Variable Order Harmonic Sinusoidal Parameter Estimation for Speech and Audio Signals

Christensen, Mads G.; Jensen, Søren Holdt

Published in:
Rec. Asilomar Conference on Signals, Systems, and Computers

Publication date:
2006

Document Version
Accepted author manuscript, peer reviewed version

VARIABLE ORDER HARMONIC SINUSOIDAL PARAMETER ESTIMATION FOR SPEECH AND AUDIO SIGNALS

Mads Græsbøll Christensen and Søren Holdt Jensen

Dept. of Electronic Systems
Aalborg University, Denmark
{mgc,shj}@kom.aau.dk

ABSTRACT

In this paper, a computationally efficient method for the estimation of the parameters of harmonic sinusoidal signals, including the order, which is of particular importance, for speech and audio signals is presented. The signal is modeled as a sum of harmonically related sinusoids in colored Gaussian noise. Aside from the order, the proposed method estimates the noise parameters, the fundamental frequency and the phase and amplitude of the individual harmonics. For the special case of white Gaussian noise, the method becomes particularly simple. The application of the proposed estimator to analysis of voiced speech and music signals is illustrated using real-life signals.

1. INTRODUCTION

Speech and audio signals are often modeled as sums of sinusoids in colored Gaussian noise, e.g., [1, 2], and recent audio coding standards, such as MPEG-4, include such parametric signal models as a basis for data compression. The fundamental signal model is often described as consisting of a deterministic part, the sinusoids, and a stochastic part, a noise term. The deterministic part is, in most cases, synthesized such that the reconstructed waveform matches the original while the stochastic part is typically generated such that the spectral envelope and sometimes also a coarse temporal envelope match that of the original signal. In other words, the stochastic part is only described in terms of its second order statistics. This is motivated by the notion that the human auditory system cannot distinguish two realizations of the same stochastic process. Signals produced by many musical instruments as well the human speech production system exhibit strong local periodicities. These signals can, for a particular segment, be modeled as a sum of harmonically related sinusoids, meaning that the frequencies of the individual sinusoids are integer multiples of a fundamental frequency. This fundamental frequency is an important feature in many signal processing applications ranging from speech coding to automatic music transcription. The task of extracting this fundamental frequency or the pitch period is known as fundamental frequency estimation or pitch estimation. The former is usually preferred since the latter may easily be confused with the related task of determining the perceived pitch. We present, in this paper, a method for estimation of the parameters of a set of harmonically related sinusoids, including the number of harmonics also known as the order, in colored Gaussian noise. Additionally, the covariance matrix of the Gaussian noise and its inverse are also estimated in a computationally efficient manner and we discuss the involved tradeoffs. The proposed estimator is based on the well-known principle of maximum likelihood which is combined with an order-dependent penalty term (see, e.g., [3]). It must be emphasized that even when only a subset of the parameters is of interest, such as the fundamental frequency, the rest have to be estimated, implicitly, anyway to yield correct estimates. For example, the background noise, which may or may not be produced by the instrument or the speaker, may be white or colored Gaussian noise and the number of sinusoidal components may vary [4]. Often, the order is assumed known (e.g. [5, 6, 7]). However, it is important to estimate the order for several reasons. Firstly, if the order is not estimated (or chosen) correctly, the fundamental frequency may erroneously be estimated at, for example, half or double of the true value [8, 9]. This is easy to see for the nonlinear least-squares estimator. Secondly, it can be seen from the Cramér-Rao bound for the fundamental frequency (see [5, 8]) that it is desirable to include as many of the harmonics as possible as this increases the accuracy. Therefore the order should be estimated on a segment to segment basis. The fundamental frequency estimation problem can be defined as follows. Consider a harmonic signal with the fundamental frequency ω_0 in additive complex circularly symmetric Gaussian noise, $e(n)$, i.e.,

$$x(n) = \sum_{l=1}^{k} A_l e^{j(\omega_0 l n + \phi_l)} + e(n), \tag{1}$$

This work is supported by the Intelligent Sound project, Danish Technical Research Council grant no. 26–04–0092.

where $A_l > 0$ and ϕ_l are the amplitude and the phase of the l'th harmonic, respectively. The frequency of the l'th harmonic is thus $\omega_0 l$, and the problem considered in this paper is to estimate the fundamental frequency ω_0, the amplitude and phase of the individual harmonics as well as the model order k and the noise covariance matrix from a set of N measured samples, $x(n)$. We note in passing that the complex model used here also is valid for real signals through the use of the down-sampled discrete-time analytic signal, provided that there is no signal of interest near 0 and 2π.

The remaining part of this paper is organized as follows. First, we present, in Section 2, the fundamentals of the proposed estimator, namely the principle of the maximum likelihood estimator. In Section 3, we then proceed to discuss how to evaluate the log-likelihoods in a computationally efficient manner. Specifically, we discuss how to, for a particular candidate fundamental frequency, obtain amplitude and noise covariance matrix estimates for various orders. We then treat the special case of white Gaussian noise in Section 4, and, in Section 5, we give some examples of the application of the proposed estimator to analysis of speech and audio signals. Finally, Section 6 concludes on the work.

2. MAXIMUM LIKELIHOOD ESTIMATOR

We will now present the signal model and the proposed algorithm. The algorithm operates on a signal sub-vector at time n $\mathbf{x}(n) \in \mathbb{C}^M$, defined as

$$\mathbf{x}(n) = [\ x(n)\ \cdots\ x(n+M-1)\ ]^T, \tag{2}$$

which is constructed from the observed signal $x(n)$. For many speech and audio signals, such a sub-vector can be modeled as a sum of k harmonically related complex sinusoids $\hat{\mathbf{x}}(n) \in \mathbb{C}^M$, in colored Gaussian noise $\mathbf{e}(n) \in \mathbb{C}^M$ having covariance matrix $\mathbf{Q}$, i.e.,

$$\begin{aligned}\mathbf{x}(n) &= \hat{\mathbf{x}}(n) + \mathbf{e}(n) && (3)\\ &= \mathbf{Z}_k(n)\mathbf{a}_k + \mathbf{e}(n), && (4)\end{aligned}$$

with $\mathbf{a}_k = [\ A_1 e^{j\phi_1}\ \cdots\ A_k e^{j\phi_k}\ ]^T$ being a vector containing the complex amplitudes and $(\cdot)^T$ denotes the transpose. Furthermore, $\mathbf{Z}_k(n)$ is a Vandermonde matrix at time n

$$\mathbf{Z}_k(n) = [\ \ \mathbf{z}_1(n)\ \ \ \cdots\ \ \ \mathbf{z}_k(n)\ \ ], \tag{5}$$

where the m'th entry of the column vector $\mathbf{z}_k(n) \in \mathbb{C}^M$ is defined as $[\mathbf{z}_k(n)]_m = e^{j\omega_0 k(n+m-1)}$. Since $\mathbf{x}(n)$ has length M and we have N observations of $x(n)$, we can thus construct a set of $G = N - M + 1$ different sub-vectors $\{\mathbf{x}(n)\}_{n=0}^{G-1}$. Next, we introduce the signal and noise parameter vector $\boldsymbol{\theta}$ containing the fundamental frequency ω_0, the complex amplitudes $\{A_l e^{j\phi_l}\}$ and thereby implicitly the order k and the noise covariance matrix $\mathbf{Q}$ of the model in (4). The likelihood function of the observed signal sub-vector $\mathbf{x}(n)$ can then be written as

$$p(\mathbf{x}(n);\boldsymbol{\theta}) = \frac{1}{\pi^M \det(\mathbf{Q})} e^{-\mathbf{e}^H(n)\mathbf{Q}^{-1}\mathbf{e}(n)}, \tag{6}$$

with $(\cdot)^H$ denoting the conjugate transpose and the noise vector being found as $\mathbf{e}(n) = \mathbf{x}(n) - \hat{\mathbf{x}}(n)$ by estimating the deterministic part of the signal model $\hat{\mathbf{x}}(n)$. Now, assuming that the deterministic part $\hat{\mathbf{x}}(n)$ is stationary and $\mathbf{e}(n)$ is independent and identically distributed over n, the likelihood of the observed set of vectors $\{\mathbf{x}(n)\}_{n=0}^{G-1}$ can be written as

$$\begin{aligned}p(\{\mathbf{x}(n)\};\boldsymbol{\theta}) &= \prod_{n=0}^{G-1} p(\mathbf{x}(n);\boldsymbol{\theta}) = \frac{1}{\pi^{MG}\det(\mathbf{Q})^G}\\ &\quad\times e^{-\sum_{n=0}^{G-1}\mathbf{e}^H(n)\mathbf{Q}^{-1}\mathbf{e}(n)}.\end{aligned} \tag{7}$$

Although the approach of splitting the signal into sub-vectors $\mathbf{x}(n)$ is inherently suboptimal since it ignores inter-vector dependencies, it is required in order to estimate signal and noise covariance matrices. Taking the logarithm of (7), we get the so-called log-likelihood function, i.e.,

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) &= \sum_{n=0}^{G-1} \ln p(\mathbf{x}(n);\boldsymbol{\theta}) = -GM\ln\pi\\ &\quad - G\ln\det(\mathbf{Q}) - \sum_{n=0}^{G-1}\mathbf{e}^H(n)\mathbf{Q}^{-1}\mathbf{e}(n).\end{aligned} \tag{8}$$

The maximum likelihood estimates of the parameters $\boldsymbol{\theta}$ are then $\arg\max \mathcal{L}(\boldsymbol{\theta})$. However, it is well-known that this will result in a preference for more complicated models, i.e. higher order. Instead, we find the parameters, with $\widehat{\cdot}$ denoting estimates, as

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} 2\mathcal{L}(\boldsymbol{\theta}) - \nu(|\boldsymbol{\theta}|, GM), \tag{9}$$

where $\nu(|\boldsymbol{\theta}|, GM)|$ is an order-dependent penalty term (see, e.g., [3]) and $|\boldsymbol{\theta}|$ is the number of parameters to be estimated. Here, we have used the minimum description length (MDL) for model order selection, which is identical to the Bayesian information criterion under certain conditions (see [10, 3]), i.e.,

$$\nu(|\boldsymbol{\theta}|, GM) = |\boldsymbol{\theta}| \ln(2GM). \tag{10}$$

Note that the factor 2 is due to the complex case being considered here. Furthermore, we have assumed that all model orders are equally probable. The Bayesian information criterion is consistent meaning that the probability of correct detection tends to 1 as the number of samples grow. The main difficulty in evaluating the cost function in (9) is that the fundamental frequency is a nonlinear parameter and that the noise covariance matrix $\mathbf{Q}$ generally is unknown. The amplitudes and phases, on the other hand, can be seen in (4) to be linear complex parameters that can easily be found given the fundamental frequency and the noise covariance matrix. Defining $\mathbf{Z}_k = \mathbf{Z}_k(0)$ and assuming that the phases of the harmonics are independent and uniformly distributed on the interval $(-\pi, \pi]$, the covariance matrix $\mathbf{R} \in \mathbb{C}^{M\times M}$ of the signal in (4) can be written as (see [3, 11]),

$$\mathbf{R} = \mathrm{E}\left\{\mathbf{x}(n)\mathbf{x}^H(n)\right\} = \mathbf{Z}_k\mathbf{P}_k\mathbf{Z}_k^H + \mathbf{Q} \tag{11}$$

where $\mathrm{E}\{\cdot\}$ denotes the statistical expectation and

$$\mathbf{P}_k = \mathrm{E}\left\{\mathbf{a}_k\mathbf{a}_k^H\right\} = \mathrm{diag}\left(\left[\begin{array}{ccc} A_1^2 & \cdots & A_k^2 \end{array}\right]\right). \tag{12}$$

Given the noise covariance matrix $\mathbf{Q}$, the complex amplitudes $\mathbf{a}_k$ can be found, for a certain candidate fundamental frequency, as

$$\hat{\mathbf{a}}_k = \left(\sum_{n=0}^{G-1}\mathbf{Z}_k^H(n)\mathbf{Q}^{-1}\mathbf{Z}_k(n)\right)^{-1}\sum_{n=0}^{G-1}\mathbf{Z}_k^H(n)\mathbf{Q}^{-1}\mathbf{x}(n), \tag{13}$$

which for the case where $\mathbf{Z}_k(n)$ does not depend on n can be interpreted as a weighted least-squares fit of the signal model to the sample mean vector $\frac{1}{G}\sum_{n=0}^{G-1}\mathbf{x}(n)$.

3. EFFICIENT COMPUTATIONS

We will now discuss how to construct a practical estimator. In particular, we will discuss how to evaulate the likelihood function (7) from observed data in a computationally efficient manner. First of all, the noise covariance matrix $\mathbf{Q}$ is unknown and an estimate has to be obtained. Here, we will do this based on the signal covariance model in (11). In practice, the signal covariance matrix in (11) too is unknown and is replaced by an estimate, the sample covariance matrix, i.e.,

$$\widehat{\mathbf{R}} = \frac{1}{G}\sum_{n=0}^{G-1}\mathbf{x}(n)\mathbf{x}^H(n). \tag{14}$$

There are some inherent tradeoffs in the choices of N and M and thereby G. The number of observations N should be chosen appropriately such that the signal $x(n)$ can be assumed to be stationary while M should be chosen such that all significant non-zero correlations in the covariance matrix are modeled. On the other hand,M should also not be chosen higher than strictly necessary, since the goodness of the signal covariance matrix estimate in (14) depends on G being as high as possible. Additionally, since the evaluation of the likelihood function requires the existence (and calculation) of the inverse of the noise covariance matrix, it is required that $\mathrm{rank}\left(\mathbf{Q}\right) = M$. This in turn implies that $G \geq M$ and hence $M \leq \frac{N}{2}$. For a particular candidate fundamental frequency, we can construct the order k Vandermonde matrix $\mathbf{Z}_k$ in (11). However, we also need an estimate of the sinusoidal amplitudes in $\mathbf{P}_k$ to obtain a noise covariance matrix estimate. An obvious approach is to use the estimated signal covariance matrix instead of the noise covariance matrix resulting in a Capon-like amplitude estimator or one of the other amplitude estimates proposed and discussed in [11]. However, in order to minimize the computational complexity, we here use the following asymptotically efficient estimate (see [11]) of the complex amplitudes for approximating the likelihood for various fundamental frequencies and orders:

$$\hat{\mathbf{a}}_k = \left(\sum_{n=0}^{G-1}\mathbf{Z}_k^H(n)\mathbf{Z}_k(n)\right)^{-1}\sum_{n=0}^{G-1}\mathbf{Z}_k^H(n)\mathbf{x}(n), \tag{15}$$

which can be approximated for large N as

$$\hat{\mathbf{a}}_k \approx \frac{1}{MG}\sum_{n=0}^{G-1}\mathbf{Z}_k^H(n)\mathbf{x}(n). \tag{16}$$

This is a number of phase-shifted FFTs. In fact, (16) can be evaluated using just one FFT of $x(n)$ for $n = 0, \ldots, N-1$. If desired, refined estimates may be obtained using one of the above estimators once the order and fundamental frequency have been estimated. Having found the complex amplitudes associated with a certain candidate fundamental frequency, we can now estimate the noise covariance matrix from the k'th order sinusoidal model as

$$\widehat{\mathbf{Q}}_k = \widehat{\mathbf{R}} - \mathbf{Z}_k\widehat{\mathbf{P}}_k\mathbf{Z}_k^H = \widehat{\mathbf{R}} - \sum_{l=1}^{k}\hat{A}_l^2\mathbf{z}_l\mathbf{z}_l^H, \tag{17}$$

with $\widehat{\mathbf{P}}_k = \mathrm{diag}\left(\left[\begin{array}{ccc} \hat{A}_1^2 & \cdots & \hat{A}_k^2 \end{array}\right]\right)$. The inverse noise covariance matrix is needed for the evaluation of the likelihood function, and direct inversion for different model orders and fundamental frequencies poses a computational burden. Therefore, it is advantagous to compute it using the matrix inversion lemma as follows:

$$\begin{aligned}\widehat{\mathbf{Q}}_k^{-1} &= \widehat{\mathbf{R}}^{-1} + \widehat{\mathbf{R}}^{-1}\mathbf{Z}_k\widehat{\mathbf{P}}_k^{\frac{1}{2}} \\ &\times\left(\mathbf{I} - \widehat{\mathbf{P}}_k^{\frac{1}{2}}\mathbf{Z}_k^H\widehat{\mathbf{R}}^{-1}\mathbf{Z}_k\widehat{\mathbf{P}}_k^{\frac{1}{2}}\right)^{-1}\widehat{\mathbf{P}}_k^{\frac{1}{2}}\mathbf{Z}_k^H\widehat{\mathbf{R}}^{-1},\end{aligned} \tag{18}$$

which is computationally less demanding than direct inversion of $\widehat{\mathbf{Q}}_k$ since $k < M$, or, iteratively for $k = 1, 2, \ldots$ as

$$\widehat{\mathbf{Q}}_k^{-1} = \widehat{\mathbf{Q}}_{k-1}^{-1} + \widehat{\mathbf{Q}}_{k-1}^{-1}\frac{\hat{A}_k^2\mathbf{z}_k\mathbf{z}_k^H}{1 - \hat{A}_k^2\mathbf{z}_k^H\widehat{\mathbf{Q}}_{k-1}^{-1}\mathbf{z}_k}\widehat{\mathbf{Q}}_{k-1}^{-1}, \tag{19}$$

with $\widehat{\mathbf{Q}}_0^{-1} = \widehat{\mathbf{R}}^{-1}$. Note that the stochastic parts of speech and audio signals are often modeled as auto-regressive Gaussian noise, in which case the matrix inversion may also be computed using Gohberg-Semencul's formula [3].

4. WHITE NOISE CASE

In some cases, the stochastic part of the signal model in (1) can be assumed to be white. Then, the computational complexity of the proposed estimator can be reduced significantly. First of all, the structure of the noise covariance matrix is now known, i.e., it becomes a scaled diagonal matrix $\mathbf{Q} = \sigma^2\mathbf{I}$ where σ_2 is the variance of the noise. This has the consequence that we no longer need to estimate a full covariance matrix but only the variance, and, therefore, there is no need to split the observed signal into sub-vectors, i.e., we can simply set $M = N$ and thus $G = 1$. For notational simplicity, we in the following set $\mathbf{e}(0) = \mathbf{e}$ and $\mathbf{x}(0) = \mathbf{x}$. The log-likelihood function can now be written as

$$\mathcal{L}(\boldsymbol{\theta}) = -N \ln \pi - 2N \ln \sigma - \frac{1}{\sigma^2}\|\mathbf{e}\|_2^2. \qquad (20)$$

Since the noise variance is generally unknown, we need to form an estimate of it. Like the noise covariance matrix estimate, this estimate will depend on the order k. For a particular fundamental frequency candidate and k harmonics, the maximum likelihood noise variance estimate is

$$\hat{\sigma}_k^2 = \frac{1}{N}\|\mathbf{x} - \mathbf{Z}_k\left(\mathbf{Z}_k^H\mathbf{Z}_k\right)^{-1}\mathbf{Z}_k^H\mathbf{x}\|_2^2, \qquad (21)$$

$$\approx \frac{1}{N}\|\mathbf{x} - \frac{1}{N}\mathbf{Z}_k\mathbf{Z}_k^H\mathbf{x}\|_2^2, \qquad (22)$$

where the last step follows from $N \gg 1$. We see that the amplitude once again can be found for various fundamental frequencies and orders using one FFT, this time, however, we are not ignoring the noise color. Inserting this variance estimate into (20), we get

$$\mathcal{L}(\boldsymbol{\theta}) = -N \ln \pi - 2N \ln \hat{\sigma}_k - N, \qquad (23)$$

which then has to be evaluated for different fundamental frequencies and orders. Now we can also clearly see the problem that was pointed out in Section 2. As the order k is increased, the log-likelihood too is increased and therefore the maximum likelihood estimator will lead to the choice of the highest possible order. Therefore, the log-likelihoods have to be combined with an order-dependent penalty terms as was done in (9). Specifically, we get the estimator

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} 2\mathcal{L}(\boldsymbol{\theta}) - |\boldsymbol{\theta}| \ln(2N). \qquad (24)$$

It is worth noting that, because the white noise case is so much simpler than the colored case, the estimator presented in this section may be preferred over the colored noise estimator in real-time applications even if the noise is known not to be completely white as it may still provide adequately accurate estimates.

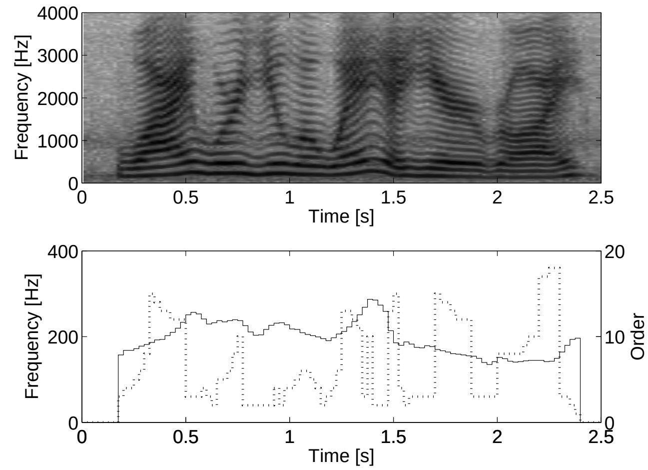

Fig. 1. Spectrogram of voiced speech signal in AR(2) noise (top) and estimated fundamental frequencies (solid) and orders (dotted) (bottom) for an SNR of 40 dB.

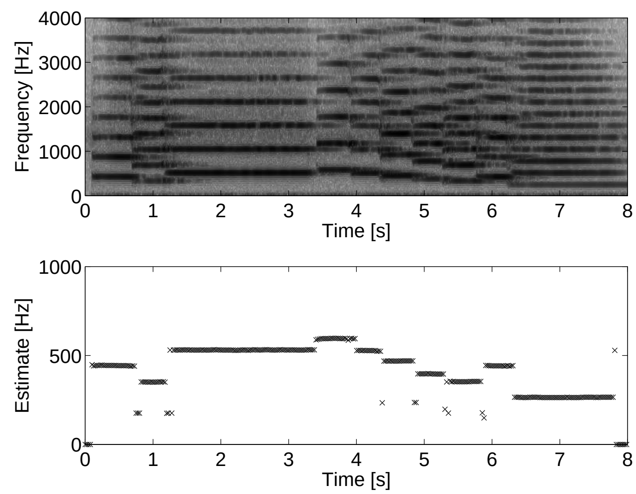

Fig. 2. Spectrogram of music signal, a trumpet, in AR(2) noise (top) and estimated fundamental frequencies (bottom) for an SNR of 40 dB.

5. SOME EXAMPLES

We start out by briefly reporting on some observations that we have made regarding the use of the colored noise estimator versus the much simpler white noise estimator. It was observed that the estimator based on the white noise assumption generally works well even for the colored noise case. This can be explained by some results on general sinusoidal frequency estimation for colored Gaussian noise; in [12] it was shown that the estimator based on the white noise assumption, the nonlinear least-squares estimator, is actually asymptotically efficient also for the colored noise case. Therefore, one would expect the same to hold for the fundamental frequency estimation problem. However, we have observed the order estimate to be erroneous when the white noise estimator is applied to colored noise. Particularly, the white noise estimator was observed to consistently report orders different from zero when no harmonic content could be observed in the signal. This can be explained from the following: the joint estimation of the signal model parameters and the order requires that the likelihoods be calculated and this depends on the inverse noise covariance matrix. We will now proceed to illustrate the application of the proposed estimator to a speech signal. Speech signals can be modeled using the model in (4) with the harmonics representing voiced speech and the colored noise representing unvoiced speech. In the top panel of Figure 1, a spectrogram of the speech signal in auto-regressive Gaussian noise, here a 2nd order auto-regressive process added at a signal-to-noise ratio (SNR) of 40 dB, is shown while at the bottom, the estimated fundamental frequency and the order is depicted. Here, the SNR is defined as $10\log_{10}(\bar{\sigma}^2/\sigma^2)$, with $\bar{\sigma}^2$ and σ^2 being the power of the speech and noise signals, respectively. When some of the mid-frequency harmonics are missing, or buried in noise, the estimator estimates a low order but the correct fundamental frequency. In Figure 2 the fundamental frequencies as estimated by our method are shown in the bottom panel for the music signal having the spectrogram in the top panel. As before, a 2nd order auto-regressive noise is added to the harmonic signal, a trumpet, with an SNR of 40 dB. As can be seen, the proposed method estimates the expected fundamental frequency in stationary regions. Interestingly, the method can be observed to fail in transition regions where, due to the impulse response of the room, multiple fundamental frequencies can be observed simultaneously. This can generally be attributed to the signal model in (4) being invalid for multiple harmonic sources.

6. CONCLUSION

A method for estimation of the parameters of a set of harmonically related sinusoids in colored Gaussian noise has been presented. The method, which is based on maximum likelihood, estimates the number of sinusoids, the fundamental frequency, the amplitude and phase of the individual harmonics, the noise covariance matrix and its inverse in a computationally efficient manner. Specifically, the inverse of the noise covariance matrix is computed recursively using the matrix inversion lemma. For the white noise case, the method has a very efficient implementation requiring only one FFT per segment. Examples of the application of the proposed method to analysis of speech and music signal have been given. These show that the fundamental frequency can be estimated in an accurate and robust manner using the proposed method.

7. REFERENCES

[1] R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34(4), pp. 744–754, Aug. 1986.

[2] E. B. George and M. J. T. Smith, "Analysis-by-synthesis/overlap-add sinusoidal modeling applied to the analysis-synthesis of musical tones," *J. Audio Eng. Soc.*, vol. 40(6), pp. 497–516, 1992.

[3] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Pearson Prentice Hall, 2005.

[4] T. D. Rossing, *The Science of Sound*, Addison-Wesley Publishing Company, 2nd edition, 1990.

[5] A. Nehorai and B. Porat, "Adaptive comb filtering for harmonic signal enhancement," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34(5), pp. 1124–1138, Oct. 1986.

[6] K. W. Chan and H. C. So, "Accurate frequency estimation for real harmonic sinusoids," *IEEE Signal Processing Lett.*, vol. 11(7), pp. 609–612, July 2004.

[7] H. Li, P. Stoica, and J. Li, "Computationally efficient parameter estimation for harmonic sinusoidal signals," *Signal Processing*, vol. 80, pp. 1937–1944, 2000.

[8] M. G. Christensen, A. Jakobsson, and S. H. Jensen, "Joint high-resolution fundamental frequency and order estimation," *IEEE Trans. on Audio, Speech and Language Processing*, Apr. 2006, submitted.

[9] M. G. Christensen, S. H. Jensen, S. V. Andersen, and A. Jakobsson, "Subspace-based fundamental frequency estimation," in *Proc. European Signal Processing Conf.*, 2004, pp. 637–640.

[10] P. M. Djuric, "Asymptotic MAP criteria for model selection," *IEEE Trans. Signal Processing*, vol. 46, pp. 2726–2735, Oct. 1998.

[11] P. Stoica, H. Li, and J. Li, "Amplitude estimation of sinusoidal signals: Survey, new results and an application," *IEEE Trans. Signal Processing*, vol. 48(2), pp. 338–352, Feb. 2000.

[12] P. Stoica, A. Jakobsson, and J. Li, "Cisiod parameter estimation in the coloured noise case: Asymptotic cramér-rao bound, maximum likelihood, and nonlinear least-squares," in *IEEE Trans. Signal Processing*, Aug. 1997, vol. 45(8), pp. 2048–2059.