08-230

SUB-BAND IMPLEMENTATION OF THE HARMONIC MUSIC ALGORITHM

Johan Xi Zhang[1], *Mads Græsbøll Christensen*[1], *Joachim Dahl*[1],
Søren Holdt Jensen[1], *Marc Moonen*[2]

[1]Dept. of Electronic Systems (ES-MISP), Aalborg University, Aalborg, Denmark
[2]Dept. of Electrical Engineering (ESAT-SCD), Katholieke Universiteit Leuven, Leuven, Belgium
{jxz,mgc,joachim,shj}@es.aau.dk, marc.moonen@esat.kuleuven.be

ABSTRACT

In this paper, we present a novel method for joint estimation of the order and fundamental frequency of a set of harmonically related sinusoids. This method uses a subband based approach to estimate the involved parameters using subspace techniques, and the resulting algorithm is termed Frequency-selective Harmonic MUSIC (F-HMUSIC). The performance of F-HMUSIC is evaluated and compared to both Harmonic MUSIC (HMUSIC) and Cramér-Rao lower bound (CRLB). Especially, in a low signal-to-noise ratio (SNR) with the colored noise scenarios, where F-HMUSIC outperforms HMUSIC. The F-HMUSIC is concluded to be more computationally efficient and more robust against the colored noise than other subspace based fundamental frequency estimators.

Index Terms— Fundamental frequency, pitch, sub-band, subspace, orthogonality.

1. INTRODUCTION

The problem of estimating fundamental frequencies or pitch periods has been of interest to the signal processing society throughout the years, and many sophisticated solutions have been suggested. It is especially important in many speech and audio processing applications such as linear prediction based speech coding, coding of speech and audio using a harmonic sinusoidal model, and musical information retrieval. The problem is described by considering a set of harmonic signals with fundamental frequency ω_0 embedded in noise. The fundamental frequency estimation problem can then be stated as follow:

$$y(t) = \sum_{l=1}^{L} \beta_l e^{j\omega_0 l t} + e(t), \qquad \beta_l = \alpha_l e^{j\theta_l}, \tag{1}$$

for $t = 0, \ldots, N-1$, where α_l is the real-valued amplitude of the complex exponential, ω_0 is the fundamental frequency, L is the model order, θ_l is the phase, and $e(t)$ is the complex symmetric white Gaussian noise. The estimation problem associated with the real case can be cast as (1) by the use of analytic signals, which is valid when there is little or no spectral content of interest near 0 and π.

The work of Johan Xi Zhang is supported by the Marie Curie EST-SIGNAL Fellowship (http://est-signal.i3s.unice.fr), contract no. MEST-CT-2005-021175.

The work of Mads Græsbøll Christensen is supported by the Parametric Audio Processing project, Danish Research Council for Technology and Production Sciences, grant no. 274060521.

Fundamental frequency estimators are typically time-domain techniques based on autocorrelation, cross-correlation and the average magnitude difference function [1]. Recently, a fundamental frequency estimator based on subspace techniques termed Harmonic MUltiple SIgnal Classification (HMUSIC) has been suggested in [2], based on the MUSIC algorithm used in spectral analysis by imposing the assumed harmonic structure in (1) on the MUSIC criterion. HMUSIC is able to jointly estimate the order and the fundamental frequency where good statistical performance has been shown compared to both Markov-like weighted least squares estimator (WLS) and Cramér-Rao lower bound (CRLB) [2]. However, the computational complexity as well as the sensitivity to colored noise are still considered as major drawbacks of HMUSIC. Traditionally prewhitening is one of the standard methods to reshape colored noise, but in speech and audio signals the noise is usually non-stationary, which makes pre-whiteneing based on the noise characteristic hard to achieve. Recently, few papers have shown advantages on the computational efficiency in subspace based frequency estimators by processing the signal using a frequency-selective (FS) data model instead of the traditional covariance matrix model [3, 4, 5, 6].

In this paper we will further develop the concept used in the HMUSIC, where we propose a new joint order and fundamental frequency estimation algorithm termed Frequency-selective Harmonic MUSIC (F-HMUSIC) using ideas from frequency-selective MUSIC (F-MUSIC) [5]. The F-HMUSIC is a subband based approach where the signal spectrum is divided into Q equally spaced subbands using the FS matrix model and considering each band as an individual subproblem. This approach gives a more computationally efficient algorithm than subspace decomposition directly on the covariance matrix used in HMUSIC [5]. Furthermore, by averaging the estimated fundamental frequency from subbands will possibly be more robust to the colored noise.

2. SOME PRELIMINARIES

In this section, we present the fundamentals of the FS data model and introduce some useful vector and matrix notations. We start by defining the FS data model which is formulated using the formulations defined in [7], where the samples from a discrete Fourier transform (DFT) are used as input data. The given signal sequence (1) is first Fourier transformed using N points FFT. Let us then assume that the component of interest lie in a prespecified subband composed of the following Fourier frequencies:

$$\left\{ \tfrac{2\pi}{N} k_{1,m} \quad \tfrac{2\pi}{N} k_{2,m} \quad \ldots \quad \tfrac{2\pi}{N} k_{M,m} \right\}, \tag{2}$$

where m denotes the subband index of Q equally divided subbands, and the $\{k_{1,m} \ldots k_{M,m}\}$ are M given consecutive integers. The num-

ber of components L_m of (1) lying in the subband specified by (2) is assumed to be $L_m \leq L$.

For the derivation of the FS data model and the orthogonality principle, following notations will be used:

$$w_k = e^{j\frac{2\pi}{N}k}, \quad k = 0, 1, \ldots, N-1 \tag{3}$$

$$\mathbf{u}_k = \begin{bmatrix} w_k & \ldots & w_k^s \end{bmatrix}^T \tag{4}$$

$$\mathbf{v}_k = \begin{bmatrix} 1 & w_k & \ldots & w_k^{N-1} \end{bmatrix}^T \tag{5}$$

$$\mathbf{y} = \begin{bmatrix} y(0) & \ldots & y(N-1) \end{bmatrix}^T \tag{6}$$

$$Y_k = \mathbf{v}_k^* \mathbf{y}, \quad k = 0, 1, \ldots, N-1 \tag{7}$$

where $\mathbf{u}_k$ is the phase shift vector, $\mathbf{v}_k$ is the Fourier vector, $\mathbf{y}$ is the signal vector, $*$ is the hermitian transpose, T is the vector transposition operator, and s is a user parameter.

The FS data model for a subband with index m is decomposed using either Singular value decomposition (SVD) or Eigen value decomposition (EVD) which is given as [7, 3]:

$$\mathbf{Y}_m \mathbf{\Pi}_m^{\perp} = \mathbf{H}_m \mathbf{\Lambda}_m \mathbf{V}_m. \tag{8}$$

with $\mathbf{Y}_m \in \mathbb{C}^{s \times M}$, and $\mathbf{\Pi}_m^{\perp} \in \mathbb{C}^{M \times M}$. The involved matrices in (8) are given as:

$$\mathbf{Y}_m = \begin{bmatrix} \mathbf{u}_{k_{1,m}} Y_{k_{1,m}} & \ldots & \mathbf{u}_{k_{M,m}} Y_{k_{M,m}} \end{bmatrix} \tag{9}$$

$$\mathbf{\Pi}_m^{\perp} = \mathbf{I} - \mathbf{U}_m^*(\mathbf{U}_m \mathbf{U}_m^*)^{-1}\mathbf{U}_m \tag{10}$$

$$\mathbf{U}_m = \begin{bmatrix} \mathbf{u}_{k_{1,m}} & \ldots & \mathbf{u}_{k_{M,m}} \end{bmatrix}, \tag{11}$$

where $\mathbf{Y}_m$ is the compact matrix form of DFT data samples (7) for a given k multiplied with their corresponding phase shift vectors (4), and $\mathbf{\Pi}_m^{\perp}$ is the projection matrix which projects onto the nullspace of $\mathbf{U}_m$ [3, 7]. The decomposed matrix $\mathbf{H}_m$ in (8) is written as

$$\mathbf{H}_m = \begin{bmatrix} \mathbf{h}_{1,m} & \mathbf{h}_{2,m} & \ldots & \mathbf{h}_{s,m} \end{bmatrix}, \tag{12}$$

where the columns of $\mathbf{H}_m$ containing the singular vectors of the signal and the noise subspaces, and $\mathbf{\Lambda}_m$ is a diagonal matrix containing the corresponding singular values. Furthermore, let $\mathbf{S}_m$ and $\mathbf{G}_m$ be the orthonormal subspaces denoted as follows:

$$\mathbf{S}_m = \begin{bmatrix} \mathbf{h}_{1,m} & \mathbf{h}_{2,m} & \ldots & \mathbf{h}_{L_m,m} \end{bmatrix} \tag{13}$$

$$\mathbf{G}_m = \begin{bmatrix} \mathbf{h}_{L_m+1,m} & \mathbf{h}_{L_m+2,m} & \ldots & \mathbf{h}_{s,m} \end{bmatrix}, \tag{14}$$

with $\mathbf{S}_m$ being the signal subspace associated with L_m principal singular values, and $\mathbf{G}_m$ being the orthonormal noise subspace associated with $s - L_m$ singular values. To model the singular vectors of the signal subspace a Vandermonde vector $\mathbf{a}(\omega_k)$ is introduced, and given as:

$$\mathbf{a}(\omega_k) = \begin{bmatrix} e^{j\omega_k} & \ldots & e^{js\omega_k} \end{bmatrix}^T, \tag{15}$$

which is orthogonal to $\mathbf{G}_m$ for frequencies $\omega_k = \omega_0 l$ where $l = 1, ..., L_m$. The cost function of F-MUSIC is formed as [4]:

$$P(\omega) = |\mathbf{a}^*(\omega)\mathbf{G}_m\mathbf{G}_m^*\mathbf{a}(\omega)|^2, \tag{16}$$

where the estimates are found on every ω_k of (16) which is orthogonal to $\mathbf{G}_m$.

3. PROPOSED METHOD

Herein, we will extend the cost function of F-MUSIC for jointly estimate the fundamental frequency and the order of harmonic signals distributed over several subbands. The spectrum of the signal is divided into Q number of equally spaced subbands where the number of subbands containing harmonics goes from $m = 1, \ldots, Q'$, and $Q - Q'$ are the remaining bands without harmonics. In this paper the number of band Q' containing the harmonics are assumed to be known due to the limitation of the MUSIC algorithm where the limit is $L \geq 1$. The estimation of Q' is a simple detection problem that will not be discussed here. The number of harmonics L distributed into subbands, is defined by the model as:

$$L = \sum_{m=1}^{Q'} L_m(\omega_0), \tag{17}$$

where the function $L_m(\omega_0)$ stands for number of harmonics in the band m with respect to the frequency ω_0. The function $L_m(\omega_0)$ is given as:

$$L_m(\omega_0) = \left\lfloor \frac{\omega_m}{\omega_0} - \sum_{i=1}^{m-1} \left\lfloor \frac{\omega_i}{\omega_0} \right\rfloor \right\rfloor, \tag{18}$$

with $\omega_m = \frac{2\pi}{N} k_{M,m}$ being the highest frequency in the subband m.

The Vandermonde matrix $\mathbf{A}_m \in \mathbb{C}^{s \times L_m(\omega_0)}$ modeling the harmonics in the desired subband with index m consist of (15) applied on different frequencies, and given as follows:

$$\mathbf{A}_m = \begin{bmatrix} \mathbf{a}(\omega_0 + \varphi) & \ldots & \mathbf{a}(\omega_0 L_m(\omega_0) + \varphi) \end{bmatrix}, \tag{19}$$

where $L_0(\omega_0) = 0$, and the term $\varphi = \omega_0 L_{(m-1)}(\omega_0)$ being a frequency offset off harmonics on the frequency ω_0 with respect to the previous band. The offset term is important in order to unify the estimation of ω_0 in subbands containing harmonics. In each band the signal and the noise subspaces are orthogonal, and defined as:

$$\left\| \mathbf{A}_m^H \mathbf{G}_m \right\|_F = 0, \tag{20}$$

for frequency ω_0, with $\|\cdot\|_F$ being the Forbenius norm.

The two dimensional cost function for the joint order and fundamental frequency estimator is defined as:

$$J(\omega_0, L) = \frac{1}{Q'} \sum_{m=1}^{Q'} \frac{\|\mathbf{A}_m^H \mathbf{G}_m\|_F^2}{L_m(\omega_0) s (s - L_m(\omega_0))}, \tag{21}$$

where the denominator is a scaling factor which makes the noise floor of the cost function invariant to the changing matrix dimensions of $\mathbf{A}_m$ and $\mathbf{G}_m$ [2].

The order L and the fundamental frequency ω_0 are the desired estimates when values for which the Vandermonde matrix is closest to being orthogonal to the noise subspace, formulated as:

$$\hat{\omega} = \arg \min_{\omega_0 \in \Omega} \min_{L \in \mathcal{L}} J(\omega_0, L), \tag{22}$$

with Ω being the searching space for the fundamental frequency, and $\mathcal{L}$ being the space for the order estimation. It should be noted that due to the knowledge of Q' the candidates of $\mathcal{L}$ can be reduced to $L_{Q'-1}(\omega_0) \leq L \leq L_{Q'}(\omega_0)$ which is an advantage in order to reduce the computational complexity related to computations of the cost function. The resolution of the algorithm is mainly dependent on the parameters such as the data length N, the number of subband Q, and the user parameter s. Previous experience of similar

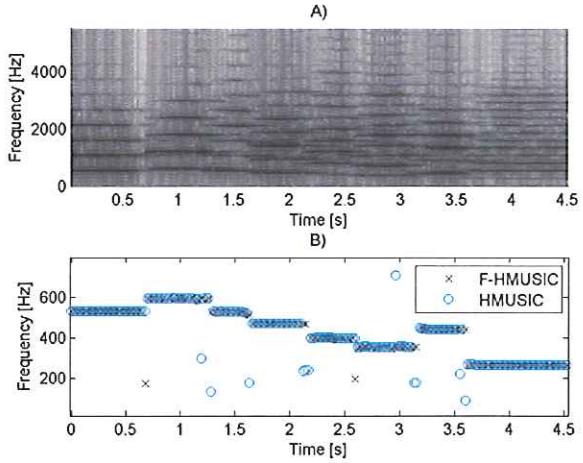

Fig. 1. *A) Spectrogram of the trumpet signal. B) Fundamental frequencies estimated using F-HMUSIC and HMUSIC.*

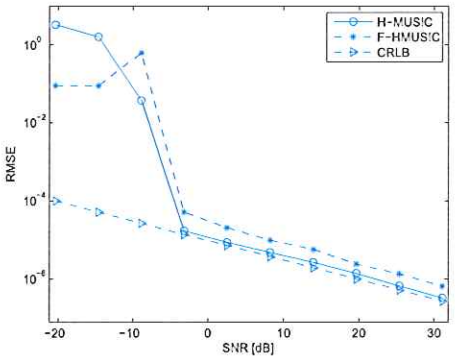

Fig. 2. *The calculated RMSE of F-HMUSIC compared with HMUSIC and CRLB, versus different SNR with the white noise.*

approaches show that this parameter s may be selected as large as possible to increase the number of linearly independent vectors in the noise subspace, but still less than M in order to still achieve a correct estimate of the FS data model [7].

Like HMUSIC our proposed method can also be efficiently implemented using FFT based method described in [2]. For applications that require a very accurate estimates for a given L a gradient search algorithm with a slightly modification of the method described in [2] can be used.

4. NUMERICAL EXAMPLES

First, an audio example using a trumpet signal is evaluated where the spectrogram of the signal is shown in Fig. 1A). The sampling frequency was $f_s = 11025$Hz and segments with $N = 512$ were used to calculate the time-frequency representation of the signal. The model order of HMUSIC was set to $\lfloor 0.85N \rfloor$, and for F-HMUSIC $s = \lfloor 0.85M \rfloor$. In the proposed method two subband was selected where $Q' = 1$. The cost function was evaluated on a 0.5 Hz grid from 80 to 1000 Hz. The estimated fundamental frequencies is

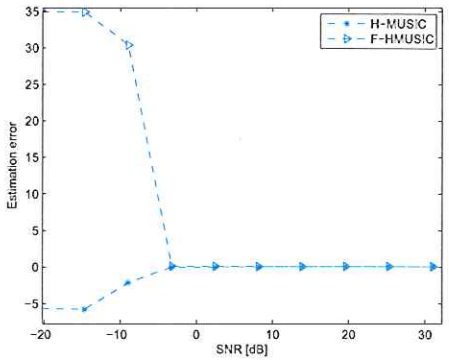

Fig. 3. *The order estimation errors calculated using F-HMUSIC and HMUSIC with the white noise.*

shown in Fig. 1 B), both methods are able to provide good estimates of the signal.

The statistical properties of the proposed method will be evaluated next using the Monte-Carlo simulation. The signal setup used in the following examples will consist of $L = 13$ complex exponentials embedded in noise with a fundamental frequency of $\omega_o = 0.15$, and amplitudes $\alpha_l = 1\ \forall l$, and observed data is a sequence of $N = 1024$ samples. Furthermore the spectrum of region 0 to π is divided into $Q = 2$ subbands with $M = \frac{N}{2Q}$, and the harmonics is occupied by both bands ($Q' = 2$). The cost function in (21) is evaluated for candidates in the interval Ω. In both methods first a coarse estimate of the fundamental frequency and the order are given using the FFT-based method [2]. Later, a refined estimate based on the gradient search method using [2]. The calculated root mean square estimation error (RMSE) of F-HMUSIC is compared to both HMUSIC and exact CRLB. We use the signal-to-noise ratio (SNR) defined as,

$$SNR = 10 \log_{10} \sum_{l=1}^{L} \frac{\alpha_l^2}{\phi(\omega_l)}, \tag{23}$$

where the function $\phi(\omega_l)$ is the power spectrum of the noise at frequency $\omega_l = \omega_0 l$. The exact CRLB is calculated for both the white and the colored noise cases using equations stated in [2, 8, 9]. The order error is defined as the difference between the estimated order subtracted on the true order.

In the first example the noise is a symmetric white Gaussian noise with SNR calculated as (23) with the $\phi(\omega_l)$ being the variance of the white noise. The candidates for ω_0 was set to $\Omega \in [0.06, 0.4]$ for both algorithms. Note that this interval includes $2\omega_0$ and $\frac{1}{2}\omega_0$, so any potential problems with spurious estimates at these frequencies would show up in the statistical evaluation. For each possible fundamental frequency $\omega_0 \in \Omega$, the model orders considered were $\mathcal{L} \in [5, \lfloor \pi/\omega_0 \rfloor - 1]$. The estimated RMSE versus SNR, and the associated order errors are plotted in Fig. 2 and 3. As expected, due to the computational savings the performance of F-HMUSIC is a bit worse than HMUSIC in case of the white noise, shown in Fig. 2. The order estimation performance of both algorithms are shown in Fig 3, where F-HMUSIC has the tendency to overestimate the order in low SNR conditions on region when the algorithm performance breaks down. From numerous simulations it can be concluded that F-HMUSIC is sensitive to the noise if lower part of the searching space Ω is selected to be too low. The performance of the estima-

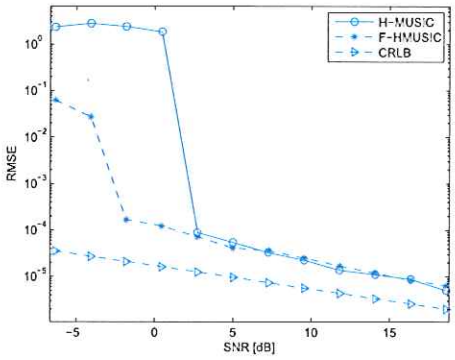

Fig. 4. *The calculated RMSE of F-HMUSIC compared with HMUSIC and CRLB, versus different SNR with the colored noise.*

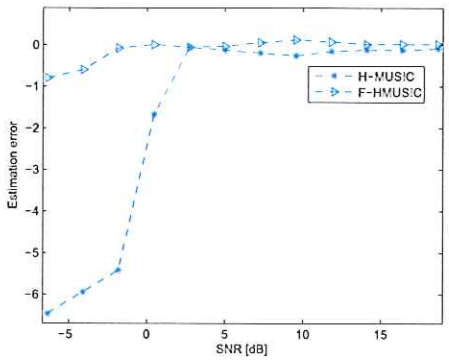

Fig. 5. *The order estimation errors calculated using F-HMUSIC and HMUSIC with the colored noise.*

tor can be significantly increased if Ω is moved higher up in the frequency scale or select s larger, but the RMSE evaluated on each SNR will still be a bit worse than HMUSIC in the white noise scenario. Fortunately the upper part of Ω will not significantly influence the estimation performance. However for many applications the degraded performance might still be attractive because of the computational savings due to subspace decomposition on the individual subbands [4].

Another example shown here will demonstrate the robustness to colored noise. The goal here is to show that by formulating the problem into subbands, and estimating parameters from the averaged cost function of involving bands will reduce the influence of the colored noise. The same signal setup is used except that white noise is filtered with a second order AR process ($\frac{1}{1+0.3z^{-1}+0.8z^{-2}}$), where the power of the colored noise is mainly concentrated on the subband containing higher frequencies. The SNR of the colored noise is calculated using (23) where function $\phi(\omega_l)$ is the power spectrum of the filtered noise at frequency ω_l. A new regions is used for both algorithms where $\Omega \in [0.1, 0.4]$, and the order space was calculated as the previous example. The evaluated RMSE of F-HMUSIC is compared with HMUSIC and CRLB, and the order estimates are shown in respectively Fig. 4 and 5. It shows that HMUSIC fails under "low" SNR with the colored noise conditions while F-HMUSIC still can achieve good estimates close to CRLB. The performance drop down of HMUSIC can be explained by the phenomena termed subspace swapping where part of the noise subspace erroneously determined as the signal subspace while estimates provided by F-HMUSIC is from the averaged mean cost function between the involving subbands. Therefore the errors caused due to the subspace swapping are reduced.

5. CONCLUSION

In this paper, a joint order and fundamental frequency estimator termed F-HMUSIC has been proposed. This algorithm is a frequency domain based estimator using subspaces decomposed from the FS data model to efficiently estimate the fundamental frequency, where a subband based approach is addopted to reduce the sensitivity to the colored noise and increase the computational efficiency. The performance of F-HMUSIC has been evaluated and compared to both HMUSIC and CRLB. From simulations it has been shown that F-HMUSIC is more robust against colored noise in low SNR scenarios compared to HMUSIC. In general the price to be paid for the reduced computational complexity and increased robustness to the colored noise is a slightly reduction in the estimation accuracy.

6. REFERENCES

[1] A. de Cheveigne and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *J. Acoust. Soc. of Am.*, vol. 111(4), pp. 1917–1930, 2002.

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

[3] T. McKelvey and M. Viberg, "A robust frequency domain subspace algorithm for multi-component harmonic retrieval," in *Rec. of 34th Asilomar Conference on Signals, Systems and Computers*, 2001.

[4] J. X. Zhang, M. G. Christensen, J. Dahl, S. H. Jensen, and M. Moonen, "Frequency-domain parameter estimations for binary masked signals," in *Proc. Interspeech*, 2008.

[5] J. X. Zhang, M. G. Christensen, J. Dahl, S. H. Jensen, and M. Moonen, "Robust implementation of the MUSIC algorithm," in *IEEE Int. Conf. on Acoust., Speech, Signal Processing*, 2009, Submitted.

[6] M. G. Christensen A. Jakobsson and S. H. Jensen, "Frequency selective sinusoidal order estimation," *IEEE Electronic Letters*, vol. 43(21), pp. 1164–1165, 2007.

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

[8] J. M. Francos and B. Friedlander, "Bounds for estimation of complex exponentials in unknown colored noise," *IEEE Trans. Signal Process.*, vol. 43, pp. 2176–2185, 1995.

[9] P. Stoica, A. Jakobsson, and J. Li, "Cisoid parameter estimation in the colored noise case: asymptotic cramer-rao bound, maximum likelihood, and nonlinear least-squares," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 45, no. 8, pp. 2048–2059, 1997.