Two stage DOA and Fundamental Frequency Estimation based on Subspace Techniques

Zhou, Zhenhua; Christensen, Mads Græsbøll; So, Hing-Cheung

Published in:
International Conference on Signal Processing. Proceedings

DOI (link to publication from Publisher):
10.1109/ICoSP.2012.6491637

Publication date:
2012

Document Version
Early version, also known as pre-print

Two Stage DOA and Fundamental Frequency Estimation Based on Subspace Techniques

Zhenhua Zhou†, M. G. Christensen‡, H. C. So†
†Department of Electronic Engineering, City University of Hong Kong, Hong Kong SAR, China
‡Department of Architecture, Design and Media Technology, Aalborg University, Aalborg, Denmark
E-mail: sjtu.zzh2010@gmail.com, mgc@create.aau.dk, hcso@ee.cityu.edu.hk

Abstract—**In this paper, the problem of fundamental frequency and direction-of-arrival (DOA) estimation for multi-channel harmonic sinusoidal signal is addressed. The estimation procedure consists of two stages. Firstly, by making use of the subspace technique and Markov-based eigenanalysis, a multi-channel optimally weighted harmonic multiple signal classification (MCOW-HMUSIC) estimator is devised for the estimation of fundamental frequencies. Secondly, the spatio-temporal multiple signal classification (ST-MUSIC) estimator is proposed for the estimation of DOA with the estimated frequencies. Statistical evaluation with synthetic signals shows the high accuracy of the proposed methods compared with their non-weighting versions.**

Keywords-fundamental frequency estimation; DOA estimation; subspace method; harmonic sinusoidal signal; Markov optimum weighting

I. INTRODUCTION

The problem of fundamental frequency or pitch estimation has been of interest to the signal processing community for many years, finding applications in a wide range of areas such as audio and speech coding, classification of music, and speech analysis. Kinds of algorithms have been proposed for both the single-pitch and multi-pitch scenarios. The nonlinear least squares (NLS) estimator [1] is statistically efficient for single pitch scenario, but performs poorly in the multi-pitch scenario due to the decoupling difficulty [2]. In [2]-[3], the harmonic multiple signal classification (HMUSIC) algorithm is proposed for single-pitch and multi-pitch estimation. However, the HMUSIC estimation is not statistically efficient, and its accuracy cannot attain Cramér-Rao lower bound (CRLB) [4].

Recently, multi-channel approaches have attracted considerable attention in both single-pitch and multi-pitch scenarios [5]-[7], where both direction-of-arrival (DOA) estimation and fundamental frequency estimation are of importance. In [8], the multi-channel multi-pitch harmonic multiple signal classification (MC-HMUSIC) is presented for solving the problem of joint fundamental frequency and DOA estimation. The simulations show its increased accuracy in DOA and fundamental frequency estimation compared with other state-of-the-art method. Nevertheless, the two-dimensional search is utilized for the estimates of fundamental frequency and DOA, which makes the computation complexity high. Still, its fundamental frequency estimation is not statistically efficient.

To overcome the disadvantages of the MC-HMUSIC estimator, we propose to estimate the fundamental frequency and DOA in two stages. At first, the multi-channel optimally weighted harmonic multiple signal classification (MCOW-HMUSIC) estimator is devised, and the estimation of fundamental frequency is conducted. The MCOW-HMUSIC estimator is the extension of the HMUSIC estimator to the multi-channel scenario with optimum weighting. As mentioned, the HMUSIC estimator is not statistically efficient, and there is a gap between its mean square error and CRLB. Therefore, a kind of Markov optimum weighting is utilized to improve its accuracy. To facilitate the statistical analysis of orthogonality error, we modify the HMUSIC formulation. Then we analyze the perturbation of the orthogonality error, and derive the weighting matrix. The fundamental frequency estimates are calculated in an iterative manner. Simulation results show that the MCOW-HMUSIC estimator performs better than its non-weighting version. Especially in the single-pitch scenario, its performance can attain the CRLB. Secondly, we make use of the spatio-temporal multiple signal classification (ST-MUSIC) estimator to estimate the DOA with the estimated fundamental frequency. Since we estimate the frequency and DOA at two stages, the computationally intensive two-dimensional search in [8] is avoided.

The rest of this paper is organized as follows. The proposed estimators for fundamental frequency and DOA, MCOW-HMUSIC and ST-MUSIC, are developed in Section II. The statistical properties of the orthogonality error is analyzed, and the optimum weighting matrix is derived. In Section III, simulation results are included to show the performance of the proposed estimators by comparing with their non-weighting versions as well as CRLB. Finally, conclusions are drawn in Section IV.

II. ALGORITHM DEVELOPMENT

A. Spatio-Temporal Signal Model

Without multi-path propagation of sources, the multi-channel signal model is given as follows. The signal $x_i(n)$ received by the microphone element (channel) i arranged in a uniform linear array (ULA) configuration, $i = 1, \cdots, I$, is given by [8]

$$x_i(n) = s_i(n) + q_i(n), \tag{1}$$

$$s_i(n) = \sum_{k=1}^{K} \sum_{l=1}^{L_k} A_{l,k} e^{j(\omega_k l n + \alpha_k l (i-1) + \phi_{l,k})}, \tag{2}$$

for $n = 1, \cdots, N$, with ω_k, α_k, $\{A_{l,k}\}$ and $\{\phi_{l,k}\}$ denoting unknown fundamental frequency, scaled DOA, amplitudes and initial phases of the *k*-th source. The number of sources K and the number of harmonics L_k of each source k, are assumed to be known, or found in some way such as [9]. The $q_i(n)$ $(i = 1, \cdots, I)$ are assumed to be uncorrelated white Gaussian noise with variance σ^2. The objective is to estimate the nonlinear parameters ω_k and α_k. Once they are estimated, the remaining linear parameters can be estimated as a linear least squares (LLS) solution.

B. Fundamental Frequency Estimation

Firstly, we focus on the single-pitch case, that is, $K = 1$. Construct the data matrix for each channel i [10]:

$$\mathbf{X}_i = \mathbf{S}_i + \mathbf{Q}_i, \tag{3}$$

where $\mathbf{X}_i$ is the $M \times (N - M + 1)$ data matrix with elements $[\mathbf{X}_i]_{m,n} = x_i(m + n - 1)$, $L_1 < M < N + 1$. The matrices $\mathbf{S}_i$ and $\mathbf{Q}_i$ are the noise-free and noise components of $\mathbf{X}_i$, respectively. It is straightforward that $\mathbf{S}_i$ can be factorized as:

$$\mathbf{S}_i = \mathbf{A}\mathbf{\Gamma}_i\mathbf{H}^T, \tag{4}$$

where

$$\mathbf{\Gamma}_i = \mathrm{diag}(\gamma_1, \gamma_2, \cdots, \gamma_{L_1}), \tag{5}$$

$$\gamma_l = A_{l,1}e^{j(\alpha_1 l(i-1)+\phi_{l,1})}, \tag{6}$$

$$\mathbf{A} = [\mathbf{a}_{1,1}, \mathbf{a}_{1,2}, \cdots, \mathbf{a}_{1,L_1}], \tag{7}$$

$$\mathbf{H} = [\mathbf{h}_{1,1}, \mathbf{h}_{1,2}, \cdots, \mathbf{h}_{1,L_1}], \tag{8}$$

$$\mathbf{a}_{1,l} = \left[a_{1,l}\ a_{1,l}^2\ \cdots\ a_{1,l}^M\right]^T, \tag{9}$$

$$\mathbf{h}_{1,l} = \left[1\ h_{1,l}\ \cdots\ h_{1,l}^{N-M}\right]^T, \tag{10}$$

for $l = 1, \cdots, L_1$, and $a_{1,l} = h_{1,l} = e^{jl\omega_1}$. On the other hand, $\mathbf{X}_i$ can be decomposed using singular value decomposition (SVD) as:

$$\mathbf{X}_i = \mathbf{U}_i\mathbf{\Lambda}_i\mathbf{V}_i^H = [\mathbf{U}_{si}\ \mathbf{U}_{ni}]\begin{bmatrix}\mathbf{\Lambda}_{si} & \mathbf{0}\\ \mathbf{0} & \mathbf{\Lambda}_{ni}\end{bmatrix}[\mathbf{V}_{si}\ \mathbf{V}_{ni}]^H, \tag{11}$$

with $\mathbf{U}_{si}$ and $\mathbf{U}_{ni}$ spanning the signal and noise subspace, respectively. Comparing (4) and (11), the HMUSIC estimate of ω_1 from channel i, denoted by $\hat{\omega}_{1i}$, is obtained by minimizing the orthogonality error between $\mathbf{A}$ and $\mathbf{U}_{ni}$:

$$\hat{\omega}_{1i} = \arg\min_{\tilde{\omega}}\ P_i(\tilde{\omega}) = \arg\min_{\tilde{\omega}}\ ||\mathbf{A}^H\mathbf{U}_{ni}||_{\mathrm{F}}^2, \tag{12}$$

with $\tilde{\omega}$ the variable for ω_1, and $||\cdot||_{\mathrm{F}}$ denoting the Frobenius norm. As demonstrated in [2]-[3], this kind of estimator is not statistically efficient. To improve its accuracy, we introduce Markov optimum weighting. Due to the non-uniqueness of the columns of $\mathbf{U}_{ni}$, that is any linear combination of these vectors still spans the noise space, it is impossible to analyze their statistical properties. To overcome such difficulty, we reformulate the HMUSIC estimator to a different but equivalent form as follows:

$$\begin{aligned}\hat{\omega}_{1i} &= \arg\min_{\tilde{\omega}}\ ||\mathbf{A}^H\mathbf{U}_{ni}||_{\mathrm{F}}^2\\ &= \arg\min_{\tilde{\omega}}\ ||\mathbf{A}^H\mathbf{U}_{ni}\mathbf{U}_{ni}^H||_{\mathrm{F}}^2,\\ &= \arg\min_{\tilde{\omega}}\ \mathbf{e}_i^H\mathbf{e}_i\end{aligned} \tag{13}$$

where $\mathbf{e}_i$ is the orthogonality error vector of the form $\mathbf{e}_i = \mathrm{vec}(\mathbf{A}^H\mathbf{U}_{ni}\mathbf{U}_{ni}^H)$ with $\mathrm{vec}(\cdot)$ being the vectorization operator. According to the results of [11], and after some extra manipulations, the first-order approximation of $\mathbf{e}_i$ originating from $\mathbf{U}_{ni}\mathbf{U}_{ni}^H$ is derived as

$$\begin{aligned}\mathbf{e}_i &\approx -\mathrm{vec}(\mathbf{A}^H\mathbf{U}_{si}\mathbf{\Lambda}_{si}^{-1}\mathbf{V}_{si}^H\Delta\mathbf{X}_i^H\mathbf{U}_{ni}\mathbf{U}_{ni}^H)\\ &= -(\overline{\mathbf{U}}_{ni}\mathbf{U}_{ni}^T)\otimes(\mathbf{A}^H\mathbf{U}_{si}\mathbf{\Lambda}_{si}^{-1}\mathbf{V}_{si}^H)\mathrm{vec}(\Delta\mathbf{X}_i^H)\\ &= -(\overline{\mathbf{U}}_{ni}\mathbf{U}_{ni}^T)\otimes(\mathbf{A}^H\mathbf{U}_{si}\mathbf{\Lambda}_{si}^{-1}\mathbf{V}_{si}^H)\mathbf{T}\overline{\mathbf{q}}_i\\ &= \mathbf{D}_i\overline{\mathbf{q}}_i\end{aligned}, \tag{14}$$

where $\otimes$ stands for the Kronecker product, $\Delta\mathbf{X}_i$ is the perturbation of $\mathbf{X}_i$, and $\mathbf{q}_i$, $\mathbf{T}$ are defined as

$$\mathbf{q}_i = [q_i(1)\ q_i(2)\ \cdots\ q_i(N)]^T, \tag{15}$$

$$\mathbf{T} = \left[\mathbf{E}_1^T\ \cdots\ \mathbf{E}_M^T\right]^T, \tag{16}$$

$$\mathbf{E}_j = \left[\mathbf{0}_{N'\times(j-1)}\ \mathbf{I}_{N'\times N'}\ \mathbf{0}_{N'\times(M-j)}\right], \tag{17}$$

$N' = N - M + 1$, for $j = 1, \cdots, M$. Based on the above perturbation analysis, the Markov optimum weighting matrix for the channel i, is determined as [12]

$$\mathbf{W}_i = \left[E\{\mathbf{e}_i\mathbf{e}_i^H\}\right]^{\dagger}\sigma^2 \approx (\mathbf{D}_i\mathbf{D}_i^H)^{\dagger}, \tag{18}$$

where † stands for pseudoinverese. Then for each channel i, ω_1 is estimated as

$$\hat{\omega}_{1i} = \arg\min_{\tilde{\omega}}\ J_i(\tilde{\omega}) = \arg\min_{\tilde{\omega}}\ \mathbf{e}_i^H\mathbf{W}_i\mathbf{e}_i. \tag{19}$$

When dealing with the multi-channel signal, we stack the error vectors $\mathbf{e}_i$ as $\mathbf{e} = \left[\mathbf{e}_1^T,\ \mathbf{e}_2^T,\ \cdots,\ \mathbf{e}_I^T\right]^T$, and the fundamental frequency estimate $\hat{\omega}_1$ is computed as

$$\hat{\omega}_1 = \arg\min_{\tilde{\omega}}\ \mathbf{e}^H\mathbf{W}\mathbf{e}, \tag{20}$$

where $\mathbf{W} = \left[E\left\{\mathbf{e}\mathbf{e}^H\right\}\right]^{\dagger} = \mathrm{blkdiag}(\mathbf{W}_1, \mathbf{W}_2, \cdots, \mathbf{W}_I)$ due to the uncorrelatedness among the I channels. Then the fundamental frequency estimate $\hat{\omega}_1$ is expressed as

$$\hat{\omega}_1 = \arg\min_{\tilde{\omega}}\ \textstyle\sum_{i=1}^I \mathbf{e}_i^H\mathbf{W}_i\mathbf{e}_i. \tag{21}$$

As $\mathbf{W}_i$ $(i = 1, \cdots, I)$ are functions of the unknown ω_1, the following iterative procedure is employed to solve for $\hat{\omega}_1$:

- Step 1. Set $\mathbf{W}_i = \mathbf{I}_{ML_1}$ which is the $ML_1 \times ML_1$ identity matrix.
- Step 2. Find $\hat{\omega}_1$ by searching for the minimum of (21).
- Step 3. Compute $\mathbf{W}_i$ by (18).
- Step 4. Repeat Steps 2 and 3 until a stopping criterion is reached.

This estimator is termed as the MCOW-HMUSIC method.

We now proceed to focus on the multi-pitch case. Following the idea of the single-pitch case, $\mathbf{S}_i$ can be factorized as:

$$\mathbf{S}_i = \mathbf{A}\boldsymbol{\Gamma}_i\mathbf{H}^T, \tag{22}$$

where $\mathbf{A} = [\mathbf{A}_1\ \mathbf{A}_2\ \cdots\ \mathbf{A}_K]$, $\mathbf{A}_k = [\mathbf{a}_{k,1}, \mathbf{a}_{k,2}, \cdots, \mathbf{a}_{k,L_k}]$, $\mathbf{a}_{k,l} = \left[a_{k,l}\ a_{k,l}^2\ \cdots\ a_{k,l}^M\right]^T$, and $a_{k,l} = e^{jl\omega_k}$. $\boldsymbol{\Gamma}_i$ and $\mathbf{H}$ can be found similarly. On the other hand, the left singular vectors of the data matrix $\mathbf{X}_i$ associated with its $(M - \sum_{k=1}^{K} L_k)$ smallest singular values span the noise subspace $\mathbf{U}_{ni}$. Then the orthogonality error vector for each source k is given as $\mathbf{e}_{i,k} = \mathrm{vec}(\mathbf{A}_k^H\mathbf{U}_{ni}\mathbf{U}_{ni}^H)$. Following the similar steps like finding the initial values of fundamental frequencies, constructing weighting matrices of (18) for each of them, and searching for its closest minimum of (21), we can solve the MCOW-HMUSIC estimates of the multi-pitch fundamental frequencies in an iterative way.

C. DOA Estimation

First, the N time-domain data samples of the array output $\mathbf{x}(n) = [x_1(n)\ x_2(n)\ \cdots\ x_I(n)]^T$ are collected to form the $I \times N$ data matrix $\mathbf{X}$, which has the form of

$$\mathbf{X} = [\mathbf{x}(1)\ \mathbf{x}(2)\ \cdots\ \mathbf{x}(N)], \tag{23}$$

and can also be expressed as

$$\mathbf{X} = \mathbf{C}\left[\boldsymbol{\Phi}\mathbf{b}\ \boldsymbol{\Phi}^2\mathbf{b}\ \cdots\ \boldsymbol{\Phi}^N\mathbf{b}\right] + \mathbf{Q}, \tag{24}$$

where $\mathbf{Q} \in \mathbb{C}^{I\times N}$ is the noise part of $\mathbf{X}$, $\mathbf{C} = [\mathbf{C}_1\ \mathbf{C}_2\ \cdots\ \mathbf{C}_K]$, $\mathbf{C}_k = [\mathbf{c}_{k,1}, \mathbf{c}_{k,2}, \cdots, \mathbf{c}_{k,L_k}]$, $\mathbf{c}_{k,l} = \left[1\ c_{k,l}\ \cdots\ c_{k,l}^{(I-1)}\right]^T$, $c_{k,l} = e^{jl\alpha_k}$, $\boldsymbol{\Phi} = \mathrm{diag}(\boldsymbol{\Phi}_1, \cdots, \boldsymbol{\Phi}_K)$, $\boldsymbol{\Phi}_k = \mathrm{diag}(e^{j\omega_k}, \cdots, e^{j\omega_k L_k})$, and the vector $\mathbf{b}$ contains the complex amplitudes

$$\mathbf{b} = [\beta_{1,1}\ \cdots\ \beta_{L_1,1}\ \cdots\ \beta_{1,K}\ \cdots\ \beta_{L_K,K}]^T, \tag{25}$$

with $\beta_{l,k} = A_{l,k}e^{j\phi_{l,k}}$. The multiple sources impinge on the array with different DOAs consisting of various frequency components may, for certain frequency combinations, give the same array steering vector, which causes the matrix $\mathbf{C}$ to be rank deficient. Here, the ambiguities and the rank-deficiency are avoided by introducing temporal smoothness in order to restore the rank of $\mathbf{C}$. The temporally smoothed data matrix is obtained by stacking t times temporally shifted versions of the original data matrix, given as [8]

$$\mathbf{X}_t = \begin{bmatrix} \mathbf{C}\boldsymbol{\Phi}\ [\mathbf{b}\ \boldsymbol{\Phi}\mathbf{b}\ \cdots\ \boldsymbol{\Phi}^{N-t}\mathbf{b}] \\ \mathbf{C}\boldsymbol{\Phi}^2\ [\mathbf{b}\ \boldsymbol{\Phi}\mathbf{b}\ \cdots\ \boldsymbol{\Phi}^{N-t}\mathbf{b}] \\ \vdots \\ \mathbf{C}\boldsymbol{\Phi}^t\ [\mathbf{b}\ \boldsymbol{\Phi}\mathbf{b}\ \cdots\ \boldsymbol{\Phi}^{N-t}\mathbf{b}] \end{bmatrix} + \mathbf{E}_t, \tag{26}$$

where $\mathbf{X}_t \in \mathbb{C}^{tI\times(N-t+1)}$ is the temporally smoothed data matrix, and $\mathbf{E}_t$ is its noise part. On one hand, $\mathbf{X}_t$ can be factorized as

$$\begin{aligned}\mathbf{X}_t &= \mathbf{C}_t\mathbf{B}_t + \mathbf{E}_t \\ &= \begin{bmatrix} \mathbf{C}\boldsymbol{\Phi} \\ \mathbf{C}\boldsymbol{\Phi}^2 \\ \vdots \\ \mathbf{C}\boldsymbol{\Phi}^t \end{bmatrix} \left[\mathbf{b}\ \boldsymbol{\Phi}\mathbf{b}\ \cdots\ \boldsymbol{\Phi}^{N-t}\mathbf{b}\right] + \mathbf{E}_t,\end{aligned} \tag{27}$$

where $\mathbf{C}_t = [\mathbf{C}_{t,1}\ \mathbf{C}_{t,2}\ \cdots\ \mathbf{C}_{t,K}]$, whose submatrix for each individual source k is given by

$$\mathbf{C}_{t,k} = \begin{bmatrix} e^{j\omega_k} & \cdots & e^{jL_k\omega_k} \\ e^{j\alpha_k}e^{j\omega_k} & \cdots & e^{j\alpha_k}e^{jL_k\omega_k} \\ \vdots & & \vdots \\ e^{j(I-1)\alpha_k}e^{j\omega_k} & \cdots & e^{j(I-1)\alpha_k}e^{jL_k\omega_k} \\ \vdots & & \vdots \\ e^{j\omega_k t} & \cdots & e^{jL_k\omega_k t} \\ e^{j\alpha_k}e^{j\omega_k t} & \cdots & e^{j\alpha_k}e^{jL_k\omega_k t} \\ \vdots & & \vdots \\ e^{j(I-1)\alpha_k}e^{j\omega_k t} & \cdots & e^{j(I-1)\alpha_k}e^{jL_k\omega_k t} \end{bmatrix}. \tag{28}$$

On the other hand, $\mathbf{X}_t$ can be decomposed using SVD as:

$$\mathbf{X}_t = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H = [\mathbf{U}_s\ \mathbf{U}_n]\begin{bmatrix} \boldsymbol{\Lambda}_s & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_n \end{bmatrix}[\mathbf{V}_s\ \mathbf{V}_n]^H, \tag{29}$$

with $\mathbf{U}_s$ and $\mathbf{U}_n$ spanning the signal and noise subspace, respectively. Comparing (27) and (29), the DOA and fundamental frequency for each source k can be estimated jointly by minimizing the cost function

$$J(\tilde{\omega}_k, \tilde{\alpha}_k) = ||\mathbf{C}_{t,k}\mathbf{U}_n||_{\mathrm{F}}^2, \tag{30}$$

with $\tilde{\omega}_k$ and $\tilde{\alpha}_k$ the variable for ω_k and α_k. Now that we obtain the estimate for the fundamental frequency ω_k in the last subsection, the DOA is estimated with the frequency estimate $\hat{\omega}_k$ as

$$\hat{\alpha}_k = \arg\min_{\tilde{\alpha}_k}\ J(\hat{\omega}_k, \tilde{\alpha}_k). \tag{31}$$

Note that in the two-stage estimation, only two one-dimensional searches are required and the two-dimensional search in the joint estimation of DOA and fundamental frequency is avoided, which decreases the computation complexity. This DOA estimation method is termed as spatio-temporal multiple signal classification (ST-MUSIC) estimator.

III. Simulation Results

In this section, we perform Monte Carlo simulations to evaluate the fundamental frequency and DOA estimation performance of the proposed method. The estimation accuracy is evaluated using the mean square error (MSE): $\mathrm{MSE}_f = \frac{1}{SK}\sum_{k=1}^{K}\sum_{s=1}^{S}(\hat{\omega}_k^{(s)} - \omega_k)^2$ and $\mathrm{MSE}_a = \frac{1}{SK}\sum_{k=1}^{K}\sum_{s=1}^{S}(\hat{\alpha}_k^{(s)} - \alpha_k)^2$, with ω_k, α_k and $\hat{\omega}_k^{(s)}$, $\hat{\alpha}_k^{(s)}$ being the true fundamental frequency, DOA and their estimates, respectively, and S being the number of trials. We use the number of iterations as the stopping criterion in the MCOW-HMUSIC algorithm, which is assigned as 3. In the fundamental frequency estimation, the row number of the data matrix $\mathbf{X}_i$ for each channel i is set as $M = \lfloor N/2 \rfloor$ with $\lfloor u \rfloor$ denoting the smallest integer larger than u. In the DOA estimation, the row number of the data matrix $\mathbf{X}_t$ is set as $M' = \lfloor N/2 \rfloor I$. Such row number settings are found to result in good performance empirically. The microphone array size is set as $I = 5$. All the results are averages of 100 independent runs.

First, we provide an example of single-pitch estimation. The harmonic signal consists of $L = 4$ sinusoids with fundamental frequency of $\omega_1 = 0.5$ and DOA of $\alpha_1 = 0.4$. The parameter setting is listed in Table 1. Fig.1 shows the MSEs of fundamental frequency estimates by MCOW-HMUSIC and its non-weighting version (which is termed as MC-HMUSIC here) as well as CRLB, with $N = 50$ and 100, respectively. It is seen that the MCOW-HMUSIC estimate attains the optimum accuracy when $\mathrm{SNR} \geq 5\mathrm{dB}$. However, there is about 3 dB gap between the MSE of MC-HMUSIC and CRLB. Fig.2 shows the corresponding MSEs of DOA estimates by ST-MUSIC with frequency estimate from MCOW-HMUSIC and MC-HMUSIC as well as CRLB. For the DOA estimation, the results are the same with frequency estimates from any method, and there are about 3 dB and 5 dB gaps from CRLB when $N = 50$ and 100, respectively.

TABLE 1. SIMULATION SETTING OF SINGLE-PITCH ESTIMATION

DOA	*l*	*Frequency*	*Amplitude*	*Initial Phase*
0.4	1	0.5	2.0	1
	2	1.0	1.5	2
	3	1.5	2.5	3
	4	2.0	4.0	4

The next example is about multi-pitch estimation. The harmonic signal consists of $K = 2$ pitches, each with $L_k = 2$ tones. The parameter setting is listed in Table 2. Fig.3 and Fig.4 show the MSEs of fundamental frequency and DOA, respectively. From Fig.3, we can still see that the MCOW-HMUSIC estimator is superior to MC-HMUSIC by 5–10 dB, and its MSE can be close to CRLB. In addition, when the data length is larger, the MCOW-HMUSIC estimator performs with better accuracy. For the DOA estimation, the results are the same with frequency estimates from any method, and there are about 5 dB and 3 dB gaps from CRLB when $N = 50$ and 100, respectively.

TABLE 2. SIMULATION SETTING OF TWO-PITCH ESTIMATION

k	*DOA*	l_k	*Frequency*	*Amplitude*	*Initial Phase*
1	0.4	1	0.3	2.0	1
		2	0.6	1.0	2
2	0.6	1	0.5	2.0	3
		2	1.0	1.0	4

IV. CONCLUSION

The two stage estimation of the fundamental frequency and DOA for the multi-channel harmonic signal is proposed. In the fundamental frequency estimation, the MCOW-HMUSIC estimator is developed, where the perturbation of orthogonality error is analyzed, and the Markov optimum weighting is utilized. Then the DOA is estimated with the ST-MUSIC and the estimated frequency. The two-dimensional search in the joint DOA and fundamental frequency estimation is replaced with two one-dimenional searches, which saves the computational complexity. Simulation results show that the MCOW-HMUSIC method improves the accuracy of the conventional HMUSIC method, and its performance can attain CRLB for single-pitch estimation. Moreover, the ST-MUSIC method provides accurate estimation for the DOA. Further works include dealing with more complicated multi-pitch estimation problems, adaptive pitch estimation, and their application to speech and audio signal processing.

REFERENCES

[1] P. Stoica and A. Nehorai, “Statistical analysis of two non-linear least-squares estimators of sine waves parameters in the colored noise,” in *Proc. ICASSP*, vol.4, pp.2408-2411, Apr. 11-14, 1988, New York, New York, USA.

[2] M.G. Christensen, P. Stoica, A. Jakobsson and S.H. Jensen, “Multi-pitch estimation,” *Signal Processing*, vol.88, no.4, pp.972-983, Apr. 2008.

[3] M.G. Christensen, A. Jakobsson and S.H. Jensen, “Joint high-resolution fundamental frequency and order estimation,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol.15, no.5, pp.1635-1644, Jul. 2007.

[4] S.M. Kay, *Fundamentals of Statistical Signal Processing – Estimation Theory*, Englewood Cliffs, N.J.: PTR Prentice-Hall, 1993.

[5] L. Armani and M. Omologo, “Weighted autocorrelation-based F0 estimation for distant-talking interaction with a distributed microphone network,” in *Proc. ICASSP*, vol.1, pp.113-116, May 17-21, 2004, Montreal, Quebec, Canada.

[6] S.N. Wrigley and G.J. Brown, “Recurrent timing neural networks for joint F0-localisation based speech separation,” in *Proc. ICASSP*, vol.1, pp.157-160, Apr. 15-20, 2007, Honolulu, Hawai'i, USA.

[7] J.R. Jensen, M.G. Christensen and S.H. Jensen, “Joint DOA and fundamental frequency estimation methods based on 2-D filtering,” in *Proc. EUSIPCO*, vol.2010, pp.2091-2095, Aug. 23-27, 2010, Aalborg, Denmark.

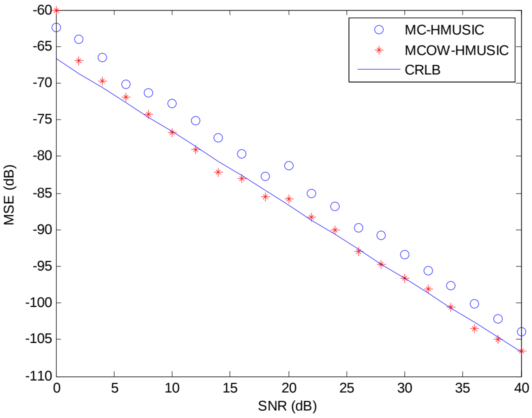
[8] J. X. Zhang, M.G. Christensen, S.H. Jensen and M. Moonen, “Joint DOA and multi-pitch estimation based on subspace techniques,” *EURASIP Journal on Advances in Signal Processing*, vol.2012, no.1, pp.1-11, Jan. 2012.

[9] M.G. Christensen and A. Jakobsson, *Multi-pitch estimation*, [San Rafael, Calif.] : Morgan & Claypool Publishers, 2009.

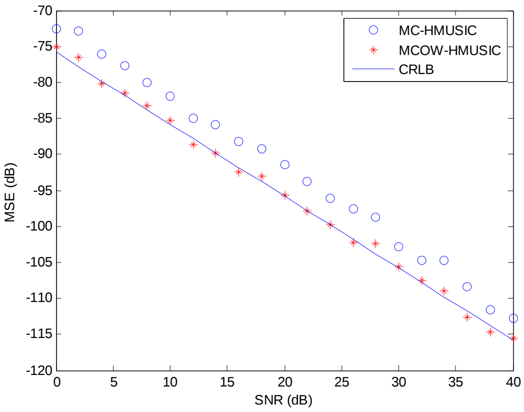
[10] P. Stoica, *Introduction to spectral analysis*, Upper Saddle River, N.J. : Prentice Hall, 1997.

[11] J. Liu, X.-Q. Liu and X.-L. Ma, “First-order perturbation analysis of singular value decomposition,” *IEEE Transactions on Signal Processing*, vol.56, no.7, pp.3044-3049, Jul. 2008.

[12] T. Soderstrom and P. Stoica, *System Identification*, Englewood Cliffs, N.J.: Prentice-Hall International, 1989.

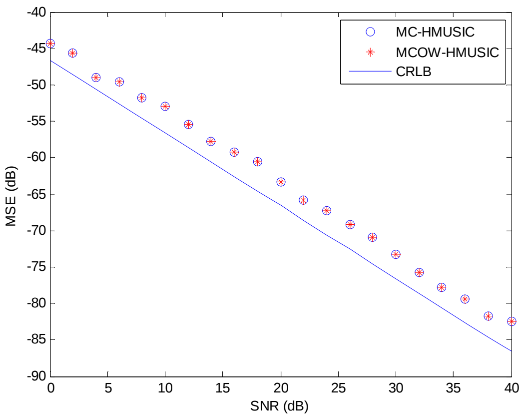


(a)

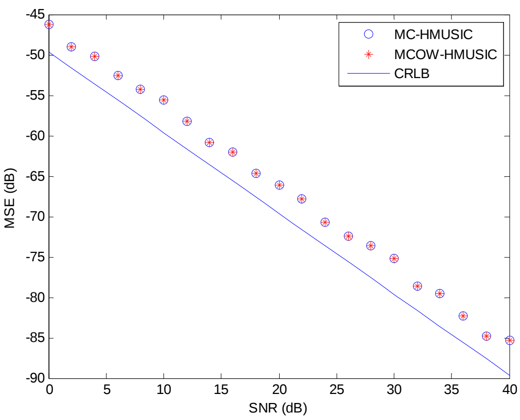


(b)

Figure 1. MSEs of single-pitch fundamental frequency estimation versus SNR for: (a) N = 50 and (b) N = 100.

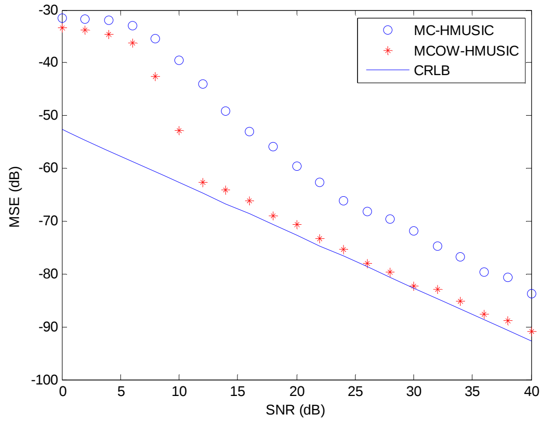


(a)

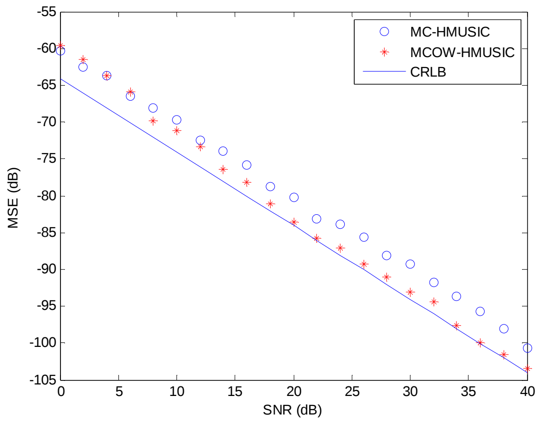


(b)

Figure 2. MSEs of single-pitch DOA estimation versus SNR for: (a) N = 50 and (b) N = 100.

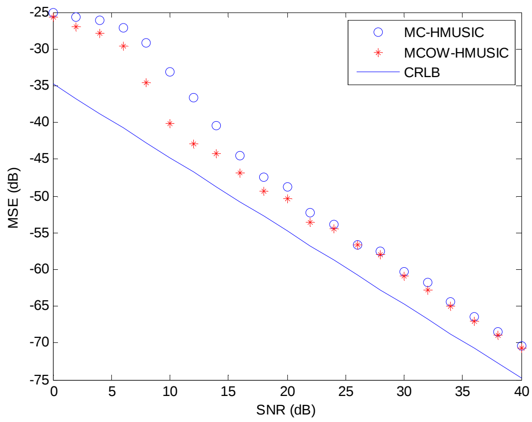


(a)

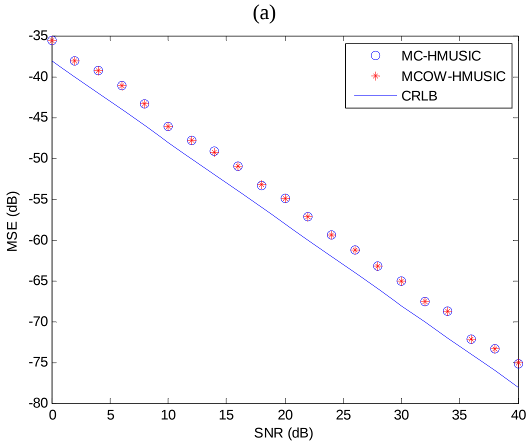


(b)

Figure 3. MSEs of two-pitch fundamental frequency estimation versus SNR for: (a) N = 50 and (b) N = 100.



(a)



(b)

Figure 4. MSEs of two-pitch DOA estimation versus SNR for: (a) N = 50 and (b) N = 100.