FUNDAMENTAL FREQUENCY ESTIMATION USING POLYNOMIAL ROOTING OF A SUBSPACE-BASED METHOD

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 Media Technology, Aalborg University
Niels Jernes Vej 14, 9220, Aalborg, Denmark
email: mgc@imi.aau.dk

ABSTRACT

We consider the problem of estimating the fundamental frequency of periodic signals such as audio and speech. A novel estimation method based on polynomial rooting of the harmonic MUltiple SIgnal Classification (HMUSIC) is presented. By applying polynomial rooting, we obtain two significant improvements compared to HMUSIC. First, by using the proposed method we can obtain an estimate of the fundamental frequency without doing a grid search like in HMUSIC. This is due to that the fundamental frequency is estimated as the argument of the root lying closest to the unit circle. Second, we obtain a higher spectral resolution compared to HMUSIC which is a property of polynomial rooting methods. Our simulation results show that the proposed method is applicable to real-life signals, and that we in most cases obtain a higher spectral resolution than HMUSIC.

1. INTRODUCTION

In many signal processing applications, it is of great importance to estimate the fundamental frequency. A specific example is in audio and speech processing. For example, the fundamental frequency is needed in parametric coding of audio and speech using a harmonic sinusoidal model. Also, many music information retrieval applications, such as automatic music transcription and musical genre classification, rely on the knowledge of the fundamental frequency. Within the last couple of decades, the problem of estimating the fundamental frequency has attracted considerable attention. This has resulted in numerous different fundamental frequency estimators. For a few examples of such estimators, we refer to [1–5].

Following, we define the fundamental frequency estimation problem. Consider a harmonic signal buried in white Gaussian noise $w(n)$, for $n = 0,\dots,N-1$,

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

where L is the model order and $\alpha_l = A_l e^{j\phi_l}$ is the complex amplitude of the lth sinusoid with $A_l > 0$ and ϕ_l being the real amplitude and the phase, respectively. In this paper, we will assume that the model order is known, hence, the problem at hand is to estimate the unknown fundamental frequency ω_0. While not considered in this paper, we refer the reader to [6] for few examples on how the model order could be estimated. In many existing methods for fundamental frequency estimation, the estimator is based on a grid search over a set of candidate fundamental frequencies [7]. This can be problematic for several reasons. For example, it can be hard to choose the resolution of the grid since the width of the peaks in the cost-function relies on, the sample size, the method, the signal-to-noise ratio (SNR), the source spacing (in multi-source scenarios), etc. Another issue is the computational complexity. Naturally, the computational complexity depends on the resolution of the grid. That is, if the peaks are narrow or if high-resolution is required, it is necessary to use a fine grid which of course increases the computational complexity. The problem of choosing the right grid can, to some extend, be relieved by introducing a gradient search. To alleviate the abovementioned issues, we consider the problem of obtaining an estimate of the fundamental frequency without having to do a grid search.

It was shown in [8] that the MUltiple SIgnal Classification (MUSIC) estimation criterion [9, 10] can be used to obtain a high-resolution estimate of the fundamental frequency. The resulting estimator, refered to as Harmonic MUSIC (HMUSIC), was shown to have a good statistical performance. In this paper, we propose an estimator which is a relaxation of the HMUSIC cost-function from the unit circle onto the whole complex plane. That is, the proposed estimator evaluates the HMUSIC cost-function using a polynomial rooting method which can be seen as a generalization of the original root MUSIC method [11]. Using polynomial rooting has two signficant advantages. First, it gives an increased spectral resolution in multi-source scenarios and, second, it will give an estimate of the fundamental frequency without using a grid search. For more on the performance of the MUSIC and root MUSIC algorithms see, e.g., [12, 13]. Through simulations we investigate the performance of the proposed method on real-life signals. Also, using synthetic data we evaluate the proposed estimator in Monte-Carlo simulations, and we compare the result with both the performance of the HMUSIC estimator and the Cramér-Rao Lower Bound (CRLB).

The rest of the paper is organized as follows. In Section 2, we make a brief introduction to the HMUSIC estimation method and we describe the proposed method. In Section 3, we evalute the performance of the proposed using both qualitative and quantitative measurements. Finally, Section 4 concludes on our work.

2. PROPOSED METHODS

In this section, we present the fundamental theory behind the HMUSIC estimator [8] and we present the proposed estimator. Consider a signal of the form (1) from which we take M consecutive samples. The samples is then used to form a signal vector

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

where $(\cdot)^T$ denotes the transpose. If we then assume that the phases of the harmonics are independent and uniformly distributed in the interval $(-\pi;\pi]$, we can write the covariance matrix $\mathbf{R} \in \mathbb{C}^{M\times M}$ as [14]

$$\mathbf{R} = \mathrm{E}\left\{\mathbf{x}(n)\mathbf{x}^H(n)\right\} \quad (3)$$

$$= \mathbf{A}\mathbf{P}\mathbf{A}^H + \sigma_w^2\mathbf{I} \ , \quad (4)$$

where $\mathrm{E}\{\cdot\}$ and $(\cdot)^H$ denotes the expectation and the conjugate transpose, respectively, σ_w^2 is the noise variance and $\mathbf{I}$ is the $M\times M$ identity matrix. The matrix $\mathbf{P}$ is a diagonal matrix containing the squared real amplitudes, i.e.,

$$\mathbf{P} = \mathrm{diag}\left(\begin{bmatrix} A_1^2 & \cdots & A_L^2 \end{bmatrix}\right) \ , \quad (5)$$

and $\mathbf{A} \in \mathbb{C}^{M\times L}$ is a full-rank Vandermonde matrix

$$\mathbf{A} = [\mathbf{a}(\omega_0) \quad \cdots \quad \mathbf{a}(L\omega_0)] \ , \quad (6)$$

with $\mathbf{a}(\omega) = \begin{bmatrix} 1 & e^{-j\omega} & \cdots & e^{-j\omega(M-1)} \end{bmatrix}^T$. Note that since we assume a harmonic model, the Vandermonde matrix $\mathbf{A}$ is only dependend on a single frequency, namely the fundamental frequency. Let us then define

$$\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H \ , \quad (7)$$

as the eigenvalue decomposition (EVD) of the covariance matrix. The matrix $\mathbf{U} = [\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_M]$ then contains the M orthonormal eigenvectors of $\mathbf{R}$ and $\mathbf{\Lambda}$ is a diagonal matrix containing the corresponding eigenvalues, λ_k. Note that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_M$. It is well known that the L most significant eigenvectors will span the signal subspace while the noise subspace is spanned by the $M-L$ least significant eigenvectors. That is, the noise subspace is spanned by $\mathbf{G}$ defined as

$$\mathbf{G} = [\mathbf{u}_{L+1} \quad \cdots \quad \mathbf{u}_M] \ . \quad (8)$$

We know that $\mathrm{range}(\mathbf{A}) = \mathrm{range}(\mathbf{S})$ where $\mathbf{S} = [\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_L]$ spans the signal subspace. Also, we know that the signal subspace is orthogonal to the noise subspace which allows us to write

$$\mathbf{A}^H\mathbf{G} = \mathbf{0} \ . \quad (9)$$

The covariance matrix, however, is most often not available in practice. Therefore, we will replace the covariance matrix in the above expression by the sample covariance matrix defined as

$$\hat{\mathbf{R}} = \frac{1}{N-M+1}\sum_{n=M-1}^{N-1}\mathbf{x}(n)\mathbf{x}^H(n) \ . \quad (10)$$

Due to estimation errors, $\mathbf{A}$ will not be exactly orthogonal to $\mathbf{G}$. Therefore, in HMUSIC, the fundamental frequency is found by

$$\hat{\omega}_0 = \arg\max_{\omega_0\in\Omega_0} \frac{1}{\|\mathbf{A}^H\mathbf{G}\|_F^2} \quad (11)$$

$$= \arg\max_{\omega_0\in\Omega_0} \underbrace{\frac{1}{\mathrm{Tr}\left\{\mathbf{A}^H\mathbf{G}\mathbf{G}^H\mathbf{A}\right\}}}_{J(\omega_0)} \ , \quad (12)$$

with $\mathrm{Tr}\{\cdot\}$ and $\|\cdot\|_F$ denoting the trace and the Frobenius norm, respectively, and Ω_0 is the set of candidate fundamental frequencies. Notice, that the HMUSIC criterion can be seen as an approximation to the angle between subspaces [15]. The minimization is done over the set Ω_0, i.e., the resolution of the estimate depends on the cardinality of Ω_0. The resolution can, however, be refined by performing a gradient search after a coarse estimate has been obtained.

Instead, we will now present how the cost-function can be evaluated using a rooting method. This has both the advantage of obtaining a solution without doing a grid search and an increased spectral resolution. Let us define a new matrix $\mathbf{C} = \mathbf{G}\mathbf{G}^H$ and rewrite the cost-function $J(\omega_0)$ by using the definition of the trace

$$J(\omega_0) = \frac{1}{\mathrm{Tr}\{\mathbf{A}^H\mathbf{C}\mathbf{A}\}} \quad (13)$$

$$= \frac{1}{\sum_{l=1}^{L}\mathbf{a}^H(l\omega_0)\mathbf{C}\mathbf{a}(l\omega_0)} \ . \quad (14)$$

As mentioned previously, the expression in the denominator will have no solutions when equated with zero. However, if we instead replace $e^{j\omega}$ in $\mathbf{a}(\omega)$ with the variable $z = |z|e^{j\arg(z)}$, we can expect that denominator will have some solutions when equated with zero. That is, we can write

$$\frac{1}{J(z)} = \sum_{l=1}^{L}\mathbf{a}^T(z^{-l})\mathbf{C}\mathbf{a}(z^l) \quad (15)$$

$$= \sum_{l=1}^{L}\sum_{k=-(M-1)}^{M-1} c_k z^{lk} \quad (16)$$

$$= \sum_{l=1}^{L} p_l(z) = p(z) = 0, \quad (17)$$

where $p_l(z)$ is the lth polynomial and c_k is the sum of entries of $\mathbf{C}$ along the kth diagonal, i.e.,

$$c_k = \sum_{m-n=l}\mathbf{A}_{mn} \ . \quad (18)$$

The expression in (17) will only be zero when all of the individual polynomials $p_l(z)$ for $l = 1,\ldots,L$ is equal to zero. This can be proven by the fact that $\mathbf{C} = \mathbf{G}\mathbf{G}^H$ is Hermitian and thereby positive semi-definite. Positive semi-definiteness implies that

$$\mathbf{x}^H\mathbf{C}\mathbf{x} \geq 0 \ , \qquad \forall\mathbf{x} \ , \quad (19)$$

which proves our statement. Therefore, we can conclude that $p(z)$ has a root close on the unit circle only when $\hat{\omega}_0$ approaches ω_0. This will only be fulfilled when $M\rightarrow\infty$ which implies that $N\rightarrow\infty$

$$\lim_{N\rightarrow\infty} p(z) = 0\Big|_{z=e^{j\omega_0}} \Leftrightarrow \lim_{N\rightarrow\infty} J(z) = \infty\Big|_{z=e^{j\omega_0}} \ . \quad (20)$$

In reality, the roots of the polynomial will not lie exactly on the unit circle since we have a limited number of samples. Instead, if N and M are sufficiently large, we can assume that the root lying closest to the unit circle will correspond to the largest peak of the HMUSIC pseudospectra. This is

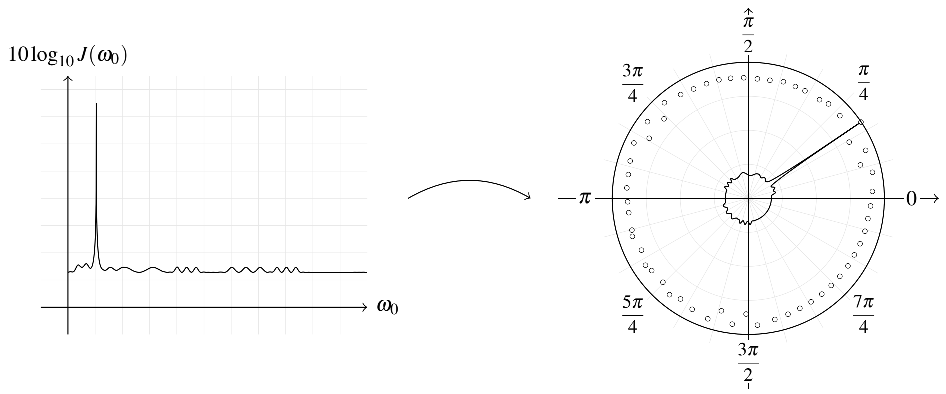

Figure 1: An example of a HMUSIC cost-function transformed into polar coordinates. The point $(0,0)$ in the right-hand plot corresponds to $J(\omega_0) = 0$ while the whole unit circle corresponds to $J(\omega_0) = \infty$. Note that $\circ$ denotes a root of $p(z)$.

also illustrated in Fig. 1 which shows an example of a HMUSIC cost-function and its related roots. The fundamental frequency can therefore be estimated as the angle of the root $\hat{r}$ being closest to the unit circle, i.e.,

$$\hat{\omega}_0 = \angle \hat{r} \ . \tag{21}$$

Notice, however, that the roots come in complex conjugate pairs so we only consider the roots within the unit circle.

3. EXPERIMENTAL RESULTS

This section contains the experimental results obtained during evaluation of the proposed method. First, we investigate the performance of the proposed method on a real-life signal. The signal used in this experiment, was a trumpet signal sampled at 8,820 Hz. In Fig. 2, the spectrogram of the trumpet signal is shown. We divided the trumpet signal into blocks of length $N = 160$ overlapping each other by 50 %. The fundamental frequency was estimated from each block with $M = 65$ and by assuming that $L = 7$. The results are depicted in Fig. 2. In the end of the signal, the proposed estimator seems to give erroneous estimates, however, it can also be seen that the model order in this part of the signal is rather five than seven. Except from the parts where there is a missmatch between the assumed model order and the true model order, the proposed estimator obtains estimates close to the true fundamental frequency. This verifies that the proposed estimator is applicable to real-life signals.

Also, we have conducted a series of Monte-Carlo simulations evaluating the statistical performance of the proposed method compared to both the original HMUSIC estimator and the CRLB [16]. In the first of these simulations, we evaluated the estimation performance with respect to the choice of M for N being fixed to 80. The signal used in this simulation, was a synthetic signal composed by $L = 3$ harmonically related complex sinusoids each with unit amplitudes with a fundamental frequency of 189.44 Hz. Complex white Gaussian noise was added to the signal such that the SNR

$$\mathrm{SNR} = 10\log_{10} \frac{\sum_{l=1}^{L} A_l^2}{\sigma_w^2} \ , \tag{22}$$

was 20 dB. Furthermore, the signal was sampled at $f_s = 2$ kHz. We then conducted 500 Monte-Carlo trials for each different M where we estimated the fundamental frequency. Also, for each different M we calculated the mean squared estimation error (MSE) defined as

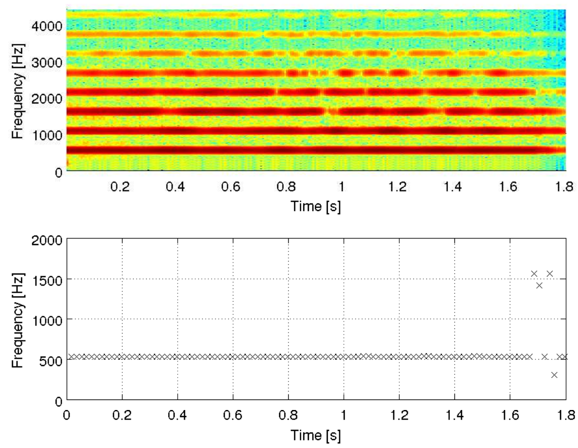

Figure 2: A spectrogram of a trumpet signal sampled at 8,820 Hz (top) and fundamental frequency estimates obtained using root HMUSIC (bottom).

$$\mathrm{MSE} = \frac{1}{S} \sum_{s=1}^{S} \left(\omega_0 - \hat{\omega}_0^{(s)} \right)^2 \ , \tag{23}$$

with ω_0 and $\hat{\omega}_0^{(s)}$ being the true fundamental frequency and its estimate in the sth Monte-Carlo trial, respectively, and S is the number of Monte-Carlo trials. The resulting MSEs for both root HMUSIC and MUSIC from this Monte-Carlo simulation are shown in Fig. 3 together with the CRLB. We calculated the CRLB by using the asymptotic expression in [8]

$$\mathrm{CRLB}(\omega_0) = \frac{6\sigma_w^2}{N(N^2-1)\sum_{l=1}^{L} A_l^2 l^2} \ . \tag{24}$$

The first observation from the first Monte-Carlo simulation is, that both root HMUSIC and HMUSIC shows similar performance independently on the choice of M. Below $M = 10$

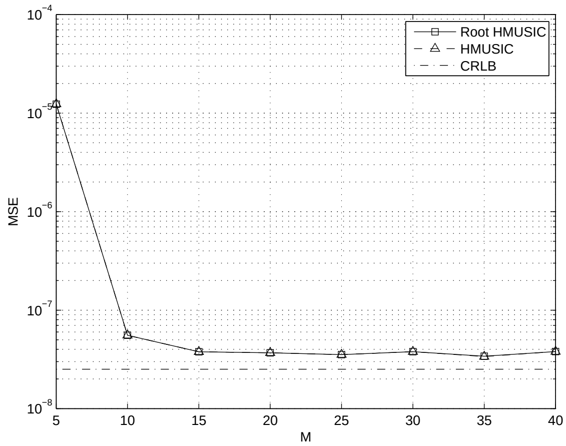

Figure 3: Plot of the asymptotic CRLB and the MSE of root HMUSIC and HMUSIC as a function of M.

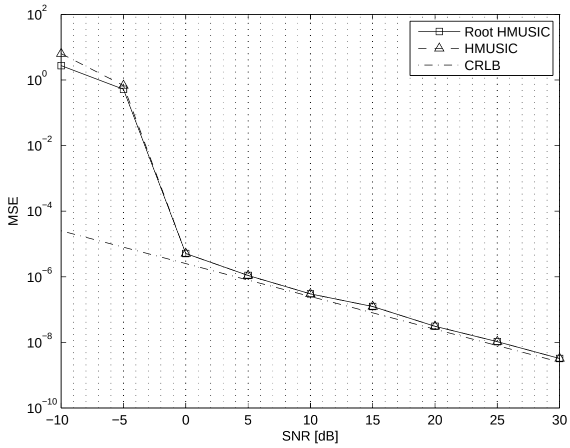

Figure 4: Plot of the asymptotic CRLB and the MSE of root HMUSIC and HMUSIC as a function of the SNR.

we see some thresholding behaviour for both methods. Also, we note that both methods are following but not reaching the CRLB as it was also reported in [8]. In another Monte-Carlo simulation, we evaluated the performance of root HMUSIC and HMUSIC with respect to the SNR. The parameters N, ω_0, L and f_s had the same values as in the previous Monte-Carlo simulation while M was fixed to $\lfloor \frac{N}{3} \rfloor$. We then ran 500 Monte-Carlo trials for each different SNR, and the results are depicted in Fig. 4 in terms of the MSE. We note that for high SNRs, the two methods show the same performance while for low SNRs root HMUSIC seems to perform slightly better than HMUSIC. Thresholding behaviour is observed around an SNR of 0 dB for this particular setup.

Also, we evaluated the performance with respect to the fundamental frequency. In this Monte-Carlo simulation, $N = 60$ samples with a sampling frequency of $f_s = 2$ kHz of a synthetic signal having $L = 3$ sinusoids with unit amplitudes was used. Complex white Gaussian noise was added such that the SNR was 40 dB. Again, M was chosen to $\lfloor \frac{N}{3} \rfloor$. We ran 500 trials for each different fundamental frequency, and the

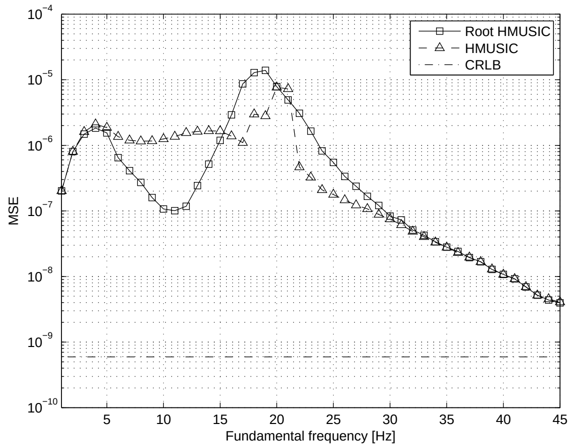

Figure 5: Plot of the asymptotic CRLB and the MSE of root HMUSIC and HMUSIC as a function of the fundamental frequency.

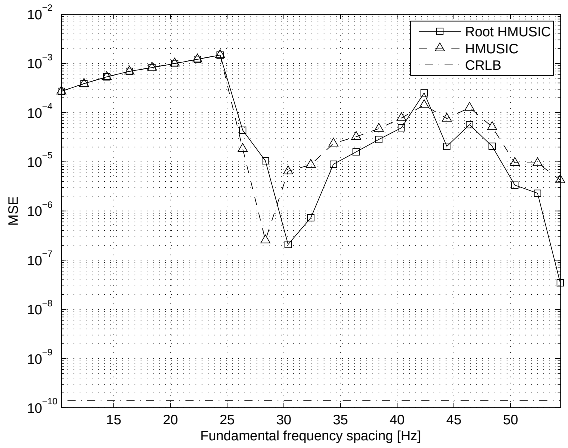

Figure 6: Plot of the asymptotic CRLB for the one source in white Gaussian noise scenario and the MSE of root HMUSIC and HMUSIC as a function of the fundamental frequency frequency spacing in a two-source scenario.

results are depicted in Fig. 5. Notice that for low fundamental frequencies, the proposed method shows a better performance compared to HMUSIC. This is also expected, since it has been reported that rooting methods have a better spectral resolution than spectral methods [11]. In the final Monte-Carlo simulation, we evaluated the performance of both root HMUSIC and HMUSIC in a two-source scenario. The sample length in this experiment was $N = 120$, M was 40 and the sampling frequency was $f_s = 2$ kHz. We generated the signal such that it was composed by two harmonic signals each with $L = 2$. The fundamental frequency of one of the harmonic signals was fixed to 114.79 Hz while the fundamental frequency of the other harmonic signal was varied. Furthermore, the SNR with respect to one harmonic signal was set to 40 dB. We ran 500 trials for each different fundamental frequency of the second harmonic source, and the outcome of this Monte-Carlo simulations is shown in Fig. 6. Using this particular setup, it can be seen that at low frequency spacings ($<$ 30 Hz), both methods show the same poor performance since they cannot resolve the sources. However, for higher frequency spacings ($>$ 30 Hz), the proposed method shows a better performance compared to HMUSIC. In this simulation, the performance of both methods are relatively far away from the CRLB which is partly explained by the fact, that the CRLB is derived for a single source scenario with white noise.

4. CONCLUSION

In this paper, we considered the fundamental frequency estimation problem. We proposed a new estimation method which is based on polynomial rooting of the known HMUSIC estimator. This has two significant advantages: 1) using the proposed method we obtain an estimate of the fundamental frequency without having to do a grid search and 2) using polynomial rooting we obtain a better spectral resolution compared to HMUSIC. We evaluated the proposed method using simulations. First, we showed that the proposed method is applicable to real-life signals, by using the root HMUSIC to correctly estimate the fundamental frequency. Second, we performed a series of statistical measurements on the proposed method. These simulations showed, that in many cases root HMUSIC will have a similar performance as HMUSIC. However, in multi-source scenarios with closely-spaced sources, the simulations showed that for most fundamental frequency spacings the proposed root HMUSIC method outperforms HMUSIC. This was also expected due to the properties of polynomial rooting methods. Like the HMUSIC method, the root HMUSIC method follows, but do not reach, the CRLB in good conditions.

REFERENCES

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

[2] K. W. Chan and H. C. So. Accurate frequency estimation for real harmonic sinusoids. *IEEE Signal Process. Lett.*, 11(7):609–612, 2004.

[3] A. de Cheveigné and H. Kawahara. YIN, a fundamental frequency estimator for speech and music. *J. Acoust. Soc. Am.*, 111(4):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*, volume 1, pages 249–252, 2007.

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

[6] P. Stoica and Y. Selen. Model-order selection: a review of information criterion rules. *IEEE Signal Process. Mag.*, 21(4):36–47, 2004.

[7] M. G. Christensen and A. Jakobsson. Multi-pitch estimation. *Synthesis Lectures on Speech and Audio Processing*, 5(1):1–160, 2009.

[8] M. G. Christensen, A. Jakobsson, and S. H. Jensen. Joint high-resolution fundamental frequency and order estimation. *IEEE Trans. Audio, Speech, and Language Process.*, 15(5):1635–1644, 2007.

[9] R. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas Propag.*, 34(3):276–280, Mar. 1986.

[10] 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*, volume 4, pages 306 – 309, Apr 1979.

[11] A. Barabell. Improving the resolution performance of eigenstructure-based direction-finding algorithms. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, volume 8, pages 336–339, Apr 1983.

[12] B. D. Rao and K. V. S. Hari. Performance analysis of root-MUSIC. *IEEE Trans. Acoust., Speech, Signal Process.*, 37(12):1939 –1949, Dec 1989.

[13] P. Stoica and A. Nehorai. MUSIC, maximum likelihood, and Cramér-Rao bound. *IEEE Trans. Acoust., Speech, Signal Process.*, 37(5):720 –741, May 1989.

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

[15] M. G. Christensen, A. Jakobsson, and S. H. Jensen. Sinusoidal order estimation using angles between subspaces. *EURASIP J. on Advances in Signal Processing*, vol. 2009:1–11, 2009.

[16] S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, Inc., 1993.