A New Analysis Method for Sinusoids+Noise Spectral Models

Guillaume Meurisse, Pierre Hanna, Sylvain Marchand

A NEW ANALYSIS METHOD FOR SINUSOIDS+NOISE SPECTRAL MODELS

Guillaume Meurisse, Pierre Hanna, Sylvain Marchand

SCRIME – LaBRI, University of Bordeaux 1
351 cours de la Libération, F-33405 Talence cedex, France
firstname.name@labri.fr

ABSTRACT

Existing deterministic+stochastic spectral models assume that the sounds are with low noise levels. The stochastic part of the sound is generally estimated by subtraction of the deterministic part: It is assumed to be the residual. Inevitable errors in the estimation of the parameters of the deterministic part result in errors – often worse – in the estimation of the stochastic part. We propose a new method that avoids these errors. Our method analyzes the stochastic part without any prior knowledge of the deterministic part. It relies on the study of the distribution of the amplitude values in successive short-time spectra. Computations of the statistical moments or the maximum likelihood lead to an estimation of the noise power density. Experimentations on synthetic or natural sounds show that this method is promising.

1. INTRODUCTION

Many representations of musical sounds are based on spectral models and consider audio signals as sums of sinusoids whose amplitudes and frequencies evolve slowly with time [1]. Sinusoids+noise models [2] decompose natural sounds into two independent parts: the deterministic part and the stochastic part. The deterministic part is a sum of sinusoids evolving slowly, whereas the stochastic part corresponds to the noisy part of the original sound. This decomposition is usually required for performing several high-quality transformations such as time stretching or pitch shifting, because it allows two different treatments for the two parts. These hybrid models considerably improve the quality of the synthesized sounds.

Spectral models usually consider the stochastic part of the signal as residual or artifacts due to the analysis errors. Most of these techniques try to eliminate this stochastic part. In this paper, we are interested in noisy sounds: The stochastic part is considered as very important from a perceptual point of view. This assumption imposes new techniques and new approaches. Our method analyzes the stochastic part without any prior knowledge of the deterministic part.

After reviewing the representations of noise in existing spectral models in Section 2 and their limitations in Section 3, we present the theory about distribution functions in Section 4. The method proposed is then detailed in Section 5. Finally, the results of experimentations are given in Section 6.

2. NOISE IN SPECTRAL MODELING

Existing hybrid spectral models are specially dedicated to natural sounds with low noise levels. The stochastic part is composed of all the signal components that have not been considered as sinusoids whose amplitudes and frequencies evolve slowly with time. It is assumed to be entirely defined by the time variations of the short-time spectral envelopes. Therefore, usual methods for the estimation of the noisy part are dependent on the analysis of the deterministic part. They require a high-precision analysis (in frequency, amplitude, and phase) of sinusoidal peaks. Detected sinusoids are then subtracted from the original sound in order to analyze the stochastic part. Limitations of these approaches appear if the frequencies and the amplitudes of the sinusoids are not precisely estimated: The errors of these estimations are added to the residual, and this part is thus badly estimated.

Recent works have shown the limitations of sinusoidal analysis methods [3]. The presence of high-level noise considerably degrades the quality of the results of the analysis methods. Furthermore, theory indicates that the precision of the frequency estimation is limited according to the Cramér-Rao lower bound [4, 5] which gives the limit of the variance on an estimator computing data that are corrupted by noise. As several real-world sounds (musical instruments, natural sounds, etc.) contain high noise levels, the analysis step cannot be precise enough. Errors cannot be avoided. Moreover, these errors result in an imprecise estimation of the stochastic part of the signal. For example, an error for the estimation of the frequency, amplitude, or phase of a sinusoid may imply the presence of this sinusoid in the residual. Furthermore, even if the sinusoid is correctly analyzed, residual analysis methods relying on a spectral subtraction [2] define the residual magnitude spectra as composed of several *holes*, at the frequency of the subtracted sinusoids. All these reasons explain why we think that analyzing the stochastic part of sounds after having estimated the deterministic part is not the most accurate technique. In our application, the stochastic part is the most important part of the analyzed sound. This part has not to be considered as a residual. We think that a new technique considering first the analysis of this stochastic part may certainly give more accurate results.

3. ANALYSIS OF THE STOCHASTIC PART

Several approaches for the extraction of the stochastic component have been proposed. These techniques rely mainly on the classification of spectral components (or peaks) into sinusoidal components or stochastic components induced by noise [6]. This decision is binary which implies that a component is always associated to a sinusoid or to noise. But a component cannot be assumed as a mix of sinusoid and noise. Moreover, the decision is made according to the values of audio descriptors (correlation with the window spectrum, duration, energy location, etc.) computed in the current analysis frame [6].

Such methods have limitations if the analyzed sound is with high noise levels: The Cramér-Rao bound theoretically indicates that errors cannot be avoided. Moreover we think that considering

only one short-time amplitude spectrum cannot be sufficient for a precise estimation of the level of the noise.

Another approach consists in considering a long-time analysis of the amplitude spectrum. Several short-time spectra are computed from several consecutive frames. The estimation thus relies on the study of the variations of the short-time amplitude spectra. The observation of successive short-time amplitude spectra shows significant differences between noise and sinusoidal spectral contributions. Amplitude spectra appear to be nearly constant in the case of sinusoidal sounds provided that frequencies and amplitudes do not highly vary over time (tremolo, vibrato). At the opposite, amplitude spectra of noisy sounds vary very rapidly with time.

Empirical methods based on these realizations have already been proposed [7] with some success. The first possibility is to consider for each Discrete Fourier Transform (DFT) bin the minimum of the amplitude spectra. In the case of noisy bins, this minimum may take values near zero whereas in the case of sinusoidal bins, this minimum approximates the amplitude of the sinusoid (slightly lower if noise exists). Thus, the noise level for each bin cannot be estimated.

Another similar idea is to consider the maximum of the amplitude spectra. Here again, this maximum approximates the amplitude of the sinusoid (slightly higher if noise exists). But if the energy of the analyzed bin is due to the presence of noise, the maximum of the amplitude spectra may have a very high value. Indeed, whatever the noise level is, there is a non-null probability that the amplitude of this bin is very high.

The last empirical method is to consider the average of the amplitude spectra. This method leads to errors in the case of noisy bins. We detail the explanations in the next section and we show how this method can be improved.

4. DISTRIBUTION OF THE AMPLITUDE SPECTRUM

The method introduced in this paper relies on a study of variations in the magnitude spectrum along the time axis. High variations seem to indicate the presence of noise whereas stationarity seems to characterize sinusoidal components. We propose here to revert to statistical considerations. The way these variations occur leads to a new analysis method for the stochastic part. We present in this section the theoretical distribution of the amplitude spectrum of noises.

4.1. Spectral Properties of Noises

Thermal noises can be described in terms of a Fourier series [8]:

$$x(t) = \sum_{n=1}^{N} [A_n \cos(\omega_n t) + B_n \sin(\omega_n t)] \tag{1}$$

where N is the number of frequencies, n is an index, and ω_n are equally-spaced frequency components. The random variables A_n and B_n are normally distributed with zero mean and variance σ^2. The magnitude spectrum computed by the Fourier transform is defined by random variables C_n:

$$C_n = \sqrt{A_n^2 + B_n^2} \tag{2}$$

The amplitudes C_n are distributed according to a Rayleigh distribution with most probable value σ.

4.2. Rayleigh Distribution

Let us consider a complex random variable whose real and imaginary parts, denoted X_r and X_i, follow a Gaussian probability distribution (PD) with a standard deviation σ. The probability of the magnitude $M = \sqrt{X_r^2 + X_i^2}$ is given by the Rayleigh PD defined by:

$$p(M) = \frac{M}{\sigma^2} e^{\frac{-M^2}{2\sigma^2}} \tag{3}$$

where σ is the most probable value.

As explained in Section 4.1, the amplitude spectrum of any colored noise is defined by this probability density function: For each DFT bin, each amplitude value is a random variable that is distributed according to this Rayleigh distribution. Here, it is important to note that the probability that the amplitude of a bin reaches a very high or a very low value is not null.

4.3. Rice Distribution

If we add a complex Gaussian noise X with standard deviation σ to a complex value $A_r + jA_i$ of module A, the probability distribution of the magnitude $M = \sqrt{(X_r + A_r)^2 + (X_i + A_i)^2}$ is a random variable distributed according to the Rice distribution [9]. This distribution is defined by:

$$p_{A,\sigma}(M) = \frac{M}{\sigma^2} e^{\frac{-\left(M^2 + A^2\right)}{2\sigma^2}} I_0\left(\frac{AM}{\sigma^2}\right) \tag{4}$$

where I_0 is the modified Bessel function of the first kind of order 0.

Figure 1 shows the Rice PD for $\sigma = 1$ and various A. The A value represents a fixed amplitude value due to the presence of a sinusoid. That is the reason why if A is zero, the Rice distribution turns into the Rayleigh distribution. At the opposite, if A is much greater than σ, the amplitudes are distributed according to a normal (Gaussian) distribution with standard deviation σ and mean A.

Whatever the noise level is, each bin amplitude is theoretically distributed according to the Rice law. Properties of this distribution can be exploited to extract information about the noise part of any signal.

5. NOISE ESTIMATION FROM AMPLITUDE DISTRIBUTION

We consider long-time stationary sounds: Noise power density and the frequencies and amplitudes of sinusoidal components are assumed to be constant in several consecutive frames.

The complex value observed at a bin of the spectrum of such sounds can be decomposed into a constant-magnitude component and a complex Gaussian noise. For each bin, the complex spectrum can be characterized with two parameters A and σ that respectively represent the magnitude induced by one or more sinusoidal peaks or side lobes, and the standard deviation induced by noise at this bin. Each bin is the realization of the Rice PD with parameters A and σ associated to this bin. When observing the magnitude on the same bin at different frames, a Rice-distributed set is obtained. This magnitude distribution may be analyzed in turn to determine the parameters A and σ associated to the studied bin.

The standard deviation σ indicates the energy of noise, while A indicates the amplitude of a sinusoid. So the noise power density of a sound can be obtained by the estimation of the standard

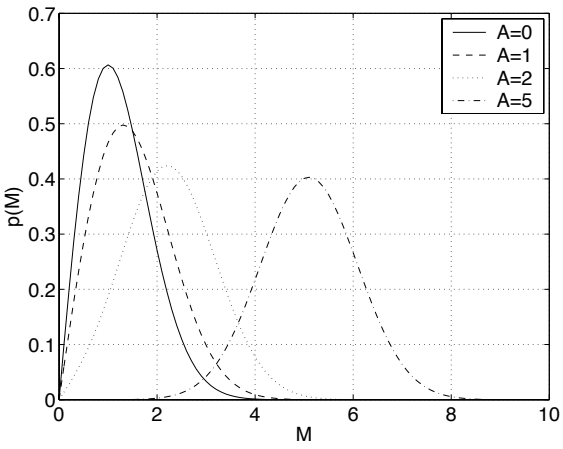

Figure 1: *The Rice probability distribution function: If the parameter A of the function is zero, the Rice distribution turns into the Rayleigh distribution, whereas if A is much greater than σ, the Rice distribution turns into the normal (Gaussian) distribution.*

deviation σ of noise at each bin. Since this distribution follows a Rice PD, two methods are proposed: It is possible to apply the relations on the moments of the Rice PD or to apply the likelihood method to estimate the standard deviation of the noise for each bin and thus obtain the noise power density.

Let us consider L non-overlapping consecutive frames. The discrete Fourier transform is computed for each frame. For each frequency bin the magnitude is computed for each frame in order to obtain a data set of L realizations per bin. The following methods are estimators for σ from a single set of realizations. The variance σ^2 for the whole bins leads to an estimation of the noise power density.

5.1. Moments Method

The first moment can be expressed in terms of the modified Bessel function:

$$E[M] = \left[\left(1+\frac{A^2}{2\sigma^2}\right) I_{e_0}\left(\frac{A^2}{4\sigma^2}\right) + \left(\frac{A^2}{2\sigma^2}\right) I_{e_1}\left(\frac{A^2}{4\sigma^2}\right)\right] \times \sigma\sqrt{\frac{\pi}{2}} \tag{5}$$

where I_{e_0} and I_{e_1} are the scaled modified Bessel functions defined by:

$$I_{e_0}(x) = e^{-x} I_0(x) \tag{6}$$
$$I_{e_1}(x) = e^{-x} I_1(x) \tag{7}$$

I_0 and I_1 being the modified Bessel functions of the first kind of orders 0 and 1, respectively.

The second moment of the Rice PD can be expressed as polynomials in A and σ:

$$E\left[M^2\right] = A^2 + \sigma^2 \tag{8}$$

The standard deviation σ is evaluated by using any pair of moments and by finding the correct value that matches these moments.

The normalized mean μ is the mean computed on the data set normalized by the square root of the second moment. The normalized mean can be expressed only in terms of the signal-to-noise ratio (SNR). In this paper, the SNR is denoted γ and is defined by:

$$\gamma = A^2/\sigma^2 \tag{9}$$

The normalized mean can be expressed as a function of γ [10] as:

$$\begin{aligned}\mu &= \frac{E[M]}{\sqrt{E[M^2]}} \\ &= \frac{\sqrt{\pi}}{2\sqrt{1+\gamma}}\left((1+\gamma) I_{e_0}(\gamma/2) + \gamma I_{e_1}(\gamma/2)\right)\end{aligned} \tag{10}$$

In order to obtain an estimation of γ, it is possible to calculate the first and second moments and to find the value of γ that makes the calculated normalized mean match the theoretical normalized mean.

Expressing the first moment as a function of γ and σ leads to:

$$\hat{\sigma} = \sqrt{\frac{2}{\pi}} E[M] \Big/ \left[(1+\gamma) I_{e_0}\left(\frac{\gamma}{2}\right) + \gamma I_{e_1}\left(\frac{\gamma}{2}\right)\right] \tag{11}$$

5.2. Maximum Likelihood Method

For a probability distribution $p_{A,\sigma}(x)$ and a set M of L realizations following the probability density function denoted p, the likelihood function is given by:

$$\mathcal{L} = \prod_{i=1}^{L} p_{A,\sigma}(M_i) \tag{12}$$

where M_i is the i-th element of M.

Since M is a set of numerical values, L depends only of p. If p is the Rice PD, L can be considered as a function of A and σ. The log-likelihood function is given by:

$$\log(\mathcal{L}) = \sum_{i=1}^{N} \frac{M_i}{\sigma^2} I_0\left(\frac{AM_i}{2\sigma^2}\right) - \frac{NA^2}{2\sigma^2} - \sum_{i=1}^{N} \frac{M_i^2}{2\sigma^2} \tag{13}$$

The amplitude and the standard deviation can be computed by maximizing the log-likelihood function [11]:

$$\{\hat{A}, \hat{\sigma}\}_{\mathrm{ML}} = \underset{A,\sigma}{\operatorname{argmax}} \quad \log(\mathcal{L}) \tag{14}$$

The maximization of the log-likelihood function on each data set gives us the parameters A and σ of the associated bin.

However, the maximization of 2-dimensional functions can be time consuming. So the maximization problem is here reduced to a 1-dimension problem by normalizing the data set M by the square root of the second moment (the second moment of the data set is an unbiased estimator of the Rice second moment). The log-likelihood function is then given by:

$$\begin{aligned}\log(\mathcal{L}) =& N\log\left(2\left(1+\gamma\right)\right) - N\gamma - (1+\gamma)\sum_{i=1}^{N} y_i \\ &+ \sum_{i=1}^{N} \log\left(I_0\left(2y_i\sqrt{\gamma\left(1+\gamma\right)}\right)\right)\end{aligned} \tag{15}$$

where

$$y_i = \frac{M_i}{\sqrt{(E[M^2])}} \tag{16}$$

The maximization of this second log-likelihood function gives us the approximate location of the parameter γ:

$$\hat{\gamma} = \underset{\gamma}{\text{argmax}} \ \log(\mathcal{L}) \tag{17}$$

A solution may be derived by solving:

$$\frac{\partial}{\partial \gamma} \log(\mathcal{L}) = 0 \tag{18}$$

5.3. Algorithms

Analysis of the theoretical distribution of the amplitude for each bin leads to the proposal of two algorithms for the estimation of the noise power density. The first method is based on the computation of two moments whereas the second one is based on the computation of the maximum likelihood.

The general algorithms for the estimation of the noise power density from L consecutive frames (of 1024 samples for example) are described here:

5.3.1. Moments Method

- L DFT with size $2N$ are computed;
- N distributions of L realizations are computed;
- For each distribution M_k ($k = 1, \cdots, N$):
  - $\mu_k = \frac{E[M_k]}{\sqrt{E[M_k^2]}}$ is computed;
  - Equation 10 is applied to compute γ_k;
  - Equation 11 is applied to compute σ_k.

5.3.2. Maximum Likelihood Method

- L DFT with size $2N$ are computed;
- N distributions of L realizations are computed;
- For each distribution M_k ($k = 1, \cdots, N$):
  - Equation 17 is applied on M_k to compute the approximate value for γ_k;
  - Finding the root of Equation 18 is applied to refine solution for γ_k;
  - Equation 11 is applied to compute σ_k.

6. EXPERIMENTAL RESULTS

The two estimation methods described previously have been compared with various SNR and number of realizations. The conclusions of these experimentations indicate that the maximum likelihood and moment-based methods lead to the same precision.

Since the moments method has a better complexity, this method is preferred during our experimentations.

6.1. Number of Samples

Experimentations show that increasing the number of realizations reduces both error and bias. When using more than 1000 realizations, the spectral envelope obtained is smooth. However, using 1000 frames cannot be acceptable. Considering 1000 observations imposes a sound duration of at least 23 seconds when the sampling frequency is 44100 Hz and the DFT size is 1024. The signal is likely to change in a significant way during such a long period. Since the estimator is unbiased for 20 or more realizations, it can be a good choice to compute distributions on 20 frames and then recursively smooth values in time, according to:

$$\sigma(k,i) = \alpha\sigma(k,i-1) + (1-\alpha)\hat{\sigma}(k,i) \tag{19}$$

It should be also possible to use a temporal recursive computation of the moments instead of smoothing the estimated $\hat{\sigma}$. There are several advantages to do so. Complexity of the computation of the moments is simplified and there is no need to save all L magnitude spectra.

Figure 2 shows the estimated value for σ on a Rice-distributed set with $\gamma = 2$ and $\sigma = 1$ according to the number of realizations.

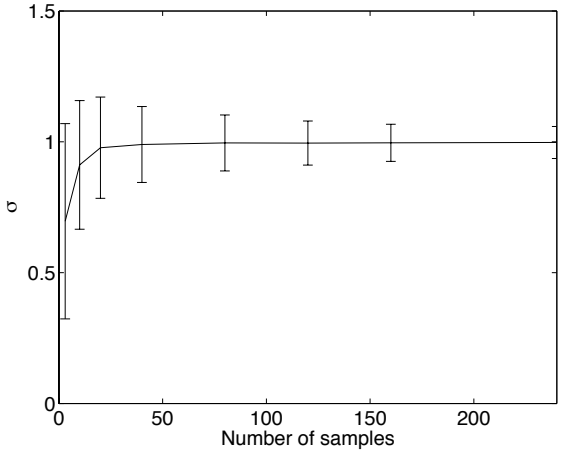

Figure 2: *Estimated σ of a Rice-distributed set (theoretical values $\sigma = 1$ and $\gamma = 2$) with the moments method. Vertical bars indicate the standard deviation of the estimation.*

6.2. Signal-to-Noise Ratio (SNR)

Experimentations show that the estimation is biased at low SNR. Figure 3 shows the estimated σ as a function of γ with different numbers of realizations. The distributions are computed with $\sigma = 1$, and A varies according to γ. For 20 samples, the estimator is biased for γ lower than 1. When $\gamma = 0$, the estimation is biased by 20%. For 1000 samples, the estimation is biased for γ lower than 0.5 and the bias is very low (5%). Therefore increasing the number of realizations reduces the bias. For low SNR, the Rice PD slightly changes. More observations are needed to fit closely the Rice PD. So, if the number of frame is not sufficient, errors in the estimation of γ are likely to occur. If the SNR is low, more observation are needed to avoid bias. If the SNR on a bin is *a priori* known to be $-\infty$ ($\gamma = 0$), the distribution follows the Rayleigh law and the computation of the first moment gives an unbiased estimator.

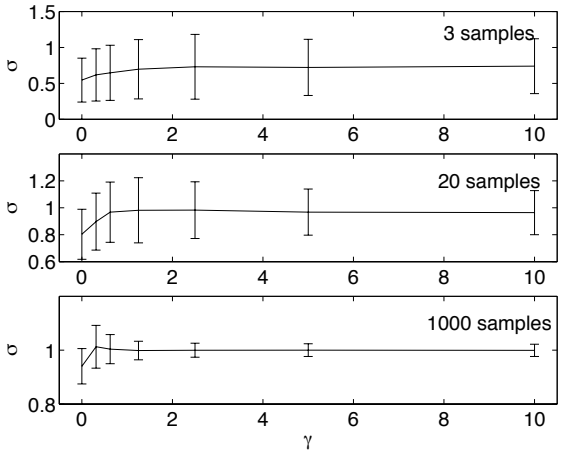

Figure 3: *Estimated σ on Rice-distributed sets (theoretical values $\sigma = 1$, γ varying from 0 to 10). Vertical bars indicate the standard deviation of the estimation.*

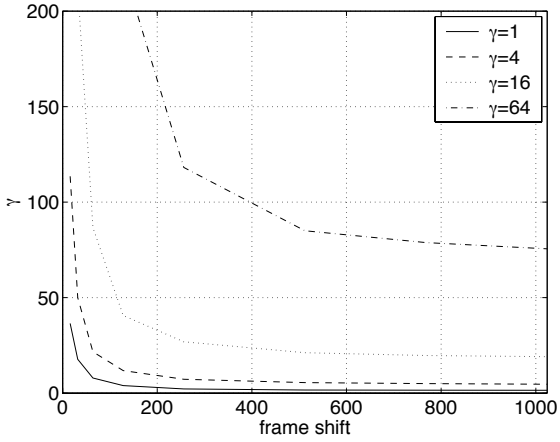

Figure 4: *Estimated γ for several Rice-distributed DFT bins with various SNR according to the overlap rate. Distributions are computed with 20 frames. For a frame shift of 1024 samples, there is no overlap. Overlap increases when the frame shift decreases.*

SNR(γ)	1	4	16	64
MSE with overlap	0.865	3.99	38.1	597
MSE without overlap	1.30	5.76	56.5	841

Table 1: *Compared MSE for the estimation of σ with 20 non overlapping frames or 39 overlapping frames, for the same duration and for various γ.*

6.3. Effect of the Overlap

The moments and likelihood methods assume that the L realizations are statistically independent. Overlapping frames breaks this assumption and induces correlation in the data set. The following tests evaluate the effects of the induced correlation in the estimation of σ. Tests are made with different values of γ and different frame shifts.

The data sets are computed using a sound sampled at 44100 Hz, composed of a white noise of standard deviation $\sqrt{1024/2}$ and a sine wave of frequency 11025 Hz whose amplitude is $\sqrt{2\gamma}$. FFT are computed on 1024 samples. Data sets are computed on bin 256. In this way, the data set values obtained when the frames are not overlapping are realizations of a Rice distribution with standard deviation 1 and amplitude $f(\gamma)$. Several data sets are computed, with different frame shifts and γ (see Figure 4).

It has been observed that the first moment computed on data sets is constant when the frame shift varies. Since the first moment of the data set remains unbiased when overlapping frames, γ and σ are affected in the same way. So the effects of the overlap are only studied on the estimation of γ.

It has been observed that overlapping changed the magnitude distribution and biased noise estimation. Overlapping frames adds correlation in the computed magnitude data sets. So A may be overestimated while σ is underestimated. However, overlapping frames by 50% seems to have a small impact on the results. Bias and mean square error on the estimation of γ have been calculated using L non-overlapping frames on the one hand, and $2L-1$ overlapping frames (50%) on the same duration on the other hand. It appears that overlapping frames reduces bias and mean square error for the estimation of γ. Computing σ on overlapping frames improves precision in time. It is even recommended to use a 50% overlap since it reduces both the bias and mean square error (see Figure 1).

It appears from our tests that using 21 overlapped frames gives the best results. Using less frames strongly degrades performances whereas increasing the number of frames improves slightly the results. Due to the under-estimation of the method at low SNR, it may be interesting to increase the number of frames if a bias at low SNR is not acceptable.

6.4. Sound Tests

Sounds have been analyzed using the moments method. Data sets are computed on 21 overlapping frames of 1024 samples. Estimated σ values are smoothed in time using Equation 19 and $\alpha = 0.9$.

6.4.1. Synthetic Sound

Figure 5 shows the spectrogram of a synthetic sound and its stochastic component. This sound is composed of a pink noise and several sinusoids with various magnitudes. Due to the under-estimation of σ at low SNR, horizontal lines appear on the spectrogram. These lines are located on the frequency bins inhabited by the sines. However this error is hardly audible.

6.4.2. Natural Sound

The moments method has been tested on a sound composed of a saxophone sound and wind noise. Due to the length of the analysis frame, variations in the color of the noise are stretched in time while the attack and the release from the saxophone disturb the magnitude distribution. When the sound is nearly stationary during the analysis frame, sinusoids are correctly removed. Due to the under-estimation at low SNR and the small amplitude modulation of the harmonics of the saxophone sound, some estimation errors appear for the frequency bins inhabited by the sinusoids. This error is not disturbing, when the amplitude modulations are limited.

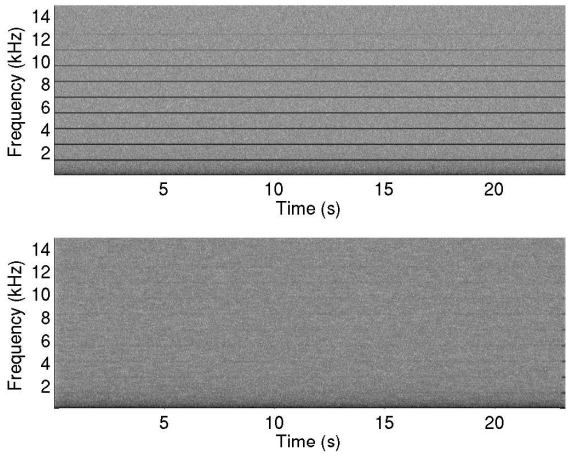

Figure 5: *Spectrogram of a synthetic sound (*23 *s) composed of* 9 *stationary sinusoids (fundamental* 1378 *Hz) with a colored noise (top) and its analyzed stochastic component (bottom).*

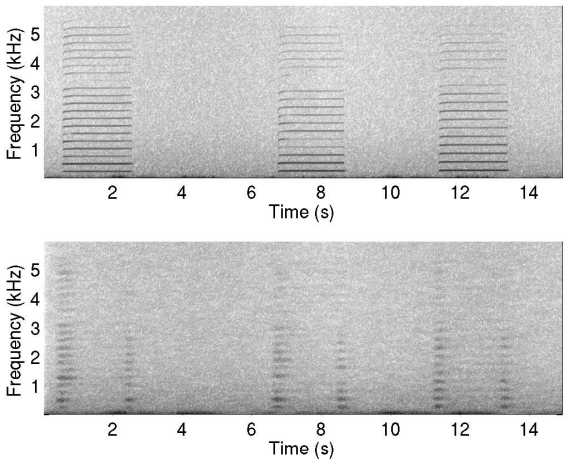

Figure 6: *Spectrograms of a natural sound (*15 *s) composed of* 3 *notes (A#4, C#4, and D4) of saxophone with a background wind noise (top) and its analyzed stochastic component (bottom).*

7. CONCLUSION

In this paper, we propose a new technique that estimates the stochastic part of the signal without having previously estimated the deterministic part. This method relies on a long-term analysis of the variation of the amplitude spectrum. It avoids errors due to the estimation of the sinusoidal parameters for the noisy bins. The limitations of this method is due to the assumption of stationarity for the analyzed signal. When sounds are nearly stationary, the method shows accurate results. Where classical methods require a short-term stationarity, our method requires that the sound is stationary over several frames. Due to the stochastic nature of noise, a single amplitude spectrum does not contain enough information to retrieve statistical properties of noise. That seems to be in agreement with perception. More time is needed to identify spectral content from noise than sinusoidal sounds. In the same way noise was ignored in early sinusoidal models, in this first approach we have neglected the variation of the sinusoids. Indeed, we are also interested in noise where sines could even be absent.

Applications of the methods proposed in this paper are numerous and concern essentially the improvement of the analysis method for the sinusoid+noise spectral models. Existing methods assume that each bin are either sinusoidal or stochastic, whereas the technique we introduce here estimates the proportion of noise – and thus the proportion of sinusoid – for each bin. In the future, improvements induced by this method on spectral model will be studied. Sound examples are available online[1].

8. REFERENCES

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

[2] X. Serra and J. O. Smith, "Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic plus Stochastic Decomposition," *Computer Music J.*, vol. 14, no. 4, pp. 12–24, 1990.

[3] F. Keiler and S. Marchand, "Survey on extraction of sinusoids in stationary sounds," in *Proc. Int. Conf. on Digital Audio Effects (DAFx-02),* Hamburg, Germany, Sept. 2002, pp. 51–58. [Online]. Available: http://citeseer.csail.mit.edu/keiler02survey.html

[4] D. C. Rife and R. R. Boorstyn, "Single-tone parameter estimation from discrete-time observations," *IEEE Trans. Information Theory*, vol. IT-20, pp. 591–598, 1974.

[5] S. M. Kay, *Fundamentals of Statistical Signal Processing – Estimation Theory*, ser. Signal Processing Series. Prentice Hall, 1993.

[6] A. Röbel, M. Zivanovic, and X. Rodet, "Signal Decomposition by Means of Classification of Spectral Peaks," in *Proc. Int. Comp. Music Conf. (ICMC'04),* Miami, USA, Nov. 2004, pp. 446–449. [Online]. Available: http://mediatheque.ircam.fr/articles/textes/Roebel04a/

[7] M. Okazaki, T. Kunimoto, and T. Kobayashi, "Multi-stage spectral subtraction for enhancement of audio signal," in *Proc. IEEE Int. Conf. Acoust., Speech, and Sig. Proc. (ICASSP'04),* Montreal, Canada, vol. II, 2004, pp. 805–808.

[8] W. M. Hartmann, *Signals, Sound, and Sensation*. Woodbury, NY: Modern Acoustics and Signal Processing, AIP Press, 1997.

[9] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 2nd ed. McGraw-Hill, 1984.

[10] K. K. Talukdar and W. D. Lawing, "Estimation of the Parameters of the Rice Distribution," *J. Audio Eng. Soc.*, vol. 89, pp. 1193–1197, 1991.

[11] J. Sijbers, A. den Dekker, D. V. Dyck, and E. Raman, "Estimation of Signal and Noise from Rician Distributed Data," in *Proc. Int. Conf. Sig. Proc. and Communications*, Feb. 1998, pp. 140–142. [Online]. Available: http://citeseer.ist.psu.edu/article/sijbers98estimation.html

[1] http://www.labri.fr/perso/hanna/Expe/sounds.html