Fundamental Frequency Estimation using the Shift-Invariance Property

Christensen, Mads Græsbøll; Jakobsson, Andreas; Jensen, Søren Holdt

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

Publication date:
2007

Document Version
Accepted author manuscript, peer reviewed version

FUNDAMENTAL FREQUENCY ESTIMATION USING THE SHIFT-INVARIANCE PROPERTY

Mads Græsbøll Christensen[∗†], *Andreas Jakobsson*[‡], *and Søren Holdt Jensen*[†]

[†] Dept. of Electronic Systems
Aalborg University, Denmark
{mgc,shj}@es.aau.dk

[‡] Dept. of Electrical Engineering
Karlstad University, Sweden
andreas.jakobsson@ieee.org

ABSTRACT

In this paper, we propose a method for the estimation of the fundamental frequency of a periodic waveform based on the shift-invariance property of the sinusoidal signal model known from ESPRIT. An advantage of the proposed method is that the model order and the fundamental frequency can be found jointly and that the method does not depend on the density of the observation noise but rather on it being approximately white. Also, the cost function is observed to very smooth as compared to that of MUSIC. The method is shown in simulations to have good performance with the root mean square estimation error approaching the Cramér-Rao lower bound.

1. INTRODUCTION

In subspace methods, the full space is divided into a subspace known as the signal subspace that spans the space of the signal of interest, and its orthogonal complement the noise subspace. The properties of these subspaces are then exploited for various estimation and identification tasks. Subspace methods have a rich history in sinusoidal parameter estimation and are among the most eloquent estimators available today. Especially for the estimation of sinusoidal frequencies, a problem occurring in spectral estimation and direction-of-arrival problems in array processing, these methods have proven successful during the past three decades. Perhaps the most prominent subspace methods for frequency estimation are the MUSIC (MUltiple SIgnal Classification) method [1, 2] and the ESPRIT (Estimation of Signal Parameters by Rotational Invariance Techniques) method of [3] while the earliest example of such methods is perhaps [4]. Fundamental frequency estimation, i.e., the problem of estimating the fundamental frequency of a set of harmonically related sinusoids, is an important component in many speech and audio processing systems. Until recently, however, this problem has not received much attention in the literature on subspace-based estimation. In [5], a fundamental frequency estimator based on the subspace orthogonality property of MUSIC was proposed and its application to analysis of speech and audio signals was demonstrated. In later publications, this method was extended to the multi-pitch case [6, 7]. The fundamental frequency estimation problem can be defined as follows. A signal consisting of a set of harmonically related complex sinusoids in additive white complex circularly symmetric noise, $w(n)$, for $n = 0, \ldots, N-1$, is considered, i.e.,

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

where $A_l > 0$ and ϕ_l are the amplitude and the phase of the l'th harmonic, respectively. The task at hand is then to estimate the fundamental frequency ω_0, or, equivalently, the pitch period, from a set of N measured samples, $x(n)$. We propose a new method for fundamental frequency estimation. It is a subspace method that exploits the structure of the signal subspace to obtain the fundamental frequency estimate. More specifically, the shift-invariance property known from the ESPRIT algorithm [3] is used. The proposed method is computationally simpler than that of [5] and the associated cost function is very simple and much smoother. The performance of the method is assessed in Monte Carlo simulations comparing the root mean square estimation error (RMSE) to those of MUSIC [5], the related WLS fitting method [8], and the Cramér-Rao lower bound for the fundamental frequency estimation, found in [9] for the real case and in [5] for the complex case. We remark that while we have assumed that the model order L is known, the method presented in this paper can be extended to include joint fundamental frequency and order estimation in a straight-forward manner as described in [5] using the principles of [10].

The remaining part of this paper is organized as follows. In Section 2, the covariance matrix model that forms the basis of this paper is briefly described. Then, the proposed estimator is presented in Section 3. In Section 4, numerical

[∗]The work of M. G. Christensen is supported by the Parametric Audio Processing project, Danish Research Council for Technology and Production Sciences grant no. 274–06–0521.

results are presented before the conclusions in Section 5.

2. COVARIANCE MATRIX MODEL

We start out by defining $\mathbf{x}(n)$ as a signal sub-vector containing M samples of the observed signal, i.e.,

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

with $(\cdot)^T$ denoting the transpose. Assuming that the phases $\{\phi_l\}$ are independent and uniformly distributed, the covariance matrix $\mathbf{R} \in \mathbb{C}^{M\times M}$ of the signal in (1) can be written as

$$\mathbf{R} = \mathrm{E}\left\{\mathbf{x}(n)\mathbf{x}^H(n)\right\} = \mathbf{A}\mathbf{P}\mathbf{A}^H + \sigma^2\mathbf{I}_M, \qquad (3)$$

where $\mathrm{E}\{\cdot\}$ and $(\cdot)^H$ denote the statistical expectation and the conjugate transpose. We here require that the sub-vector length M is chosen such that $L < M$. Furthermore, $\mathbf{P} = \mathrm{diag}\left([\, A_1^2\ \cdots\ A_L^2\,]\right)$, and $\mathbf{A} \in \mathbb{C}^{M\times L}$ a full rank Vandermonde matrix defined as

$$\mathbf{A} = \left[\ \mathbf{a}(\omega_0)\ \ \cdots\ \ \mathbf{a}(\omega_0 L)\ \right], \qquad (4)$$

where $\mathbf{a}(\omega) = [\, 1\ e^{j\omega}\ \cdots\ e^{j\omega(M-1)}\,]^T$. Also, σ^2 denotes the variance of the additive noise, $w(n)$, and $\mathbf{I}_M$ is the $M\times M$ identity matrix. We note that $\mathbf{A}\mathbf{P}\mathbf{A}^H$ has rank L. Let

$$\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H \qquad (5)$$

be the eigenvalue decomposition (EVD) of the covariance matrix. Then, $\mathbf{U}$ contains the M orthonormal eigenvectors of $\mathbf{R}$, i.e., $\mathbf{U} = \left[\ \mathbf{u}_1\ \ \cdots\ \ \mathbf{u}_M\ \right]$ and $\mathbf{\Lambda}$ is a diagonal matrix containing the corresponding eigenvalues, λ_k, with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_M$. The subspace-based methods are based on a partitioning of the eigenvectors into a set belonging to the signal subspace spanned by the columns of $\mathbf{A}$ and an orthogonal complement known as the noise subspace. Let $\mathbf{S}$ be formed from the eigenvectors corresponding to the L most significant eigenvalues, i.e.,

$$\mathbf{S} = \left[\ \mathbf{u}_1\ \ \cdots\ \ \mathbf{u}_L\ \right]. \qquad (6)$$

The subspace that is spanned by the columns of $\mathbf{S}$ we denote $\mathcal{R}(\mathbf{S})$ and is henceforth referred to as the signal subspace. Similarly, let $\mathbf{G}$ be formed from the eigenvectors corresponding to the $M-L$ least significant eigenvalues, i.e.,

$$\mathbf{G} = \left[\ \mathbf{u}_{L+1}\ \ \cdots\ \ \mathbf{u}_M\ \right], \qquad (7)$$

where $\mathcal{R}(\mathbf{G})$ is referred to as the noise subspace. Using the EVD in (5), the covariance matrix model in (3) can now be written as $\mathbf{U}\left(\mathbf{\Lambda} - \sigma^2\mathbf{I}_M\right)\mathbf{U}^H = \mathbf{A}\mathbf{P}\mathbf{A}^H$. Introducing

$$\mathbf{\Lambda}_S = \mathrm{diag}\left([\, \lambda_1 - \sigma^2\ \cdots \lambda_L - \sigma^2\,]\right)$$

we can write

$$\mathbf{S}\mathbf{\Lambda}_S\mathbf{S}^H = \mathbf{A}\mathbf{P}\mathbf{A}^H. \qquad (8)$$

From this equation, it can be seen that the columns of $\mathbf{A}$ span the same space as the columns of $\mathbf{S}$ and that $\mathbf{A}$ therefore also must be orthogonal to $\mathbf{G}$, i.e., $\mathcal{R}(\mathbf{S}) = \mathcal{R}(\mathbf{A})$ and $\mathcal{R}(\mathbf{G}) \perp \mathcal{R}(\mathbf{A})$.

3. PROPOSED ESTIMATOR

By post-multiplying (8) by $\mathbf{S}$, the following relation between the signal subspace eigenvectors and the Vandermonde matrix can be established (see [11]):

$$\mathbf{S} = \mathbf{A}\mathbf{C} \qquad (9)$$

with $\mathbf{C} = \mathbf{P}\mathbf{A}^H\mathbf{S}\mathbf{\Lambda}_S^{-1}$. Next, we define matrices the $\mathbf{A}_1$ and $\mathbf{A}_2$, constructed by removing the last and first rows of $\mathbf{A}$, i.e., $\mathbf{A}_1 = [\,\mathbf{I}_{M-1}\ \mathbf{0}\,]\,\mathbf{A}$ and $\mathbf{A}_2 = [\,\mathbf{0}\ \mathbf{I}_{M-1}\,]\,\mathbf{A}$. Similarly, we define from $\mathbf{S}$, $\mathbf{S}_1 = [\,\mathbf{I}_{M-1}\ \mathbf{0}\,]\,\mathbf{S}$ and $\mathbf{S}_2 = [\,\mathbf{0}\ \mathbf{I}_{M-1}\,]\,\mathbf{S}$. From these definitions and (9), the matrices $\mathbf{S}_1$ and $\mathbf{A}_1$ can be related through the matrix $\mathbf{C}$ as $\mathbf{S}_1 = \mathbf{A}_1\mathbf{C}$. Then, due to the particular structure of $\mathbf{A}$ known as the shift-invariance property, the following can be seen to hold:

$$\mathbf{A}_2 = \mathbf{A}_1\mathbf{D} \quad \text{and} \quad \mathbf{S}_2 = \mathbf{S}_1\mathbf{\Gamma}, \qquad (10)$$

with $\mathbf{D} = \mathrm{diag}\left([\, e^{j\omega_0}\ \cdots\ e^{j\omega_0 L}\,]\right)$. Then, the matrix relating $\mathbf{S}_1$ to $\mathbf{S}_2$ can be written as follows:

$$\mathbf{\Gamma} = \mathbf{C}^{-1}\mathbf{D}\mathbf{C}. \qquad (11)$$

Thus, $\mathbf{\Gamma}$ and $\mathbf{D}$ are related through a similarity transform.

The ESPRIT algorithm [3] is based on $\mathcal{R}(\mathbf{S}) = \mathcal{R}(\mathbf{A})$ and the so-called shift-invariance property in (10) of the matrix $\mathbf{A}$. In practice, the expectation operator in (3) is replaced by a finite sum and the right relation in (10) holds only approximately and the underlying assumptions may only be approximations of the observed phenomenon. Consequently, the sinusoidal parameters are found by constructing the matrices $\mathbf{S}_1$ and $\mathbf{S}_2$ and then solving for $\mathbf{\Gamma}$ in $\mathbf{S}_2 \approx \mathbf{S}_1\mathbf{\Gamma}$ in some sense. For instance,

$$\widehat{\mathbf{\Gamma}} = \arg\min_{\mathbf{\Gamma}} \|\mathbf{S}_2 - \mathbf{S}_1\mathbf{\Gamma}\|_F^2 = \left(\mathbf{S}_1^H\mathbf{S}_1\right)^{-1}\mathbf{S}_1^H\mathbf{S}_2, \qquad (12)$$

or using total least-squares. The sinusoidal frequencies are then found from the empirical EVD of $\widehat{\mathbf{\Gamma}}$ via the relation in (11), i.e,

$$\widehat{\mathbf{\Gamma}} = \mathbf{Q}\widehat{\mathbf{D}}\mathbf{Q}^{-1} \qquad (13)$$

with $\mathbf{Q}$ containing the empirical eigenvectors of $\widehat{\Gamma}$ and

$$\widehat{\mathbf{D}} = \mathrm{diag}\left([\, e^{j\hat{\omega}_1}\ \cdots\ e^{j\hat{\omega}_L}\,]\right), \qquad (14)$$

where $\{\hat{\omega}_l\}_{l=1}^{L}$ is a set of unconstrained frequencies. It is not clear how to estimate the fundamental frequency from these equations since the eigenvalues are not constrained to be equally spaced on the unit circle. We proceed as follows. We here assume that the eigenvalues and eigenvectors in (13) are ordered by increasing arguments, i.e., $\hat{\omega}_1 \leq \ldots \leq \hat{\omega}_L$. Using the shift-invariance property in (10) and (11), we can write $\mathbf{S}_2 = \mathbf{S}_1\mathbf{C}^{-1}\mathbf{D}\mathbf{C}$, and thus

$$\mathbf{S}_2 \approx \mathbf{S}_1\mathbf{Q}\widehat{\mathbf{D}}\mathbf{Q}^{-1}. \tag{15}$$

Defining the diagonal matrix containing the unknown fundamental frequency as

$$\bar{\mathbf{D}} = \mathrm{diag}\left([\, e^{j\omega_0} \ \cdots \ e^{j\omega_0 L} \,]\right) \tag{16}$$

we define a cost function

$$J \triangleq \|\mathbf{S}_2 - \mathbf{S}_1\mathbf{Q}\bar{\mathbf{D}}\mathbf{Q}^{-1}\|_F^2, \tag{17}$$

from which the fundamental frequency can be estimated as

$$\hat{\omega}_0 = \arg\min_{\omega_0} J, \tag{18}$$

where only $\bar{\mathbf{D}}$ depends on ω_0. Note that also the order L can be estimated using (17) (see, e.g., [10]). We see from (10) that in the ideal case, we have equality in (15). So, instead we may introduce the modified cost function as

$$J \triangleq \|\mathbf{S}_2\mathbf{Q} - \mathbf{S}_1\mathbf{Q}\bar{\mathbf{D}}\|_F^2 \tag{19}$$

$$= \|\mathbf{V} - \mathbf{W}\bar{\mathbf{D}}\|_F^2 \tag{20}$$

with obvious definitions. The minimization of this norm generally is not equivalent to minimizing (17) since $\mathbf{Q}$ is not orthogonal. The cost function in can be rewritten as follows (20)

$$J = -2\,\mathrm{Re}\left(\mathrm{Tr}\,\mathbf{V}\bar{\mathbf{D}}^H\mathbf{W}^H\right) \tag{21}$$

$$+ \mathrm{Tr}\left\{\mathbf{V}\mathbf{V}^H\right\} + \mathrm{Tr}\left\{\mathbf{W}\bar{\mathbf{D}}\bar{\mathbf{D}}^H\mathbf{W}^H\right\}, \tag{22}$$

where the last two terms can be seen to be constant. Therefore, we introduce $\mathbf{Z} = \mathbf{W}^H\mathbf{V}$ and redefine the cost function once again as

$$J \triangleq -2\,\mathrm{Re}\left(\mathrm{Tr}\left\{\mathbf{Z}\bar{\mathbf{D}}^H\right\}\right), \tag{23}$$

and then use (18) with this cost function for finding the fundamental frequency. We remark that the proposed method is superior to fitting the model matrix to the signal subspace using the relation in (9) and the MUSIC method in terms of computational complexity. One may wonder how the proposed method is different from direct fitting of the fundamental frequency to the unconstrained frequencies in (15) a la [8], i.e.,

$$\hat{\omega}_0 = \arg\min_{\omega_0} \|\hat{\mathbf{D}} - \bar{\mathbf{D}}\|_F^2, \tag{24}$$

To answer this, we consider the ideal case, where the right relation in (10) holds exactly, and we may write

$$\mathbf{S}_2 = \mathbf{S}_1\widehat{\mathbf{\Gamma}}. \tag{25}$$

Using this relation in (20), we get

$$J = \|\mathbf{S}_1\widehat{\mathbf{\Gamma}} - \mathbf{S}_1\mathbf{Q}\bar{\mathbf{D}}\mathbf{Q}^{-1}\|_F^2 \tag{26}$$

$$= \|\mathbf{S}_1\mathbf{Q}\widehat{\mathbf{D}}\mathbf{Q}^{-1} - \mathbf{S}_1\mathbf{Q}\bar{\mathbf{D}}\mathbf{Q}^{-1}\|_F^2 \tag{27}$$

$$= \|\mathbf{S}_1\mathbf{Q}\left(\widehat{\mathbf{D}} - \bar{\mathbf{D}}\right)\mathbf{Q}^{-1}\|_F^2, \tag{28}$$

which is different from the cost function in (24). The WLS method of [8] is essentially a closed-form fit of the fundamental frequency to the unconstrained frequencies in a weighted least-squares sense. However, unlike the proposed method and the MUSIC method, it requires that the amplitudes are known or are estimated.

The algorithm can be summarised as follows. First the covariance matrix is estimated from the observed signal and the EVD is calculated and partitioned into signal and noise subspaces. The first and last row of the eigenvectors belonging to the signal subspace are removed and $\widehat{\mathbf{\Gamma}}$ is found using least-squares or total least-squares. The EVD of $\widehat{\mathbf{\Gamma}}$ is calculated and subsequently the matrix $\mathbf{Z}$. Finally, the cost function (23) is evaluated on a coarse grid. As we shall see, the cost function is very smooth and thus only a few points are needed on this grid. A refined estimate can then easily be obtained using standard numerical optimization techniques.

4. EXPERIMENTAL RESULTS

We will now give a short example showing that the algorithm can indeed estimate the fundamental frequency. In Figure 1, the spectrum of the generated test signal in white Gaussian noise is shown with $\omega_0 = 0.2501$ for an SNR of 20 dB with the SNR being defined as in [5] with $N = 200$. In all of the experiments reported here, we use $M = N/2$. The cost functions that are obtained using the proposed method is shown in Figure 2 along with that of the MUSIC method of [5]. Specifically, the cost function in (20) is shown. As can be seen, the cost function has a global minimum near the true fundamental frequency. One noticeable possible advantage, compared to the MUSIC method, is that the ESPRIT cost function appears to be very simple and smooth. Also shown in the figure is the cost function for a direct fit of $\bar{\mathbf{D}}$ to $\hat{\mathbf{D}}$ with $\|\hat{\mathbf{D}} - \bar{\mathbf{D}}\|_F^2$ being shown for various fundamental frequencies. This last curve may help in understanding the differences between minimizing (17) or (20) and a direct fit.

We will now proceed to evaluate the RMSE of the estimators under various conditions. In all these experiments, we use 200 Monte-Carlo runs for each combination of parameters. The results are depicted in Figure 3 as a function of the SNR for $N = 200$ with $A_l = 1\ \forall l$ along with the Cramér-Rao lower bound (CRLB) and the RMSE for the MUSIC method. The results for a similar experiment but

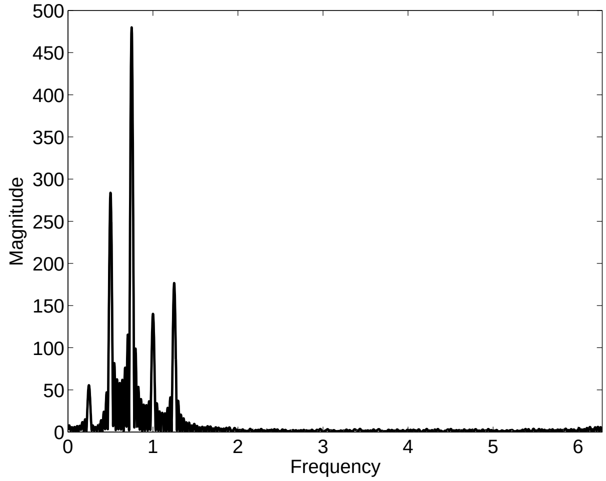

Fig. 1. Spectrum estimate, here the periodogram, of the test signal.

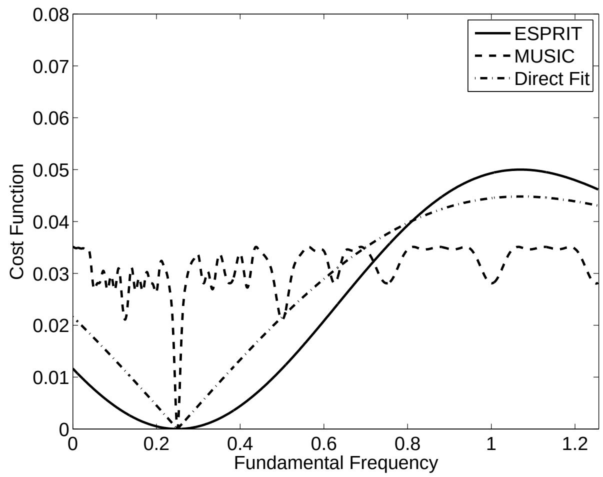

Fig. 2. ESPRIT and MUSIC cost functions (scaled) for the test signal shown in Figure 1.

now with Rayleigh distributed amplitudes are shown in Figure 4. As can be seen from the figures, the proposed method performs well for unit amplitudes, but the performance is degraded when the amplitudes are drawn from a Rayleigh distribution and it appears, like the WLS method, to be more sensitive to this than the MUSIC method with the threshold beneath which the results are not informative being increased. Above the threshold, however, the method appears to perform equivalently or a bit better than the MUSIC method. A likely explanation for this is that both the ESPRIT and WLS methods are sensitive to spurious frequency estimates for the individual frequencies, which would cause the ordering of the eigenvalues in (13) in forming the estimate to be erroneous. Such spurious estimates are likely to occur when sinusoids are closely spaced or the amplitudes are small, like for the Rayleigh distributed amplitudes, since this may cause signal and noise subspace eigenvectors of the EVD to be swapped. Lastly, we have investigated the performance of the estimators as a function of the fundamental frequency for $N = 100$ and an SNR of 40 dB. The results are shown in Figure 5. The three methods seem to exhibit similar thresholding behaviour for low fundamental frequencies. This thresholding effect is due to frequencies of the individual sinusoids coming increasingly closer as the fundamental frequency is lowered. Though, it appears that MUSIC is somewhat more robust to this than the other methods. Despite these drawbacks, the method may still be preferable to the MUSIC method in some cases since the proposed method is computationally simpler than the MUSIC-based method.

5. CONCLUSION

We have proposed a new method for the estimation of the fundamental frequency of a set of harmonically related sinusoids. The method is based on subspace techniques where bases for the signal and a noise subspaces are identified from eigenvalue decomposition of the covariance matrix. It is based on the a specific feature of the signal subspace known as the shift-invariance property which was first exploited in ESPRIT. The performance of the method has been assessed and been found to be good for unit amplitudes and less good for Rayleigh distributed amplitudes compared to other state-of-the-art methods, but the proposed method is significantly simpler than, for example, a method based on the subspace orthogonality property of MUSIC.

6. REFERENCES

[1] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagat.*, vol. 34(3), pp. 276–280, Mar. 1986.

[2] G. Bienvenu, "Influence of the spatial coherence of the background noise on high resolution passive methods," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1979, pp. 306–309.

[3] R. Roy and T. Kailath, "ESPRIT – estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37(7), July 1989.

[4] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophys. J. Roy. Astron. Soc.*, vol. 33, pp. 347–366, 1973.

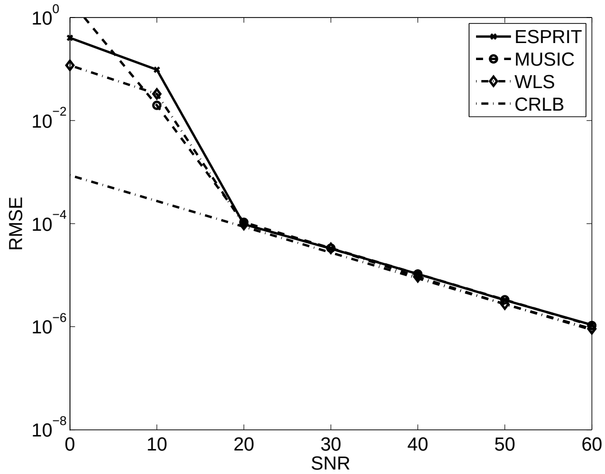

Fig. 3. RMSE as a function of the SNR for unit amplitudes and $N = 200$.

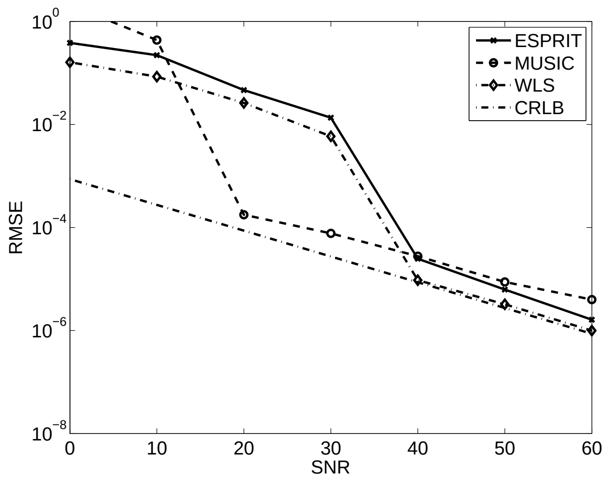

Fig. 4. RMSE as a function of the SNR for Rayleigh distributed and amplitudes with $N = 200$.

[5] 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*, vol. 15(5), pp. 1635–1644, July 2007.

[6] M. G. Christensen, A. Jakobsson and S. H. Jensen, "Multi-pitch estimation using harmonic music," in *Rec. Asilomar Conf. Signals, Systems, and Computers*, 2006, pp. 521–525.

[7] M. G. Christensen, P. Stoica, A. Jakobsson and S. H. Jensen, "Multi-pitch estimation," *Elsevier Signal Processing*, 2007, accepted.

[8] H. Li, P. Stoica, and J. Li, "Computationally efficient

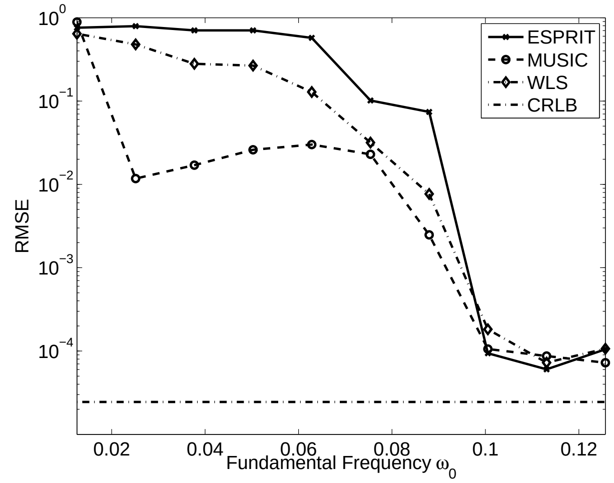

Fig. 5. RMSE as a function of the fundamental frequency with Rayleigh amplitudes, $N = 100$ and an SNR of 40 dB.

parameter estimation for harmonic sinusoidal signals," *Signal Processing*, vol. 80, pp. 1937–1944, 2000.

[9] 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.

[10] R. Badeau, B. David, and G. Richard, "A new perturbation analysis for signal enumeration in rotational invariance techniques," *IEEE Trans. Signal Processing*, vol. 54(2), pp. 450–458, Feb. 2006.

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