A Single Snapshot Optimal Filtering Method for Fundamental Frequency Estimation

Jensen, Jesper Rindom; Christensen, Mads Græsbøll; Jensen, Søren Holdt

Published in:
I E E E International Conference on Acoustics, Speech and Signal Processing. Proceedings

DOI (link to publication from Publisher):
10.1109/ICASSP.2011.5947297

Publication date:
2011

Document Version
Accepted author manuscript, peer reviewed version

A SINGLE SNAPSHOT OPTIMAL FILTERING METHOD FOR FUNDAMENTAL FREQUENCY ESTIMATION

Jesper Rindom Jensen[†], *Mads Græsbøll Christensen*[‡] *and Søren Holdt Jensen*[†]

[†] Dept. of Electronic Systems, Aalborg University
Fredrik Bajers Vej 7, 9220, Aalborg, Denmark
email: {jrj,shj}@es.aau.dk

[‡] Dept. of Architecture, Design and Media Technology., Aalborg University
Niels Jernes Vej 14, 9220, Aalborg, Denmark
email: mgc@imi.aau.dk

ABSTRACT

Recently, optimal linearly constrained minimum variance (LCMV) filtering methods have been applied for fundamental frequency estimation. Like many other fundamental frequency estimators, these methods utilize the inverse covariance matrix. Therefore, the covariance matrix needs to be invertible which is typically ensured by using the sample covariance matrix involving data partitioning. The partitioning adversely affects the spectral resolution. We propose a novel optimal filtering method which utilizes the LCMV principle in conjunction with the iterative adaptive approach (IAA). The IAA enables us to estimate the covariance matrix from a single snapshot, i.e., without data partitioning. The experimental results show, that the performance of the proposed method is comparable or better than that of other competing methods in terms of spectral resolution.

***Index Terms*—** Fundamental frequency estimation, optimal filtering, iterative adaptive approach.

1. INTRODUCTION

There exists a multitude of signal processing applications in which the fundamental frequency is an essential parameter. A few examples are, e.g., parametric coding of audio and speech, automatic music transcription, musical genre classification, tuning of musical instruments, separation and enhancement of audio and speech sources, etc. Due to the importance of knowing the fundamental frequency, numerous of approaches and methods have been proposed for estimating this parameter. For a few examples of such estimators see, e.g., [1–7] and the references therein.

We will now introduce the problem of fundamental frequency estimation. The reasoning behind describing audio and speech signals by the fundamental frequency, among other parameters, is that audio and speech signals are quasi-periodic. That is, for a limited amount of signal samples, we can safely assume that for $n = 0, \ldots, N-1$

$$x(n) = \sum_{l=1}^{L} \alpha_l e^{jl\omega_0 n} + w(n) \,, \tag{1}$$

where L is the number of harmonics, $\alpha_l = A_l e^{j\phi_l}$ with $A_l > 0$ and ϕ_l denoting the real amplitude and the phase of the lth harmonic, ω_0 is the fundamental frequency and $w(n)$ is complex noise. We assume that the model order L is known, hence, the fundamental frequency estimation problem is to estimate ω_0 from (1). While not considered in this paper, the model order assumption can easily be avoided by using a model order estimator [8, 9] or even by doing the model order and fundamental frequency estimation jointly [7].

Many of the aforementioned fundamental frequency estimators (e.g., optimal filtering techniques and subspace-based methods) utilizes the covariance matrix inverse [7], hence, in such estimators the covariance matrix must be invertible. In consequence of that, the covariance matrix must be full-rank. Typically, this is ensured by using the sample covariance matrix

$$\hat{\mathbf{R}} = \frac{1}{N-M+1} \sum_{n=M-1}^{N-1} \mathbf{x}(n)\mathbf{x}^H(n) \,, \tag{2}$$

where $\mathbf{x}(n) = \begin{bmatrix} x(n) & \cdots & x(n-M+1) \end{bmatrix}^T$ and $M < \frac{N}{2} + 1$. It is well-known that the spectral resolution depends on the sample length. That is, the resolution is decreased by the data partitioning embedded in (2).

Recently, however, the iterative adaptive approach (IAA) was proposed [10, 11], which can be used for covariance matrix and spectrum estimation. There is no data partitioning in this method, i.e., the covariance matrix is estimated iteratively from only a single snapshot. In this paper, we will propose to use a covariance matrix estimate, obtained by using the IAA, in conjunction with an optimal filtering method for fundamental frequency estimation. Note that the IAA could be used in conjunction with other covariance based fundamental frequency estimators as well. Since our method operates on a single snapshot of data, we can expect that our proposed optimal filtering method has a higher spectral resolution compared to the optimal filtering method in [7].

The remainder of the paper is organized as follows. In Section 2, we briefly review the optimal filtering method for fundamental frequency estimation and propose to use it in conjunction with the IAA. In Section 3, we present some experimental results obtained from quantitative experiments. Finally, in Section 4 we conclude on our work.

2. OPTIMAL FILTERING METHOD UTILIZING THE ITERATIVE ADAPTIVE APPROACH

2.1. Fundamental Frequency Estimation using Optimal Filtering

First, we will briefly review the concept of using an optimal filtering method for fundamental frequency estimation. This concept was introduced in [12] and is based on an optimal harmonic LCMV (hLCMV) filter. Consider M time-reversed samples from (1) in vec-

tor format

$$\mathbf{x}(n) = \begin{bmatrix} x(n) & x(n-1) & \cdots & x(n-M+1) \end{bmatrix}^T , \quad (3)$$

for $n = M-1, \ldots, N-1$. We introduce the FIR filter $\mathbf{h} = \begin{bmatrix} h(0) & \cdots & h(M-1) \end{bmatrix}^H$, from which the output is given by

$$y(n) = \mathbf{h}^H \mathbf{x}(n) . \quad (4)$$

The output power of the filter is the defined as

$$\mathrm{E}\{|y(n)|^2\} = \mathbf{h}^H \mathbf{R} \mathbf{h} , \quad (5)$$

where $\mathbf{R} = \mathrm{E}\{\mathbf{x}(n)\mathbf{x}^H(n)\}$. The optimal filter response is found, by using the LCMV principle. That is, we design the filter to have a unit gain at the harmonic frequencies while having maximum noise suppression

$$\min_{\mathbf{h}} \mathbf{h}^H \mathbf{R} \mathbf{h} \quad \text{s.t. } \mathbf{h}^H \mathbf{z}(l\omega_0) = 1 , \quad \text{for } l = 1, \ldots, L, \quad (6)$$

where $\mathbf{z}(\omega) = \begin{bmatrix} 1 & e^{-j\omega_0} & \cdots & e^{-j(M-1)\omega_0} \end{bmatrix}^T$. The well-know solution to this optimization problem is

$$\hat{\mathbf{h}} = \mathbf{R}^{-1}\mathbf{Z}(\omega_0) \left(\mathbf{Z}(\omega_0)^H \mathbf{R}^{-1} \mathbf{Z}(\omega_0) \right)^{-1} \mathbf{1} , \quad (7)$$

with $\mathbf{Z}(\omega_0) = \begin{bmatrix} \mathbf{z}(\omega_0) & \cdots & \mathbf{z}(L\omega_0) \end{bmatrix}$. We can then obtain an estimate of the fundamental frequency by inserting (7) into (5) and maximize the output power as

$$\hat{\omega}_0 = \arg\max_{\omega_0} \mathbf{1}^H \left(\mathbf{Z}^H(\omega_0) \mathbf{R}^{-1} \mathbf{Z}(\omega_0) \right)^{-1} \mathbf{1} . \quad (8)$$

The covariance matrix $\mathbf{R}$ is replaced by (2). Recall, that for $\mathbf{R}$ to be invertible, it is required that $M < \frac{N}{2} + 1$. In this paper, we propose instead to use a covariance matrix estimate obtained by using the iterative adaptive approach. In this method, the covariance matrix can be estimated from a single snapshot, i.e., we can obtain an $N \times N$ covariance matrix estimate.

2.2. Covariance Matrix Estimation using the Iterative Adaptive Approach

The iterative adaptive approach (IAA), proposed in [11], is a method for estimating the spectral amplitudes. In the estimation procedure, a WLS cost-function [13] is minimized

$$\hat{\alpha}_k = \arg\min_{\alpha_k} \left(\mathbf{x}(n) - \alpha_k \mathbf{z}(\omega_k) \right)^H \mathbf{Q}^{-1}(\omega_k) \left(\mathbf{x}(n) - \alpha_k \mathbf{z}(\omega_k) \right) \quad (9)$$

where $\mathbf{Q}(\omega_k)$ is the noise covariance matrix defined as

$$\mathbf{Q}(\omega_k) = \mathbf{R} - |\alpha_k|^2 \mathbf{z}(\omega_k) \mathbf{z}^H(\omega_k) . \quad (10)$$

In the IAA, the covariance matrix is approximated by the well-known covariance matrix model [9]

$$\tilde{\mathbf{R}} = \bar{\mathbf{Z}}(\boldsymbol{\omega}) \hat{\mathbf{P}} \bar{\mathbf{Z}}^H(\boldsymbol{\omega}) , \quad (11)$$

where $\boldsymbol{\omega} = \begin{bmatrix} 0 & 2\pi\frac{1}{K} & \cdots & 2\pi\frac{K-1}{K} \end{bmatrix}$ is the K-point frequency grid. The matrices $\bar{\mathbf{Z}}(\boldsymbol{\omega})$ and $\hat{\mathbf{P}}$ are defined as

$$\bar{\mathbf{Z}}(\boldsymbol{\omega}) = \begin{bmatrix} \mathbf{z}(\boldsymbol{\omega}(0)) & \cdots & \mathbf{z}(\boldsymbol{\omega}(K-1)) \end{bmatrix} \quad (12)$$

$$\hat{\mathbf{P}} = \mathrm{diag}\left\{ \begin{bmatrix} |\hat{\alpha}_0|^2 & \cdots & |\hat{\alpha}_{K-1}|^2 \end{bmatrix}^T \right\} , \quad (13)$$

$\hat{\alpha}_k = \dfrac{\mathbf{z}^H(\boldsymbol{\omega}(k))\mathbf{x}(n)}{N} , \qquad k = 0, \ldots, K-1$

repeat

$\tilde{\mathbf{R}} = \bar{\mathbf{Z}}(\boldsymbol{\omega}) \hat{\mathbf{P}} \bar{\mathbf{Z}}^H(\boldsymbol{\omega})$

  for $k = 0, \ldots, K-1$

    $\hat{\alpha}_k = \dfrac{\mathbf{z}^H(\boldsymbol{\omega}(k))\tilde{\mathbf{R}}^{-1}\mathbf{x}(n)}{\mathbf{z}^H(\boldsymbol{\omega}(k))\tilde{\mathbf{R}}^{-1}\mathbf{z}(\boldsymbol{\omega}(k))}$

    $\hat{P}_k = |\hat{\alpha}_k|^2$

  end

until (convergence)

$\hat{\omega}_0 = \arg\max_{\omega_0} \mathbf{1}^H \left(\mathbf{Z}^H(\omega_0) \tilde{\mathbf{R}}^{-1} \mathbf{Z}(\omega_0) \right)^{-1} \mathbf{1}$

Table 1. The optimal filtering method for ω_0 estimation utilizing the IAA

where $|\hat{\alpha}_k|^2 = \hat{P}_k$. Minimizing (9) with respect to α_k yields

$$\hat{\alpha}_k = \frac{\mathbf{z}^H(\omega_k)\mathbf{Q}^{-1}(\omega_k)\mathbf{x}(n)}{\mathbf{z}^H(\omega_k)\mathbf{Q}^{-1}(\omega_k)\mathbf{z}(\omega_k)} . \quad (14)$$

By using the matrix inversion lemma it turns out that we can simplify (14) as

$$\hat{\alpha}_k = \frac{\mathbf{z}^H(\omega_k)\tilde{\mathbf{R}}^{-1}\mathbf{x}(n)}{\mathbf{z}^H(\omega_k)\tilde{\mathbf{R}}^{-1}\mathbf{z}(\omega_k)} . \quad (15)$$

Note, however, that to estimate the covariance matrix using (11), we need an estimate of the spectral amplitudes (15) and vice versa. The estimation is therefore performed iteratively initialized by the periodogram estimate. For most applications, 15 iterations is enough [11].

2.3. Proposed Optimal Filtering Method Utilizing the Iterative Adaptive Approach

In the proposed filtering method, we use the filter design in (7) where we replace the covariance matrix with the estimate in (11). This result in the optimal harmonic IAA (hIAA) filter

$$\tilde{\mathbf{h}} = \tilde{\mathbf{R}}^{-1}\mathbf{Z}(\omega_0) \left(\mathbf{Z}(\omega_0)^H \tilde{\mathbf{R}}^{-1} \mathbf{Z}(\omega_0) \right)^{-1} \mathbf{1} . \quad (16)$$

We could also use the noise covariance matrix $\mathbf{Q}$ instead of $\mathbf{R}$ in (16) which is intuitively more correct, i.e.,

$$\tilde{\tilde{\mathbf{h}}} = \tilde{\mathbf{Q}}^{-1}(\omega_0)\mathbf{Z}(\omega_0) \left(\mathbf{Z}(\omega_0)^H \tilde{\mathbf{Q}}^{-1}(\omega_0) \mathbf{Z}(\omega_0) \right)^{-1} \mathbf{1} . \quad (17)$$

We can write the IAA-based noise covariance matrix estimate as

$$\tilde{\mathbf{Q}}(\omega_0) = \tilde{\mathbf{R}} - \mathbf{Z}(\omega_0)\hat{\mathbf{P}}_s \mathbf{Z}^H(\omega_0) , \quad (18)$$

where $\hat{\mathbf{P}}_s$ is a diagonal matrix containing the estimated powers of the harmonics. By making use of the matrix inversion lemma, it can then be shown that

$$\tilde{\mathbf{h}} = \tilde{\tilde{\mathbf{h}}} . \quad (19)$$

Since the two filter designs are identical for the problem at hand, we will just use the design in (16) which is simpler. In Table 1 it is shown how we can use the optimal hIAA filter to estimate the fundamental frequency.

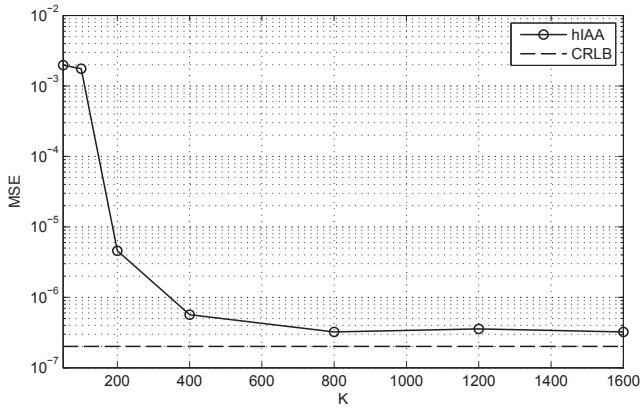

Fig. 1. Fundamental frequency estimation MSE as a function of the grid size used in estimation of the covariance matrix with $N = 40$.

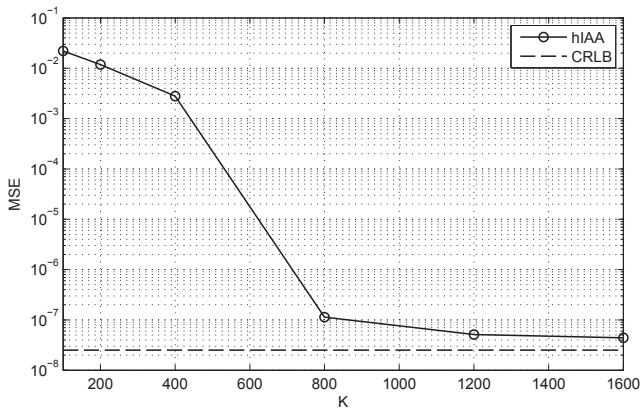

Fig. 2. Fundamental frequency estimation MSE as a function of the grid size used in estimation of the covariance matrix with $N = 80$.

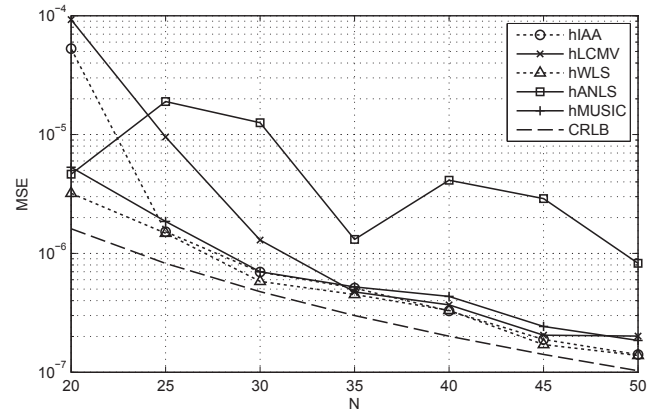

Fig. 3. Fundamental frequency estimation MSE as a function of the sample length N.

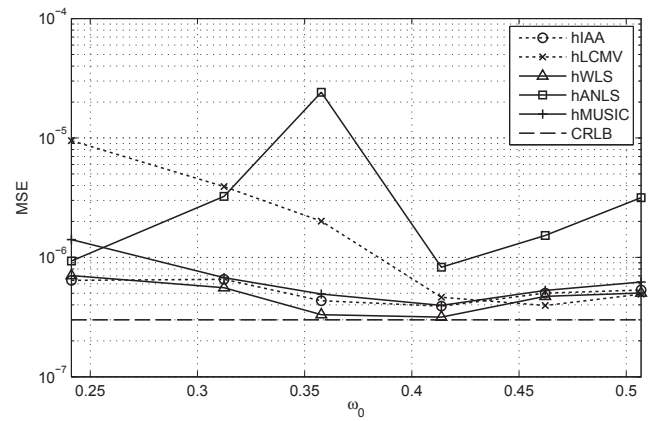

Fig. 4. Fundamental frequency estimation MSE as a function of the fundamental frequency ω_0.

As it can be seen, the estimate is obtained by maximizing the expected filter output power over a set of candidate frequencies. If a fine estimate is required, a relatively coarse set of candidate frequencies can be chosen whereupon the coarse fundamental frequency estimate is refined using a gradient search. The gradient, needed in that respect, is given by

$$g_{\omega_0} = -2\mathrm{Re}\{\mathbf{1}^H(\mathbf{Z}^H\tilde{\mathbf{R}}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^H\tilde{\mathbf{R}}\mathbf{Y}(\mathbf{Z}^H\tilde{\mathbf{R}}^{-1}\mathbf{Z})^{-1}\mathbf{1}\}\ , \quad (20)$$

where $[\mathbf{Y}]_{pq} = \left[\frac{\partial}{\partial\omega}\mathbf{Z}\right]_{pq} = -j(p-1)qe^{-j\omega_0 q(p-1)}$.

3. EXPERIMENTAL RESULTS

In this section, we describe the experimental evaluation of the proposed method. Note that in all simulations we estimate the fundamental frequency over a relatively coarse grid and refine the estimate using (20) in a steepest-descent algorithm with exact line search. First, we investigated how to choose the frequency grid size when estimating the covariance matrix using (11). To investigate this, we performed a series of Monte-Carlo simulations where we varied the frequency grid size. For each grid size we conducted 500 Monte-Carlo simulations. To evaluate the average error of doing the discretization in (11), we chose a random fundamental frequency in all simulations for a certain grid size. The random fundamental frequency was sampled from a uniform distribution $\mathcal{U}(0.4, 0.5)$. The model order was set to $L = 3$, the sample length was $N = 40$ and the SNR, defined as

$$\mathrm{SNR} = 10\log_{10}\frac{\sum_{l=1}^{L}|\alpha_l|^2}{\sigma_w^2}\ , \quad (21)$$

was 20 dB (σ_w^2 is the noise variance). The results from this series of simulations are shown in Fig. 1. From the results it can be seen that for this particular setup, a grid size of $K \approx 600$ frequency points is enough. Note also, that the MSE is following but not reaching the Cramér-Rao lower bound (CRLB). This is common, however, for the inverse covariance based methods [7]. The depicted CRLB is the asymptotic CRLB [14]

$$\mathrm{CRLB}(\omega_0) \approx \frac{6\sigma_w^2}{N^3\sum_{l=1}^{L}A_l^2 l^2}\ . \quad (22)$$

The same simulations were conducted when $N = 80$ and the results from these simulations are depicted in Fig. 2. For the case with $N = 80$, $K \approx 1000$ is enough. The important thing to note is, that when we increase the number of samples N we also need to increase the number of frequency grid points K, to achieve the maximum possible performance.

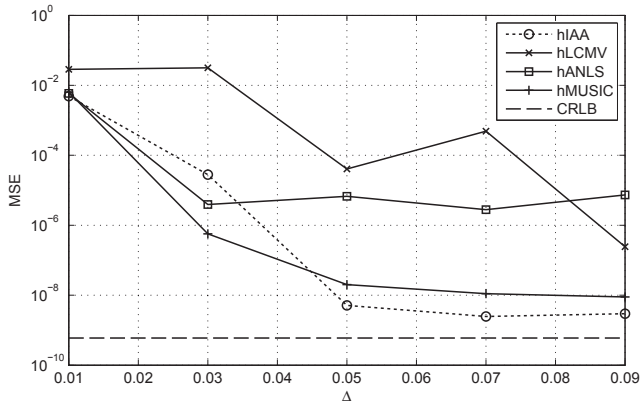

Fig. 5. Fundamental frequency estimation MSE as a function of the fundamental frequency spacing Δ in a two-source scenario.

We also compared the proposed method with the harmonic WLS (hWLS) [1], the harmonic LCMV (hLCMV) [7], the harmonic approximate NLS (hANLS) [7], and the harmonic MUSIC (hMUSIC) methods [7]. For example, we compared the methods for different sample lengths. For each sample length we conducted 500 Monte-Carlo simulations and in each simulation ω_0 was sampled randomly from $\mathcal{U}(0.42, 0.43)$. The remaining setup was: $L = 3$, SNR = 20 dB and $K = 2000$. The results from this series of simulations are shown in Fig. 3. First, we note that the hANLS method shows an erratic behaviour for these small sample lengths and is thereby outperformed by the other methods. The hIAA method outperforms the hLCMV method for all Ns, which is also expected since it has more degrees of freedom in the filter. Finally, we note that hMUSIC and hWLS performs best for $N < 25$ while for $N \geq 25$ hIAA, hWLS and hMUSIC show the same performance. Also we note, that for high Ns all methods seem to closely follow the CRLB. Then we compared the methods for different values of the fundamental frequency. A series of Monte-Carlo simulations were conducted with 500 simulations for each fundamental frequency. In each simulation K was sampled randomly from $\mathcal{U}_d(2000, 3000)$ ($\mathcal{U}_d(x_1, x_2)$ is the discrete uniform distribution taking integer values in the interval from x_1 to x_2). The remaining set up was: $N = 35$, $L = 3$ and SNR = 20 dB. The results from this experiment are depicted in Fig. 4. Again we note that hANLS is unreliable for the given setup. The hIAA shows an improvement compared to hLCMV for $\omega_0 < 0.4$. For low fundamental frequencies ($\omega_0 < 0.3$), hIAA and hWLS outperforms the other methods, while for $\omega_0 > 0.4$ all methods except hANLS show the same performance. The results indicate that the proposed method (along with hWLS and hMUSIC) has a better spectral resolution than hLCMV. Finally, we compared hIAA, hLCMV, hANLS and hMUSIC in a scenario with two harmonic sources. The two sources both had $L = 3$ harmonics each with unit amplitudes. We then conducted a series of Monte-Carlo simulations for different spacings of the fundamental frequencies of the two sources (500 simulations for each frequency spacing). In each simulation, the number of samples was $N = 80$ and the SNR was 40 dB. The results from these simulations are shown in Fig. 5. For $\Delta > 0.05$ the proposed method clearly outperforms the other methods.

4. CONCLUSION

In this paper, we proposed a new optimal filtering method for estimating the fundamental frequency of a (quasi-)periodic signal. The proposed method is an optimal LCMV filtering method which operates on single data snapshot. This is possible, because we estimate the covariance matrix using the iterative adaptive approach (IAA). By filtering on a single data snapshot rather than having to partition the data vector as in the filtering methods in [7], we obtain a better spectral resolution. The claim on increased spectral resolution was supported by the simulation results. The results showed that for small numbers of samples, low fundamental frequencies, and small frequency spacings in a two-source scenario, the proposed method clearly outperforms the optimal LCMV filtering method in [7]. This was also expected since the proposed method is an improvement of this method. Furthermore, for small number of samples and low frequencies, the proposed methods performance is comparable with that of the harmonic MUSIC and harmonic WLS methods. In a two-source scenario it outperforms all the methods in the comparison above the frequency spacing threshold.

5. REFERENCES

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

[2] K. W. Chan and H. C. So, "Accurate frequency estimation for real harmonic sinusoids," *IEEE Signal Process. Lett.*, vol. 11, no. 7, pp. 609–612, 2004.

[3] A. de Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *J. Acoust. Soc. Am.*, vol. 111, no. 4, pp. 1917–1930, 2002.

[4] V. Emiya, B. David, and R. Badeau, "A parametric method for pitch estimation of piano tones," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 2007, vol. 1, pp. 249–252.

[5] S. Godsill and M. Davy, "Bayesian harmonic models for musical pitch estimation and analysis," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 13–17 May 2002, vol. 2, pp. 1769–1772.

[6] M. G. Christensen, P. Stoica, A. Jakobsson, and S. H. Jensen, "Multi-pitch estimation," *Elsevier Signal Processing*, vol. 88(4), pp. 972–983, 2008.

[7] M. G. Christensen and A. Jakobsson, "Multi-pitch estimation," *Synthesis Lectures on Speech and Audio Processing*, vol. 5, no. 1, pp. 1–160, 2009.

[8] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, 2004.

[9] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Pearson Education, Inc., 2005.

[10] L. Du, T. Yardibi, J. Li, and P. Stoica, "Review of user parameter-free robust adaptive beamforming algorithms," *Digital Signal Processing*, vol. 19, no. 4, pp. 567 – 582, 2009.

[11] T. Yardibi, J. Li, P. Stoica, M. Xue, and A. B. Baggeroer, "Source localization and sensing: A nonparametric iterative adaptive approach based on weighted least squares," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 1, pp. 425–443, Jan 2010.

[12] M. G. Christensen, J. H. Jensen, A. Jakobsson, and S. H. Jensen, "On optimal filter designs for fundamental frequency estimation," *IEEE Signal Process. Lett.*, vol. 15, pp. 745–748, 2008.

[13] J. Li and P. Stoica, "An adaptive filtering approach to spectral estimation and SAR imaging," *IEEE Trans. Signal Process.*, vol. 44, no. 6, pp. 1469–1484, jun 1996.

[14] M. G. Christensen, A. Jakobsson, and S. H. Jensen, "Joint high-resolution fundamental frequency and order estimation," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 15, no. 5, pp. 1635–1644, 2007.