TIME-VARIANT HARMONIC SIGNAL MODELING BY USING POLYNOMIAL APPROXIMATION AND FULLY AUTOMATED SPECTRAL ANALYSIS

Miroslav Zivanovic [1] *and Johan Schoukens*[2]

[1]Dpt. IEE, Universidad Publica de Navarra, Campus Arrosadia, 31006, Pamplona, Spain
phone: + 34 948 16 90 24, fax: + 34 948 16 97 20, email: miro@unavarra.es

[1]Dpt. ELEC, Vrije Universiteit Brussel, Pleinlaan 2, B-1050, Brussels, Belgium
phone: +32 2 629 29 47, fax: +32 2 629 28 50, email: jschouk@vub.ac.be

ABSTRACT

We present a novel approach to modelling time-variant harmonic content in audio signals. We show that both amplitude and fundamental frequency time variations can be compactly captured in a single time polynomial which modulates the fundamental harmonic component. A correct estimation of the fundamental frequency is assured through the fully automated spectral analysis method (ASA).The best-fit is easily obtained by linear least-squares, given the fact that the set of equations is linear-in-parameters. In contrast to the existing methods, the proposed approach is designed to properly describe harmonic structures in audio signals under conditions of both AM and FM modulations and low signal-to noise ratios.

1. INTRODUCTION

Most audio signals are efficiently described by time-varying harmonic structures plus additive noise. The harmonic modelling, usually referred to as the detection and estimation of the harmonic parameters, is a key issue in audio signal processing. It has been widely used in signal synthesis, compression, transformation etc.

Modelling of the harmonic content of audio signals has often been carried out by least-squares (LS) solution of sets of linear equations derived from minimizing the mean square error (MSE) between the original and estimated signal [1, 2]. They first obtain the fundamental frequency estimate by some non-linear search and then compute the parameters estimate by means of linear LS. Those equations are linear-in-parameters and therefore the computational effort is small even for a large number of harmonics. The modeling errors are/or an underlying assumption of quasi-stationarity are, however, the principal drawback of these methods.

Another approach to harmonic modelling is based on the analysis of the signal's STFT or some other conveniently chosen time-frequency representation. Typically the harmonic features are detected in the spectrum on a local level by examining spectral peaks [3, 4]. Next, the harmonic parameters are estimated from the peaks considered sinusoidal [5, 6] and then some harmonic consistency over the time is enforced to obtain time-varying partials [7, 8]. The fact that there is no need for an a priori knowledge about the fundamental frequency makes these methods attractive not only for harmonic, but also for more general sinusoidal modelling. Although the estimation of the harmonic components can be done in a very efficient way, the overall performance still strongly depends on the correct detection of the harmonic components in the time-frequency representation.

A statistical approach [9] deserves to be mentioned too. It implements a Bayesian network with a particular prior structure, built from conditional probabilities which establish the relationship among the harmonic parameters. Apart from the strong computational complexity, a large number of parameters that must be a priori known make the algorithm highly application dependent.

Herein we describe a novel approach to time-variant harmonic modeling, which differs substantially from the aforementioned methods in the following aspect. Under the assumption of the fundamental frequency variation around some mean value F_0 within the analysis window, we model both AM and FM by a single time polynomial. The F_0 estimation is a critical step due to the presence of modulations and noise. It is, however, efficiently performed by the fully automated spectral analysis method (ASA) [10], based on iterative leakage reduction in the discrete spectrum of an audio signal. In addition, the ASA behaves very well in conditions of low signal-to-noise ratios, which improves substantially the harmonic approximation. Once F_0 is estimated, the rest of the harmonic parameters can be easily estimated from a linear least-squares (LS) cost function which is linear-in-parameters.

The present paper is organized as follows: in Section 2 we propose a harmonic model used to describe amplitude and frequency time-variations in audio signals. Section 3 is a summary of the fully automated spectral analysis method used to estimate the mean fundamental frequency F_0. In Section 4 we pose the estimation problem as linear LS. In Section 5 we present a comparative study among different methods together with an illustrative example. The conclusions appear in Section 6.

2. THE HARMONIC MODEL

A general model for single-source *N*-sample discrete audio signals is given by the following expression:

$$s(n) = \sum_{i=0}^{I} A_i(n)\cos[2\pi i F_0(n)n + \varphi_i] + r(n). \tag{1}$$

The deterministic part is given as a superposition of *I* harmonically related components with time-variant amplitudes and frequencies, while the additive disturbance *r*(*n*) is typically a Gaussian sequence or colored noise. The model (1) is, strictly speaking, unidentifiable because we have *N* measurements for estimating *I*(*N* + 1) + *N* parameters. We, however, assume that the harmonic parameters evolve continuously and slowly in the time domain, the fact which is satisfied for most real-world audio signals.

By applying the trigonometric identity regarding sum of angles we can rewrite (1) as follows:

$$s(n) = \sum_{i=0}^{I} a_i(n)\sin(2\pi i F_0(n)n) + b_i(n)\cos(2\pi i F_0(n)n)n, \tag{2}$$

$$a_i(n) = -A_i(n)\sin\varphi_i \ , \quad b_i(n) = A_i(n)\cos\varphi_i . \tag{3}$$

The parameters $a_i(n)$, $b_i(n)$ and $F_0(n)$ are time-variant functions and are often modelled by polynomials or sets of basis functions. As common in harmonic modelling, we will assume that the amplitude and fundamental frequency vary approximately linearly along the analysis window:

$$a_i(n) = a_0^{(i)} + a_1^{(i)}n \ , \quad b_i(n) = b_0^{(i)} + b_1^{(i)}n, \tag{4}$$

$$F_0(n) = F_0 + \beta\, n. \tag{5}$$

By substituting (4) and (5) into (2) and applying again the trigonometric identities regarding sum of angles, we obtain a linear combination of sine/cosine products with the arguments $2\pi F_0 n$ and $2\pi\beta n^2$ and corresponding amplitude weights. This form is not suitable for harmonic parameter estimation, but a simplification can be done based on the following argumentation.

In most real-world audio signals, the harmonic components generated by a combined action of AM and FM often pose serious difficulties to analysis. Among them, the vibrato signals are perhaps the most representative, as they usually contain quasi-sinusoidal AM and FM mutually coupled in an arbitrary way. For a well-performed vibrato, a typical frequency deviation is a quarter tone, i.e. 3% fundamental frequency (pitch) variation or 50 cents [11]. If we model a quarter of the vibrato FM period by the linear trend (5), we may use the following approximations:

$$\sin 2\pi\beta n^2 \approx 2\pi\beta n^2, \qquad \cos 2\pi\beta n^2 \approx 1. \tag{6}$$

By means of the last expression, we can rewrite (2) as:

$$s(n) = \sum_{i=0}^{I} p_s^{(i)}(n)\sin(2\pi F_0 n) + p_c^{(i)}(n)\cos(2\pi F_0 n), \tag{7}$$

$$p_s^{(i)}(n) = a_0^{(i)} + a_1^{(i)}n - 2\pi b_0^{(i)}\beta n^2 - 2\pi b_1^{(i)}\beta n^3, \tag{8.a}$$

$$p_c^{(i)}(n) = b_0^{(i)} + b_1^{(i)}n + 2\pi a_0^{(i)}\beta n^2 + 2\pi a_1^{(i)}\beta n^3. \tag{8.b}$$

Regarding the last expressions, a few important things need to be stressed. In spite of modelling the amplitude and frequency variations as separate polynomials (4), the last expressions describe them in a more compact way through a single polynomial. In addition, the expressions (8) can be given a deeper interpretation through the concept of covariance for signals.

Let us recall that the covariance between two arbitrary time-variant signal parameters *x* = *x*(*t*) and *y* = *y*(*t*) is defined as the following combination of time averages [12]:

$$C_{xy} = \overline{x(t)y(t)} - \overline{x}(t)\overline{y}(t) \ , \tag{9}$$

In (9) all the averages are calculated with respect to signal's normalized energy time density. In a view of (9) and by assigning the time origin to the geometric centre of the analysis window, we can rewrite (8) as:

$$p_S^{(i)}(n) = \overline{a} + \frac{1}{\sigma_T^2}\left(C_{at}n - 2\pi\overline{b}C_{ft}n^2 - 2\pi C_{bf}n^3\right), \tag{10.a}$$

$$p_c^{(i)}(n) = \overline{b} + \frac{1}{\sigma_T^2}\left(C_{bt}n + 2\pi\overline{a}C_{ft}n^2 + 2\pi C_{af}n^3\right). \tag{10.b}$$

We observe that each polynomial captures the harmonic parameters' dynamics by modelling its coefficients through the covariance terms describing the amplitude time-variations (C_{at}, C_{bt}), frequency time-variations (C_{ft}) and mutual amplitude-frequency variations (C_{af}, C_{bf}). The parameter σ_T^2 is the duration of the analysis window. These polynomials, hence, provide a rough estimate of the variation trends in the signal. This property can be very useful in applications where only an overall measure of variation is needed without an a priori knowledge about the harmonic parameters.

The models (4) and (5) can be extended to 2-order polynomials. The approximation (6) remains valid and accordingly, the polynomials (8) increase in order. It can be shown that the new coefficients can also be expressed as in terms of covariance similar to (10). We will, however, omit this step because it is not necessary in the context of the present application.

3. F_0 ESTIMATION – THE *ASA* METHOD

There are various strategies that aim to estimating the fundamental frequency of quasi-harmonic signals in either time or frequency domain e.g. [13]. They all try to evaluate the periodicity hypothesis in the search range. However, they often shift from the optimal performance due to noise or/and subharmonic errors. Non-linear F_0 estimates [10, 14] are potentially good candidates for our goal. They use different non-linear search procedures to yield the desired estimate. We have opted for the fully automatic spectral analysis (ASA) method [10]. The choice was motivated by the fact that the ASA performs the F_0 estimation independently of the rest of the harmonic parameters. Although it has been developed for periodic signals, it has been proven heuristically to work well for most real-world audio signals which are inherently non-periodic. We summarize in the following paragraphs only the main features of the algorithm.

3.1 Initial F_0 estimate

The initial estimate is performed through the classical correlation-based method. The goal is to detect the distance between the successive peaks in the autocorrelation sequence of the audio signal. In case of a wideband signal with a flat spectrum, the estimate will be correct. However, this approach will fail in many special cases regarding narrow-band signals. In order to handle properly this kind of signals, the correlation method has been refined as follows.

3.2 Improved F_0 estimate

From the initial F_0 estimate, we know that the analysis window cover more than M periods of the signal $s(n)$. In addition, the signal itself is not strictly periodic, because of the time-varying harmonic parameters. Consequently, the DFT analysis of the signal will produce a discrete spectrum $S(k)$ with leakage. Nevertheless, we make use of this spectrum to define a particular measure of periodicity through the following cost function:

$$V(F_0)=\frac{\sum_{k=1}^{N}\left|S(Mk+1,F_0)\right|^2+\left|S(Mk-1,F_0)\right|^2}{\sum_{k=1}^{N}\left|S(Mk,F_0)\right|^2}\ , \qquad (11)$$

where F_0 is the fundamental frequency of the signal to be estimated. The function $V(F_0)$ expresses the ratio of the power at the harmonic and non-harmonic frequencies. Accordingly, the estimate of F_0 is defined as:

$$\hat{F}_0=\arg\min_{F_0} V(F_0). \qquad (12)$$

The minimization problem in (12) is non-linear in F_0, hence a non-linear search is used. Since $V(F_0)$ can have various local minima, the search is split in coarse search by scanning the cost function around the initial guess. The final estimate is obtained by a fine search based on parabolic interpolation. The algorithm converges rapidly in usually only a few iterations. More details on the algorithm performance can be found in [10].

4. LS PARAMETER ESTIMATION

Once F_0 is estimated, the coefficients of the polynomials $p_s^{(i)}$ and $p_c^{(i)}$ are estimated by LS as they are linear-in-parameters:

$$p_{\mathrm{s}}^{(\mathrm{i})}(n)=\sum_{\mathrm{k}=0}^{P}p_{sk}^{(i)}n^k\ ,\quad p_{\mathrm{s}}^{(\mathrm{i})}(n)=\sum_{\mathrm{k}=0}^{P}p_{sk}^{(i)}n^k\ . \qquad (13)$$

Assuming that the data is available at times (n - N, ..., n), the following regression equation results:

$$\boldsymbol{s}=\boldsymbol{\Psi}\boldsymbol{p}+\boldsymbol{r}\ , \qquad (14)$$

$$\boldsymbol{p}=\left(p_{s0}^{(0)}\cdots p_{sP}^{(0)}\cdots p_{s0}^{(I)}\cdots p_{sP}^{(I)}\,p_{c0}^{(0)}\cdots p_{cP}^{(0)}\cdots p_{c0}^{(I)}\cdots p_{cP}^{(I)}\right)^T, \qquad (15)$$

$$\boldsymbol{\Psi}=\begin{pmatrix}\boldsymbol{\psi}^T\left(s(n),c(n)\right)\\ \vdots\\ \boldsymbol{\psi}^T\left(s(n-N-1),c(n-N-1)\right)\end{pmatrix}, \qquad (16)$$

$$\boldsymbol{\psi}^T\left(s(n),c(n)\right)=\left(1\ldots s(In)1\ldots c(In)\right)\ , \qquad (17)$$

$$s(n)=\sin 2\pi F_0 n\ ,\quad c(n)=\cos 2\pi F_0 n\,. \qquad (18)$$

The vectors $\boldsymbol{s}$ and $\boldsymbol{r}$ contain the input data and additive noise respectively. Accordingly, the LS estimate $\hat{p}^{LS}$ is given by:

$$\hat{\boldsymbol{p}}^{LS}=\left(\boldsymbol{\Phi}^T\boldsymbol{\Phi}\right)^{-1}\boldsymbol{\Phi}^T\boldsymbol{s}\,. \qquad (19)$$

5. EXPERIMENTAL RESULTS

In this section we quantitatively evaluate the efficiency of the proposed harmonic modelling approach through a comparative study. The comparison reference methods are chosen to be [1] (from now on the Time-variant LS) and [3] (from now on the Peak selection method), as they represent completely different approaches to harmonic modelling.

The Time-variant LS models the harmonic component of an audio signal by means of a two-level LS algorithm. In the first level the fundamental frequency variation within the analysis window is estimated. In the second level this variation is incorporated into the harmonic model and the harmonic parameters are estimated. The Peak selection method is based on spectral peak selection in the STFT of an audio signal. First a harmonic model is defined and the harmonic parameters are estimated for each peak in the STFT. From those parameters the corresponding harmonic component is generated and its spectral peak calculated. Finally, the original and estimated peak are compared through a complex correlation and if the maximum of the correlation if above a certain threshold, the component is considered harmonic.

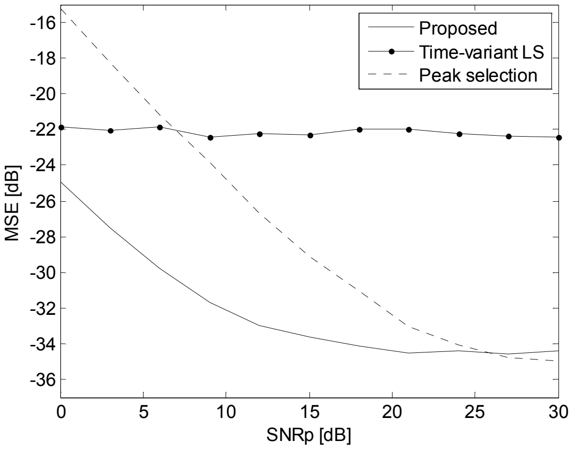

Figure 1 – Harmonic approximation mean square error (MSE). The SNRp is expressed with reference to the smallest harmonic.

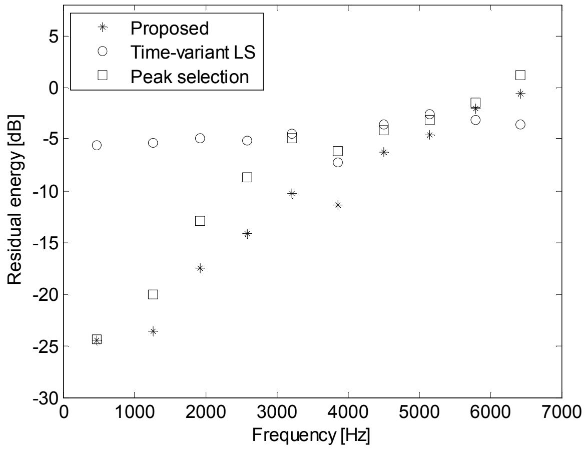

Figure 2 – Residual energy in the harmonic subbands.

From all possible modulation laws we have chosen the sinusoidal amplitude and frequency modulation for the test signal. The choice is motivated by the fact that a wide range of FM and AM conditions can be covered. For example, if the window size is significantly smaller than the modulation period then our model creates approximately linear FM and AM. Otherwise, we get a vibrato-like signal component. The phase relationship between the modulation laws for the real-world vibrato signals will in general be arbitrary [11, 15]. Because part of the AM modulation is induced by the FM and the resonator filter of the sound source, the dominant AM rate may either be the same as the FM rate, or twice as high. As the letter case is more critical, we chose it for our model. Accordingly, the test signal $s_t(n)$ is defined as follows:

$$s_t(n) = \sum_{i=0}^{10} A_i(n)\cos\left[2\pi i F_0 n + \varphi_i(n)\right] + r(n), \qquad (20)$$

$$\varphi_i(n) = iA_{FM}\sin\left(2\pi F_{FM} n + \alpha\right) + \gamma_i\ , \qquad (21)$$

$$A_i(n) = \frac{A_0}{i}\left[1 + A_{AM}\cos\left(2\pi F_{AM} n + \delta\right)\right]\ . \qquad (22)$$

According to the aforementioned discussion, we set $F_{AM} = 2F_{FM}$. In order to guarantee the correct operation of the Peak selection method we must assure the presence of the dominant mainlobe at the harmonic frequencies in the STFT. If L is the size of the analysis window, then this constraint is accomplished by letting $A_{AM} = 0.5$, $A_{FM} = 2$ and $F_{FM} = (4L)^{-1}$ for arbitrary combinations of α, γ and δ in the range $(-\pi, \pi)$. The noise impact in (20) is controlled in an intuitive way through the *Peak Signal-to-Noise ratio (*SNR$_\text{p}$) which we define as the peak power of the smallest harmonic above the neighbouring noise floor. The DFT size N_{FFT} is chosen in such a way to assure that the Picket-Fence effect has minimal impact on a peak representation in the discrete spectrum. The remaining parameters do not have any impact on the result, so they can be chosen arbitrary (e.g. L = 20ms, N_{FFT} = 4096, F_o = 1kHz*,* Sampling rate = 44 kS/s).

In order to evaluate the performance of the methods we let the SNR$_\text{p}$ vary in range [0, 30]dB and for each value we calculate the MSE of the approximation over 100 realizations. The resulting curves are plotted on Figure 1. We observe that the proposed and Peak selection method follow a similar trend but the proposed method is clearly better for low SNR$_\text{p}$. This is due to the fact that the Peak selection method is no longer able of making a clear distinction between harmonic and noise peaks. The Time-variant LS produces an approximately constant MSE along the analysis range. This comes from the fact that amplitude modulation is not accounted for in this method. Hence, we have a superposition of modelling and measuring errors.

Next we want to examine the approximation error along the frequency grid. For SNR$_\text{p}$ = 0dB we divide the spectrum corresponding to the harmonic band in 10 subbands. Then for each subband delimited by the DFT indices (k_i, k_{i+1}) we calculate the residual energy $E_R^{(i)}$ as follows:

$$E_R^{(i)} = \left| \sum_{k=k_i}^{k_{i+1}} \left|H(k)\right|^2 - \sum_{k=k_i}^{k_{i+1}} \left|\hat{H}(k)\right|^2 \right| \Bigg/ \sum_{k=k_i}^{k_{i+1}} \left|H(k)\right|^2\ , \qquad (23)$$

where $H(k)$ and $\hat{H}(k)$ are the DFTs of the harmonic part in (20) and its approximation respectively. Accordingly, the mean residual energy variation over 100 realizations is shown on Figure 2. The Peak selection method reaches the performance of the proposed method only in the lowest subband. The residual error is maximal at the highest subband for both methods, due to the maximal modulation and minimal SNR$_\text{p}$. The time-variant LS approximation is dominated by the modelling errors and therefore the residual energy is fairly uniform over the analysis bandwidth.

A representative example of the performance of the proposed method is a japanese flute, whose time record contains a strong convolutive noise (air flow) and very complex frequency and amplitude variations over time. The analysis results are shown on Figure 3. The proposed method captures the harmonic time-variation trend, even at the time instants where the parameters change abruptly. In addition, the perceived approximation quality is high, as the listening confirms that there are no audible artefacts in the residual signal.

6. CONCLUSIONS

We have proposed a compact and computationally efficient description of time-variant harmonicities in audio signals through a single polynomial that captures both amplitude and frequency variations and an equivalent stationary fundamental harmonic component. The estimation of the mean fundamental frequency is performed by a noise-robust algorithm which significantly improves the original LS estimator. The performance of the proposed method has been tested against techniques belonging to different approaches to harmonic modelling. According to the experimental results, the proposed method achieves the best performance for vibrato-like signals in noisy conditions. Its performance has also been tested for a variety of real-world audio signals, among which an illustrative example is shown.

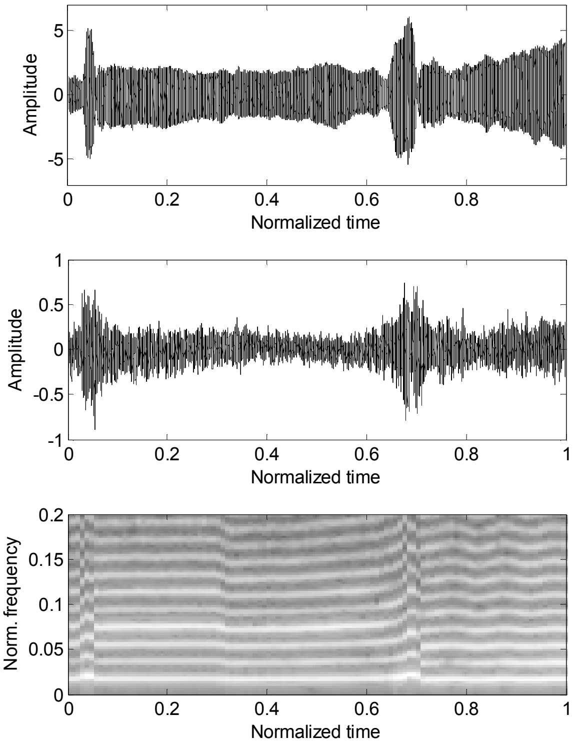

Figure 3 – Japanese flute signal (above); the residual after subtracting the estimated harmonic part (middle); spectrogram of the estimated harmonic part (below).

REFERENCES

[1] Q. Li, L. Atlas, “Time-variant least-squares harmonic modelling”, *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol.2, Hong-Kong, April 2003

[2] Y. Mahgoub, R. Dansereu, “Time domain methods for precise estimation of sinusoidal model parameters of co-channel speech”, *Research Letters in Signal Processing,* Article ID 364674, 5 pages, 2008

[3] M. Lagrange et al, “Sinusoidal Parameter Estimation in a Non-Stationary Model”, *Proceedings of the 5th International Conference on Digital Audio Effects*, Hamburg, Germany, September 2002

[4] X. Rodet, “Musical sound signal analysis/synthesis: Sinusoidal + residual and elementary waveform models,” *in Proceedings IEEE Time-Frequency and Time-Scale Workshop* 97, (TFTS’97), 1997

[5] S. Marchand, P. Depalle, “Generalization of the derivative analysis method to non-stationary sinusoidal modelling”, *Proceedings of the 11th International Conference on Digital Audio Effects*, Espoo, Finland, September 2008

[6] M. Robine, R. Strandh, S. Marchand, “Fast additive sound synthesis using polynomials”, *Proceedings of the 9th International Conference on Digital Audio Effects*, Montreal, Canada, September 2006

[7] G. Peeters, X. Rodet “SINOLA: A New Analysis/Synthesis using Spectrum Peak Shape Distortion, Phase and Reassigned Spectrum”, *Proceedings of the International Computer Music Conference*, Bejing, China, 1999

[8] M. Raspaud, S. Marchand, L. Girin, “A generalized polynomial and sinusoidal model for partial tracking and time stretching”, *Proceedings of the 8th International Conference on Digital Audio Effects*, Madrid, Spain, September 2005

[9] S. Godsill, M. Davy, “Bayesian harmonic models for musical pitch estimation and analysis”, *Proceedings of International Conf. Acoustic Speech and Signal Processing*, pp. 1769-1772, Orlando, USA, May 2003

[10] J. Schoukens, Y. Rolain, G. Simon, R. Pintelon, “Fully automated spectral analysis of periodic signals”, *IEEE Transactions on Instrumentation and Measurement*, Vol. 52, NO. 4, pp. 1021-1024, August 2003

[11] I. Arroabarren, X. Rodet, A. Carlosena, “On the measurement of the instantaneous frequency and amplitude of partials in vocal vibrato”, *IEEE Transactions on Speech and Audio Processing,* Vol. 14, NO. 4, pp.1413-1421, 2006

[12] L.Cohen, “Time-frequency analysis”, Prentice Hall, 1995

[13] A. Cheveigne, H. Kawahara, “YIN, a fundamental frequency estimator for speech and music”, *Journal Acoustics Society of America*, 111(4), April 2002

[14] R. Pintelon, J. Schoukens, “An improved sine-wave fitting procedure for characterizing data acquisition channels”, *IEEE Transactions on Instrumentation and Measurement*, Vol. 2, No. 45, pp. 588-593, April 1996

[15] V. Verfaille, C. Guastavino, P. Depalle, “Perceptual evaluation of vibrato models”, *Proc. of the Conf. on Interdisciplinary Musicology (CIM’05)*, 2005