DETECTION OF SINUSOIDAL COMPONENTS IN SOUNDS USING STATISTICAL ANALYSIS OF INTENSITY FLUCTUATIONS

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

Myriam DESAINTE-CATHERINE
SCRIME - LaBRI
Université de Bordeaux 1
F-33405 Talence Cedex, France
myriam@labri.fr

ABSTRACT
This article presents a signal characterization method to detect the sinusoidal components of any signal. This is based on the statistical studies of the fluctuations of the temporal envelope power. Two algorithms are proposed, one more efficient, the other more precise. The main advantages are the use of only one analysis window and especially, very small sensitivity to noise. Some experiments and results on synthetic sounds illustrate the high quality of this analysis with noisy sounds.

1. INTRODUCTION
Many representations of musical sounds, based on spectral models, consider signals as a sum of sinusoids whose amplitude and frequency evolve slowly with time. The existing analysis methods are often developed in order to extract sinusoidal components [1]. However, noisy sounds normally used in electro-acoustic music are not composed of periodic pulses only. When such analysis methods are applied to sounds whose noisy part is very intense, they extract artificial frequency components, which are not representative. Indeed with one spectrum, calculated from a short-time analysis window (a few milliseconds) using the Fourier transform, it is very difficult to differentiate peaks which correspond to real periodicities from the ones related to noise or transients. Moreover when the noise amplitude gets close to the signal amplitude (signal-to-noise ratio low), some basic analysis methods cannot even detect the real sinusoidal components.

Hence we propose in this paper a new approach, mainly based on the study of the temporal representation of the sound and particularly on the fluctuations of the dynamic envelope. This approach leads to a signal characterization method which finds out the parts of spectrum where pure sinusoids are located. We also present some experiments and results with synthetic sounds.

2. EXISTING METHODS
In this section we give details about a few existing methods to detect sinusoidal components in any sound. We also present their main advantages and limits.

2.1. Basic method
The most classical method consist in testing every local maximum from a spectrum \( S(k) \) so that \( S(k-1) < S(k) > S(k+1) \). Obviously, peaks are rapidly chosen, but if sinusoids are mixed with noise whose amplitude is high (SNR < 0), peak amplitudes corresponding to sinusoids will not always be higher than peaks related to noise. Furthermore more peaks than the real number of pure sinusoids may be detected.

2.2. IRCAM measure of sinusoidality
The characterization proposed in [2] calculates the complex correlation between the frequency shifted Fourier transform of the analysis window and each peak of the discrete Fourier transform of the signal. A coefficient gives a value between 0 and 1. The value 0 indicates the presence of noise while the value 1 indicates a sinusoidal component. Theoretically it can be used with any type of sounds. However, it has certain drawbacks which are presented in details in [3]:

Stationarity of the signal: The precision of this measure relies on the assumption that every properties of the analyzed signal (frequency, amplitude) are constant in the analysis window. With real-world sounds, this assumption is rarely true and this leads to detection errors.

Detection of non-sinusoidal components: The measure applied to peaks corresponding to side-lobes of sinusoidal components indicates pure sinusoids.

Time versus frequency resolution: The use of short time Fourier transform lends to the usual problem of temporal versus frequency resolution. To calculate the measure, good frequency precision (large analysis window) is necessary whereas the stationarity assumption requires good temporal precision (small analysis window).

2.3. PSOLA method
Peeters proposed in [3] some improvements of the IRCAM measure.

Normalization of fundamental frequency: Assuming that the frequency modulation of signals are correlated to that of fundamental frequency, signal is re-sampled to minimize the influence of these frequency modulations.

Phase variations measure: This new measure is based on the phase difference at two instants. This method leads to different and complementary results. For example spurious side-lobes peaks are no more detected.

Neighboring peak subtraction: Each peak previously identified as a sinusoidal component is subtracted from the studied spectrum to cancel its influence on other peaks.
These improvements are efficient but related to the assumption that sounds are either harmonic or pseudo-harmonic, and that they have low noise. Moreover they require a previous knowledge about the sinusoidality of the spectrum’s components or about phase variations. This implies working with successive signal windows.

3. ANALYSIS OF THE FLUCTUATIONS OF THE TEMPORAL ENVELOPE

In particular the sounds we propose to characterize are noisy and not necessarily harmonic. Therefore we first define noise and then we use the consequences of this definition on the fluctuations of the temporal envelope.

3.1. Definition of noise

Thermal noises have been described in terms of a Fourier series [4]: \( X(t) = \sum_{n=1}^{N} C_n \sin(\omega_n t + \Phi_n) \) where \( N \) is the number of frequencies, \( n \) is an integer, \( \omega_n \) are the pulsations which are equally spaced, \( C_n \) are random variable distributed according to a Rayleigh distribution and \( \Phi_n \) are uniformly distributed random variables.

3.2. Temporal envelope

3.2.1. Definition

A pure sinusoid, whose amplitude and frequency are fixed, is perceived as a steady sound whereas it may have many cycles per second. This stability is described by the fluctuations of temporal envelope.

The temporal envelope \( E(t) \) can be defined [5] as:

\[
E(t) = \sum_{n=1}^{N} C_n e^{i(\omega_n t + \Phi_n)}
\]

We can easily extract the envelope of any signal from an inverse-Fourier transform by removing the negative frequency components from the spectrum.

The auditory system cannot detect too fast fluctuations of an envelope. Hence, usual models take into account this property: the fluctuations are attenuated by a low-pass filter [4]. However, in our approach, this filter is not really necessary because of the frequency resolution. The frequency bandwidth we want to analyze is very narrow to be as precise as possible. The frequency differences in this bandwidth, which define the envelope power [6], are indeed very small. So filtering high frequencies may not be very useful.

3.2.2. Fluctuations of the envelope power

In this paper we study, in particular, statistical variations of the power of the dynamic envelope. The mean of the variance of this power is described as a function of the number \( N \) of the frequency components in the signal and the bandwidth \( W \) of the signal [4]:

\[
\langle V(E^2) \rangle = f(N, W)
\]

Especially, if \( N \) increases, \( \langle V(E^2) \rangle \) also increases. This variance is the measure we use in this article.

3.2.3. Application

The previous equation shows that the measure we defined depends only on the frequency distribution and mainly on the parameter \( N \) [6].

By considering the studied sound as a mix between some sinusoids and some white noise, and by applying the definition of noise proposed above (part 3.1), our signal can be described as a sum of only sinusoids. But one can show that the amplitude corresponding to the frequency component of noise is lower than that of the sinusoids (even with signal-to-noise ratios negative). So we can approximate, in a narrow band containing one pure sinusoid, \( N \) as 1. This implies that the variance of the power envelope will be smaller.

3.3. Approximation methods

In equation 1, the variance of the envelope power of the dynamic envelope is the mean average. This mean is temporal but also depends on the phase random values (see the definition part 3.1). However, the variance that we can easily measure is nothing but the variance of one instance of the signal. Experiments show that the variance values fluctuate a lot around their mean which gives less precise results. We propose here two improvements. The first method involves averaging the variance by distributing component phases. The second one calculates the largest variance. Each one has advantages and drawbacks but they are complementary.

3.3.1. Phase distribution

Many realizations can be obtained from the original signal by modifying phases. From the definition of noise, phases are supposed to be uniformly distributed. After having calculated the Fourier transform, we can replace the values of the phase spectrum with random values, uniformly distributed between 0 and \( 2\pi \). Then we can synthesize one instance of any signal by inverse Fourier transform, and we can eventually do it again as many times as necessary. For a more efficient algorithm, it is important to note that the spectrum of any signal is the same (for its positive part) as the one of its temporal envelope. Hence only one Fourier transform has to be performed: the temporal envelope can be calculated and the phase spectrum can be modified simultaneously (see the algorithms in the part 3.4).

3.3.2. Setting phase constants

One can show that for equal values of the phase of each frequency component, the largest fluctuations of envelope power are obtained. In the same way as the average variance described above, this maximum variance depends on the number \( N \) of components. Hence, in theory, by choosing every phase value equal to a constant value (for example 0), variations in envelope fluctuations can be measured by modifying \( N \). This is how sinusoids can be detected in a noisy sound.

3.3.3. Frequency filtering

To detect one pure sinusoid in a band of any spectrum, we have to study only a small part of this spectrum. We have to analyze not only a small temporal window but also a small frequency window. We take into account the possibility of filtering a signal from its spectral representation. Once the Fourier transform is performed, we modify the spectrum to extract the narrow band we want to study and then, by inverse Fourier transform, we only calculate the envelope power of this band.
3.4. Algorithm and implementation

In this section, we present the two general algorithms to calculate the fluctuations of the dynamic envelope. Experiments show that some difficulties appear when implementing the two methods. Indeed if the analysis window is too small (for example, 5.8 ms), phase spectrum is not precise and the calculated phase values are not significant enough to synthesize other realizations of the same signal. That’s why it is necessary to differentiate two analysis windows: the first one will be large enough (for example, 23ms) to give a good average variance or to modify phase values, while the second one will be smaller to have a good time precision and to calculate the measure of the variance of the envelope.

3.4.1. Algorithm for average variance

\textbf{Data} : Signal samples

\textbf{begin}

Take a large signal window;

\textbf{for} the \textbf{number of instances chosen do}

Calculate FFT window;

Reset negative parts;

Take frequency window (band-pass filter);

Draw phase values randomly between 0 and 2π;

Modify the spectrum with the phase values;

Calculate IFFT spectrum;

Take new small signal window;

Calculate variance;

\textbf{end}

Average calculated variances for all instances chosen;

\textbf{end}

3.4.2. Constant phase values

The algorithm is same as the previous one except that there is only one instance and the phases are set to a constant value (generally 0).

4. EXPERIMENTS AND RESULTS

We have made several experiments with synthetic sounds (white noise is computed with random samples). We propose here only three examples chosen in the critical cases. Each point of the presented curves is calculated with only 3 frequency values (bins). This is the minimum number to detect a local maximum. Here, all the possible points have been calculated which is not always necessary. For each experiment we compare the two presented methods; plain lines denote random phases and dotted lines denote constant phases. We use a 1024 samples long window at the first step and then only 256 samples are used to compute the variance. For the random-phase method, we calculate 10000 instances which takes a few minutes (20 minutes in our examples). However, the constant-phase method takes only a very few seconds (5 seconds in our examples).

4.1. Harmonic sinusoids (different SNR values)

For the first example (figure 1), we have mixed 7 harmonic sinusoids ($k \times 2930$Hz) with white noise. The signal-to-noise ratios is decreasing (SNR= 0, -3, -6, -9, -12, -18, -21). Other experiments show that one pure sinusoid is perfectly detected for SNR= $-15dB$.

4.2. Sinusoids with amplitude modulation

The second example (figure 2) has been made with 7 harmonic sinusoids ($k \times 2930$Hz) whose amplitude $A_k$ is modulated at 20Hz: $A_k + \frac{3}{4} \sin(2\pi 20t)$. The cases of linearly increasing or decreasing amplitude have also been tested and have given good results.

4.3. Sinusoid with frequency modulation

The third example (figure 3) has been made with 7 harmonic sinusoids ($k \times 2930$Hz) whose frequency $F_k$ is modulated at 20Hz: $F_k + \frac{3}{4} \sin(2\pi 20t)$ (R sample rate and $N = 256$). This sinusoid has been mixed with white noise. The cases of linearly increasing
or decreasing frequency have also been tested and the results are good.

5. DISCUSSION

After having presented some examples of experiments and results, we explain in this part the advantages and drawbacks of the two methods that we propose in this paper.

5.1. Efficiency versus precision

Our technique to measure the sinusoidality can be performed in two different ways: by setting the phases at the same value or by averaging and choosing random phase values with a uniform distribution. It is obvious that the second method costs more time. Indeed a large number of instances of the analyzed signal have to be synthesized in order that the average is significant. But, as an inverse Fourier transform is performed each time to calculate the variance of the envelope power, this operation becomes slower and slower, and the calculation can take up to a few hours for a very large window. By comparison, setting the phases at the same value is very fast.

But it is also important to take into account the precision given by each method. Indeed the constant-phase method appears to be less precise than the average method at times. This can be seen, for example, in figure 2, where spurious peaks are detected.

The efficiency of each method is also determined by the size of the large analysis window. The method becomes slower as this size becomes larger, though experiments show that increasing this window size increases the precision of the results. It depends of the nature of the studied sounds: very noisy sounds require large window size whereas noiseless harmonic sounds require small window size.

5.2. Advantages and drawbacks

The results presented just above show that the proposed methods seem to improve the existing techniques because of special properties:

High frequency precision: Basing our method on the fluctuations of the temporal envelope gives the opportunity to study a frequency window whose width is only three bins. Indeed only three bins are necessary to define significant envelope fluctuations. By comparison, methods based on the form of the amplitude spectrum around the analyzed frequency require more than three bins in most cases.

Only one analysis window: Our detection method needs only one analysis window, as opposed to the usual techniques based on the phase variations. Here, it is not necessary to know the phase spectrum of the future window to establish the presence or the absence of a pure sinusoid.

Accuracy in noisy environment: Different experiments show that the main advantage of our method is that it can detect sinusoids mixed with noise, especially when the noise amplitude is higher than that of the sinusoidal component.

Of course, our methods have a few limits:

Stationarity assumption: The precision is high when analyzing sinusoids whose amplitude is modulated but it decreases when the frequency is modulated.

Efficiency: As already explained, although the random-phase method is very precise, it can be very slow. However if one want a high precision analysis, this technique may be used.

6. CONCLUSION

We have presented a new approach to determine from a spectrum of any type of sounds, even very noisy sounds, which peak corresponds really to a sinusoidal component and which one is related to noise or transients. The main application of these methods is to preanalyze signals and identify part of the spectrums where sinusoids are located. This input can then be effectively used by other techniques like DFT which uses the signal derivate [1] is very precise to extract precise partial parameters (frequency, amplitude).

7. REFERENCES


