ROBUST PITCH ESTIMATION USING AN OPTIMAL FILTER ON FREQUENCY ESTIMATES

Sam Karimian-Azari ⋆*, Jesper Rindom Jensen* †*, and Mads Græsbøll Christensen*

Audio Analysis Lab, AD:MT, Aalborg University, email: {ska, jrj, mgc}@create.aau.dk

ABSTRACT

In many scenarios, a periodic signal of interest is often contaminated by different types of noise, that may render many existing pitch estimation methods suboptimal, e.g., due to an incorrect white Gaussian noise assumption. In this paper, a method is established to estimate the pitch of such signals from unconstrained frequency estimates (UFEs). A minimum variance distortionless response (MVDR) method is proposed as an optimal solution to minimize the variance of UFEs considering the constraint of integer harmonics. The MVDR filter is designed based on noise statistics making it robust against different noise situations. The simulation results confirm that the proposed MVDR method outperforms the state-of-the-art weighted least squares (WLS) pitch estimator in colored noise and has robust pitch estimates against missing harmonics in some time-frames.

Index Terms— Audio signal, harmonic model, pitch estimation, minimum variance distortionless response (MVDR), maximum likelihood (ML).

1. INTRODUCTION

In audio analysis, the pitch value is closely related to the primary frequency of vibration (fundamental frequency), e.g., the vocal cord vibration for voiced speech signals. Pitch estimation is a challenging problem in audio signal processing, e.g., [1–6], and it is important in many speech and audio applications such as coding, enhancement, and separation. We can model many parts of audio signals using a constrained harmonic model which consists of a fundamental frequency, or pitch as it is often referred to, and integer multiples of the fundamental frequency. Observed signals consisting of such a harmonic signal are commonly contaminated by noise. With the assumption of white Gaussian noise, the maximum likelihood (ML) and maximum a posteriori (MAP) estimation methods are commonly used for pitch and model order estimation, e.g., [1, 4, 7–9]. We here focus on a two stage procedure, where the pitch is estimated from an unconstrained set of frequency estimates to match with the constrained harmonic model. This approach is fast and statistically efficient even with inefficient unconstrained frequency estimates (UFEs). Different unconstrained estimators of frequencies have been investigated in [10], e.g., the MUSIC [11], ESPRIT, [12], NLS [13, 14], and Capon [15, 16].

The Markov-like weighted least squares (WLS) pitch estimator in [1] is computationally efficient with good statistical performance when the noise is white. In this method, the weights of UFEs relate to the magnitude estimates of the unconstrained frequencies. The pitch estimate used in the WLS method of [1] is not optimal for a nonidentical noise variance across harmonics, e.g., for colored noise. Furthermore, often in practice, the results of this method suffer from large errors in the pitch estimation when we have spurious frequency estimates and missing harmonics in the unconstrained frequencies [16].

In this paper, we propose a filtering method to obtain the pitch from UFEs even when the noise is not white. Since an additive Gaussian noise signal is equivalent to an additive angular noise for a high signal-to-noise ratio (SNR) [17], the UFEs can be modeled like multiple random variables with a joint probability density function (pdf), which each have a normal distribution with a constrained expected value. Thus, an initial concept of the filtering method is applied to the UFEs to align the constrained random values with a least variability. If we do not know the local SNRs, we can assume identical white Gaussian noise across harmonics, and the minimum-variance unbiased (MVU) estimator [18] is a fixed filter. These normally distributed random variables may have different variances related to the reciprocal of the narrowband SNRs [17], and herein, we can estimate the filter coefficients adaptively based on the noise characteristics with a linear constraint to satisfy the integer relationship between the UFEs. Intuitively, we estimate the covariance matrix of the UFEs, and design an optimal filter to minimize the noise variance on the UFEs, that is known as the minimum variance distortionless response (MVDR) estimator. For the particular case of white Gaussian noise, we design the maximum likelihood (ML) estimator from the concept of the MVDR with identical noise power spectrum across all frequencies. We see that the ML estimator is equivalent to the WLS estimator in [1]. As a result, because we design the constrained MVDR filter using the noise variance estimates of the UFEs, the proposed method estimates pitch optimally in the presence of dif-

This research was funded by: ⋆the Villum Foundation, and †the Danish Council for Independent Research, grant ID: DFF 1337-00084.

ferent types of Gaussian noise. Moreover, the MVDR method is robust against missing some harmonics, since this phenomena is encompassed in the statistics estimates.

The rest of this paper is organized as follows. In Section 2, we introduce the harmonic signal model. Then, we propose the MVDR filter to estimate the fundamental frequency from unconstrained frequency estimates in Section 3. Later on, in Section 4, experimental results are reported. In closing, the work is concluded in Section 5.

2. PROBLEM FORMULATION

2.1. Signal model

We model harmonic signals as the sum of analytic sinusoids which can be applied on real signals through the Hilbert transform. A harmonic signal consists of L sinusoids with the fundamental frequency $\omega_0 \in (0, \pi]$, real magnitudes $\mathbf{a} = [\,\alpha_1,\, \alpha_2,\, \ldots,\, \alpha_L\,]^T$, and phases $\varphi_l \in (-\pi, \pi]$ for $l = 1, \ldots, L$ like

$$s(n, \boldsymbol{\theta}) = \sum_{l=1}^{L} \alpha_l \, e^{\,j\,(l\omega_0 n + \varphi_l)}, \tag{1}$$

where the superscript ^{T} is the transpose operator, and $j = \sqrt{-1}$. The harmonic signal is parameterized by the vector $\boldsymbol{\theta} = [\,\omega_1,\, \alpha_1,\, \varphi_1,\, \ldots,\, \omega_L, \alpha_L,\, \varphi_L\,]^T$. We only consider the harmonic frequencies which can be seen that

$$\boldsymbol{\Omega} = [\,\omega_1,\, \omega_2,\, \ldots, \omega_L\,]^T = \mathbf{d}_L\, \omega_0, \tag{2}$$

where $\omega_l = l\omega_0$, and $\mathbf{d}_L = [\,1, 2,\, \ldots, L\,]^T$ is the constraint vector. We assume that the observed signal $x(n)$ of the signal source $s(n, \boldsymbol{\theta})$ is contaminated by Gaussian noise $v(n)$ with complex value and zero mean, i.e.,

$$x(n) = s(n, \boldsymbol{\theta}) + v(n). \tag{3}$$

For white Gaussian noise, the real and imaginary components of $v(n)$ are uncorrelated and have an equivalent variance $\sigma^2/2$. At a high narrowband SNR, i.e., $\mathrm{SNR}(\omega_l) = \alpha_l^2/\sigma^2 \gg 1$, the harmonic frequency ω_l is perturbed with a real angular noise $\Delta\omega_l(n) = \frac{v(n)}{\alpha_l} \sin(l\omega_0 n + \varphi_l)$, which has a normal distribution with zero mean and variance $\mathrm{E}\{(\Delta\omega_l)^2\} = 1/(2\,\mathrm{SNR}(\omega_l))$ [17], where $\mathrm{E}\{\cdot\}$ denotes the statistical expectation. Therefore, we can approximate the complex signal model (3) like

$$x(n) \approx \sum_{l=1}^{L} \alpha_l \, e^{\,j\,(l\omega_0 n + \Delta\omega_l(n) + \varphi_l)}. \tag{4}$$

Ideally, white Gaussian noise has a homogeneously distributed power spectrum $\Phi_\omega = \sigma^2$ across frequencies $\omega \in [0, \pi]$. On the other hand, colored noise has an inhomogeneous distributed power spectrum Φ_ω, that results in different angular noise across harmonics with the variances

$$\mathrm{E}\{(\Delta\omega_l)^2\} = \frac{1}{2\,\mathrm{SNR}(\omega_l)} = \frac{\Phi_{\omega_l}}{2\,\alpha_l^2}. \tag{5}$$

We assume that we have a set of unconstrained frequency estimates (UFEs) $\hat{\boldsymbol{\Omega}} = [\,\hat{\omega}_1,\, \hat{\omega}_2,\, \ldots,\, \hat{\omega}_L\,]^T$, which are estimated from the signal vector $[\,x(n), x(n+1), \ldots, x(n+N-1)\,]^T$. We model these UFEs with the equivalent linear constrained harmonic model in (2) that is contaminated by Gaussian noise. Therefore, the harmonic frequency estimates can be written like

$$\hat{\boldsymbol{\Omega}} = \boldsymbol{\Omega} + \Delta\boldsymbol{\Omega} = \mathbf{d}_L\, \omega_0 + \Delta\boldsymbol{\Omega}, \tag{6}$$

where $\Delta\boldsymbol{\Omega} = [\,\Delta\omega_1,\, \Delta\omega_2,\, \ldots,\, \Delta\omega_L\,]^T$ are additive angular noise across harmonics. For a high number of samples, the UFEs are asymptotically unbiased, i.e., $\lim_{N\to\infty} \mathrm{E}\{\hat{\boldsymbol{\Omega}}\} = \boldsymbol{\Omega}$, with the covariance matrix

$$\begin{aligned} \boldsymbol{\Phi}_{\Delta\boldsymbol{\Omega}} &= \mathrm{E}\{\Delta\boldsymbol{\Omega}\,\Delta\boldsymbol{\Omega}^T\} \\ &= \mathrm{E}\{(\hat{\boldsymbol{\Omega}} - \mathrm{E}\{\hat{\boldsymbol{\Omega}}\})(\hat{\boldsymbol{\Omega}} - \mathrm{E}\{\hat{\boldsymbol{\Omega}}\})^T\}. \end{aligned} \tag{7}$$

Independent UFEs would be implicitly uncorrelated, i.e., $\mathrm{E}\{\Delta\omega_i \Delta\omega_k\} = 0$ for $i \neq k$, when harmonics are not close to each other and the narrowband SNRs are high enough. Consequently, the covariance matrix of the angular noise vector in such situations can be shown to be

$$\boldsymbol{\Phi}_{\Delta\boldsymbol{\Omega}} = \mathrm{diag}\Big\{\Big[\,\frac{\Phi_{\omega_1}}{2\,\alpha_1^2},\, \frac{\Phi_{\omega_2}}{2\,\alpha_2^2},\, \ldots,\, \frac{\Phi_{\omega_L}}{2\,\alpha_L^2}\,\Big]\Big\}, \tag{8}$$

where $\mathrm{diag}\{\cdot\}$ denotes the diagonal matrix formed with the vector input along its diagonal. For close harmonics, which are not well-separated, with low narrowband SNRs, nevertheless, there is a cross-correlation between harmonics, i.e., $\mathrm{E}\{\Delta\omega_i \Delta\omega_k\} \neq 0$.

3. PROPOSED METHOD

We have already formulated the relationship between the UFEs of a harmonic signal and the pitch (fundamental frequency) ω_0. In this sense, pitch can be interpreted as the slope of the line that fits the UFEs. We propose a filtering method to seek an unbiased pitch estimate with a minimum variance via the line fitting with the determined constrained vector $\mathbf{d}_L$. Hence, we apply a filter $\mathbf{h} \in \mathbb{R}^L$ to estimate the pitch value from the set of UFEs as

$$\begin{aligned} \hat{\omega}_0 &= \mathbf{h}^T \hat{\boldsymbol{\Omega}} \\ &= \mathbf{h}^T \mathbf{d}_L\, \omega_0 + \mathbf{h}^T \Delta\boldsymbol{\Omega}. \end{aligned} \tag{9}$$

With the distortionless constraint that $\mathbf{h}^T \mathbf{d}_L = 1$, the mean squared error (MSE) of the unbiased estimator is given by

$$\begin{aligned} \mathrm{MSE}[\,\hat{\omega}_0\,] &= \mathrm{E}\{(\hat{\omega}_0 - \omega_0)^2\} \\ &= \mathrm{E}\{(\mathbf{h}^T \Delta\boldsymbol{\Omega})(\Delta\boldsymbol{\Omega}^T \mathbf{h})\} \\ &= \mathbf{h}^T \boldsymbol{\Phi}_{\Delta\boldsymbol{\Omega}} \mathbf{h}. \end{aligned} \tag{10}$$

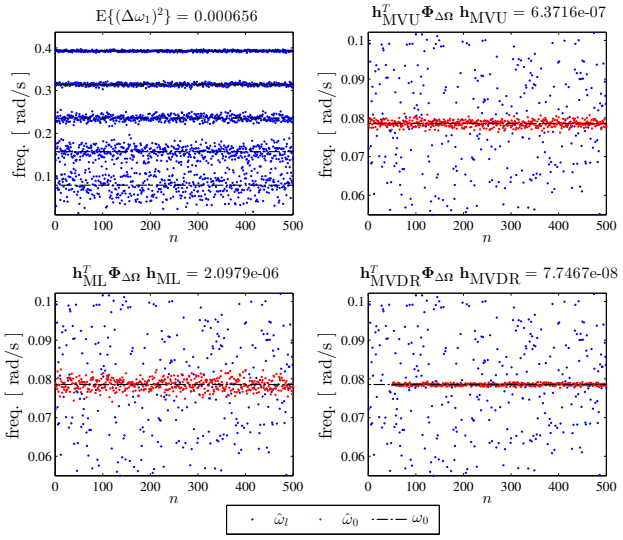

Fig. 1. An example of the fundamental frequency estimate using the MVU, ML, and MVDR filtering methods (red dots) from UFEs (blue dots) with $L = 5$ at the presence of colored Gaussian noise.

To minimize the variance of the pitch estimates, the optimal filter is given as the solution to the following problem:

$$\begin{aligned} \min_{\mathbf{h}} \quad & \mathbf{h}^T \mathbf{\Phi}_{\Delta\mathbf{\Omega}} \mathbf{h} \\ \text{subject to} \quad & \mathbf{h}^T \mathbf{d}_L = 1. \end{aligned} \tag{11}$$

Using the method of Lagrange multipliers, the minimum variance distortionless response (MVDR) filter is then given by [15]

$$\mathbf{h}_{\text{MVDR}} = \mathbf{\Phi}_{\Delta\mathbf{\Omega}}^{-1} \mathbf{d}_L (\mathbf{d}_L^T \mathbf{\Phi}_{\Delta\mathbf{\Omega}}^{-1} \mathbf{d}_L)^{-1}, \tag{12}$$

and inserting the proposed optimal filter in the MSE expression in (10) yields

$$\mathbf{h}_{\text{MVDR}}^T \mathbf{\Phi}_{\Delta\mathbf{\Omega}} \mathbf{h}_{\text{MVDR}} = \frac{1}{\mathbf{d}_L^T \mathbf{\Phi}_{\Delta\mathbf{\Omega}}^{-1} \mathbf{d}_L}. \tag{13}$$

If we assume identical narrowband SNRs across the harmonics, we can obtain a simple and signal independent, minimum-variance unbiased (MVU) filter design:

$$\mathbf{h}_{\text{MVU}} = \mathbf{d}_L (\mathbf{d}_L^T \mathbf{d}_L)^{-1}. \tag{14}$$

In the particular case of statistically independent UFEs, which are corrupted by white Gaussian noise with the variance σ^2, we can express the inverse of the diagonal covariance matrix simply as

$$\mathbf{\Phi}_{\Delta\mathbf{\Omega}}^{-1} = \frac{2}{\sigma^2} \text{ diag}\{ \alpha_1^2, \alpha_2^2, \ldots, \alpha_L^2 \}, \tag{15}$$

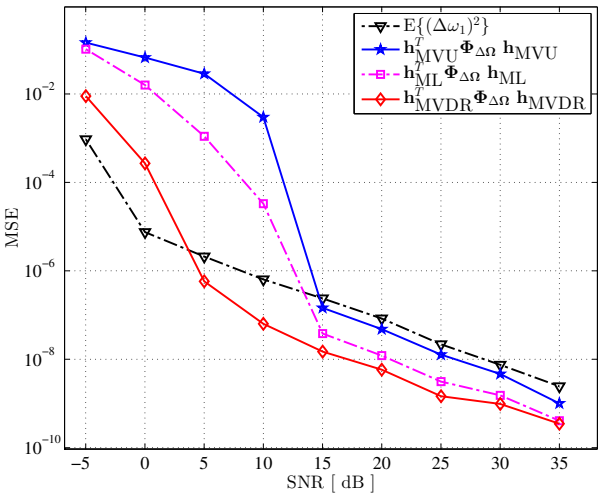

Fig. 2. MSEs of pitch estimates of a harmonic signal versus SNR levels of colored noise in dB.

in (12). As a result, the maximum likelihood (ML) estimator is

$$\mathbf{h}_{\text{ML}} = \frac{1}{\sum_{l=1}^{L} (l\alpha_l)^2} [\, \alpha_1^2, \, 2\alpha_2^2, \, \ldots, \, L\alpha_L^2 \,]^T, \tag{16}$$

which is the same as the weighted least squares (WLS) method in [1].

4. SIMULATION RESULTS

We evaluate the performance of the proposed filtering method to estimate the fundamental frequency from a set of unconstrained frequency estimates, and measure the MSE by averaging the squared error in multiple trials. First, we simulate a set of unconstrained estimates of a harmonic signal with nonidentical magnitudes, which are contaminated by colored Gaussian noise with exponentially decreasing variances, and then compare the results of the MVU, ML, and MVDR filters. In the other experiments, we estimate the unconstrained frequencies of sinusoids from a simulated harmonic signal using the subspace orthogonality method [4][1] based on the MUSIC method [11], and then compare the results in different SNRs and number of harmonics. Finally, we evaluate the proposed method to estimate the fundamental frequency from UFEs of a real noisy trumpet signal to show the applicability of the proposed method on real signals.

In practice, the covariance matrix $\hat{\mathbf{\Phi}}_{\Delta\mathbf{\Omega}}(n)$ in a time instance n is derived from M number of UFEs $\hat{\mathbf{\Omega}}(n-m)$, where $m = 0, 1, \ldots, M-1$, and the expectation is estimated by ensemble averaging. Indeed, the fundamental frequency is assumed stationary along M time frames. Moreover, we assume that the covariance matrix is full rank in the optimal

[1]The MATLAB implementation of the method is available online at [19].

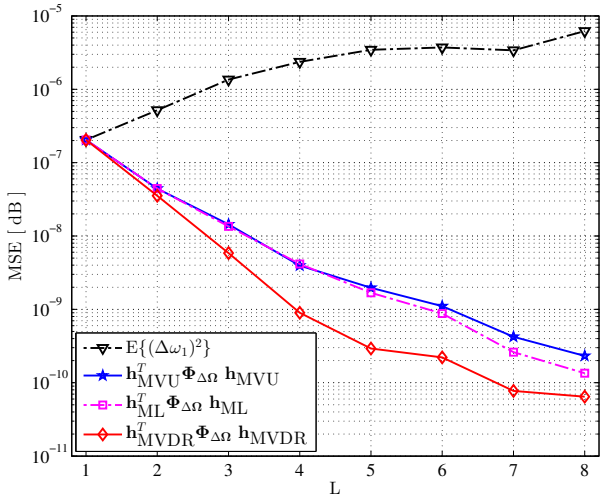

Fig. 3. MSEs of pitch estimates of a harmonic signal, contaminated by colored noise in SNR $=$ 20 dB, versus number of harmonics.

filter design (12), and this can be guaranteed by choosing a minimum number of time frames as $M \geq L$. Figure 1 shows the results of 500 fundamental frequency estimates. The results show that the MVDR filter outperforms the MVU and ML filters in colored noise, although we do not have initial estimates during $n = 1, 2, \ldots, M$ for the MVDR pitch estimator, i.e., $M = 50$.

In Figure 2, we show the MSE of the pitch estimates, when conducting 200 simulations, using a synthetic signal consisting of $L = 5$ complex sinusoids with $\omega_0 = 0.1250\pi$ and magnitudes $\mathbf{a} = [1, 1.5, 2, 1.5, 1]^T$ and uniformly distributed random phases, contaminated by colored noise, which is generated by passing a complex white Gaussian noise with zero mean and unit variance through an autoregressive (AR) filter given by $1/(1 - 0.1z^{-1} + 0.3z^{-2})$ in Z-transform. In the subspace orthogonality method, the method which we applied to estimate UFEs, the model order is assumed to be known, and the sampling window and the length of the discrete Fourier transform (DFT) respectively are $N = 128$ and $F = 65{,}536$ samples. In the SNR $=$ 20 dB, we also evaluate the results of the pitch estimates of the same signal with different number of harmonics, and we choose magnitudes like $\mathbf{a} = \text{hann}(L) + \mathbf{1}_L$ that is the function of the Hanning window plus all-ones column vector of size L. Although with a high number of harmonics, the SNR of the first harmonic is decreased, assuming the other harmonics as noise, Figure 3 indicates that pitch estimates will be more robust for a signal with a high number of harmonics. In general, the results of the constrained-harmonic methods perform better than the results of the lowest frequency estimate $\hat{\omega}_1$, and the MVDR filter outperforms the ML, which is the same as the WLS method [1].

We also conduct an experiment on a trumpet signal which

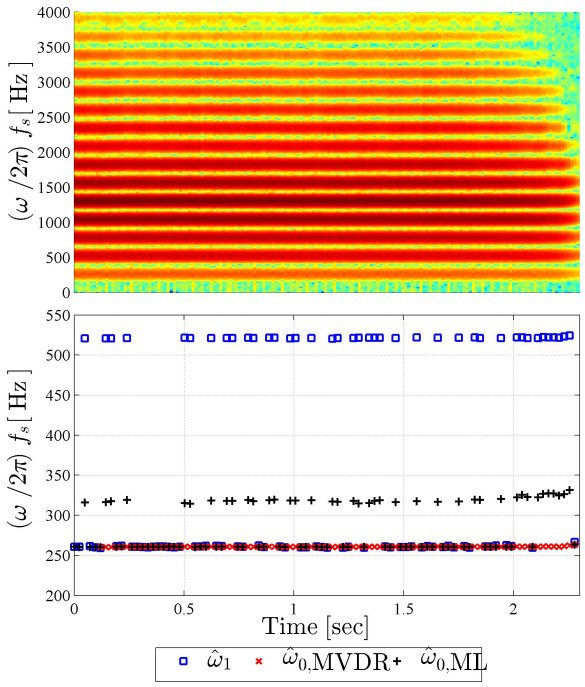

Fig. 4. Spectrogram of a trumpet signal contaminated by colored noise (top), and pitch estimates using the first element of UFE, and the ML and MVDR methods (bottom).

is contaminated by colored noise in SNR $=$ 30 dB, and the sampling frequency f_s is 8.0 kHz. We estimate model order using the subspace orthogonality method [4], and then estimate UFEs as in the previous simulations. Figure 4 indicates that the first harmonic estimate is missed in some time frames and the second harmonic is selected instead. As a result, the ML pitch estimates are wrong in these time frames, while the proposed MVDR method estimates pitch right.

5. CONCLUSION

The WLS method has been proposed as an optimal solution for the pitch estimation in [1] with the assumption of white Gaussian noise which is not often valid in real scenarios, e.g., in colored noise. In this paper, we have proposed the MVDR pitch estimator by imposing the constraint of integer harmonics to apply on UFEs. We have presented that the MVDR estimator is designed depending on narrowband SNRs of harmonics, which can be estimated from statistics of UFEs. For white Gaussian noise, we have derived the ML estimator from the MVDR estimator, that is the same as the WLS pitch estimator. Then, we evaluated the performance of the proposed method in simulations by using synthetic and real harmonic signals. The results show that the proposed method outperforms the WLS method in colored noise, even when some estimates of harmonics are missed.

6. REFERENCES

[1] H. Li, P. Stoica, and J. Li, "Computationally efficient parameter estimation for harmonic sinusoidal signals," *Elsevier Signal Process.*, vol. 80(9), 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, pp. 1917–1930, Apr. 2002.

[4] 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, pp. 1635–1644, Jul. 2007.

[5] J. R. Jensen, M. G. Christensen, and S. H. Jensen, "Nonlinear least squares methods for joint DOA and pitch estimation," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 21, pp. 923 –933, May 2013.

[6] B. Doval and X. Rodet, "Fundamental frequency estimation and tracking using maximum likelihood harmonic matching and hmms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 1, pp. 221–224 vol.1, April 1993.

[7] M. G. Christensen, J. L. Højvang, A. Jakobsson, and S. H. Jensen, "Joint fundamental frequency and order estimation using optimal filtering," *EURASIP J. on Applied Signal Process.*, vol. 2011, pp. 1–18, Jun. 2011.

[8] P. Djuric, "A model selection rule for sinusoids in white gaussian noise," *IEEE Trans. Signal Process.*, vol. 44, pp. 1744–1751, Jul 1996.

[9] J. Tabrikian, S. Dubnov, and Y. Dickalov, "Maximum a-posteriori probability pitch tracking in noisy environments using harmonic model," *IEEE Trans. Speech Audio Process.*, vol. 12, pp. 76 – 87, Jan. 2004.

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

[11] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, pp. 276–280, Mar. 1986.

[12] R. Roy and T. Kailath, "ESPRIT - estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 984–995, Jul. 1989.

[13] P. Stoica and A. Nehorai, "Statistical analysis of two non-linear least-squares estimators of sine waves parameters in the colored noise," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, pp. 2408–2411 vol.4, 1988.

[14] P. Stoica, A. Jakobsson, and J. Li, "Cisoid parameter estimation in the colored noise case: asymptotic Cramér-Rao bound, maximum likelihood, and nonlinear least-squares," *IEEE Trans. Signal Process.*, vol. 45, pp. 2048–2059, Aug. 1997.

[15] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, pp. 1408–1418, Aug. 1969.

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

[17] S. Tretter, "Estimating the frequency of a noisy sinusoid by linear regression (corresp.)," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 832–835, 1985.

[18] S. M. Kay, "Fundamentals of statistical signal processing: detection theory," 1998.

[19] M. G. Christensen, "Multi-pitch estimation toolbox." [Online]. http://www.create.aau.dk/audio, 2009.