Journal of Signal and Information Processing, **2013, 4, 51-56**
http://dx.doi.org/10.4236/jsip.2013.41006 Published Online February 2013 (http://www.scirp.org/journal/jsip)

Short-Term Sinusoidal Modeling of an Oriental Music Signal by Using CQT Transform

Lhoucine Bahatti[1], Mimoun Zazoui[1], Omar Bouattane[2], Ahmed Rebbani[1]

[1]Faculté des Sciences et Techniques, Mohammedia, Morocco; [2]Ecole Normale Supérieure d'Enseignement Technique, Université Hassan II Mohammedia, Mohammedia, Morocco.
Email: lbahatti@gmail.com

Received October 16th, 2012; revised November 28th, 2012; accepted December 10th, 2012

ABSTRACT

In this paper, we propose a method for characterizing a musical signal by extracting a set of harmonic descriptors reflecting the maximum information contained in this signal. We focus our study on a signal of oriental music characterized by its richness in tone that can be extended to 1/4 tone, taking into account the frequency and time characteristics of this type of music. To do so, the original signal is slotted and analyzed on a window of short duration. This signal is viewed as the result of a combined modulation of amplitude and frequency. For this result, we apply short-term the non-stationary sinusoidal modeling technique. In each segment, the signal is represented by a set of sinusoids characterized by their intrinsic parameters: amplitudes, frequencies and phases. The modeling approach adopted is closely related to the slot window; therefore great importance is devoted to the study and the choice of the kind of the window and its width. It must be of variable length in order to get better results in the practical implementation of our method. For this purpose, evaluation tests were carried out by synthesizing the signal from the estimated parameters. Interesting results have been identified concerning the comparison of the synthesized signal with the original signal.

Keywords: Oriental Music Signal; Short Time Fourier Transform; Constant Q Transform; Modulation; Sinusoidal Modeling; Weighting Window; 1/4 Tone

1. Introduction

In the field of the music signal transcription, the extraction of the parameters remains of paramount importance. This transcription requires a model which best reflects the signal to be studied. In this regard, the sinusoidal analysis, based on the decomposition in Fourier series, is used from the outset, for the processing of sounds and their generation.

The sinusoidal model is very suitable for modeling harmonic signals and their superposition where their changes in frequency and amplitude are low.

The sinusoidal modeling is also the best way able to analyze and synthesize audio sounds. It is therefore an invertible representation of the sound, if all parameters are kept. However, due to variations of the characteristics of sound signals according to time, it is irrelevant to consider that the parameters of the model components are constant throughout the duration of the signal. Also to be valid even for highly variable signals, transient signals, and non-stationary signals, the sinusoidal modeling has been perfected, leading to the noisy sinusoids model [1]. The latter model is both simplest and most general to represent a musical signal where the sinusoids are characterized by a set of parameters to be estimated. For this, the method QIFFT (quadratic interpolation FFT) [2], used mainly due to its simplicity and accuracy, should be studied and improved, especially for an oriental music signal whose tonality can be up to 1/4 tone.

In our case, we will exploit the sinusoidal model to extract a set of parameters of a signal from an Arabic lute. The extracted parameters will allow us to synthesize another signal that will be compared to the original one. The first section of this manuscript will address the basics of sinusoidal modeling, short term and long term aspects modeling. The analysis of the short term estimation and the signal parameters modeling will be discussed in third section. Section 4 presents the commented experimental results for a real music signal. The final section gives some concluding, remarks and future perspectives.

2. Sinusoidal Modeling

The sinusoidal modeling is, initially, an application of Fourier's theorem that shows that any periodic signal can be represented by a sum of sinusoids with different frequencies and amplitudes. In the real context, the audio signals (music in particular) are characterized by vibrations. Also this vibrating signals aspect can be effectively modeled by a sum of generalized sinusoids whose amplitudes and phases may change over the time. (Equation (1) or Equation (2))

$$s(t)=\sum_{i=1}^{P}A_i(t)\cdot\exp\left(j\Phi_i(t)\right) \quad (1)$$

$$s(t)=\sum_{i=1}^{P}A_i(t)\cos\left(\Phi_i(t)\right) \quad (2)$$

where:

$s(t)$: the signal to be analyzed,

P : number of partials,

$A_i(t)$: instantaneous amplitude of partial *i*, and

$\Phi_i(t)$: instantaneous phase of partial *i*.

However, in most cases where the signals are highly variable, or transitional, and also in order to take into account the non-deterministic part [1], the model of Equation (1) is insufficient, then the signal is a superposition of a quasi harmonic part followed by a noise, in according to the following equation:

$$s(t)=\sum_{i=1}^{P}A_i(t)\exp\left(j\Phi_i(t)\right)+n(t) \quad (3)$$

where $n(t)$ represents the non-deterministic residual of the signal $s(t)$.

Certainly, the forms assigned to the partial amplitudes $A_i(t)$ and the phases $\Phi_i(t)$ have a very important role regarding the performance of sinusoidal modeling. This model is then characterized by a series of parameters to be estimated, and whose number depends on the expressions of $A_i(t)$ and $\Phi_i(t)$.

2.1. Short-Term Modeling

Short-term modeling is especially designed to obtain a stationary modelIndeed to valid this model, the signal to be analyzed must be slotted into small fragments where the signal parameter variations will be considered small. Then, in each segment, of duration *T*, starting at $t=n\cdot\Delta T$, the signal can be represented by a plurality of sinusoids of the form:

$$S_n(t)=\sum_{i=1}^{N}S_n^i(t) \quad (4)$$

$$S_n^i(t)=a_n^i\cos\left(2\pi f_n^i(t-n\Delta T)+\Phi_n^i\right) \quad \text{for } n\Delta T<t<n\Delta T+T \quad (5)$$

Non stationary extensions of the signal can be envisaged to follow faithfully the signal variations along the viewing window that can last up to 32 ms [3].

2.2. Long-Term Modeling

For quasi-periodic signal sounds, correlations between the parameters of the sinusoids issued from successive frames can be exploited. Then, it is useful, and required, to consider a long-term sinusoidal model where the amplitudes and frequencies of the sinusoids change slowly and continuously over time, in order to keep and insure continuity of phase [1].

$$s(t)=\sum_{i=1}^{P}A_k(t)\cos\left(\Phi_k(t)\right) \quad (6)$$

$$\Phi_k(t)=\Phi_k(0)+2\pi\int_0^t F_k(u)\,\mathrm{d}u \quad (7)$$

The parameters F_k, A_k and Φ_k the frequencies, amplitudes and phases of instantaneous partial P_k respectively, are estimated instantly using the short term model.

3. Short Term Analysis

The short-term sinusoidal analysis consists of two tasks: The first, consists of detecting the presence of a sinusoidal components in the analyzed signal (peaks in the Fourier spectrum). The second task is used to estimate the signal parameters (amplitude, frequency and phase).

This analysis process can be represented by the following algorithm:

```
For $n = 1$ to number of frames do
  Begin
    -Isolate the frame of index $n$
    -Select the spectral peaks corresponding to the
     partial of signal
    -Model the signal concerning each partial
    -Estimate the model parameters
  end
```

Each step of this algorithm is described below.

3.1. Frame Isolation

The frame isolation, known as the windows weighting, is considered to isolate a frame $S_n(t)$ of index *n* and its width *T*. To do so we take:

$$S_n(t)=s(t)\cdot w(t-nT) \quad (8)$$

$w(t-nT)$ is a weighting window such that:

$$w(t-nT)=0 \text{ when } t<\left(n-\frac{1}{2}\right)T \text{ and } t>\left(n+\frac{1}{2}\right)T \quad (9)$$

It is of symmetrical shape so that its phase spectrum is zero. In addition to extracting a signal frame, the window weighting should allow best estimates of the model parameters assigned to that frame. The window depends strongly on the signal to be analyzed according to its temporal features and especially frequency characteristics. It is completely defined by its type (expression) and length (size).

3.2. Window Sizing

A longer window generally increases the bias, while a short window is useless in the steady state of an audio signal. Therefore the same kind of window; with the same length along the entire signal is study case to be avoided. Several solutions are possible, for example:

- A single window kind of variable frequency
- Window of the same type with variable length, or
- Windows belonging to different classes of signals.

In [4], to properly estimate the instantaneous frequency of a signal, the solution used is to select a window w_i from a finite set W: $w_i \in [w_1 \quad w_2 \quad \cdots \quad w_m]$ according to a criterion named Maximum Correlation Criterion (MCC).

In our approach, we choose one kind of window having variable length by using on the CQT Transform (constant Q transform), in which the temporal resolution increases with frequency. So, a large analysis window is used at low frequencies and when frequency increases, the window size will decrease.

The basic tool of the sinusoidal modeling is the short term Fourier transform (STFT) as follow:

$$S_w(t,f) = \int_{-\infty}^{+\infty} s(\tau)\, w(\tau - t)\, \mathrm{e}^{-j2\pi f\tau}\, \mathrm{d}\tau \qquad (10)$$

In the case of initial STFT, w is the window of fixed length: $L = \dfrac{N}{F_e}$, (N: fixed size and F_e sample rate), while in the opposite case, the length of the window CQT becomes variable.

In the case of a time signal $s(n)$ sampled at the frequency Fe, The constant Q transform (CQT) can be directly determined by:

$$S_w(k) = \sum_{m=0}^{N_k - 1} w(m,k)\, s(m)\, \mathrm{e}^{-j2\pi m f_k} \qquad (11)$$

where $S_w(k)$ is the kth CQT component, and the analysis window $w(m,k)$, of the size N_k, depending on frequency ("bin" k). The frequencies corresponding to the CQT bins are geometrically spaced, related to the Oriental musical scale: So if we denote $f_{\min}$ the starting frequency analysis, the other frequencies are derived from the relation: $f_k = A^k \cdot f_{\min}$ With:

A: ratio for the resolution 1/4 tone: $A = \dfrac{37}{36} = 1.027$ [5]

For the CQT form, the ratio $Q = \dfrac{f_k}{\Delta f_k}$ is constant [6], where, $\Delta f_k = f_{k+1} - f_k$.

In the Oriental range we have:

$$Q = \frac{A^k \cdot f_{\min}}{A^{k+1} \cdot f_{\min} - A^k \cdot f_{\min}} = \frac{1}{A-1} = 37.$$

And the size of the analysis window is determined by: $N_k = Q \cdot \dfrac{F_e}{f_k}$.

3.3. Window Kind Determination

To isolate a frame $s_k(t)$ of index k and width T_k, we use the following expression: $s_k(t) = s(t) \cdot w(t - k \cdot T_k)$. So in the frequency domain, we have: $|S_k(f)| = |S(f)| * |W(f)|$. This convolution must cause the minimum possible strain on $|S(f)|$. To do so, the window $w(t)$ must largely decrease too its lobe sides and increase the selectivity of the main lobe.

The analysis window is also conditioned by the adopted model QIIFT [2] (the phase is of a quadratic form) to the frame signal as:

$$s(t) = \sum_{k=1}^{P} A_k(t) \cos(\Phi_k(t)) \qquad (12)$$

where: $\Phi_k(t) = \Phi_k + \omega_k \cdot t + \dfrac{\psi_k}{2} \cdot t^2$. (13)

Arbitrarily, the rectangular window is adopted, but, for more accuracy, the Gaussian window is considered as a reference in the literature, since it allows accurate estimation of model parameters QIFFT [2]. However, the Hann window, in addition to be of C^∞ (infinitely continuous and differentiable), remains a good candidate for estimating sinusoidal parameters [7]. (See **Figure 1** and **Table 1**). Thus it is adopted in our approach, its basic expression is:

$$w_h(t) = \begin{cases} 0.5 + 0.5\cos\left(2\pi \dfrac{t}{T}\right) & \text{Pour} -\dfrac{T}{2} \le t \le \dfrac{T}{2} \\ 0 & \text{elsewhere} \end{cases} \qquad (14)$$

3.4. Selection of Spectral Peaks and Modeling

After isolating the frame $s_n(t)$, we proceed by determining its amplitude spectrum. The selection of the spectral peak, and filtering the frame $s_n(t)$ around this peak, aims to reduce the signal to a single partial (example $k = 1$ for the fundamental signal). For music signal of an Arabic lute, the fundamental frequencies of the different basic notes, subject of study and analysis, in the first octave, are summarized in **Table 2**.

The filter used is a pass band having a variable cutoff frequency to track the fundamental frequency of the detected note.

Under these conditions, each partial can be represented by several models [3]. The chosen model is of the form:

$$s(t) = \exp\left(\underbrace{(\lambda_0 + \mu_0 \cdot t)}_{\lambda(t) = \log(a(t))} + j \underbrace{\left(\Phi_0 + \omega_0 \cdot t + \frac{\psi_0}{2} \cdot t^2 \right)}_{\Phi(t)} \right) \qquad (15)$$

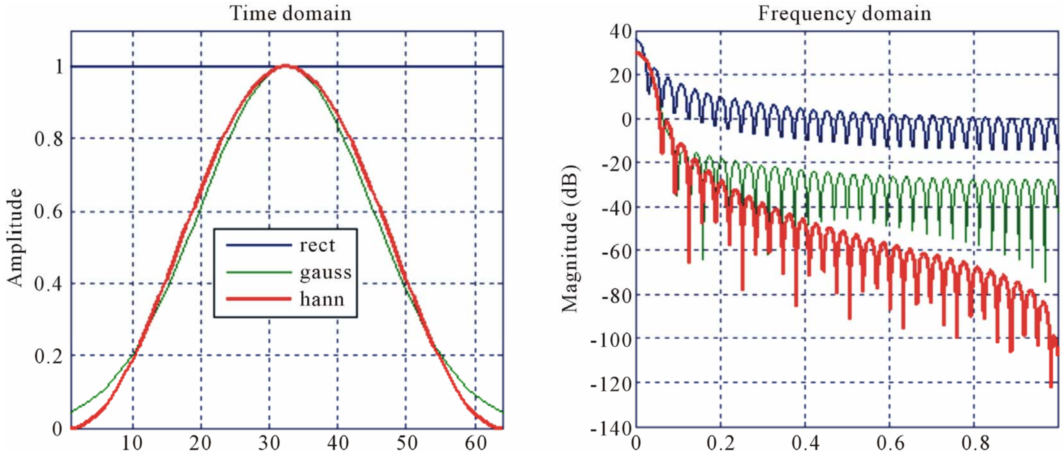


Figure 1. Hann, Gauss, and rectangular windows.

Table 1. Characteristics of the Hann window.

Window	Width of the main lobe	Amplitude of the side lobes	Side lobe attenuation
Hann	4/N	−32 dB	−18 db/Octave

N: number of samples per window.

Table 2. Frequency notes of RAST range, first octave.

Note	C	D	Eƀ	F	G	A	Bƀ	C
Freq (Hz)	65	73	79	87	98	110	120	131
Code	1	3/4	3/4	1	1	3/4	3/4	
Gap (Hz)	8	6	8	11	12	10	11	

This model is more realistic since it assumes that the frequency variations are combined with variations in amplitude and may best reflect the temporal evolution of a musical note. The different parameters to be estimated are:

- μ_0 (amplitude modulation parameter) which is the derivative of $\lambda(t)$ (the log-amplitude).
- ω_0 (pulsation), and ψ_0 (frequency modulation parameter) are the first and second derivatives of the instantaneous phase $\Phi(t)$ respectively.

The amplitude and phase are modeled by polynomials of degrees 1 and 2, respectively [8]. These polynomial models can be considered either as: an expansions of more complicated modulation amplitude and frequency, or as an extension of the stationary case where $\mu_0 = 0$ and $\psi_0 = 0$.

Notice that: $a_0 = \exp(\lambda_0)$ and Φ_0 are the initial amplitude and the initial phase of the signal respectively.

3.5. Estimation of Model Parameters

After the porcessing case phase of Section 3, where the music signal, is reduced to Equation (15), we estimate its parameters using [9] which proposes a generalization of the reassignment method [8] based on a non-stationary model.

$\tilde{\omega}$ frequency and time t are first estimated by the method described in [10]:

$$\tilde{\omega} = \omega - \underbrace{\mathrm{Im}\frac{S_{\omega'}(t,\omega)}{S_{\omega}(t,\omega)}}_{-\Delta\omega} \tag{16}$$

and

$$\tilde{t} = t + \underbrace{\mathrm{Re}\frac{S_{t\omega}(t,\omega)}{S_{\omega}(t,\omega)}}_{-\Delta t} \tag{17}$$

Modulation parameters of the amplitude $\tilde{\mu}$ and frequency $\tilde{\psi}$ are obtained by generalizing the method proposed in [11]

$$\tilde{\mu} = \frac{\partial}{\partial t}\mathrm{Re}\left(\log\left(S_{\omega}(t,\omega)\right)\right) = -\mathrm{Re}\left(\frac{S_{\omega'}(t,\omega)}{S_{\omega}(t,\omega)}\right) \tag{18}$$

$$\tilde{\psi} = \frac{\partial\tilde{\omega}}{\partial\tilde{t}} = \frac{\dfrac{\partial\tilde{\omega}}{\partial t}}{\dfrac{\partial\tilde{t}}{\partial t}} \tag{19}$$

$$\tilde{\psi} = \frac{\mathrm{Im}\left(\dfrac{S_{\omega''}(t,\omega)}{S_{\omega}(t,\omega)}\right) - \mathrm{Im}\left(\left(\dfrac{S_{\omega'}(t,\omega)}{S_{\omega}(t,\omega)}\right)^{2}\right)}{\mathrm{Re}\left(\dfrac{S_{t\omega}(t,\omega)\cdot S_{\omega'}(t,\omega)}{S_{\omega}^{2}(t,\omega)}\right) - \mathrm{Re}\left(\dfrac{S_{t\omega'}(t,\omega)}{S_{\omega}(t,\omega)}\right)} \qquad (20)$$

All these results are given with: $S_{\omega}(t,\omega)$ $S_{\omega'}(t,\omega)$ and $S_{\omega''}(t,\omega)$: the short-term Fourier transform of the signal $s(t)$ using the window $\omega'(t) = \frac{\mathrm{d}\omega(t)}{\mathrm{d}t}$ and $\omega''(t) = \frac{\mathrm{d}^2\omega(t)}{\mathrm{d}t^2}$ respectively. $S_{t\omega}(t,\omega)$ is the short-term Fourier transform of signal $s(t)$ using the window $\omega(t)$ weighted by the time axis: $t\omega(t)$.

Once the parameters ψ and ω are estimated, taking as exemple the pitch signal, the evolution of the fundamental during the time, and over a choseen window, can be extracted by a frequency demodulation processes (Equation (15)). Its expression is:

$$\omega(t) = \omega_0 + \psi_0 t\,. \qquad (21)$$

This frequency demodulation technique is a good alternative to the method presented in [12] which is based on the maximum likelihood and considers the musical signal as a pseudoperiodical sound.

4. Experimental Results and Comments

The first part of the experimental results by applying the modeling technique is to extract a sinusoidal signal $s(t)$ issued from Equation (15) and perturbed by an additive noise, with a lower S/N ration (10 dB) (**Figure 2**). The duration of the observation window is 23 ms, in order to remain in the context of short-term. The estimated parameters correspond to baseline signal except for ψ_0 (FM modulation term), having a negligible influence as the weighting window is short. Overall, the correct extraction of the signal $s(t)$ demonstrates the reliability and robustness of the short term sinusoidal modeling.

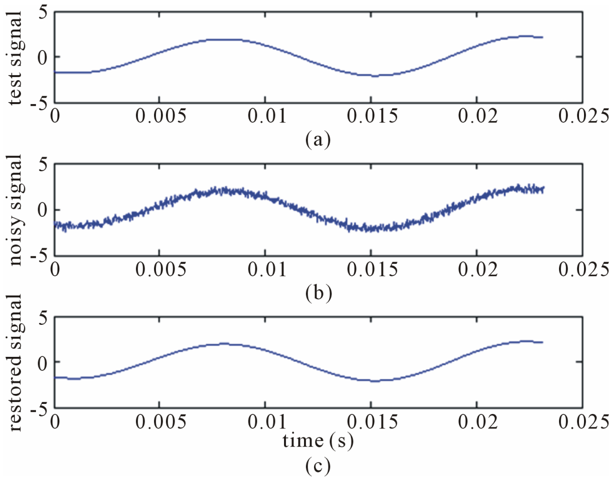


Figure 2. test Sinusoidal modeling short-term: duration of the window: 23 ms. (a) Test signal a_0 = 2 μ_0 = 10, phi = π/2, f_0 = 440 Hz, ψ_0 = 100; (b) Noisy signal with S/N = 10 dB; (c) Restored signal with estimated parameters a_0 = 1.9882; μ_0 = 12.2410, phi = 1.5631, f_0 = 439.2728, ψ_0 = 462.9664.

Figure 3 is the result of the application of the short-term sinusoidal modeling for the extraction of the fundamental frequency (pitch) by frequency demodulation.

The signal under test has a very strong attack. This explains the presence of peaks on the curve pitch for each onset.

Figure 4 illustrates the application of our method to a real signal issued from an arabic lute. In **Figure 4(a)**, notice the residual noise (difference between the original signal and the synthesized one), presents a high level at the note starting time (the transient state) In **Figures 4(b)** and **(c)**. The spectral difference between the two signals can be clearly seen through the two spectrograms where the synthesized signalin **Figure 4(a)** presents a finite and limited number of partials.

5. Conclusion and Perspectives

The most convenient approch to represent a musical signal is clearly the sinusoidal modeling long term. However, its parameters are deduced by using the short term approch. The estimation of model parameters by the short-term reallocation technique leads to the determination of the pitch (to identify the note), and all needed parameters to the analysis and a good synthesis of musical sounds.

The result of the short term sinusoidal modeling is closely related to the kind of the weighting window. In this work the Hann window is the most switable. However, the use of the other types such as sigmoid, that is largely used with proven results in image processing, can be exploited. The sinusoidal modeling method presented in this paper is based on an "open loop" strategy. As a perspective of this work, the obtained results can be enhanced and improved by introducing a cost function to be

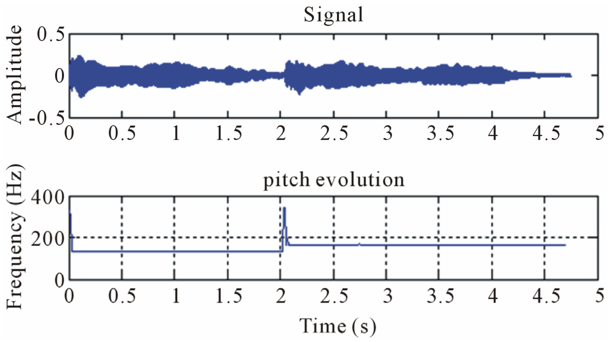
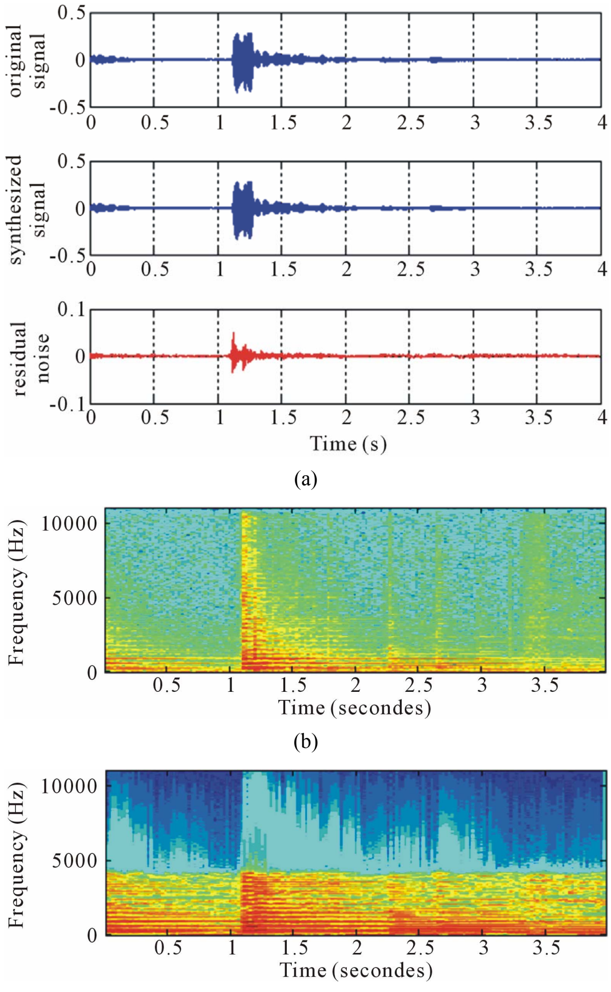


Figure 3. Application to a real luth signal containing two musical notes (evolution of pitch).



(a)

(b)

(c)

Figure 4. Real signal analysis (origin) and signal synthesis. (a) Temporal forms; (b) Spectrogram of the original signal; (c) Spectrogram of the synthesized signal.

minimised. This leads us to considered this improvement as an optimisation problem to be solved. Since, the oriental music is well known by its richness in melody, the proposed perspective task requires more investigation and exploration. This proposal will be discuted largely in the futur work.

REFERENCES

[1] X. Serra, "Musical Sound Modeling with Sinusoids Plus Noise," In: C. Roads, S. Pope, A. Picialli, G. De Poli, Eds., *Musical Signal Processing*, Swets & Zeitlinger Publishers, Lisse, 1997.

[2] M. A. J. O. Smith, "AM/FM Rate Estimation for Time-Varying Sinusoidal Modeling," ICASSP 2005.

[3] M. Betser, "Modélisation Sinusoïdale et Applications à l'Indexation Audio," Thèse Doctorat, Telecom ParisTech, Laboratoire LTCI, 2008

[4] H. K. Kwok and D. L. Jones, "Improved Instantaneous Frequency Estimation Using an Adaptive Short-Time Fourier Transform," *IEEE Transactions on Signal Processing*, Vol. 48, No. 10, 2000, pp. 2964-2972. doi:10.1109/78.869059

[5] B. Marzouki, "Application de l'Arithmétique et Des Groupes Cycliques à la Musique," Département de Mathématiques et Informatique Faculté des Sciences, Oujda, 2010.

[6] J. C. Brown, "Calculation of a Constant Q Spectral Transform," *Journal Acoustical Society of America*, Vol. 89, No. 1, 1991, pp. 425-434. doi:10.1121/1.400476

[7] S. Marchand, "Sound Models for Computer Music (Analysis, Transformation, Synthesis)," PhD Thesis, University of Bordeaux, Talence, 2000.

[8] C. de Villedary, K. Kodera and R. Gendrin, "A New Method for the Numerical Analysis of Time-Varying Signals with Small BT Values," *IEEE Transactions on Acoustics*, *Speech and Signal Processing*, Vol. 26, No. 1, 1978, pp. 64-76. doi:10.1109/TASSP.1978.1163047

[9] S. Marchand and P. Depalle, "Generalization of the Derivative Analysis Method to Non-Stationary Sinusoidal Modeling," Proceedings of the Digital Audio Effects (DAFx) Conference Digital Audio Effects (DAFx) Conference, Espoo Finlande, 2008, pp. 281-288.

[10] F. Auger and P. Flandrin, "Improving the Readability of Time-Frequency and Time-Scale Representations by the Reassignment Method," *IEEE Transactions on Signal Processing*, Vol. 40, No. 5, 1993, pp. 1068-1089.

[11] S. W. Hainsworth, "Techniques for the Automated Analysis of Musical Audio," Technical Report, 2003

[12] B. Doval and X. Rodet, "Estimation of Fundamental Frequency of Musical Sound Signals," *International Conference on Acoustics*, *Speech*, *and Signal Processing*, Toronto, 14-17 April 1991, pp. 3657-3660.