Iterative Estimation of Sinusoidal Signal Parameters

Yannis Pantazis, Olivier Rosec, and Yannis Stylianou

***Abstract*—While the problem of estimating the amplitudes of sinusoidal components in signals, given an estimation of their frequencies, is linear and tractable, it is biased due to the unavoidable, in practice, errors in the estimation of frequencies. These errors are of great concern for processing signals with many sinusoidal like components as is the case of speech and audio. In this letter, we suggest using a time-varying sinusoidal representation which is able to iteratively correct frequency estimation errors. Then the corresponding amplitudes are computed through Least Squares. Experiments conducted on synthetic and speech signals show the suggested model's effectiveness in correcting frequency estimation errors and robustness in additive noise conditions.**

***Index Terms*—Amplitude estimation, frequency estimation, sinusoidal modeling, time-varying models, estimation theory.**

I. Introduction

SINUSOIDAL models are widely used in signal processing for analysis, modeling and manipulation of time-series from speech and audio to radar and sonar (see [1], [2] and the references therein). In the literature, many techniques for the estimation of the sinusoidal parameters have been proposed. A simple and fast method for sinusoidal parameter estimation uses Fourier spectrum where the locations of the peaks of the spectral magnitude are the estimated frequencies and the values of the Fourier transform at these frequencies are the estimated complex amplitudes [1], [3]. Although FFT-based spectral estimation is asymptotically unbiased and efficient, it is biased for finite length data [4]. Extensions of the basic FFT method such as quadratically interpolated FFT (QIFFT) [5], [6] and reassigned spectrogram [7] have been proposed in the literature. A survey study by Keiler and Marchand [8] compares various FFT-based amplitude and frequency estimators. FFT-based methods can be considered as local since the parameters of each sinusoidal component is estimated without taking into account the possible influence of neighboring sinusoidal components. For large windows with sufficiently narrow main frequency lobe, this influence is negligible. However, using long windows the stationarity hypothesis for the analyzed signal may be violated. Consequently, the frequency estimation provided by Fourier transform-based methods is rather unreliable in nonstationary environments.

Another approach for sinusoidal parameter estimation is to use global estimation methods, through the minimization of a Least Squares (LS) criterion. For sinusoidal signals, such a criterion is highly nonlinear when the frequencies of the sinusoidal components are unknown. Hence, the estimation procedure is usually split into two steps: i) estimation of the frequencies and ii) estimation the complex amplitudes given the estimated frequencies [9], [4], [10], [11]. A major disadvantage of splitting the estimation into two subproblems is that the estimation of the complex amplitudes are severely biased when the estimation of the frequencies is not accurate. In practice, errors in frequency estimation inevitably occur when sources of interference such as noise or closely-spaced sinusoids are present.

In this paper, we consider a time-varying model for the sinusoidal parameter estimation which is immune to small frequency estimation errors. The proposed model was initially introduced by Laroche [12] for audio analysis of percussive sounds. In this model, a complex polynomial is used in order to capture fast variations within each frequency component, providing access to the instantaneous amplitudes but also—and more interestingly—to the instantaneous frequencies. Focusing on the first order model, we showed in [13] that this model is equivalent to a time-varying quasi-harmonic representation (referred to as QHM for Quasi-Harmonic Model). We showed that by proper decomposition of QHM parameters, errors in the initially estimated frequencies of sinusoidal components of the signal can be identified and then corrected. Thus, an algorithm is suggested which iteratively improves the estimated frequencies and provides unbiased amplitude estimates. The performance of the proposed estimation method is studied and boundaries of frequency errors are provided that ensure the convergence of the frequency correction algorithm. Frequency estimation experiments using short analysis windows show that the proposed method outperforms traditional FFT-based methods, especially in the case of closely-spaced sinusoids, thus underlying the ability of the suggested signal representation to perform a high resolution frequency analysis. Finally, the robustness to noise is established since the obtained estimators asymptotically reach their Cramer–Rao lower bounds even in adverse noisy conditions.

The paper is organized as follows. In Section II, the QHM model as well as its main properties are presented. Section III establishes the conditions upon which the QHM model can be used to estimate sinusoidal component. Section IV illustrates the robustness of the proposed method in additive noise, while Section V presents results on speech signals. Finally, Section VI concludes the paper.

II. Quasi-Harmonic Model

Let us consider a signal $x(t)$ consisting of K complex sinusoids:

$$x(t) = \sum_{k=1}^{K} c_k e^{j2\pi f_k t} \tag{1}$$

where f_k and c_k denote the frequency and complex amplitude, respectively, of the kth sinusoid. In order to compute the complex amplitudes through LS, we need to have estimates of frequencies $\{\hat{f}_k\}_{k=1}^{K}$ of the sinusoidal components. Such estimates

Manuscript received November 05, 2009; revised January 27, 2010. First published February 17, 2010; current version published March 26, 2010. This work was supported by Orange Labs through Industrial Contract NÂ 200141932 with FORTH. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Mark Gales.

Y. Pantazis and Y. Stylianou are with Computer Science Department, University of Crete and ICS-FORTH, Heraklion, Greece (e-mail: pantazis@csd.uoc.gr; yannis@csd.uoc.gr).

O. Rosec is with Orange Labs TECH/ASAP/VOICE, Lannion, France, (e-mail: olivier.rosec@orange-ftgroup.com).

Digital Object Identifier 10.1109/LSP.2010.2043153

can be obtained by spectrograms, subspace methods or any other frequency estimation method.

Given a set of frequency estimates, we may write

$$f_k = \hat{f}_k + \eta_k. \tag{2}$$

If the frequency error, η_k, is high, then the estimation of the complex amplitudes c_k through LS will be severely biased. To cope with this problem, we suggest in [13] to use the QHM for the representation of the input signal

$$s(t) = \sum_{k=1}^{K} (a_k + tb_k) e^{j2\pi \hat{f}_k t} \tag{3}$$

where a_k and b_k respectively denote the complex amplitude and complex slope of the kth component. Parameters $\{a_k, b_k\}_{k=1}^{K}$ are computed through the minimization of the LS criterion $\sum_{t=-T}^{T}((x(t) - s(t))w(t))^2$, where $w(t)$ is the analysis window defined on a time interval $[-T, T]$ [10], [12]. It is worth noting from (1) and (3) that in case $f_k = \hat{f}_k$ then the LS solution will give $b_k = 0$. On the contrary, if $f_k \neq \hat{f}_k$, then $b_k \neq 0$, meaning that b_k may also carry information related to the phase evolution of the kth component.

To reveal this information, we consider the Fourier transform of $s(t)$ given by

$$S(f) = \sum_{k=1}^{K} \left(a_k W(f - \hat{f}_k) + j\frac{b_k}{2\pi} W'(f - \hat{f}_k) \right) \tag{4}$$

where $W(f)$ is the Fourier transform of the analysis window, $w(t)$, and $W'(f)$ is the derivative of $W(f)$ over f. For simplicity, we will only consider the kth component of $S(f)$

$$S_k(f) = a_k W(f - \hat{f}_k) + j\frac{b_k}{2\pi} W'(f - \hat{f}_k). \tag{5}$$

In [13], it was suggested to project b_k onto a_k according to

$$b_k = \rho_{1,k} a_k + \rho_{2,k} j a_k \tag{6}$$

where ja_k denotes the perpendicular (vector) to a_k while $\rho_{1,k} = (a_k^R b_k^R + a_k^I b_k^I)/(|a_k|^2)$ and $\rho_{2,k} = (a_k^R b_k^I - a_k^I b_k^R)/(|a_k|^2)$.

Then, the kth component is written as

$$S_k(f) = a_k \Big[W(f - \hat{f}_k) - \frac{\rho_{2,k}}{2\pi} W'(f - \hat{f}_k) + j\frac{\rho_{1,k}}{2\pi} W'(f - \hat{f}_k) \Big]. \tag{7}$$

Considering the Taylor series expansion of $W(f - \hat{f}_k - (\rho_{2,k})/(2\pi))$

$$\begin{aligned} W\left(f - \hat{f}_k - \frac{\rho_{2,k}}{2\pi}\right) &= W(f - \hat{f}_k) - \frac{\rho_{2,k}}{2\pi} W'(f - \hat{f}_k) \\ &\quad + O(\rho_{2,k}^2 W''(f - \hat{f}_k)) \\ &\approx W(f - \hat{f}_k) - \frac{\rho_{2,k}}{2\pi} W'(f - \hat{f}_k) \end{aligned} \tag{8}$$

it follows from (7) and (8) that

$$S_k(f) \approx a_k \left[W\left(f - \hat{f}_k - \frac{\rho_{2,k}}{2\pi}\right) + j\frac{\rho_{1,k}}{2\pi} W'(f - \hat{f}_k) \right] \tag{9}$$

which is written in the time domain as

$$s(t) \approx \sum_{k=1}^{K} a_k \left[e^{j(2\pi \hat{f}_k + \rho_{2,k})t} + \rho_{1,k} t e^{j2\pi \hat{f}_k t} \right] w(t). \tag{10}$$

From (10), it can be observed that $\rho_{2,k}/2\pi$ accounts for the mismatch between the frequency of the kth component, f_k, and the analysis frequency, $\hat{f}_k$. Thus, an estimate of the frequency mismatch for the kth component is

$$\hat{\eta}_k = \rho_{2,k}/2\pi. \tag{11}$$

Also, from (10), we see that $\rho_{1,k}$ accounts for the amplitude slope of the k component. Thus, for the specific signal in (1), $\rho_{1,k} = 0$ for each k.

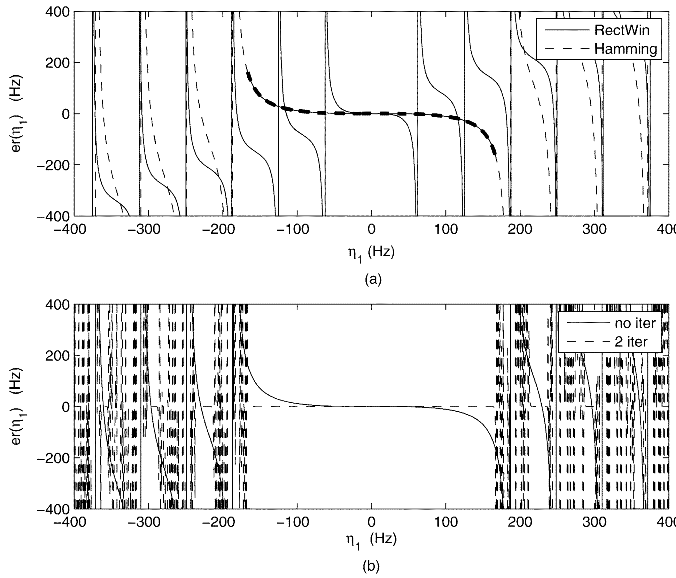

Fig. 1. Upper panel: The error for (solid line) a rectangular and (dashed line) Hamming window. Lower panel: Error using the Hamming window (as in a) without (solid line) and with two iteration (dashed line).

III. Validity of the QHM for Frequency Estimation

The error of the approximation in (8) depends on the characteristics of the analysis window as well as on the amount of frequency mismatch. In order to get further insight on the proposed estimator, we first present results on a mono-component signal analyzed using a window of length 16 ms ($T = 8$ ms). Fig. 1(a)[1] shows the error between the true frequency mismatch and the estimated one (i.e., $er(\eta_1) = \eta_1 - \hat{\eta}_1$) for a rectangular window and a Hamming window of duration 16 ms. The error is small (i.e., $|\eta_1 - \hat{\eta}_1| < \eta_1$) if the frequency mismatch is smaller than 50 Hz for a rectangular window and smaller than 135 Hz in the case of a Hamming window. The advantage of using a Hamming window (vs. a rectangular window) is twofold. First, the second order derivative $W''(f)$ is much smaller for a Hamming window than for a rectangular window and consequently, the approximation (8) is more accurate. Second, it can be shown that the LS estimates for $\{a_1, b_1\}$ involves the reciprocal of the Fourier transform of the squared analysis window evaluated at the frequency mismatch. Consequently, the larger the main lobe the larger the allowed frequency mismatch. For a squared Hamming window the main lobe is three times larger than the one of the rectangular window, which may explain why the region of small error is about three times larger for a Hamming window. After testing a variety of windows, we conclude that the allowable frequency mismatch should be less than one third of the bandwidth of the squared analysis window.

[1]MATLAB code that generates the figures is available at www.csd.uoc.gr/~pantazis/source/IterSinEst_FigsCode.zip.

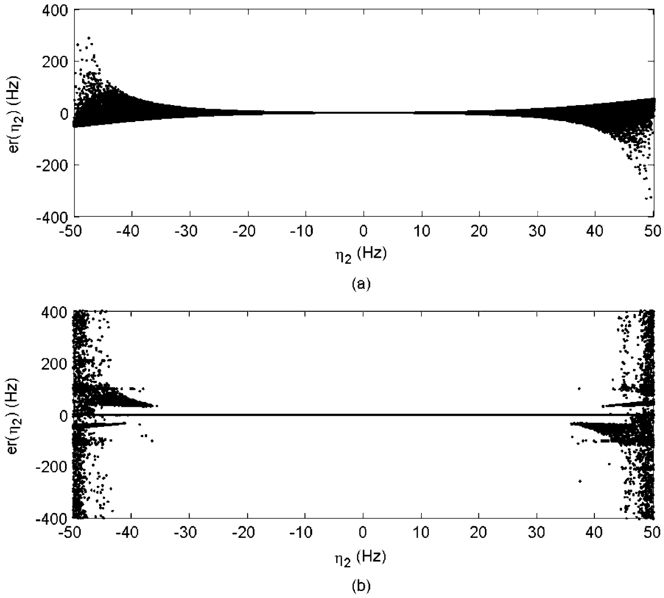

Fig. 2. Frequency mismatch estimation in the case of three sinusoidal components: error on the second component (upper panel) without iteration and (lower panel) with three iterations.

TABLE I
PARAMETERS OF A SYNTHETIC SINUSOIDAL SIGNAL WITH FOUR COMPONENTS AND INTERVALS OF ALLOWED FREQUENCY MISMATCH PER COMPONENT

Sinusoid	1st	2nd	3rd	4th
Frequency (Hz)	100	200	1000	2000
Amplitude	$e^{j\pi/10}$	$e^{j\pi/4}$	$e^{j\pi/3}$	$e^{j\pi/5}$
Mismatch int. (Hz)	$[-10, 10]$	$[-10, 10]$	$[-100, 100]$	$[-100, 100]$

Furthermore, if the error is small (i.e., $|\eta_1 - \hat{\eta}_1| < \eta_1$), then the estimated frequency mismatch can be used to correct the initial frequency estimates. In such a case, the new frequency mismatch decreases; thus, we suggest the following iterative frequency correction procedure.

1) Initialization
   i) Get an initial estimate of frequencies, $\{\hat{f}_k\}_{k=1}^K$.
   ii) Estimate $\{a_k, b_k\}_{k=1}^K$ given $\{\hat{f}_k\}_{k=1}^K$.
2) Do iterations
   i) For each kth component:
      a) Estimate $\hat{\eta}_k$ using (11).
      b) Update frequencies: $\hat{f}_k \leftarrow \hat{f}_k + \hat{\eta}_k$.
   ii) Reestimate $\{a_k, b_k\}_{k=1}^K$ given $\{\hat{f}_k\}_{k=1}^K$

Fig. 1(b) shows that this iterative procedure converges in only two iterations in this simple case.

Let us now consider multicomponent signals. It is well known that the closer the frequencies of the sinusoidal components, the more difficult the estimation becomes. Consequently, the analysis window should be adapted to the minimum spacing between adjacent sinusoids denoted by Δf. Considering harmonically related sinusoids, we will use a window of length $2T$ with $T = 1/\Delta f$. If we further assume that the minimum frequency to be estimated is greater than Δf, then the maximum frequency mismatch should be proportional to Δf. Note also that the frequency mismatch should be smaller than $\Delta f/2$, otherwise, the estimation problem may become ill-posed. In the following experiment we consider three components of equal strength respectively located at 900, 1000 and 1100 Hz (i.e., $\Delta f = 1/T = 100$ Hz). Using a Hamming window of length $2T$, we run Monte-Carlo simulations as follows: for each frequency the mismatch is drawn from a uniform distribution on $[-\Delta f/2, \Delta f/2]$ in such a way that the spacing between frequencies is constrained to be smaller than $\Delta f/2$. Fig. 2 depicts the results of the frequency mismatch estimation for the second component using 10^6 simulations. It can be clearly seen that if the mismatch is below 35 Hz, then the frequency estimation is very accurate, since the remaining mismatch can be reduced by subsequent iterations of the updating process. Similar results were obtained for the other components. More generally, by testing the algorithm on a wide range of multicomponent sinusoidal signals, it was observed that frequency correction occurs if the frequency mismatch is below 35% of the minimum frequency spacing. Thus, even though the presence of adjacent sinusoids has contributed to reduce the range of admissible mismatch as compared to the single component case, the proposed method is still able—under reasonable conditions—to provide an accurate estimation of the sinusoidal component frequencies.

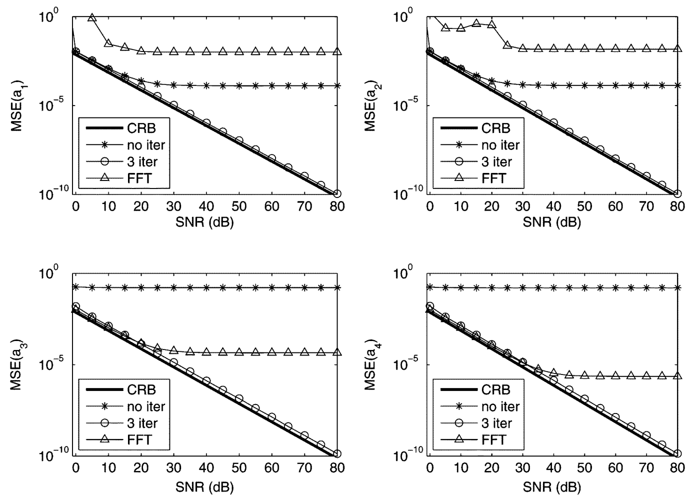

Fig. 3. MSE of the four amplitudes as a function of SNR.

IV. ROBUSTNESS IN NOISE

In this section, the performance of QHM is assessed for the case when a signal with multiple sinusoidal components is contaminated by white noise. Concisely, the ability of the proposed model to improve the accuracy of the estimation of the frequencies—hence the accuracy of the amplitudes—is tested. The signal consists of 4 sinusoids and it is corrupted by noise while window's duration is 17 ms ($T = 8.5$ ms) and sampling frequency 8000 Hz (i.e., so the duration of the window duration in samples is 137). In Table I, the frequency and the amplitude of each component are given. Two closely-spaced sinusoids and two well-separated sinusoids are considered. Monte Carlo simulations are used for the assessment of the robustness of the proposed method. For each simulation, the frequency mismatch of each sinusoid is sampled uniformly on the intervals defined in Table I.

Figs. 3 and 4 respectively depict the mean squared error (MSE) of the amplitudes and frequencies of each component after 10^5 Monte Carlo simulations. For comparison purposes, QIFFT frequency and amplitude estimation as well Cramer–Rao bounds (CRB) [1], [4] are presented. Please note that the number of FFT bins is set to 2048, thus, the zero-padding factor of QIFFT is about 15 (2048/137), which is sufficient for the accurate estimation of amplitudes and

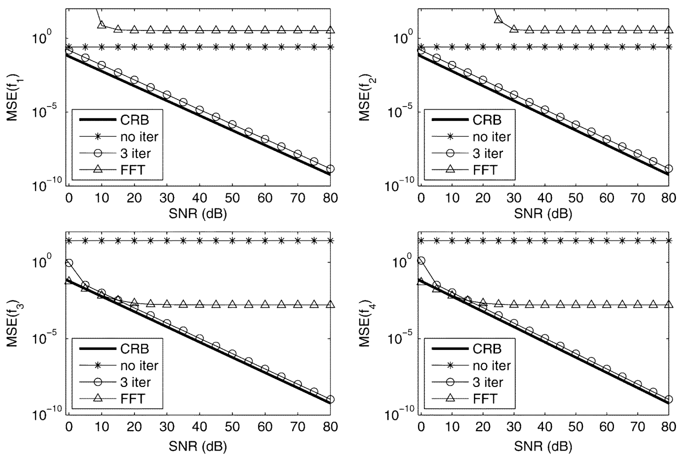

Fig. 4. MSE of the four frequencies as a function of SNR.

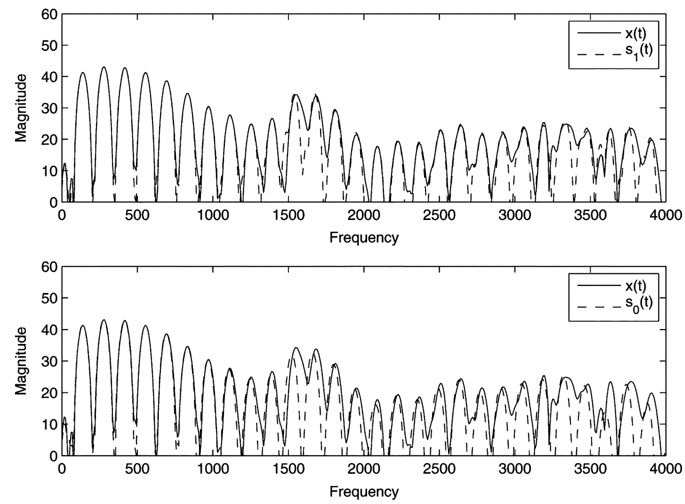

Fig. 5. (Solid line) Magnitude spectra and (dashed line) estimated spectra by (upper panel) QHM (upper panel) and (lower panel) the stationary sinusoidal model. The estimated fundamental frequency used by both models was 138.4 Hz.

frequencies [5]. Results clearly show that the proposed estimation scheme outperforms the FFT-based approach while the estimations asymptotically reach the CRB.

V. Speech Signal Analysis

There is a vast literature in coding, synthesis, modification, etc. where both speech and music signals are modeled frame-by-frame as a sum of harmonically related sinusoids. However, looking at the magnitude spectrum of short-term Fourier transform it is easily seen that the local maxima (peaks) are not exactly at the integer multiples of the fundamental frequency. This inharmonicity—also called detuning—induces biased estimation of the complex amplitudes. Furthermore, even if the frequencies of the real signals were perfect harmonics, errors may occur in the estimation of the fundamental frequency, hence once again, bias is introduced in the amplitude estimation.

To alleviate these problems, a model that is able to move to the correct frequency of each sinusoid is of great interest. To illustrate the performance of the proposed algorithm, we compare it to a classic harmonic model [10]. For that purpose we select a 30 ms frame from a reasonably stationary section of speech. The magnitude spectra computed by FFT and estimated using the classic harmonic representation as in [10] as well as the proposed model are shown in Fig. 5. Interestingly, the harmonics between 1.5 kHz and 2 kHz where the second formant takes place are greatly detuned and are missed by a purely harmonic model. By contrast, the suggested approach provides a better spectral estimation. In terms of Signal-to-Reconstruction Error-Ratio (SRER), the improvement is 4.1 dB. These observations were consistent by testing more than five minutes of voiced speech from both male and female voices where the average SRER was found to be 4.3 dB.

VI. Conclusion

We suggest an iterative approach in correcting initial frequency estimations in the context of sinusoidal modeling. We provide details on the frequency mismatch intervals where the suggested iterative frequency correction algorithm converges and we show that it is robust against additive white noise. Experiments with synthetic signals show that the suggested algorithm outperforms FFT-based frequency approaches. More specifically it performs high-resolution frequency estimations which reach asymptotically the Cramer–Rao lower bound. Experiments with speech signals show that the suggested model is more adequate for modeling voiced speech than simple stationary harmonic representations.

References

[1] S. M. Kay, *Modern Spectral Estimation: Theory and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1988.

[2] L. S. Marple Jr., "Digital spectral analysis with applications," in *Signal Processing Series*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[3] R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, pp. 744–754, 1986.

[4] P. Stoica, H. Li, and J. Li, "Amplitude estimation of sinusoidal signals: Survey, new results, and an application," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 338–352, Feb. 2000.

[5] M. Abe and J. O. Smith III, Design criteria for the quadratically interpolated FFT method (I): Bias due to interpolations Stanford Univ., Stanford, CA, Tech. Rep. STAN-M-114, Oct. 2004.

[6] M. Abe and J. O. Smith, III, CQIFFT: Correcting bias in a sinusoidal parameter estimator based on quadratic interpolation of FFT magnitude peaks Stanford Univ., Stanford, CA, Tech. Rep. STAN-M-117, Oct. 2004.

[7] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," *IEEE Trans. Signal Process.*, vol. 43, pp. 1068–1089, 1995.

[8] F. Keiler and S. Marchand, "Survey on extraction of sinusoids in stationary sounds," in *Int. Conf. on Digital Audio Effects (DAFx-02)*, Hamburg, Germany, Sept. 2002.

[9] S. M. Kay, *Funtamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[10] Y. Stylianou, "Harmonic plus noise models for speech, combined with statistical methods, for speech and speaker modification," Ph.D. thesis, Ecole Nationale Sup'erieure des T'el'ecommunications, Paris, France, 1996.

[11] W. D'haes, "Automatic estimation of control parameters for musical synthesis algorithms," Ph.D. thesis, Univ. Antwerp, Antwerp, Belgium, 2004.

[12] J. Laroche, "A new analysis/synthesis system of musical signals using Prony's method," in *Application to Heavily Damped Percussive Sounds. In ICASSP*, Glasgow, U.K., May 1989, pp. 2053–2056.

[13] Y. Pantazis, O. Rosec, and Y. Stylianou, "On the properties of a time-varying quasi-harmonic model of speech," in *Interspeech*, Brisbane, Australia, Sep. 2008, pp. 1044–1047.