation:

$$X(t) = \sum_{n=1}^{W_s/2} C_n \sin\left(2\pi n \frac{F_s}{W_s} t + \Phi_n\right). \tag{2}$$

The necessary number of sinusoids needs to be discussed. Even if experiments confirm that the number implicitly used when computing an inverse Fourier transform appears to be sufficient [6], the question is to know if it is necessary. Another question would be to know if it is necessary to define in a random way the amplitudes of sinusoids. Here again, experiments seem to indicate that fixed amplitudes do not introduce audible artifacts [20].

3.2. Perception of the spectral density

3.2.1. Psychoacoustic experiments

For psychoacoustic experiments, noise can be synthesized by two different ways. The first one filters white noise computed by random distribution according to a normal or a uniform law [5]. The second one requires the desired noise spectrum and synthesizes sound by summing sinusoids whose amplitudes depend on this desired noise spectrum [19]. The spectral method is based on the thermal noise model. It is generally preferred because it directly controls the spectrum [21]. This approach raises the question of the necessary number of sinusoids to generate a noise which cannot be discriminated from noises synthesized by random distribution of samples. Gerzso did the first experiments in [17]. These experiments have been improved by Hartmann et al. [18] to study the human ability to discriminate bands of noise composed of different numbers of sinusoids.

Results of these experiments are numerous. In the case of a narrowband of noise, the mechanism of discrimination is related to the intensity fluctuations [21]. In the case of broadband of noise, it is related to the spectral resolution [18]. This result leads to the fact that humans would perceive the energy variations in the short-time spectrum corresponding to wide intervals between two neighbor frequencies. Figure 1 shows two synthetic sounds: The second one is characterized by a low spectral density that is indicated by black points corresponding to spectral gaps.

These experiments show that human auditory system is sensitive to the number of sinusoids used to synthesize bands of noise. In the following, this number is thus assumed to be a perceptual characteristic of sounds. It is related to spectral gaps or intensity fluctuations. The control of the number of sinusoids is thus related to perception of sounds.

Moreover, these experiments confirm that a spectral approach to noise synthesis is possible. Indeed, it is now possible to compute an adequate number of oscillators to synthesize white noise whose spectral density and bandwidth are at their maximum. This case corresponds to the highest computational cost.

3.2.2. Definition of the spectral density

Psychoacoustic experiments show that the spectral density is used by the auditory system to discriminate bands of noise. Nevertheless, giving an exact and complete definition of the spectral density is difficult. Gerzso related the spectral density to the ratio of the number of sinusoids by the width of the frequency band [17]. For a band of noise of width ΔF with N sinusoids, the spectral density ρ is defined as

$$\rho = \frac{N}{\Delta F}. \tag{3}$$

We believe that this definition is not strictly correct, because it does not take into account the distribution of the sinusoidal component frequencies [20] and the duration of the band of noise. Perception of pitch depends on the duration of sounds [22]. The experiments we have done confirm that the use of successive short-time windows may cancel the sensation related to a low spectral density.

Indeed, the difference between a thermal noise and a harmonic sound comes from the value of the difference between successive frequencies. This difference corresponds to the fundamental frequency. Periodic sound waves can have a pitch only if it has a sufficient duration. This duration depends on the periodicity. Psychoacoustic experiments indicate that the number of cycles necessary lies in the range of tens of cycles [23].

This observation is also confirmed by the usual method based on the inverse Fourier transform. This method considers the number of sinusoids as a function of the number of samples, and thus as a function of the duration of the synthesized sound.

In the following, we consider two independent parameters: the number of sinusoids and the duration of the sound.

3.3. Statistical model

The spectral model we propose is based on the thermal noise model. This model defines the frequencies of the sinusoids as equally spaced. The study of the perception of the spectral density shows that humans can perceive spectral gaps or intensity fluctuations. These phenomena can be due to one or more missing sinusoids. We thus propose to define frequencies as random variables which are controlled by mathematical parameters. The random property of the frequencies is justified because the ear is not sensitive to the precise information about the intensity fluctuations or the spectral gaps, but to their statistical properties. It is useful to represent the probabilities, but it seems useless to retain exact informations about these properties.

Moreover, the study of the intensity fluctuations shows that they are dependent on the distribution of the phases of sinusoids [20]. This dependency is illustrated by the two limits: phases with same values and uniformly distributed values. In the first case, intensity fluctuations grow as the number of sinusoids increases [19], because the corresponding waveforms are composed of one or more intensity peaks that are audible. Conversely, uniformly distributed phases correspond to the thermal noise model and lead to weak intensity fluctuations. Therefore, it appears useful to control this phase distribution in order to modify the audible properties related to the intensity variations.

The thermal noise model considers the amplitudes of the sinusoids as random variables distributed according to a Rayleigh law. Practically, fixed amplitudes lead to bands of noise that cannot be discriminated from bands of noise synthesized with sinusoids whose amplitudes are randomly determined [19]. Moreover, we haven't managed to relate the distribution of the amplitudes to a perceptual property. For these reasons, we restrain our spectral model to fixed amplitudes determined from spectral envelopes.

3.4. Mathematical justification

The distribution of the frequencies and the phases of sinusoids that compose bands of noise are perceptually relevant. The synthesized signal can thus be described from (1) by

$$X_k = \sum_{n=1}^{N} a_n \sin\left(2\pi F_n \frac{k}{F_s} + \Phi_n\right), \tag{4}$$

where F_n and Φ_n are random variables, and a_n are fixed values.

We can mathematically show that this spectral and statistical approach, based on the thermal noise model, defines a white noise in the case of constant amplitudes (for all n, $a_n = a_0$). White noise satisfies the following equation:

$$\forall n, \forall \tau \neq 0, \quad \mathcal{E}\left(X_n X_{n+\tau}\right) = 0, \tag{5}$$

where $\mathcal{E}$ denotes the expectation [24].

By writing for all (p, q) the product of the expectation of X_p and $X(p+q)$, we have

$$\begin{aligned}
&\mathcal{E}\left(X_p X_{p+q}\right) \\
&\quad = \mathcal{E}\left(\sum_{n=1}^{N} a_n \sin\left(2\pi F_n \frac{p}{F_s} + \Phi_n\right) \sum_{l=1}^{N} a_l \sin\left(2\pi F_l \frac{p+q}{F_s} + \Phi_l\right)\right), \\
&\mathcal{E}\left(X_p X_{p+q}\right) \\
&\quad = \sum_n \sum_l a_n a_l \mathcal{E}\left(\sin\left(2\pi F_n \frac{p}{F_s} + \Phi_n\right) \sin\left(2\pi F_l \frac{p+q}{F_s} + \Phi_l\right)\right), \\
&\mathcal{E}\left(X_p X_{p+q}\right) \\
&\quad = \frac{1}{2} \sum_n \sum_l a_n a_l \Bigg[\mathcal{E}\left(\cos\left(2\pi F_n \frac{p}{F_s} + 2\pi F_l \frac{p+q}{F_s} + \Phi_n + \Phi_l\right)\right) \\
&\qquad + \mathcal{E}\left(\cos\left(2\pi F_n \frac{p}{F_s} - 2\pi F_l \frac{p+q}{F_s} + \Phi_n - \Phi_l\right)\right)\Bigg].
\end{aligned} \tag{6}$$

Since this expectation is defined by integrating over the phases, which are assumed to be uniformly distributed in the interval $[0; 2\pi)$, the equation reduces to

$$\mathcal{E}\left(\cos\left(2\pi F_n \frac{p}{F_s} + 2\pi F_l \frac{p+q}{F_s} + \Phi_n + \Phi_l\right)\right) = 0 \tag{7}$$

and, for $l \neq n$,

$$\mathcal{E}\left(\cos\left(2\pi F_n \frac{p}{F_s} - 2\pi F_l \frac{p+q}{F_s} + \Phi_n - \Phi_l\right)\right) = 0. \tag{8}$$

It leads to

$$\mathcal{E}\left(X_p X_{p+q}\right) = \frac{1}{2} \sum_n a_n^2 \mathcal{E}\left(\cos\left(2\pi F_n \frac{q}{F_s}\right)\right). \tag{9}$$

We conclude that for all integers (p, q),

(i) $\mathcal{E}(X_p X_{p+q}) = 0$ if $q \neq 0$,
(ii) $\mathcal{E}(X_p X_{p+q}) = (1/2) \sum_n a_n^2$ if $q = 0$.

These equations correspond to the definition of white noise.

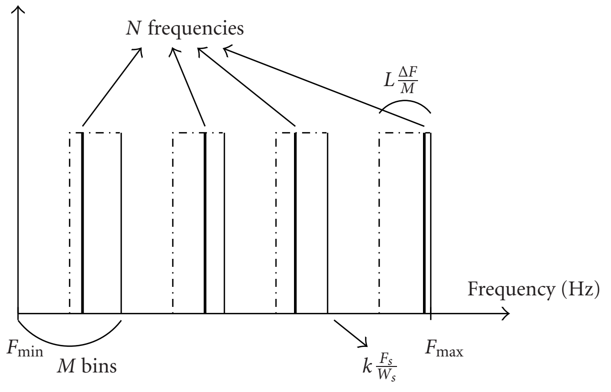

Figure 2: Parameters for the control of the frequency distribution: the number of bins defines the edge of the bins and the bin width is determined by the parameter L.

Concerning the white noise, two assumptions are thus imposed by this definition. The first one concerns phases which have to be uniformly distributed between 0 and 2π. The second one concerns frequencies which also have to be uniformly distributed all over the audible frequency range.

4. THE CNSS MODEL

In the previous section, we have shown that a band of noise can be represented by a sum of sinusoids. This representation is the starting point of the statistical and spectral model we present in this paper: the CNSS model.

4.1. Definition

The CNSS model (*colored noise by sum of sinusoids*) defines sounds as random processes X_k. They are represented by a fixed sum of sinusoids whose amplitudes a_n are fixed and whose phases Φ_n and frequencies F_n are random variables. Phases Φ_n are distributed according to a uniform law in the interval $[0;2\pi)$ and frequencies F_n are distributed in a band whose width is denoted ΔF. Therefore, (4) defines the CNSS model.

4.2. Short-time frames

Practically, the sound is analyzed and synthesized by overlapping and adding two (or more) temporal *frames*. Each frame is defined by different sets of parameters. This approach does not appear in the definition of the thermal noise model. However, it can be justified. Indeed, as previously seen, the number of sinusoids depends on the size of the synthesis window. This number can be reduced by considering successive short-time signals: the duration is too short for ears to perceive the low spectral density. Furthermore, real-time synthesis requires successive short-time windows in order to enable modifications of the parameters from frame to frame.

4.3. Parameters

The CNSS model represents sounds by analyzing successive temporal frames. Each frame is modeled by many mathematical parameters. The duration of these frames is denoted by W_s. It is a positive integer and is expressed in samples. Concerning the distribution of frequencies, M bins ($M \leq N$), equally spaced, divide the frequency bandwidth. In each successive frame, N frequency values are drawn into these bins from a uniform distribution. These parameters are illustrated by Figure 2 and are detailed below.

4.3.1. Bandwidth

The signal represented is supposed to be a band of noise. One of the parameters of the CNSS model is the width of this band. It defines the interval of the probability density function of the frequencies. It is denoted by ΔF and is defined by a maximum frequency $F_{\max}$ and a minimum frequency $F_{\min}$:

$$\Delta F = F_{\max} - F_{\min}. \tag{10}$$

Since we assume $\Delta F > 0$, we also assume $F_{\min} < F_{\max}$.

This parameter is obviously constrained by the resolution of the auditory system ($20 - 20\,000$ Hz). However, due to the Nyquist criteria (sample rate $F_s = 44\,100$ Hz), the interval is $[0; F_s/2 = 22\,050]$ Hz. It also corresponds to the interval implicitly considered when an inverse Fourier transform is computed.

4.3.2. Number of bins

In order to describe the probability density function of frequencies, we propose to define bins, whose sizes are constant, covering the entire bandwidth ΔF. The number of bins is a parameter and is denoted by M.

Each bin, denoted by B_i ($i \in [0; M-1]$), defines an interval of the bandwidth ΔF. The width of every bin, denoted by ΔB, is constant:

$$\Delta B = B_{i+1} - B_i = \frac{\Delta F}{M}, \quad 0 \leq i < M - 1. \tag{11}$$

The interval I_{B_i} defined by each bin B_i is

$$I_{B_i} = \left(F_{\min} + i\frac{F_{\max} - F_{\min}}{M}; F_{\min} + (i+1)\frac{F_{\max} - F_{\min}}{M}\right], \quad 0 \leq i \leq M-1. \tag{12}$$

The number of bins is positive and is not bounded:

$$M \in [1; \infty]. \tag{13}$$

Inside each bin, at most one frequency is randomly chosen according to a distribution law defined by the other parameters N and L.

4.3.3. Number of sinusoids

The number of sinusoids, denoted by N, appears in (4) of the CNSS model. This number is linked to the number of bins M. It is not possible to define more than one frequency in the same bin. At the opposite extreme, at least one sinusoid composes the signal:

$$1 \leq N \leq M. \tag{14}$$

As previously seen in Section 3.2, the influence of the number of sinusoids is not perfectly understood yet. It is linked to the duration of the synthesis frames [20]. Nevertheless, we present approximations about the linear variations of this number as a function of the bandwidth and the duration.

If N equally spaced frequencies are defined in a band whose width is $F_s/2$ Hz, the difference between successive frequencies is $F_s/2N$ Hz. In order to be perceived, the minimum duration is $2N/F_s$ seconds, which corresponds to $2N$ samples. Therefore, $W_s/2$ sinusoids have to be used to define a noise with maximum spectral density. It is important to note that this value is the number of sinusoids used when computing an inverse Fourier transform of size W_s. The usual technique for the synthesis of the stochastic part of hybrid models [3] requires a number of sinusoids corresponding to the maximum spectral density. Therefore, filtered white noises that can be synthesized applying this technique are always characterized by a maximum spectral density.

In the case of white noise (bandwidth $F_s/2$ Hz), the maximum number is half the synthesis frame duration W_s:

$$N \leq \frac{W_s}{2}. \tag{15}$$

As a conclusion, a band of noise defined by a width ΔF, a duration W_s, and a maximum spectral density, is represented by

$$N_{\max} = \frac{\Delta F \cdot W_s}{F_s}. \tag{16}$$

The number N of sinusoids is thus defined in the interval $[0; \Delta F \cdot W_s/F_s]$.

4.3.4. Width of the PDF of frequencies

Inside each selected bin, one frequency is randomly determined according to a uniform law. One parameter, denoted by L, defines the relative width of this law. Its value is in the interval $[0; 1]$. When it is null, the probability density function is a delta function, and the frequency is the upper boundary of the bin. In the bin B_i, the frequency would be $F_{\min}+(i+1)((F_{\max} - F_{\min})/M)$. At the opposite extreme, if the parameter L is 1, the probability density function is a rectangular function: all the frequencies have the same probability to be chosen.

In the case when the number of sinusoids equals the number of bins $N = M$, we write

$$\begin{aligned} &\text{if } L = 0, \quad F_i = F_{\min} + (i+1)\frac{F_{\max} - F_{\min}}{M}, \\ &0 < L \leq 1, \quad F_i \in \left(F_{\min} + (i+1-L)\frac{F_{\max} - F_{\min}}{M};\right. \\ &\qquad\qquad \left. F_{\min} + (i+1)\frac{F_{\max} - F_{\min}}{M}\right]. \end{aligned} \tag{17}$$

The probability density function, denoted by ρ and associated to the bin B_i, is, for $L \neq 0$,

$$\begin{gathered} \forall i \in [0; N-1], \quad \forall F_i \in \left(F_{\min} + (i+1-L)\frac{\Delta F}{M}; F_{\min} + (i+1)\frac{\Delta F}{M}\right], \\ \rho(F_i) = \frac{1}{L(\Delta F/M)}. \end{gathered} \tag{18}$$

This parameter defines the regularity of the differences between the frequencies composing modeled sounds. For example, when $L = 0$ and $N = M$, all frequencies are equally spaced:

$$F_{i+1} - F_i = \frac{\Delta F}{M}, \quad 0 \leq i \leq N-2. \tag{19}$$

4.3.5. Width of phase PDF

Thermal noise model defines phases of each sinusoid composing sounds as random variables distributed in the interval $[0; 2\pi)$ according to a uniform law. The CNSS model allows the modification of this law by limiting the interval $[0; 2\pi)$. The relative width of the probability density function is a real number in the interval $[0; 1]$ and is denoted by P. When it is null, the probability density function is a delta function and all the phases are the same. This results in a intensity peak occurring at periodic times and depending on the duration of the frames. When this parameter P is 1, the phases are uniformly distributed according to the thermal noise model.

By considering the phases $\Phi_i^{t_0}$ of the sinusoids at the time t_0, the relation between the parameter P and these phases is

$$\forall i \in [0; N-1], \quad \Phi_i \in \left[\frac{\pi}{2} - P\pi; \frac{\pi}{2} + P\pi\right], \quad 0 \leq P \leq 1. \tag{20}$$

We thus write the phases of sinusoids at the time $t = 0$, as a function of their frequency F_i:

$$\Phi_i \in \left[\frac{\pi}{2} - \frac{\pi F_i W_s}{F_s}; \frac{\pi}{2} + 2P\pi - \frac{\pi F_i W_s}{F_s}\right], \quad 0 \leq P \leq 1. \tag{21}$$

4.3.6. Color

The color is a parameter already used in other spectral representations of sounds (SMS [6], STN [14], etc.) and refers to the spectral envelope. The SAS model [7] also introduces this parameter. Its name is due to the analogy between audible and visible spectra [8].

In the CNSS model, the color is defined by smoothed spectral envelopes. It is denoted by C and represents the variations of the amplitude as a function of the frequency. It is theoretically a continuous function, but it is modeled as a finite number of points. This representation allows manipulations that are more intuitive than the manipulations of filters [3].

Here, the main point is the independence between the spectral envelope and the spectral density. Existing models only consider spectral envelope, and the information related to the spectral density is contained in the spectral envelope or is not taken into account. The CNSS model allows independent manipulations of the spectral envelope and the spectral density.

4.4. Generalization of the filtered white noise models

The CNSS model is essentially a generalization of the filtered white noise models. The mathematical parameters of the model enables control of the frequency distribution. Nevertheless, it is possible to define frequencies of sinusoids as fixed values, according to the existing models. By choosing a band of frequency whose width is half of the sample rate ($\Delta F = F_s/2 = 22\,050\,\text{Hz}$), with the number of frequencies N equal to the number of bins M and with the relative width L null, the frequencies are no longer random variables. They also are equally spaced:

$$F_i = i\frac{F_s}{W_s}, \quad 0 \leq i \leq N-1. \tag{22}$$

When the number of frequencies is half of the length of the frame,

$$N = M = \frac{W_s}{2}, \qquad L = 0, \qquad \Delta F = \frac{F_s}{2}, \tag{23}$$

it is equivalent to an inverse Fourier transform.

5. ANALYSIS

The CNSS model represents noisy sounds with two perceptual parameters: the spectral density and the spectral envelope. The analysis stage consists of approximating these two parameters and estimating the related mathematical parameters that are described in the previous section.

5.1. Approximation of the spectral density

As previously seen, psychoacoustic experiments show that energy gaps in the spectrum of noisy sounds are perceptually relevant. These energy gaps are related to intensity fluctuations [21]. We have proposed an original method [25] to analyze these properties. It is based on the statistical study of these fluctuations.

5.1.1. Limitations of the usual techniques

The study of the energy distribution is based on the use of the short-time Fourier transform (STFT) [26]. Two main limitations lead us to choose another way. The first limitation concerns the resolution of this discrete transform and the usual problem of the tradeoff of time versus frequency. The second one is related to the analysis algorithm. One basic idea would be to detect gaps in the amplitude spectrum. But the determination of thresholds is needed, and this determination must rely on psychoacoustic research. Furthermore, approximations of the short-time Fourier transform lead to amplitude gaps that are due to the analysis windows applied to the sound [27]. For these reasons, we applied another method based on the statistical analysis of the intensity fluctuations.

5.1.2. Statistical analysis of the intensity fluctuations

The intensity fluctuations have been studied and modeled in order to explain the ability for humans to discriminate noises with different spectral densities [18]. Another theoretical study of these intensity fluctuations leads to comparable results [28, 29]. We relate the variance of the envelope power of any signal to the number of sinusoidal components composing this signal. We define V_{NEP} as the ratio of this variance to the average envelope power:

$$V_{\text{NEP}} = \frac{V(E^2)}{\langle E^2 \rangle^2}. \tag{24}$$

We consider a narrow frequency band. This condition allows us to assume that the amplitudes of the sinusoidal components that compose the signal within this band are equal. We show that V_{NEP} is directly linked to the variations of sinusoidal components of the analyzed signal. In this case, the theoretical relation between the measure V_{NEP} and the number of sinusoidal components is

$$V_{\text{NEP}} = \frac{V(E^2)}{\langle E^2 \rangle^2} = 1 - \frac{1}{N}. \tag{25}$$

The method we propose consists of producing several values obtained by successively computing the measure V_{NEP} on filtered signal. The consecutive calculations of V_{NEP} lead to an approximation of the intensity fluctuations and thus to the presence of energy gaps. Indeed, if the analyzed band is composed of noise that is modeled by several sinusoids ($N \gg 1$), the measure V_{NEP} is high. At the opposite end, if the band is composed of a very few sinusoidal components $N \approx 1$, the measure V_{NEP} becomes low.

The analysis method consists of the following operations.

1. Bandpass filtering: this first stage is basic and consists of bandpass filtering the original sound in order to generate signals for the estimation of the intensity fluctuations.
2. Calculation of the measure V_{NEP}: the measure of V_{NEP} is done using the envelope power of the signal, as given in (25).
3. Thresholding: once the values of V_{NEP} have been computed, the next stage consists of counting the number $\mathcal{N}$ of values of V_{NEP} inside a frequency band which are below the selected threshold t_h. This threshold t_h is one parameter of this analysis method. After this stage, a number $\mathcal{N}$ is associated to each frequency value F, center of each studied frequency band.
4. Maximization: an iteration on the width of the frequency band used to compute V_{NEP} leads to the maximum value of $\mathcal{N}$. This maximum is assumed to be the difference between two sinusoids inside the analyzed band or, to say it differently, the size of the spectral gap in the analyzed band [20].

This method leads to an approximation of the differences between sinusoids as a function of the frequency. Several experimental examples are shown in [20].

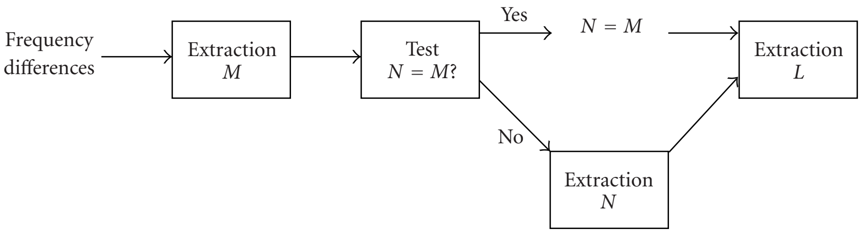

Figure 3: Principle of the analysis of the statistical parameters of the model.

5.1.3. Assumptions

The method relies on two main assumptions. The first one concerns the variations of the spectral envelope. The analysis stage of this spectral envelope consists of smoothing it using lowpass filter and compressing it. For this reason, we assume that in a narrow frequency band that the spectral envelope does not vary enough to introduce variations for the values of V_{NEP} inside this band. The spectral envelope can be assumed as constant. Nevertheless, this assumption obviously depends on the width of the frequency band used. Choosing the width too large would result in variations of the spectral envelope and thus in errors concerning the estimation of V_{NEP}.

The second assumption concerns the approximation of the spectral density of the frequency bands studied. We assume that the spectral density is constant in this frequency band. Of course, the validity of this assumption depends on the width of the frequency bands studied: the larger this band, the weaker the probability for the spectral density to be constant.

5.2. Extraction of the parameters of the model

We represent the distribution of the frequencies with the parameters N, M, and L, where N is the number of sinusoids, M is the number of bins, and L is the width of the PDF of frequencies. We estimate these parameters by analyzing the approximation of the spectral density using the previously discussed method. Therefore, this stage consists of linking the approximation of the spectral density to these mathematical parameters. The different steps of this part of the analysis algorithm are illustrated in Figure 3. The first part consists of estimating the number of bins M because it is directly related to the maximum of the probability density function of the difference between frequencies. The second part tests if the number of frequencies N is different from the number of bins M. If it is different, this number of sinusoids is estimated. Then the width of the probability density function inside each bin is approximated.

5.2.1. Estimation of the probability density function of the frequency difference

We denote by q the probability density function of the difference between two successive frequencies (or the width of a spectral gap). This function is obtained from the results of the approximation of the variations of the spectral density as a function of frequency. These variations have the same characteristics as the probability density function q. The properties of this function give the estimation of the parameters of the CNSS model.

5.2.2. Estimation of the number of bins M

Figure 4 shows an experimental probability density function of the difference between two successive frequencies. It has been computed with a high number (10000) of outcomes of frequency drawings. Different values of M have been chosen. This figure shows that the most probable value corresponds to the value $\Delta F/M$. The number of bins is thus directly linked to the most probable difference between two successive frequencies:

$$q^{\max} = \frac{\Delta F}{M}. \tag{26}$$

5.2.3. Equality between N and M

Once the number of bins has been determined, the analysis method uses the symmetry of the cumulative function extracted. Indeed, theory shows that the probability density function is symmetric around the value $\Delta F/M$ in the case of $N = M$, as opposed to the case $M > N$. These two cases are represented in Figure 4.

The algorithm that is proposed to determine whether the number of frequencies N and the number of bins M are the same, tests the maximum value of q that is not zero. We denote this value by q^m. If $N = M$, this value is slightly less than twice the number of bins M. Otherwise, the number of frequencies N is less than the number of bins M:

$$q^m > 2M. \tag{27}$$

5.2.4. Estimation of the number of frequencies N

In the case when N is different from M, an algorithm is proposed to estimate the number of frequencies needed. This algorithm calculates the number of q which is greater than a fixed frequency difference. Indeed if M is greater than N, the number of wide intervals between neighbor frequencies becomes higher. For a fixed number of bins M, the higher the number of frequencies, the higher this number of wide intervals. However, this method is slightly more complex than the ones used for the determination of M and the equality between M and N, because it requires a calibration stage [20]. Indeed, this calibration stage is necessary because the number of wide intervals detected between neighbor frequencies depends on the analysis parameters.

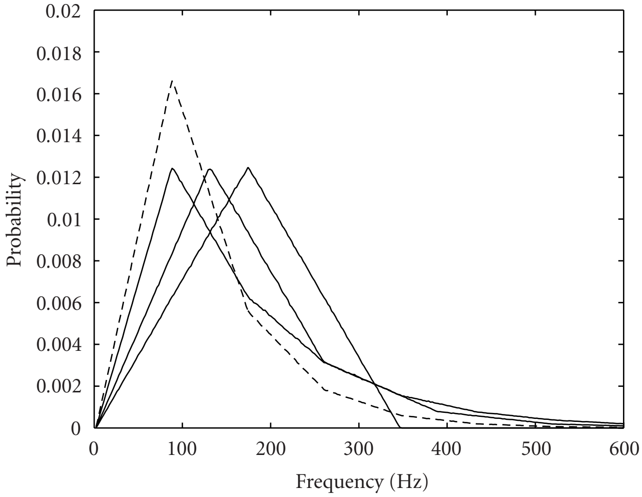

FIGURE 4: Experimental illustration ($\Delta F = F_s/2$ Hz) of the probability density function of the difference between two successive frequencies in the cases of $N = 128$ and $M = 128$ (right), $M = 171$ (center), and $M = 256$ (left), and in the cases of $N = 171$ and $M = 256$ (dotted lines). The maximum is $dF_{\max} = \Delta F/M$ (resp., 2, 3, and 4 bins, which are equivalent to 86 Hz, 129 Hz, and 172 Hz). Only the curve corresponding to the case $N = M$ is symmetric.

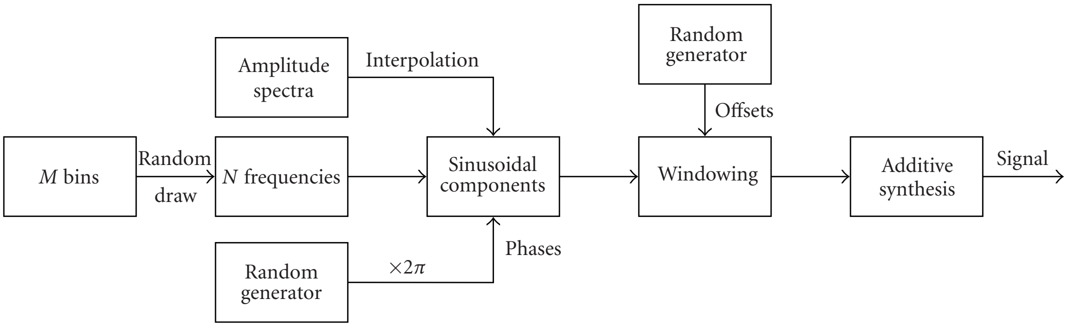

FIGURE 5: Synthesis block diagram.

5.2.5. *Estimation of the harmonicity coefficient* L

The parameter L is correlated to the harmonicity of the sound and thus to its periodicity. A low value L (near 0) imposes a fixed difference between successive frequencies whereas a high value (near 1) implies a distribution of the differences between 0 and $\Delta F/M$. For this reason, we compute the autocorrelation function to extract the value of L from the analyzed sound. We measure the ratio of the second maximum to the first point (zero-lag peak) of the autocorrelation function (which is the total energy of the signal and the maximum). This ratio is used as a discrimination function of the voiced and unvoiced sounds for speech [30].

5.2.6. *Extraction of the spectral envelope*

The spectral envelope (also named the color [7]) is estimated using the same process used in the other spectral models [3]. A short-time Fourier transform is performed and the classical methods usually applied to residual (spline interpolation, line-segment approximation, etc.) can be computed to find a function that matches the amplitude spectrum.

The CNSS model needs an adapted analysis method to be able to estimate the spectral density of natural noisy sounds. The limitations of the short-time Fourier transform necessitate the use of new approaches. The proposed method has been successfully tested on synthetic and natural sounds [20, 25]. It is difficult to compare its accuracy because, to our knowledge, there are no comparable alternatives. The technique proposed is still in experimentation and will certainly be improved in the future. But, for now, it is the only method that allows the estimation of the spectral density of the sinusoidal components and enables the extraction of the CNSS parameters.

6. SYNTHESIS

In this part, we present the algorithms used to synthesize sounds from the statistical and the mathematical parameters. The first part consists of generating the oscillators for each successive frame. The frequency, amplitude, and phase of each sinusoid are computed using the model parameters.

Then the temporal samples are generated in each successive frame to produce the synthesized sound. Figure 5 shows the general diagram for the synthesis.

6.1. Determination of sinusoids

6.1.1. Frequencies

The frequency values of the sinusoidal components of each frame have to be computed from the statistical parameters. The first step consists in defining M bins (denoted by B_i) from the bandwidth values:

$$B_i = \left[i\frac{F_{\max} - F_{\min}}{M}; (i+1)\frac{F_{\max} - F_{\min}}{M} \right). \tag{28}$$

Then N frequencies are determined from these M bins. N bins have to be drawn from the M possibilities. A statistically correct algorithm to choose these bins is the classic algorithm to randomly define a permutation. One bin i is randomly drawn, then bins $M-1$ and i are interchanged. Another bin is randomly chosen from the $M-1$ last bins. This algorithm repeats until all N bins are chosen.

This algorithm has a large cost if M is large compared to N because one large array has to be initialized and manipulated in each frame. But experiments show that defining M more than 100 times N amounts to a uniform draw of frequencies:

$$f_i = F_{\min} + \text{rand}\left(F_{\max} - F_{\min}\right). \tag{29}$$

We could consider the special case where there are the same number of frequency bins as sinusoids ($N = M$). The calculation would be more efficient in that case, because we could directly associate sinusoid i with bin B_i ($i \in \{0,\ldots,N-1\}$). But we know that most of the time spent in the synthesis is spent in partial synthesis, so improving the algorithm for that special case may not pay off enough.

Once the bins have been chosen, frequency values have to be determined from the parameter L. Another uniform draw is made in a band which is defined by the upper bound of the bin B_j ($j \in \{0,\ldots,M-1\}$) and whose length is L multiplied by the bin length $(F_{\max} - F_{\min})/M$:

$$f_i = F_{\min} + \left(j + 1.0 - \text{rand}(L)\right)\frac{F_{\max} - F_{\min}}{M}. \tag{30}$$

Therefore, the following operations have to be done in sequence for each frame of the temporal signal.

```
(1) Define an array $b$ of integers $[0, M-1]$.
(2) For $n \in [0, N-1]$,
    (a) draw an integer $k$ from the last $M-n$,
    (b) draw a real $r$ in $[0; L]$,
    (c) calculate $f_n = F_{\min} + ((1-r) + b[k]) * (F_{\max} + F_{\min})/M$,
    (d) set $b[k] = b[M-n-1]$.
```

6.1.2. Determination of phases

The model of thermal noise described in [19] imposes on each component its phase to be uniformly distributed:

$$\phi_i = \text{rand}(2\pi). \tag{31}$$

However, noise synthesized with sinusoids with equal phases results in intensity peaks. These peaks can be periodic depending on the length of the synthesis window. Such noises are described as *impulsive* noises. By changing the width of the probability density function of phases, users can control the amplitude of these peaks. The proposed synthesis model introduces a new parameter by controlling the relative width P ($P \in \mathbb{R}$, $0 \le P \le 1$) of the probability density function of the phase:

$$\phi_i = \text{rand}(2\pi P). \tag{32}$$

6.1.3. Determination of amplitudes

The amplitudes are simply defined from the frequency values and the spectral envelope by linearly interpolating the smooth spectral envelope extracted from the synthesis model. However, other types of interpolation (splines, LPC, etc.) are possible.

6.2. Additive synthesis of frames

Once the frequency, amplitude, and phase values are calculated, temporal samples are generated with additive synthesis. An efficient algorithm is presented in [31]. This algorithm can generate approximatively 2 partials per sample for each MHz of CPU clock speed.

The algorithms we present have been implemented to create a real-time sound synthesizer. The most CPU consumption is in the case of white noise (or filtered white noise). Synthesizing sounds with more sinusoidal components is useless, because the difference cannot be heard by increasing N. For this reason, we define a maximum value for N depending on the synthesis window size. N cannot be greater than half of the synthesis window size (W_s in samples). This limit corresponds to the inverse Fourier transform technique [32, 33, 34, 35]:

$$N \le \frac{W_s}{2}. \tag{33}$$

6.3. OLA of frames

Spectral synthesis techniques often use the overlap-add method. The resulting temporal signal does not taper to 0 at the boundaries of each frame because of the random values of phase spectrum. This may also be the case when analyzed sounds are transformed. This is the reason why the synthesis method uses the OLA technique. But in the case of noise synthesis, both experiments and theory show that this method introduces intensity fluctuations which result in audible artifacts [36]. Indeed, the statistical moments are not preserved. We have proposed new methods to avoid these variations. We next describe a method which involves time shifting the sinusoids.

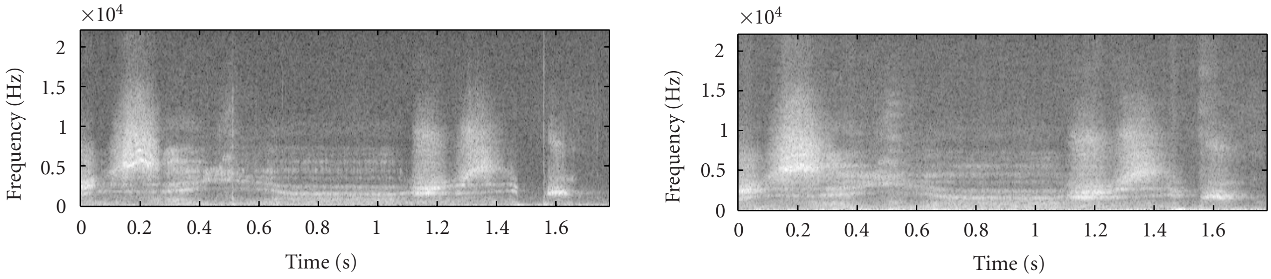

Figure 6: Spectrographic plots of sounds: (a) an original whispered French sentence; (b) sentence analyzed then synthesized using the CNSS model.

This method is applied to N sinusoids (denoted by s^n) with random phase. It consists in shifting the start of each sinusoid in each frame in order to distribute the intensity variations introduced by the weighting windows. Thus each component starting time (d_n with $n \in \{0, \dots, N-1\}$) is set to different values before being multiplied by the weighting window. The resulting signal x' can be written as

$$x'(t) = \sum_{l=0}^{L-1} \sum_{n=0}^{N-1} s_l^n(t - lH - d_n) w(t - lH - d_n), \qquad (34)$$

where s^n are sinusoids. By choosing d_n equally spaced over the half window, one can show that this sum of sinusoids leads to noise with constant statistical properties.

There are many ways to determine the offset values. For example, they can be randomly drawn according to a uniform distribution. But this method may lead to artifacts because many offsets may have the same value, which introduces variance fluctuations [36]. To avoid these probabilities, we prefer choosing these values by dividing the half window in bins.

As a conclusion, the following operations have to be done in sequence for each partial of each frame of the temporal signal:

(1) draw an offset *off*,
(2) synthesize the current partial,
(3) multiply it by weighting window,
(4) offset output buffer with *off*,
(5) add partial buffer to the output buffer.

7. APPLICATION AND CONCLUSION

This noise model is still being tested, especially with respect to the analysis method. We now present applications and details about the implementation.

7.1. Implementation

In order to test the real-time capabilities of the model, we developed the synthesis part of the model on one of the existing free software tools for real-time audio. The objective was to control all the synthesis parameters as fast as possible, while the sound is rendered. The first target was jMax (see http://www.ircam.fr/jmax) because we have used it with the SAS sound model [7] successfully for a long time.

The libcnss library and its jMax extension are free software developed on the GNU/Linux platform. They are available at http://scrime.labri.fr.

7.2. Applications

We can consider two uses for the CNSS model. The first one is the synthesis by using composer-specified control parameters. The synthesis-based applications do not apply to the analysis process. They consists in changing parameters in real time in order to modify synthesized synthetic sounds. This approach will be very useful to understand the perception of noisy sounds. Applications for this noise model have been developed. One of the major application is the pedagogic tool Dolabip [37]. The application uses the two sound models SAS and CNSS to help children to understand sound phenomena.

The second use for the CNSS model is the analysis followed by synthesis, with or without modification. Parameters of the CNSS model are extracted from natural noisy sounds. The main interest is to be able to perform original transformations on analyzed sounds. Users can then modify analyzed before resynthesizing transformed original sounds. Figure 6 shows an example of whispered voice that is analyzed and then resynthesized. The transformations allowed by the CNSS model are perceptually and musically relevant.

(i) *Time scaling.* In [38] we presented a method to perform time transformations without changing the statistical properties of noises. The first experimentations we have done show the limitations of the analysis methods. We hope to considerably improve these results by further development of better analysis methods.

(ii) *Spectral density transformations.* A key original aspect of the model we have developed is the ability to control of the spectral density by modifying parameters such as the number of sinusoids and the distribution of these sinusoids.

(iii) *Harmonicity.* For sounds with low spectral density, users can control the difference between successive frequencies and thus the periodicity of the temporal envelope. This characteristic is perceptually relevant.

(iv) *Color.* As in existing spectral models, the spectral envelope can be modified.

Another application is a musical tool for electroacoustic composers. Composers use the software developed under jMax based on the CNSS model to synthesize original sounds which can be incorporated in musical pieces.

7.3. Future work

The approach we described here is original because we analyze a new parameter, the spectral density, which has been experimentally determined to be perceptually essential for noises. The analysis method is very complex and the approach we present can certainly be improved in the future. But it can already permit the development of psychoacoustic experiments on the perception of the spectral density, which is not completely understood. The results of these experiments and, in particular, the resolution of the human auditory system will give important data to improve the model.

The analysis method proposed here is limited to sounds which do not contain any transient or sinusoid whose amplitude and frequency vary slowly with time. We are developing methods in order to detect fast energy variations (transients) and stable sinusoids. Several methods for transient detection have been proposed (e.g., [39] or [40]). These methods will soon be incorporated in the analysis stage to prevent extracting information related to transients or sinusoids, and which is now assumed as linked to the noise part of the analyzed signal.

Furthermore, we limit the model to the analysis and synthesis of one band of noise. However, a polyphonic signal can be composed of several bands. Each band can be analyzed, transformed, and synthesized independently. One of the improvements of the analysis method is to be able to discriminate noisy bands which are independent: their perceptual properties (spectral density, harmonicity, etc.) may be totally different.

7.4. Conclusion

In this paper, we propose the study of the representation of noisy sounds with short-time sinusoids. No complete justification has been proposed for this representation, whereas many models apply it implicitly. This study leads to the CNSS model, a spectral and statistical model for the analysis, the musical transformation, and the synthesis of noisy sounds. It is appropriate for representing the noisy part of natural sounds, and it allows new high-fidelity transformations (e.g., modifications of the spectral density). The quality of the classical transformations are also at least as good as transformations performed with the existing models (e.g., the time-scale operations [38]). For now, the CNSS model assumes that the modeled sound does not contain any stable sinusoid and any transient. This may lead to audible artifacts in the case of transformations of complex sounds, at the contrary of models such as [12], for example. The CNSS model we have developed is still in experimentation: the values of the parameters of the model have to be refined using psychoacoustic tests. But the model already shows considerable promise for musical creation, psychoacoustic experimentation, and pedagogy. Several sound examples can be found at http://www.labri.fr/Perso/hanna/sounds.html.

ACKNOWLEDGMENTS

This research was carried out in the context of the SCRIME[1] project which is funded by the DMDTS of the French Culture Ministry, the Aquitaine Regional Council, the General Council of the Gironde Department, and IDDAC of the Gironde Department. SCRIME project is the result of a cooperation convention between the Conservatoire National de Région of Bordeaux, ENSEIRB (School of Electronic and Computer Science 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, diffusion, and pedagogy.

REFERENCES

[1] P. Schaeffer, *Traité des objets musicaux*, Seuil, 1966.

[2] B. Edler and H. Purnhagen, "Parametric audio coding," in *Proc. 5th International Conference on Signal Processing (WCCC-ICSP '00)*, vol. 1, pp. 21–24, Beijing, China, August 2000.

[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] M. Goodwin, "Residual modeling in music analysis-synthesis," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '96)*, vol. 2, pp. 1005–1008, Atlanta, Ga, USA, May 1996.

[5] F. R. Moore, *Elements for Computer Music*, Prentice Hall, Englewood Cliffs, NJ, USA, 1990.

[6] X. Serra, "Musical sound modeling with sinusoids plus noise," in *Musical Signal Processing*, pp. 91–122, Roads Swets and Zeitlinger, Lisse, The Netherlands, 1997.

[7] S. Marchand, *Sound models for computer music: analysis, transformation, synthesis of musical sound*, Ph.D. thesis, LaBRI, Université Bordeaux I, Talence, France, 2000.

[8] M. Desainte-Catherine and S. Marchand, "Structured additive synthesis: towards a model of sound timbre and electroacoustic music forms," in *Proc. International Computer Music Conference (ICMC '99)*, pp. 260–263, Beijing, China, October 1999.

[9] H. Purnhagen, "Advances in parametric audio coding," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA '99)*, pp. 31–34, New Paltz, NY, USA, October 1999.

[10] E. B. George and M. J. T. Smith, "Speech analysis/synthesis and modification using an analysis-by-synthesis/overlap-add sinusoidal model," *IEEE Trans. Speech Audio Processing*, vol. 5, no. 5, pp. 389–406, 1997.

[11] M. W. Macon and M. A. Clements, "Sinusoidal modeling and modification of unvoiced speech," *IEEE Trans. Speech Audio Processing*, vol. 5, no. 6, pp. 557–560, 1997.

[12] K. Fitz and L. Haken, " Bandwidth enhanced sinusoidal modeling in lemur," in *Proc. International Computer Music Conference (ICMC '95)*, pp. 154–157, Banff Centre, Alberta, Canada, 1995.

[13] X. Serra, *A system for sound analysis/transformation/synthesis based on a deterministic plus stochastic decomposition*, Ph.D.

[1] Studio de Création et de Recherche en Informatique et Musique Électroacoustique, http://www.scrime.u-bordeaux.fr.

thesis, CCRMA, Stanford University, Stanford, Calif, USA, 1989.

[14] T. S. Verma and T. H. Y. Meng, "Extending spectral modeling synthesis with transient modeling synthesis," *Computer Music Journal*, vol. 24, no. 2, pp. 47–59, 2000.

[15] S. Levine, *Audio representations for data compression and compressed domain processing*, Ph.D. thesis, CCRMA, Stanford University, Stanford, Calif, USA, 1998.

[16] H. Purnhagen and N. Meine, "HILN-the MPEG-4 parametric audio coding tools," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS '00)*, vol. 3, pp. 201–204, Geneva, Switzerland, May 2000.

[17] A. Gerzso, "Density of spectral components: preliminary experiments," report, Ircam, 1978.

[18] 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.

[19] W. M. Hartmann, *Signals, Sound, and Sensation, Modern Acoustics and Signal Processing*, AIP Press, New York, NY, USA, 1997.

[20] P. Hanna, *Modélisation statistique de sons bruités: étude de la densité spectrale, analyse, transformation musicale et synthèse*, Ph.D. thesis, LaBRI, Université Bordeaux I, Talence, France, 2003, http://www.labri.fr/Perso/∼hanna/these.html.

[21] W. M. Hartmann, "Temporal fluctuations and the discrimination of spectrally dense signals by human listeners," in *Auditory Processing of Complex Sounds*, W. A. Yost and C. S. Watson, Eds., pp. 126–135, Erlbaum, Hillsdale, NJ, USA, 1987.

[22] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*, Springer-Verlag, New York, NY, USA, 1999.

[23] J. Pierce, "Introduction to pitch perception," in *Music, Cognition, and Computerized Sound*, P. R. Cook, Ed., chapter 5, pp. 57–70, MIT Press, Cambridge, Mass, USA, 1999.

[24] S. J. Orfanidis, *Introduction to Signal Processing*, Prentice Hall, Upper Saddle River, NJ, USA, 1996.

[25] P. Hanna and M. Desainte-Catherine, "Analysis method to approximate the spectral density of noises," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA '03)*, pp. 201–204, New Paltz, NY, USA, October 2003, Institute of Electrical and Electronics Engineers (IEEE).

[26] J. Allen, "Short term spectral analysis, synthesis, and modification by discrete Fourier transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 25, no. 3, pp. 235–238, 1977.

[27] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, USA, 1989.

[28] P. Hanna and M. Desainte-Catherine, "Detection of sinusoidal components in sounds using statistical analysis of intensity fluctuations ," in *Proc. International Computer Music Conference (ICMC '02)*, pp. 100–103, Göteborg, Sweden, September 2002.

[29] P. Hanna and M. Desainte-Catherine, "Using statistical analysis of the intensity fluctuations to detect sinusoids in noisy signals," Tech. Rep., LaBRI, University of Bordeaux 1, Talence, France, 2003, http://www.labri.fr/Labri/Publications/Rapports-internes/.

[30] A. Zolnay, R. Schluter, and H. Ney, "Extraction methods of voicing feature for robust speech recognition," in *Proc. European Conference on Speech Communication and Technology (EUROSPEECH '03)*, vol. 1, pp. 497–500, Geneva, Switzerland, September 2003.

[31] R. Strandh and S. Marchand, "Real-time generation of sound from parameters of additive synthesis," in *Proc. Journées d'Informatique Musicale (JIM '99)*, pp. 83–88, Paris, France, May 1999.

[32] R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, no. 4, pp. 744–754, 1986.

[33] R. J. McAulay and T. F. Quatieri, "Computationally efficient sine-wave synthesis and its application to sinusoidal transform coding," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '88)*, vol. 1, pp. 370–373, New York, NY, USA, April 1988.

[34] M. Tabei and M. Ueda, "FFT multi-frequency synthesizer," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '88)*, vol. 3, pp. 1431–1434, New York, NY, USA, April 1988.

[35] X. Rodet and P. Depalle, "A new additive synthesis method using inverse Fourier transform and spectral envelopes," in *Proc. International Computer Music Conference (ICMC '92)*, pp. 410–411, San Jose, Calif, USA, October 1992.

[36] P. Hanna and M. Desainte-Catherine, "Adapting the overlap-add method to the synthesis of noise," in *Proc. International Conference on Digital Audio Effects (DAFx '02)*, pp. 101–104, University of the Federal Armed Forces, Hamburg, Germany, September 2002.

[37] M. Desainte-Catherine, G. Kurtag, S. Marchand, C. Semal, and P. Hanna, "Playing with sounds as playing video games," *Computers in Entertainment*, vol. 2, no. 2, pp. 16–38, 2004.

[38] P. Hanna and M. Desainte-Catherine, "Time scale modification of noises using a spectral and statistical model," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '03)*, vol. 6, pp. 181–184, Hong Kong, China, April 2003, Institute of Electrical and Electronics Engineers (IEEE).

[39] P. Masri and A. Bateman, "Improved modelling of attack transients in music analysis-resynthesis," in *Proc. International Computer Music Conference (ICMC '96)*, pp. 100–103, Hong Kong, China, August 1996.

[40] X. Rodet and F. Jaillet, "Detection and modeling of fast attack transients," in *Proc. International Computer Music Conference (ICMC '01)*, pp. 30–33, Havana, Cuba, September 2001.

Pierre Hanna was born in Mont de Marsan, France, in 1974. He studied computer science at the University Bordeaux 1. He received the Ph.D. degree from the University Bordeaux 1 in 2003. His research interests are in computer music and digital sound theory. He is particularly involved in the areas of sound analysis, transformation, synthesis, and classification. He is a Member of SCRIME (Studio for Creation and Research in Music and Computer Science, www.scrime.u-bordeaux.fr).

Myriam Desainte-Catherine was born in Royan, France, in 1958. She studied computer science at the University Bordeaux 1, especially combinatorics with G.X. Viennot. She learned the practice of piano, singing, and harmony. She obtained her Ph.D. thesis and her first Associate Professor position in 1983 at ENSEIRB (www.enseirb.fr). Now, she has a Professor position and she is the Head of the Computer Science Department, ENSEIRB. She is also the Head of the research team in computer music at the LaBRI (computer science laboratory of Bordeaux). Her research interests are in sound and music modeling and their application to music creation. In 1997, she created the SCRIME (Studio for Creation and Research in Music and Computer Science, www.scrime.u-bordeaux.fr) with the composer Christian Eloy. The SCRIME is a place located at the University Bordeaux 1, for composers and computer scientists to collaborate.