Joint Pitch and DOA Estimation Using the ESPRIT method

Wu, Yuntao; Amir, Leshem; Jensen, Jesper Rindom; Liao, Guisheng

Published in:
I E E E Transactions on Audio, Speech and Language Processing

DOI (link to publication from Publisher):
10.1109/TASLP.2014.2367817

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

Joint Pitch and DOA Estimation Using the ESPRIT Method

Yuntao Wu, *Member, IEEE*, Leshem Amir, *Senior Member, IEEE*, Jesper Rindom Jensen, *Member, IEEE*, and Guisheng Liao

***Abstract*—In this paper, the problem of joint multi-pitch and direction-of-arrival (DOA) estimation for multichannel harmonic sinusoidal signals is considered. A spatio-temporal matrix signal model for a uniform linear array is defined, and then the ESPRIT method based on subspace techniques that exploits the invariance property in the time domain is first used to estimate the multi pitch frequencies of multiple harmonic signals. Followed by the estimated pitch frequencies, the DOA estimations based on the ESPRIT method are also presented by using the shift invariance structure in the spatial domain. Compared to the existing state-of-the-art algorithms, the proposed method based on ESPRIT without 2-D searching is computationally more efficient but performs similarly. An asymptotic performance analysis of the DOA and pitch estimation of the proposed method are also presented. Finally, the effectiveness of the proposed method is illustrated on a synthetic signal as well as real-life recorded data.**

***Index Terms*—DOA estimation, ESPRIT, performance analysis, pitch estimation.**

I. Introduction

THE problem of estimating the pitch of a periodic signal is a fundamental problem that has received considerable attention for many years due to its wide application in areas such as audio and speech coding, compression, enhancement and classification of music, and speech analysis [1]–[5]. Many methods for parameter estimation in both single-pitch and multi-pitch scenarios have been reported (see [6]–[8] for an overview of existing pitch estimation methods). Direction-of-arrival (DOA) estimation with an array of spatially separated sensors is another key research topic in the field of signal processing [10], [11], which has been widely studied in the past decades. Its application areas include radar [12], sonar [13], radio astronomy [14], geophysics [15], and speech processing with a microphone array [16]. In recent years, the problem of joint estimation of pitch and DOA of audio and speech signals received by multiple sensors has been extensively discussed for both single-pitch and multi-pitch scenarios in the literature [17]–[36]. Some of the key examples of applications that could benefit from this includes teleconferencing, surveillance applications, hands-free communication, and hearing aids, and so on.

The problem of single-pitch estimation has been researched thoroughly, but the multi-pitch scenario occurs more often in real-life applications. The performance of multi-pitch estimators designed for single channel recording is seriously degraded in the presence of frequency overlapping of sources, reverberation, and strong noise environments. For these reasons, the problem can be solved using a multi-channel receiver approach to obtain an accurate pitch estimation in such cases. Some robust multi-pitch estimation methods have been proposed by using joint spatio-temporal processing [17], [18], where joint parameter estimation is performed, and hence the sources are still resolvable even in the case of a similar source parameter in one dimension for both sources; i.e., with nearly same DOA or pitch. For example, if the pitch of two periodic sources, such as many audio and voiced speech, is similar, we can use the distinct spatial information to first estimate the DOA and then extract the pitch from the spatially separated sources [33]. Most of the methods for joint DOA and pitch estimation focus on the frequency and the DOA estimation of a single sinusoid defined in two dimensions, such as the state-space realization approach [31], the 2-D minimum variance distortionless response (MVDR) method [22], [23], [24], the 2-D ESPRIT method [40], and so on. Generally, these methods are based on the assumption of a narrow-band signal model, therefore, these methods are not suitable for the speech and audio processing since the problem of parameter estimation of audio and speech recorded using a microphone array is considered as a broadband case. Other methods can be considered as a generalization of the previous methods by assuming that desired sources are modelled as sums of harmonically related sinusoids, which is a good fit for many musical instruments and voiced speech. In other words, the traditional broadband case is transformed into multiple narrow-band cases. Some classic methods include the ML-based method [28], subspace-based methods [27], and the linearly constrained minimum variance (LCMV) beamformer

Manuscript received September 11, 2014; accepted October 19, 2014. Date of publication November 06, 2014; date of current version January 14, 2015. This work was supported in part by a grant from the National Natural Science Foundation of China under Project 61172156, in part by the program for New Century Excellent Talents in University under Grant NCET-13-0940, the Natural Science Foundation of Hubei Provinceunder Grant 2014CFB791, in part by the Research Plan Project of Hubei Provincial Department of Education under Grant T201206, and in part by the Danish Council for Independent Research under Grant DFF-1337-00084. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Emmanuel Vincent.

Y. Wu is with the School of Computer Sciences and Engineering, Wuhan Institute of Technology, Wuhan 430205, China (e-mail: ytwu@sina.com).

L. Amir is with the Faculty of Engineering, Bar-Ilan University, Ramat Gan 52900, Israel (e-mail: amir2@hotmail.com).

J. R. Jensen is with the Audio Analysis Lab, AD:MT, Aalborg University, 9200 Aalborg, Denmark (e-mail: jrj@create.aau.dk).

G. Liao is with the National Key Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China (e-mail: gsliao@xidian.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASLP.2014.2367817

[32]; however, these are focused on time-delay and pitch estimation, and the time-delay estimates are closely related to DOA estimation. To solve the problem of DOA and pitch estimation in the broadband signal case, one common form of processing is to use band-pass filtering of the broadband signals into multiple subbands. By doing so, the narrow-band estimators can be applied to each subband [20], [21], [40]. Since the harmonic signals consist of sinusoidal components, an alternative method is to model each source as a multiple narrow-band signal with distinct frequencies with the same DOA. In recent years, based on the harmonic model, some methods for joint pitch and DOA estimation for single-pitch and multi-pitch scenarios have been proposed. Although the maximum likelihood (ML) estimator [28] provides an optimum parameter estimation, its computational complexity is extremely demanding. A non-linear least squares (NLS) method for joint DOA and single-pitch estimation was recently reported in [36], which is also equivalent to the ML estimator under the assumption of white Gaussian noise and an anechoic environment. To avoid the multidimensional grid search requirement of both the ML and NLS-based methods, a simpler but suboptimal solution can be achieved by the subspace-based approach, which relies on the singular value decomposition or eigen-decomposition of the observation space into signal and noise subspaces. However, conventional subspace-based techniques, such as the multi-channel multi-pitch harmonic multiple signal classification (MC-HMUSIC) method for joint DOA and pitch estimation still requires a 2-D grid search [34]. In order to reduce the computational complexity of the two-dimensional searching used in the MC-HMUSIC method, an improved subspace-based technique known as multi-channel optimally weighted harmonic multiple signal classification by using a Markov-based eigen-analysis (MCOW-HMUSIC) was recently proposed in [35], where the 2-D searching for each source in the MC-HMUSIC method is converted into two 1-D searching computations. Furthermore, many of the above-mentioned methods involving grid searches require reasonable initial values of the pitch or DOA estimates or a fine estimation grid to get an accurate estimation. This technique still suffers from the problem of finding the correct coupling between the DOA parameters and the pitch parameters. Furthermore, it is well known that the 2D MUSIC cannot easily cope with multi-path.

The aim of this paper is to develop a computationally efficient pitch and DOA estimation algorithm by utilizing the structure of a spatio-temporal signal model. The conventional ESPRIT method is exploited to estimate the pitch and DOA of interest for the single-pitch and multi-pitch scenarios, which is also a subspace scheme but does not involve any searching, and hence has a lower computational complexity than methods based on MC-HMUSIC [34] with only negligible performance loss. Moreover, an important contribution of this paper is the introduction of spatial smoothing techniques, that allow us to resolve multi-path propagation. This makes the proposed technique much more practical in real life environments as we demonstrate. In the presence of multi-path propagation, we propose to generalize the spatial smoothing technique [39] to spatio-temporal smoothing which facilitates the resolvability of both pitch frequencies in the multi-path case.

The rest of the paper is organized as follows. The problem formulation is firstly described in Section II. In Section III, the spatio-temporal data model for the single-pitch case is defined and the single-pitch and DOA estimation method based on the ESPRIT technique is given. Then, the generalization of the proposed method to the case of joint multi-pitch and DOA estimation is presented. This is followed by the development of a joint pitch and DOA estimation algorithm in the presence of multi-path propagation. Section IV summarizes the proposed methods, and the asymptotic performance analysis follows in the subsequent section. Finally, simulation results and real-world signal data tests are discussed to illustrate the effectiveness of the proposed method. Conclusions are drawn in Section VII.

Notation: The following notations are used in this paper: bold, lowercase letters refer to column vectors and uppercase bold letters refer to matrices. The inverse, pseudo-inversion, transpose, Hermitian transpose and conjugation of a matrix are denoted by $(\cdot)^{-1}$, $(\cdot)^{\#}$, $(\cdot)^{T}$, $(\cdot)^{H}$ and $(\cdot)^{*}$, respectively. The statistical expectation operator is referred to as $E\{\cdot\}$. The matrix Kronecker product is denoted by $\otimes$ and the vectorized representation of a matrix is given by $vec(\cdot)$ respectively. The diagonal matrix consisting of only the diagonal entries of a square matrix is given by diag $\{\cdot\}$. $\mathrm{Re}\{\cdot\}$ stands for the real part of a complex number and $\angle(\cdot)$ denotes the phase of $(\cdot)$, respectively. $\mathbf{I}$ is the identity matrix and 0 is a matrix of zeros. Estimated parameters are denoted by a hat, $\hat{(\cdot)}$.

II. Problem Formulation

To demonstrate the proposed solution without complicating the notation, we begin by demonstrating how the joint DOA and pitch can be estimated when the received signals are incoherent, i.e., there is no multi-path propagation. The multi-channel signal model is presented as follows. Assume that the signal $x_r(n)$ received by the r-th microphone element arranged in a uniform linear array (ULA) configuration with inter-element spacing d, for $r = 0, \cdots, R-1$, is given by [34], [36]:

$$x_r(n) = s_r(n) + q_r(n), n = 0, 1, \cdots, N-1, \tag{1}$$

$$s_r(n) = \sum_{p=1}^{P} \sum_{l=1}^{L_p} \alpha_{p,l} e^{j(\omega_p l n + \phi_p l r)}, \tag{2}$$

where N denotes the number of samples, and P and L_p are the number of spatial sources and the model order (harmonic number) of the p-th spatial harmonic source, respectively. Moreover, ω_p denotes the pitch of the p-th harmonic source, $\alpha_{p,l}$ is the complex-valued amplitude of the l-th harmonic of the p-th source and $q_r(n)$ stands for additive white Gaussian noise with variance σ^2 received at the r-th sensor. The phase shift caused by the propagation time-delay between array elements is defined by $\phi_p = \omega_p f_s \frac{d}{c} sin(\theta_p)$, where c is the speed of propagation, f_s is the signal sample frequency, and θ_p is the DOA (direction-of-arrival) of the p-th spatial harmonic source. In this paper, the number of sources P and the number of harmonics L_p of each source p are assumed to be known a priori, or determined with some available methods such as [6],

[11]. The objective of this paper is to estimate the nonlinear parameters $(\omega_p, \theta_p), p = 1, 2, \cdots, P$ from the received data $x_r(n)$. It is worth noting that parameter ϕ_p is dependent on the fundamental frequency ω_p. Meanwhile, when the multi-path propagation is present, the model in (1)-(2) remains the same except that we allow identical pitch with $\omega_p = \omega_q$ for $p \neq q$.

III. Proposed Method

A. Spatial-Temporal Signal Model in the Single-Pitch Case

To present the proposed method clearly, we first focus on the problem of parameter estimation in the single-pitch case, that is where $P = 1$ in the signal model (1). In the signal model (1), for $P = 1$, the received array data can be written as

$$\begin{aligned} x_r(n) &= s_r(n) + q_r(n), n = 0, 1, \cdot, N-1, \\ s_r(n) &= \sum_{l=1}^{L} \alpha_l e^{j\omega l n + \phi l r} \end{aligned} \tag{3}$$

In this case, only a pair of unknown parameters (ω, ϕ) is estimated from the observed signal. Based on the above model, a spatio-temporal matrix signal model is first defined as follows

$$\mathbf{X}(n) = \mathbf{S}(n) + \mathbf{Q}(n), n = 0, 1, \cdots, N - M, \tag{4}$$

where

$$\mathbf{X}(n) = \begin{bmatrix} x_0(n) & \cdots & x_0(n+M-1) \\ \vdots & \ddots & \vdots \\ x_{R-1}(n) & \cdots & x_{R-1}(n+M-1) \end{bmatrix}, \tag{5}$$

and $\mathbf{S}(n)$ and $\mathbf{Q}(n)$ are defined similar to $\mathbf{X}(n)$. In addition, M is a given positive integer, which is smaller than N and satisfies $MR > L$. The signal matrix of interest $\mathbf{S}(n)$ can be rewritten by (3) as

$$\mathbf{S}(n) = \sum_{l=1}^{L} \beta_l(n) \mathbf{a}_s(l\phi) \mathbf{a}_t^T(l\omega), \tag{6}$$

where

$$\beta_l(n) = \alpha_l e^{jl\omega n}, \tag{7}$$

$$\mathbf{a}_s(\phi) = [1 e^{j\phi} \cdots e^{j(R-1)\phi}]^T, \tag{8}$$

$$\mathbf{a}_t(\omega) = [1 e^{j\omega} \cdots e^{j(M-1)\omega}]^T. \tag{9}$$

Define a new vector model $\mathbf{x}(n)$ by stacking the columns of $\mathbf{X}(n)$ as

$$\begin{aligned} \mathbf{x}(n) &= vec\{\mathbf{X}(n)\} \\ &= \mathbf{s}(n) + \mathbf{q}(n) = \mathbf{A}\mathbf{z}(n) + \mathbf{q}(n), \end{aligned} \tag{10}$$

and hence we have

$$\mathbf{s}(n) = vec\{\mathbf{S}(n)\} \tag{11}$$

$$\mathbf{q}(n) = vec\{\mathbf{Q}(n)\} \tag{12}$$

$$\mathbf{A} = [\mathbf{a}(\omega, \phi), \cdots, \mathbf{a}(L\omega, L\phi)], \tag{13}$$

$$\mathbf{a}(l\omega, l\phi) = \mathbf{a}_t(l\omega) \otimes \mathbf{a}_s(l\phi), \tag{14}$$

$$\mathbf{z}(n) = [\alpha_1 e^{j\omega n}, \cdots, \alpha_L e^{jL\omega n}]^T. \tag{15}$$

Based on the data model (10), a nonlinear least squares method for DOA and pitch estimation in the case of single pitch was recently reported in [36], where the 2-D search computation as well as relatively fine candidate grids for the unknown parameters are required to obtain accurate DOA and pitch estimates, which thus results in a high computational load. In this following Section III-B, a lower-complexity ESPRIT-based method is derived based on the spatio-temporal model in the subsection, and a generalization of the proposed method to the multi-pitch case is suggested in the Section III-C.

B. Joint Estimation of DOA and Pitch in the Single-pitch Case: Multichannel Single-pitch ESPRIT (MSP-ESPRIT)

For $n = 0, 1, \cdots, N - M$, one first defines the following matrix with $\mathbf{x}(n)$,

$$\begin{aligned} \mathbf{Y}_{RM \times (N-M+1)} &= [\mathbf{x}(0)\mathbf{x}(1) \cdots \mathbf{x}(N-M)] \\ &= \mathbf{A}[\mathbf{z}(0) \cdots \mathbf{z}(N-M)] \\ &\quad + [\mathbf{q}(0) \cdots \mathbf{q}(N-M)] \end{aligned} \tag{16}$$

Under the condition of full column rank of matrix $\mathbf{A}$, one can obtain the signal subspace from the singular-value decomposition (SVD) or equivalent eigenvalue decompositions of the covariance matrix of $\mathbf{Y}$. The signal subspace matrix $\mathbf{U}_s$ is then composed of the L singular vector associated with the largest singular values of $\mathbf{Y}$. It is well-known that the signal subspace spanned by the column vectors of $\mathbf{U}_s$ is the same subspace as all the column vectors in matrix $\mathbf{A}$; therefore, there exists a unique nonsingular matrix $\mathbf{T}$ such that $\mathbf{U}_s = \mathbf{A}\mathbf{T}$.

Using the inherent structure of $\mathbf{a}_t(\omega)$ and $\mathbf{a}_s(\phi)$, the shift invariance property of $\mathbf{A}$ can be found so that the classic ESPRIT method [19] is exploited to estimate the DOA and pitch parameters, respectively.

In so doing, we first define the following four selection matrices

$$\mathbf{W}_1 = \begin{bmatrix} \mathbf{I}_{M-1} \\ 0_{1 \times (M-1)} \end{bmatrix} \otimes \mathbf{I}_R,$$

$$\mathbf{W}_2 = \begin{bmatrix} 0_{1 \times (M-1)} \\ \mathbf{I}_{M-1} \end{bmatrix} \otimes \mathbf{I}_R,$$

$$\mathbf{W}_3 = \mathbf{I}_M \otimes \begin{bmatrix} \mathbf{I}_{R-1} \\ \mathbf{0}_{1 \times (R-1)} \end{bmatrix},$$

$$\mathbf{W}_4 = \mathbf{I}_M \otimes \begin{bmatrix} 0_{1 \times (R-1)} \\ \mathbf{I}_{R-1} \end{bmatrix},$$

where $\mathbf{0}$ is the vector of zeros.

The submatrices $\mathbf{A}_i$ of $\mathbf{A}$ are further defined by the above selection matrix $\mathbf{W}_i$ as

$$\mathbf{A}_i = \mathbf{W}_i^T \mathbf{A}, i = 1, 2, 3, 4. \tag{17}$$

With the above definition and the structure of $\mathbf{A}$, it is easily shown that we have the following equations

$$\mathbf{A}_1 \mathbf{\Psi}_t(\omega) = \mathbf{A}_2, \tag{18}$$

$$\mathbf{A}_3 \mathbf{\Psi}_s(\phi) = \mathbf{A}_4, \tag{19}$$

where

$$\begin{aligned} &\mathbf{\Psi}_t(\omega) = diag\{e^{j\omega}, \cdots, e^{jL\omega}\} \\ \text{and} \quad &\mathbf{\Psi}_s(\phi) = diag\{e^{j\phi}, \cdots, e^{jL\phi}\} \end{aligned} \tag{20}$$

Let $\mathbf{U}_1, \mathbf{U}_2$ and $\mathbf{U}_3, \mathbf{U}_4$ be the submatrices formed from $\mathbf{U}_s$ in the same way as the $\mathbf{A}_i$'s are formed from $\mathbf{A}$. Following the rationale of the ESPRIT method, the diagonal elements of $\mathbf{\Psi}_t(\omega)$ and $\mathbf{\Psi}_s(\phi)$ are the corresponding eigenvalues of $\mathbf{\Phi}_t =$

$\mathbf{T}\Psi_t\mathbf{T}^{-1}$ and $\mathbf{\Phi}_s = \mathbf{T}\Psi_s\mathbf{T}^{-1}$, which satisfy the following equations

$$\mathbf{U}_2 = \mathbf{U}_1\mathbf{\Phi}_t \quad \text{and} \quad \mathbf{U}_4 = \mathbf{U}_3\mathbf{\Phi}_s \tag{21}$$

Let μ_l and ν_l for $l = 1, 2, \cdots, L$ be the corresponding eigenvalues of $\mathbf{\Phi}_t$ and $\mathbf{\Phi}_s$. Thus the estimates of ω and θ can be given as follows

$$\hat{\omega} = \frac{2}{L(L+1)}\sum_{l=1}^{L}\angle\mu_l,$$
$$\hat{\theta} = \arcsin\left(\frac{2c\sum_{l=1}^{L}\angle\nu_l}{L(L+1)\hat{\omega}f_s d}\right) \tag{22}$$

It can be seen from (20) that the estimation accuracy of $\hat{\theta}$ depends on that of $\hat{\omega}$.

C. Spatio-temporal Signal Model in the Presence of Multi-pitch

To solve the problem of joint estimation of the DOA and pitch in the multi-pitch case, a multi-channel and multi-pitch harmonic multiple signal classification (MC-HMUSIC) method was presented in [34], where a two dimensional search is required to obtain the estimates of the pitch and DOA. To further reduce the computational complexity of the 2-D search in the MC-HMUSIC method, a two-stage DOA and pitch estimation method based on a subspace approach, i.e., MCOW-HUMUSIC, was recently proposed in [35]; however, the two 1-D searches and an iterative procedure for the refined estimation are still required for the MCOW-HMUSIC method. Furthermore, the coupling of the pitch and DOA parameters is still required and poses a significant problem. In this subsection, a generalization of the previously proposed ESPRIT method for joint DOA and pitch estimation of the single pitch case to the multi-pitch case is derived.

Following a procedure similar to the single pitch case, with model in (1), for a given integer M and satisfying $RM > \sum_{p=1}^{P} L_p$, one can construct the following data vector

$$\mathbf{X}(n) = \mathbf{S}(n) + \mathbf{Q}(n), n = 0, 1, \cdots, N - M, \tag{23}$$

where

$$\mathbf{X}(n) = \begin{bmatrix} x_0(n) & \cdots & x_0(n+M-1) \\ \vdots & \ddots & \vdots \\ x_{R-1}(n) & \cdots & x_{R-1}(n+M-1) \end{bmatrix}. \tag{24}$$

The part of the signal of interest $\mathbf{S}(n)$ in $\mathbf{X}(n)$ can be rewritten as

$$\mathbf{S}(n) = \sum_{p=1}^{P}\sum_{l=1}^{L_p}\beta_{p,l}(n)\mathbf{a}_s(l\phi_p)\mathbf{a}_t^T(l\omega_p), \tag{25}$$

where

$$\beta_{p,l}(n) = \alpha_{p,l}e^{jl\omega_p n},$$
$$\mathbf{a}_s(\phi_p) = [1 e^{j\phi_p} \cdots e^{j(R-1)\phi_p}]^T,$$
$$\mathbf{a}_t(\omega_p) = [1 e^{j\omega_p} \cdots e^{j(M-1)\omega_p}]^T. \tag{26}$$

Stacking the columns of $\mathbf{X}(n)$ into a new vector model $\mathbf{x}(n)$ gives

$$\mathbf{x}(n) = vec\{\mathbf{X}(n)\} = \mathbf{s}(n) + \mathbf{q}(n) = \mathbf{A}\mathbf{z}(n) + \mathbf{q}(n), \tag{27}$$

with

$$\mathbf{A} = [\mathbf{A}^{(1)}(\omega_1, \phi_1), \cdots, \mathbf{A}^{(P)}(\omega_P, \phi_P)], \tag{28}$$
$$\mathbf{A}^{(p)} = [\mathbf{a}(\omega_p, \phi_p), \cdots, \mathbf{a}(L_p\omega_p, L_p\phi_p)], p=1, \cdots, P, \tag{29}$$
$$\mathbf{a}(l_p\omega_p, l_p\phi_p) = \mathbf{a}_t(l\omega_p) \otimes \mathbf{a}_s(l\phi_p), \tag{30}$$
$$\mathbf{z}(n) = [\alpha_{1,1}e^{j\omega_1 n}, \cdots, \alpha_{1,L_1}e^{jL_1\omega_1 n}, \cdots, \alpha_{P,1}e^{j\omega_P n}, \cdots, \alpha_{P,L_P}e^{jL_P\omega_P n}]^T$$

For $n = 0, 1, \cdots, N - M$, we first define the following vector with $\mathbf{x}(n)$,

$$\mathbf{Y}_{RM\times(N-M+1)} = [\mathbf{x}(0)\mathbf{x}(1)\cdots\mathbf{x}(N-M)] = \mathbf{A}[\mathbf{z}(0)\cdots\mathbf{z}(N-M)] + [\mathbf{q}(0)\cdots\mathbf{q}(N-M)]. \tag{31}$$

Thus, the above spatio-temporal signal model for the multi-pitch case can be reduced to the case of a single pitch in the previous section when $P = 1$ in (23).

D. Joint Estimation of DOA and Pitch in the Multi-pitch Case Using ESPRIT: Multichannel Multi-Pitch ESPRIT (MMP-ESPRIT)

To use the ESPRIT method for joint estimation of DOA and pitch in the multi-pitch case, the singular value decomposition (SVD) of the data matrix $\mathbf{Y}$ is expressed by

$$\mathbf{Y} = \mathbf{U}\Sigma\mathbf{V}^H = [\mathbf{U}_s\mathbf{U}_n]\begin{bmatrix}\Sigma_s & \mathbf{0} \\ \mathbf{0} & \Sigma_n\end{bmatrix}\mathbf{V}^H, \tag{32}$$

where the column vectors of $\mathbf{U}_s$ and $\mathbf{U}_n$ span the signal subspace and noise subspace, respectively. The diagonal elements of Σ_s are composed of the $P_0 = \sum_{p=1}^{P} L_p$ largest singular values of $\mathbf{Y}$, and one also has that $\mathbf{U}_s = \mathbf{A}\mathbf{T}$ for a nonsingular position matrix $\mathbf{T}$.

Using the shift invariance structure of $\mathbf{a}(l\omega_p, l\phi_p)$ in $\mathbf{A}$, we can obtain the following two equations,

$$\mathbf{A}_1\mathbf{\Psi}_t = \mathbf{A}_2, \tag{33}$$
$$\mathbf{A}_3\mathbf{\Psi}_s = \mathbf{A}_4, \tag{34}$$

where $\mathbf{A}_i = \mathbf{W}_i^T\mathbf{A},\ i = 1, 2, 3, 4$. $\mathbf{\Psi}_t$ and $\mathbf{\Psi}_s$ are defined as

$$\mathbf{\Psi}_t = diag(\mathbf{\Psi}_{t_1}, \cdots, \mathbf{\Psi}_{t_P}) \tag{35}$$
$$\mathbf{\Psi}_{t_p} = diag(e^{j\omega_p}, \cdots, e^{j\omega_p L_p}),$$
$$\mathbf{\Psi}_s = diag(\mathbf{\Psi}_{s_1}, \cdots, \mathbf{\Psi}_{s_P})$$
$$\mathbf{\Psi}_{s_p} = diag(e^{j\phi_p}, \cdots, e^{j\phi_p L_p}). \tag{36}$$

Accordingly, one can also obtain the following equations by the subspace property,

$$\mathbf{U}_2 = \mathbf{U}_1\mathbf{\Phi}_t \quad \text{and} \quad \mathbf{U}_4 = \mathbf{U}_3\mathbf{\Phi}_s, \tag{37}$$

where

$$\mathbf{U}_1 = \mathbf{W}_1\mathbf{U}_s \quad \text{and} \quad \mathbf{U}_2 = \mathbf{W}_2\mathbf{U}_s,$$

$$\mathbf{U}_3 = \mathbf{W}_3\mathbf{U}_s \quad \text{and} \quad \mathbf{U}_4 = \mathbf{W}_4\mathbf{U}_s \tag{38}$$

Thus, the diagonal elements of $\mathbf{\Psi}_t$ and $\mathbf{\Psi}_s$ can be estimated from the corresponding eigenvalues of $\mathbf{\Phi}_t = T\Psi_t T^{-1}$ and $\mathbf{\Phi}_s = T\Psi_s T^{-1}$, and then the estimates of pitch and DOA can be found from the eigenvalues $\mu_l^{(p)}$, $\nu_l^{(p)}(l = 1, \cdots, L_p, p = 1, \cdots, P)$ of $\mathbf{\Phi}_t$ and $\mathbf{\Phi}_s$, i.e.,

$$\hat{\omega}_p = \frac{2}{L_p(L_p+1)} \sum_{l=1}^{L_p} \angle\mu_l^{(p)},$$
$$\hat{\theta}_p = \arcsin\left(\frac{2c \sum_{l=1}^{L_p} \angle\nu_l^{(p)}}{L_p(L_p+1)\hat{\omega}_p f_s d} \right). \tag{39}$$

It is worthy to note that the estimated eigenvalues $\mu_l^{(p)}, p = 1, 2, \cdots, P$ need be paired with the eigenvalues $\nu_l^{(p)}, p = 1, 2, \cdots, P$ so that we can determine the pairing parameters $(\hat{\omega}_p, \hat{\theta}_p), p = 1, 2, \cdots, P$ in the case of multi-pitch, however, some available methods reported in [37], [38] can be easily used to finish this pairing.

E. Spatio-Temporal Signal Model for Single-Pitch in the Presence of Multi-path Propagation

In this subsection, the joint estimation of DOA and pitch in the presence of multi-path propagation will be considered for the single pitch case. Based on the spatio-temporally smoothed data matrix model introduced in [40], it is shown that the ESPRIT method can still be used to estimate the DOA and pitch in the multi-path propagation case.

In the signal model (2), when the multiple sources impinge on the array with different DOAs, for certain frequency combinations, the steering vector $a(l\omega_p, l\phi_p)$ as a function of parameter pair (ω_p, ϕ_p) may be the same for different p, especially when the multiple sources have the same pitch but distinct DOAs (i.e., for the multi-path propagation case), which causes matrix $\mathbf{A}$ to be rank-deficient in these cases. Therefore, we need to use the spatio-temporal smoothing technique [40] to restore the rank of $\mathbf{A}$. The signal model (2) for single pitch case in the presence of P multi-path propagations (i.e., $\omega = \omega_1 = \cdots = \omega_P$) can be expressed as

$$x_r(n) = s_r(n) + q_r(n), n = 0, 1, \cdots, N-1, \tag{40}$$
$$s_r(n) = \sum_{p=1}^{P}\sum_{l=1}^{L} \alpha_{p,l} e^{j(\omega l n + \phi_p l r)}. \tag{41}$$

The array output $x(n)$ is defined as

$$x(n) = [x_0(n)x_2(n)\cdots x_{R-1}(n)]^T = \mathbf{A}\mathbf{\Phi}_t(n)b + q(n), \tag{42}$$

where

$$\mathbf{A} = [\mathbf{A}^{(1)}, \cdots, \mathbf{A}^{(P)}], \tag{43}$$
$$\mathbf{A}^{(p)} = [\mathbf{a}(\theta_p, \omega), \cdots, \mathbf{a}(\theta_p, L\omega)], \tag{44}$$
$$\mathbf{\Phi}_t(n) = [\,\Phi'_1(n), \quad \cdots, \quad \Phi'_P(n)\,] \tag{45}$$
$$\Phi'_p(n) = \mathrm{diag}([\,e^{j\omega n} \quad \cdots \quad e^{jL_p\omega n}\,]) \tag{46}$$
$$\mathbf{b} = [\beta_{1,1}\cdots\beta_{1,L}\cdots\beta_{P,1}\cdots\beta_{P,L}]^T, \tag{47}$$

with the array steering vector $a(\theta_p, \omega)$ given by

$$a(\theta_p, \omega) = [1 e^{j\omega f_s \frac{d}{c}\sin(\theta_p)}, \cdots, e^{j\omega f_s \frac{d}{c}(R-1)\sin(\theta_p)}]^T. \tag{48}$$

For N given samples of the array output, we form the following data matrix

$$\mathbf{X} = [x(0)\cdots, x(N-1)]. \tag{49}$$

Using the structure of harmonic components, the above data matrix is expressed as

$$\mathbf{X} = \mathbf{A}[\mathbf{b}, \mathbf{\Phi}_t\mathbf{b}, \cdots, \mathbf{\Phi}_t^{N-1}\mathbf{b}] + \mathbf{Q}, \tag{50}$$

where $\mathbf{Q} \in \mathbb{C}^{R\times N}$ is a matrix containing N samples of the noise vector of the array output.

For a given integer $M(< N)$, the M-times temporally smoothed data matrix is firstly constructed by stacking M times temporally shifted versions of the original data matrix as follows [34]

$$\mathbf{X}_M = \begin{bmatrix} \mathbf{x}(0) & \mathbf{x}(1) & \cdots & \mathbf{x}(N-M) \\ \mathbf{x}(1) & \mathbf{x}(2) & \cdots & \mathbf{x}(N-M+1) \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{x}(M-1) & \mathbf{x}(M) & \cdots & \mathbf{x}(N-1) \end{bmatrix} \tag{51}$$

With the equation (47), $\mathbf{X}_M$ can also be written as

$$\mathbf{X}_M = \begin{bmatrix} \mathbf{A}[\mathbf{b}, \mathbf{\Phi}_t\mathbf{b}, \cdots, \mathbf{\Phi}_t^{N-M}\mathbf{b}] \\ \mathbf{A}\mathbf{\Phi}_t[\mathbf{b}, \mathbf{\Phi}_t\mathbf{b}, \cdots, \mathbf{\Phi}_t^{N-M}\mathbf{b}] \\ \vdots \\ \mathbf{A}\mathbf{\Phi}_t^{M-1}[\mathbf{b}, \mathbf{\Phi}_t\mathbf{b}, \cdots, \mathbf{\Phi}_t^{N-M}\mathbf{b}] \end{bmatrix} + \mathbf{Q}_M \tag{52}$$

where $\mathbf{X}_M \in \mathbb{C}^{MR\times N-M+1}$ is the temporally smoothed data matrix and the part of the noise $\mathbf{Q}_M$ is constructed from $\mathbf{Q}$ in a similar way as $\mathbf{X}_M$ in Eq. (51).

On the other hand, $\mathbf{X}_M$ can be factorized as

$$\begin{aligned}\mathbf{X}_M &= \begin{bmatrix} \mathbf{A} \\ \mathbf{A}\mathbf{\Phi}_t \\ \vdots \\ \mathbf{A}\mathbf{\Phi}_t^{M-1} \end{bmatrix} [\mathbf{b}, \mathbf{\Phi}_t\mathbf{b}, \cdots, \mathbf{\Phi}_t^{N-M}\mathbf{b}] + \mathbf{Q}_M \\ &= \mathbf{A}_M\mathbf{B}_M + \mathbf{Q}_M, \end{aligned} \tag{53}$$

where $\mathbf{A}_M = [\mathbf{A}^T(\mathbf{A}\mathbf{\Phi}_t)^T, \cdots, (\mathbf{A}\mathbf{\Phi}_t^{M-1})^T]^T$ and $\mathbf{B}_M = [\mathbf{b}, \mathbf{\Phi}_t\mathbf{b}, \cdots, \mathbf{\Phi}_t^{N-M}\mathbf{b}]$.

Similarly, the spatially smoothed data matrix is further constructed from the above data matrix $\mathbf{X}_M$. If we assume the number of sensors for each subarray to be R_0, then we can separate the array into $R - R_0 + 1$ subarrays, and the spatio-temporal smoothed data matrix $\mathbf{X}_{MR_0} \in \mathbb{C}^{MR_0\times(R-R_0-1)(N-M+1)}$ is given by

$$\mathbf{X}_{M,R_0} = [\mathbf{X}_{M,1}, \cdots, \mathbf{X}_{M,(R-R_0+1)}], \tag{54}$$

where

$$\mathbf{X}_{M,r} = \mathbf{J}_r\mathbf{X}_M = \mathbf{J}_r\mathbf{A}_M\mathbf{B}_M + \mathbf{Q}_{M,r}, \tag{55}$$

and $\mathbf{J}_r \in \mathbb{R}^{MR_0\times MR}$ is the selection matrix corresponding to the rth subarray for data matrix $\mathbf{X}_M$.

It is easily shown with the shift invariance structure in $\mathbf{A}$ that the following equation holds

$$\mathbf{A}_M^{(R_0)} = \mathbf{J}_r\mathbf{A}_M = \mathbf{J}_1\mathbf{A}_M{\mathbf{\Phi}_s}^{r-1}, \tag{56}$$

where

$$\mathbf{\Phi}_s = diag\{[e^{j\phi_1}, \cdots, e^{jL\phi_1}, e^{j\phi_P}, \cdots, e^{jL\phi_P}]\}. \quad (57)$$

Thus, the matrix $\mathbf{X}_{M,R_0}$ can be written as

$$\mathbf{X}_{M,R_0} = \mathbf{J}_1\mathbf{A}_M[\mathbf{B}_M, \mathbf{\Phi}_s\mathbf{B}_M, \cdots, \mathbf{\Phi}_s^{R-R_0+1}\mathbf{B}_M] + \mathbf{Q}_{M,R_0}, \quad (58)$$

where $\mathbf{J}_1 = \mathbf{I}_M \otimes [\mathbf{I}_{R_0}\mathbf{0}]$.

F. Joint DOA and Single Pitch Estimation Using ESPRIT in the presence of Multi-path Propagation: Multichannel Single-pitch Multi-path ESPRIT (MSP-MP-ESPRIT)

Based on the constructed spatio-temporal smoothed data matrix $\mathbf{X}_{M,R_0}$, the noise subspace and signal subspace can be obtained from the SVD of $\mathbf{X}_{M,R_0}$, which is defined as

$$\mathbf{X}_{M,R_0} = [\mathbf{U}_s\mathbf{U}_n]\Sigma\mathbf{V}^H, \quad (59)$$

where the columns of $\mathbf{U}_s$ and $\mathbf{U}_n$ are the singular vectors associated with the $P_0(= LP)$ largest values and the remaining $MR_0 - P_0$ least singular values, respectively. Using the noise subspace of $\mathbf{X}_{M,R_0}$, a MUSIC-based method for joint multi-pitch and DOA estimation was presented in [34], where a 2-D grid search is required to obtain the estimates of parameter pairs $(\omega, \theta_p), p = 1, 2, \cdots, P$. In this section, using the spatio-temporal shift invariance structure of $\mathbf{A}_M^{(R_0)}$, the ESPRIT-based method is exploited to give the estimates of $\{\omega, \theta_p\}$. One can obtain the following two equations with the signal subspace matrix $\mathbf{U}_s$

$$\mathbf{U}_2 = \mathbf{U}_1\mathbf{\Phi}_t \quad \text{and} \quad \mathbf{U}_4 = \mathbf{U}_3\mathbf{\Phi}_s, \quad (60)$$

where

$$\begin{aligned} \mathbf{U}_1 = \mathbf{W}_1\mathbf{U}_s \quad &\text{and} \quad \mathbf{U}_2 = \mathbf{W}_2\mathbf{U}_s, \\ \mathbf{U}_3 = \mathbf{W}_3\mathbf{U}_s \quad &\text{and} \quad \mathbf{U}_4 = \mathbf{W}_4\mathbf{U}_s. \end{aligned} \quad (61)$$

where the selection matrices $\mathbf{W}_i, i = 1, 2, 3, 4$, are defined as follows

$$\begin{aligned} \mathbf{W}_1 &= \begin{bmatrix} \mathbf{I}_{M-1} \\ 0_{1\times(M-1)} \end{bmatrix} \otimes \mathbf{I}_{R_0}, \\ \mathbf{W}_2 &= \begin{bmatrix} 0_{1\times(M-1)} \\ \mathbf{I}_{M-1} \end{bmatrix} \otimes \mathbf{I}_{R_0}, \\ \mathbf{W}_3 &= \mathbf{I}_M \otimes [\,\mathbf{I}_{R_0-1}\mathbf{0}_{1\times(R_0-1)}\,], \\ \mathbf{W}_4 &= \mathbf{I}_M \otimes [\,0_{1\times(R_0-1)}\mathbf{I}_{R_0-1}\,]. \end{aligned}$$

Thus, the estimates of ω and θ_p can be obtained from the eigenvalues μ_l and $\nu_l^{(p)}$ of $\mathbf{\Phi}_t$ and $\mathbf{\Phi}_s$ as

$$\begin{aligned} \hat{\omega} &= \frac{2}{L(L+1)}\sum_{l=1}^{L}\angle\mu_l \\ \hat{\theta}_p &= \arcsin\left(\frac{2c\sum_{l=1}^{L}\angle\nu_l^{(p)}}{L(L+1)\hat{\omega}f_s d}\right), \\ p &= 1, 2, \cdots, P. \end{aligned} \quad (62)$$

TABLE I

MSP-ESPRIT	1. Align the observed matrix $\mathbf{Y}$ in Eq.(16) 2. Compute the SVD of $\mathbf{Y}$ 3. Compute the eigenvalues of the two matrices $\mathbf{\Phi}_t$ and $\mathbf{\Phi}_s$ in Eq.(21) 4. Estimate the pitch and DOA using Eqs.(22)
MMP-ESPRIT	1. Align the observed matrix $\mathbf{Y}$ in Eq.(31) 2. Compute the SVD of $\mathbf{Y}$ in Eq.(32) 3. Compute the eigenvalues of the two matrices $\mathbf{\Phi}_t$ and $\mathbf{\Phi}_s$ in Eq.(37) 4. Estimate the pitch and DOA using Eqs.(39)
MSP-MP-ESPRIT	1. Align the observed matrix $\mathbf{X}_{M,R_0}$ in Eq.(58) 2. Compute the SVD of matrix $\mathbf{X}_{M,R_0}$ in Eq.(59) 3. Compute the eigenvalues of the two matrices $\mathbf{\Phi}_t$ and $\mathbf{\Phi}_s$ in Eq.(60) 4. Estimate the pitch and DOA using Eqs.(62)

IV. Algorithm Summarization and Computational Complexity Analysis

In this section we analyze the computational complexity of the proposed algorithms. To that end, Table I provides a detailed description of the three algorithms.

In contrast to the NLS-based method for single pitch and the available MC-HMUSIC method for multi-pitch case, the main computational load of the proposed method is dominated by the computation of both the SVD of data matrix $\mathbf{Y}$ and the EVDs of the matrix pair $(\mathbf{\Phi}_s, \mathbf{\Phi}_t)$, which is around $O((RM)^2(N - M + 1) + (L_1 + L_2 + \cdots + L_P)^3)$ for the MMP-ESPRIT method and $O((R_0M)^2(N-M+1)+(LP)^3)$ for the MSP-MP-ESPRIT method, respectively, while the computation of the NLS-based method requires $O(NRL^2 + L^3)$ for per point in the search grid of 2-D parameters. Therefore, the whole computational load of the 2-D searching computation in the NLS-based method for single-pitch case is $O((NRL^2 + L^3)g_\omega g_\theta)$, where g_ω and g_θ are the numbers of searches conducted along the frequency axis and DOA axis. The first part of the computational load of the MC-HMUSIC method is about $O((RM)^2(N - M + 1))$, which is similar to the SVD computation of the proposed MSP-MP-ESPRIT method, but in addition, it needs P 2-D searching computations to obtain the estimates of pitch and DOA for multi-pitch case. The 2-D searches in the MC-MUSIC method are of orders $O((RM)^2 g_\omega g_\theta)$. Thus, the proposed method without 2-D searching has a lower computational load than both the NLS-based method and the MC-HMUSIC method.

V. Asymptotic Performance Analysis of Pitch and DOA Estimates

In this section, the asymptotic performance of the proposed ESPRIT-based method for multi-pitch and DOA estimation is analyzed. With the similar results derived in [42], it is easily shown that the estimations of pitch and DOA parameters are unbiased. The derivation of this result is based on the same analysis procedure for DOA estimation using the ESPRIT method in [41], while it can be directly generalized to the case of multi-path propagation scenarios, therefore, The derivation of this result is based on the same analysis procedure for DOA estimation using the ESPRIT method in [41].

We provide the variances of the estimated pitch and DOA parameters for the multi-pitch case in the absence of multi-path propagation, which are denoted by $E\{(\delta\omega_p)^2\}$ and $E\{(\delta\theta_p)^2\}$.

TABLE II
SIMULATION SETTING OF THE MSP-ESPRIT, MC-HMUSIC AND THE NLS-BASED METHOD

p	L	Pitch	DOA
1	5	243	15

The variances are based on the covariance of the eigenvalues of $\mu_l^{(p)}$, $\nu_l^{(p)}, p = 1, \cdots, P$.

The detailed derivations are found in the Appendix, and yield

$$E\{(\delta\omega_p{}^2)\} = \frac{1}{L_p^2} E\left\{\left[\sum_{l=1}^{L_p} \frac{real\{j\mu_l^{*(p)}\delta\mu_l^{(p)}\}}{l}\right]^2\right\}, \quad (63)$$

and

$$E\{(\delta\theta_p)^2\} = \frac{1}{(L_p\omega_p f_s \frac{d}{c}\cos\theta_p)^2} E\{[\sum_{l=1}^{L_p}\frac{1}{l} real\{ -j\nu_l^{*(p)}\delta\nu_l^{(p)} + j\mu_l^{*(p)} f_s \frac{d}{c}\sin\theta_p \delta\mu_l^{(p)}\}]^2\}. (64)$$

VI. SIMULATION RESULTS

To evaluate the performance of the proposed method for joint estimation of the pitch and DOA, a set of Monte Carlo simulations is presented using synthetic data followed by experiments on real-life data as presented in the final subsection.

A. Synthetic Data Results

The first evaluation was conducted on a set of simulations composed of synthetic data. In all the simulations, it is assumed that the array is a uniform linear array with inter-element spacing $d = c/f_s$, where the speed of sound is set as $c = 343.2$ m/s and the sampling frequency is $f_s = 8$ kHz. All the amplitudes of the sinusoids are assumed to be units, i.e., $\beta_{p,l} = 1$ for each (p, l). The parameter $M = N/2$ was used in all the experiments. The estimation accuracy was evaluated using the root mean square error (RMSE): $\text{RMSE}_\omega = \sqrt{\sum_{q=1}^{Q}(\hat{\omega}_q - \omega)^2/Q}$ and $\text{RMSE}_\theta = \sqrt{\sum_{q=1}^{Q}(\hat{\theta}_q - \theta)^2/Q}$, where ω and θ are the true pitch and DOA, and $\hat{\omega}$ and $\hat{\theta}$ are their estimates, respectively. Furthermore, $Q = 200$ is the number of independent trials. In all the simulations, the signal-to-noise ratio (SNR) was defined as

$$SNR = 10log10\frac{\sum_{p=1}^{P}\sum_{l=1}^{L_p}|\beta_{p,l}|^2}{\sigma^2}$$

In the first experiment, we first provide an example of single-pitch estimation. The number of microphones was set to $R = 4$ and the number of samples was $N = 80$. The other experimental parameter setting is listed in Table II. For the purpose of comparison, the results obtained using the NLS-based method reported in [36] as well the MC-HMUSIC method, and the corresponding CRLB were also included in all the simulation results. Since it was shown in [36] that the performance of the NLS-based method is optimal out of all the existing methods such as LCMV [32] and MC-ML method [28], we only compare the performance of the proposed method with those of both the NLS-based method and the MC-HMUSIC.

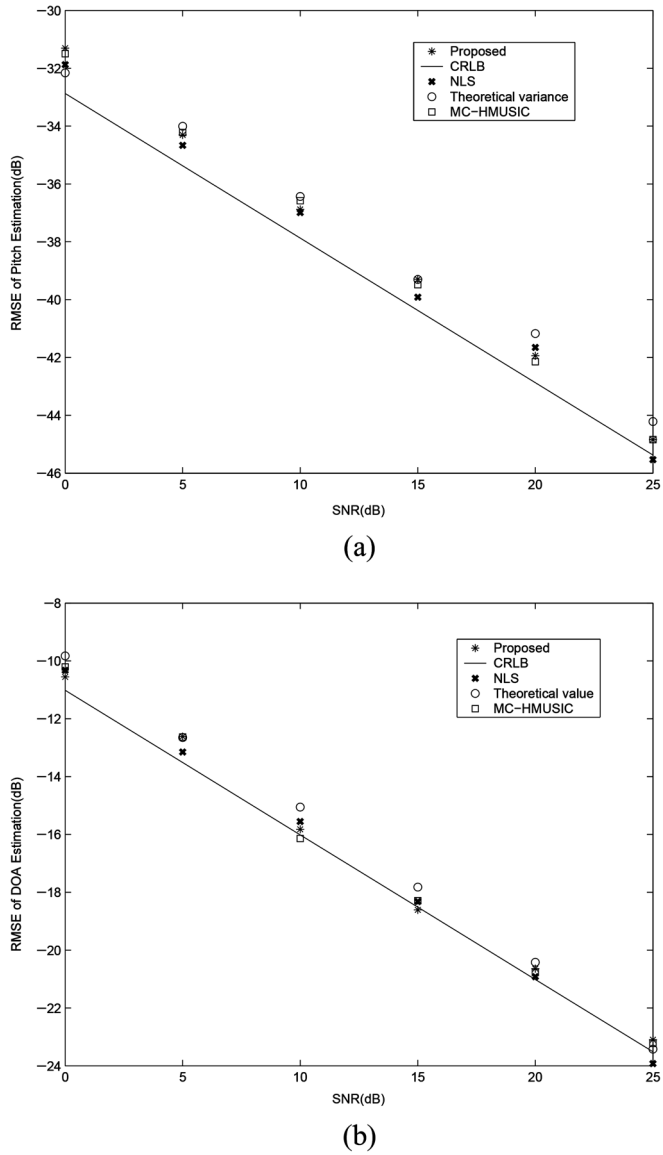

Fig. 1. Comparison of RMSEs for the estimated (a) pitch and (b) DOA parameters versus SNR.

The RMSEs of the proposed estimators (MSP-ESPRIT) versus different SNRs were compared to those of both the NLS-based method and the MC-HMUSIC method. Fig. 1 show that the performance of the proposed method is very similar to the latter two methods and also follows the corresponding CRLB in all SNRs cases, and that the corresponding theoretical values of the proposed estimation fit the simulation results. We also observe from Fig. 1 that the performance in terms of the RMSE, seem to be below the CRB in some cases (SNR = 25 dB). However, this is due to the limited number of Monte Carlo trials, making the measured RMSE fluctuate around the CRB. In addition, we fix the number of samples as $N = 80$, and check the average elapsed time of three different methods for single run versus different number of sensors. It is seen from Fig. 2 that the simulation result also confirms the significant reduction in computational complexity obtained with the proposed approach.

In the second experiment, we evaluated estimations in the multi-pitch case. The harmonic signal consisted of $P = 2$ pitches, each with $L_1 = L_2 = 2$ tones. The other experimental

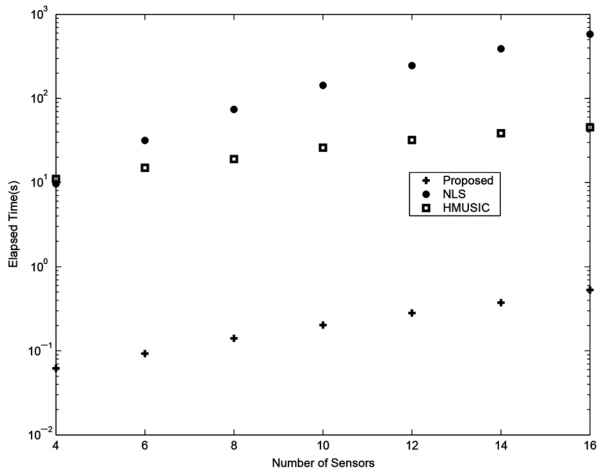

Fig. 2. Comparison of the average elapsed time of single run versus the number of sensors.

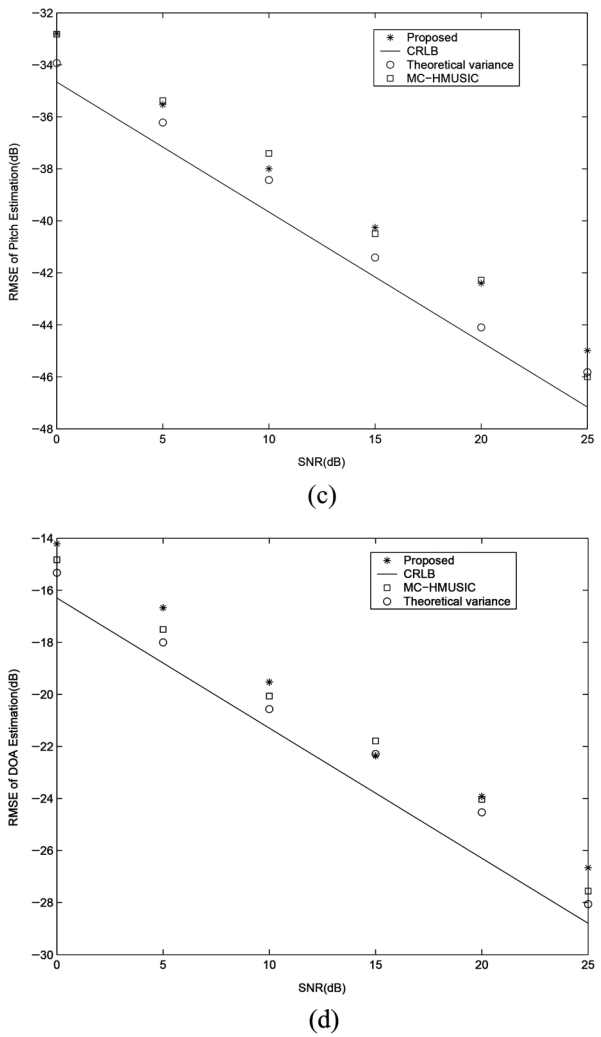

Fig. 3. Comparison of RMSE for the estimated (c) pitch and (d) DOA parameters of source 1.

parameter setting is listed in Table III. The number of sensors was $R = 8$ and the number of samples was $N = 120$, respectively. In the experiment, the performance of the recently presented MC-HMUSIC method [34] was compared to that of

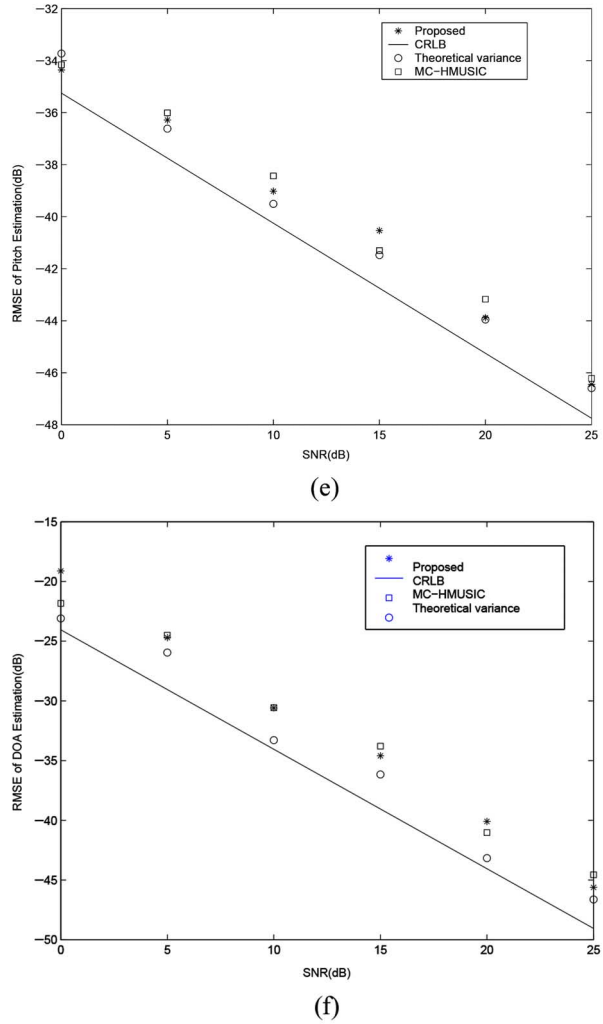

Fig. 4. Comparison of RMSE for the estimated (e) pitch and (f) DOA parameters of source 2.

TABLE III
SIMULATION SETTING OF BOTH THE MMP-ESPRIT AND THE MC-HMUSIC METHOD

p	L_p	Pitch	DOA
1	2	343	45
2	2	116	15

the proposed method. Figs. 3–4 shows the RMSEs for pitch and DOA estimates of the two methods in the different SNR cases. Figs. 3–4 indicates that the performance of the proposed method is similar to that of the MC-HMUSIC estimator and the RMSEs of the two methods are also close to the corresponding CRLB for all SNRs. Furthermore, the simulated results are consistent with those of the theoretical variances in (63) and (64).

For the purpose of further comparison, the resolvability performance in frequency as well as the angle of the two methods was compared. The SNR was set to 25 dB and the number of samples was $N = 120$. The other parameters in this experiment were the same as those in the second experiment. Firstly, both the parameters of source 1 and the DOA of source 2 were fixed, while the pitch of source 2 varied with different $\Delta\omega = |\omega_1 - \omega_2|$. Fig. 5 shows that the resolvability performance of

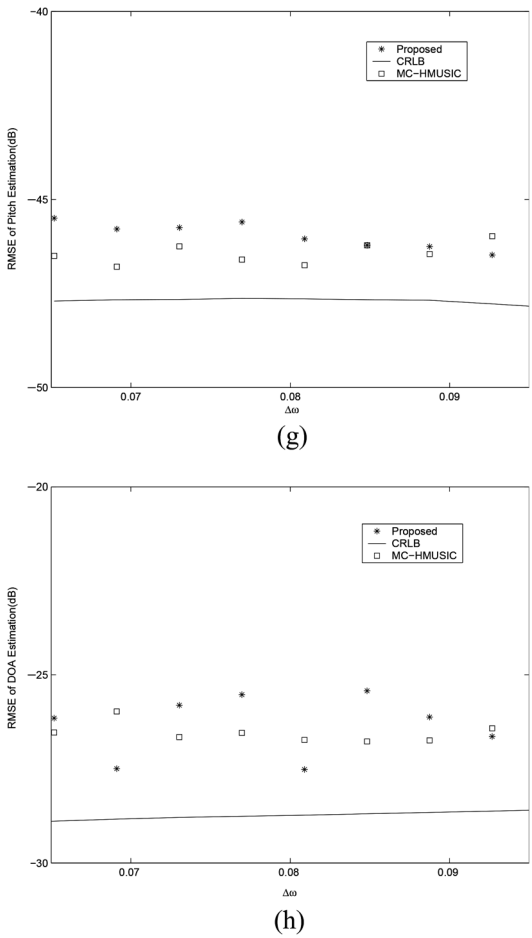

Fig. 5. Comparison of RMSE for the estimated (g) pitch and (h) DOA parameters of source 1.

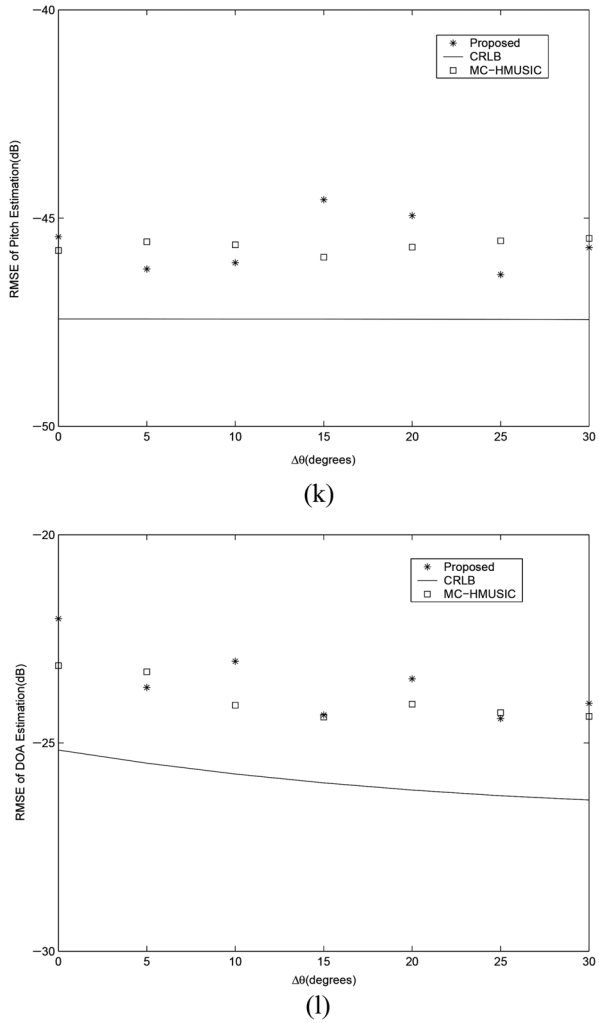

Fig. 6. Comparison of RMSE for the estimated (k) pitch and (l) DOA parameters of source 1.

the two methods is similar, but there is about a 2 dB gap between their RMSEs and the corresponding CRLB.

Similar to the above experiment on resovlability performance in frequency, both parameters of source 1 and the pitch of source 2 were now fixed in this experiment, while the DOA of source 2 varied with different $\Delta\theta = |\theta_1 - \theta_2|$. It is seen from the Fig. 6 that the resolvability performance in spatial angle of the two methods is also similar, but with a 2 dB gap between their RMSEs and the corresponding CRLB. In addition, it is worthy to note that both the MC-HMUSIC estimator and the proposed estimators can still perform well when the DOAs of the two sources are identical, i.e., $\Delta\theta = 0$, due to the fact that the two sources have different pitches and the estimations of DOA are dependent on the pitch parameters. However, the pitch estimates of the two sources are directly obtained from the eigenvalues of constructed matrices; therefore, the two sources cannot be resolved in the case of the same pitch (i.e., $\Delta\omega = 0$) without using the spatio-temporal smoothing technique to restore the rank of $\mathbf{A}$.

In the last simulation experiment, we show an example of joint DOA and single-pitch estimation in the presence of multi-path propagation. The harmonic signal consisted of L sinusoids with a pitch of $f_0 = 343$ Hz and assumed that there were 3 multi-path signals with DOAs $= [-60^o, 10^o, 45^o]$ received at the array, which the parameter setting is listed in Table IV. The number of microphones was set $R = 16$, SNR was set as SNR $= 30$ dB, and the number of samples was $N = 120$. The smoothed parameters in the spatio-temporal domain were taken as $M = 80$ and $R_0 = 8$. The results for 50 independent trials are shown in Fig. 7, which shows that the proposed method can still give accurate DOA and pitch estimation in the multi-path propagation case, and that the result for $L = 5$ is better than the result for $L = 2$ in same conditions.

B. Real-life Data Results

In this subsection, an experiment was conducted to also evaluate the performance of the proposed method on real-life signals. The real-life data were also used to measure the performance of the NLS-based method [36] for the single pitch case.

The experiment took place in a meeting room and the experimental environment was depicted in [36]. In the experiment, the sampling frequency was $f_s = 44.1$ kHz, the speed of sound was assumed to be $c = 343.2$ m/s, and the room reverberation time at 1 kHz was $T_{60} \approx 0.53$ s, the receiving array was a uniform and linear array with an inter-element spacing of $d = 4$ cm and the number of sensors was $R = 8$. We assumed that there was no attenuation across the sensors such that $\alpha_{p,l} = 1$, and the desired signal was assumed to consist of $L = 8$ harmonics. In addition, $N = 100$ was used for each parameter estimation.

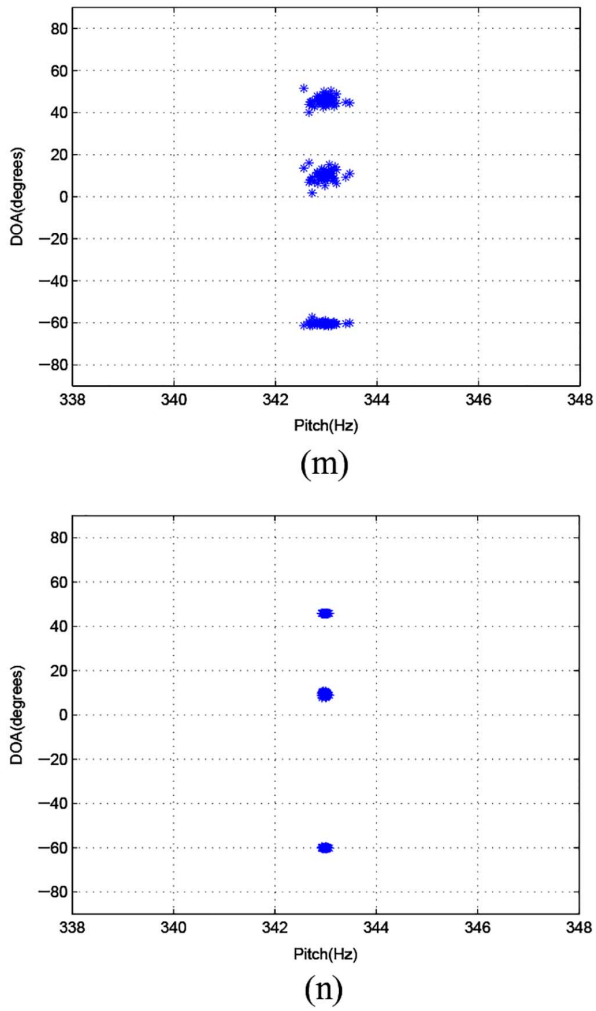

Fig. 7. Comparison of the estimated pitch and DOA parameters in multi-path case for L=2 (m), and L=5 (n), respectively.

TABLE IV
SIMULATION SETTING OF THE MSP-MP-ESPRIT

p	L_p	Pitch	DOA
1	2 5	343	45
2	2 5	343	−60
3	2 5	343	10

The estimates in the experiment were set up as in the previous simulations with synthetic data.

In the part of the real-life experiment, the single pitch case was tested using a played-back trumpet note "S1" in [36]. The anechoic trumpet signal was generated by concatenating anechoic trumpet signal excerpts (see http://theremin.music.uiowa.edu/MIS.html) [43]. The played back trumpet signal was recorded using the ULA to obtain a multichannel trumpet signal with slight reverberation, and the DOA of the single trumpet signal was about $\theta_1 = -13^\circ$. From the recorded data, we estimated the pitch and the DOA of the trumpet signal via the proposed method and the NLS-based method. The results obtained from this experiment are shown in Fig. 9. It can be seen that the two applied estimators for the pitch estimation yielded identical estimates in the recorded time interval, but that the estimation accuracy of the pitch using the NLS-based method was better than the proposed method,

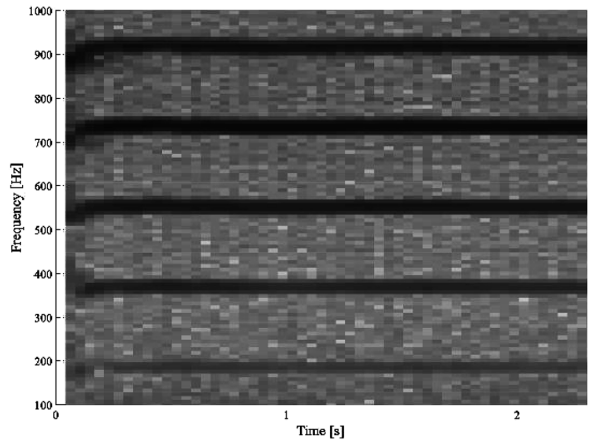

Fig. 8. Spectrogram of real recorded data from Sensor 1.

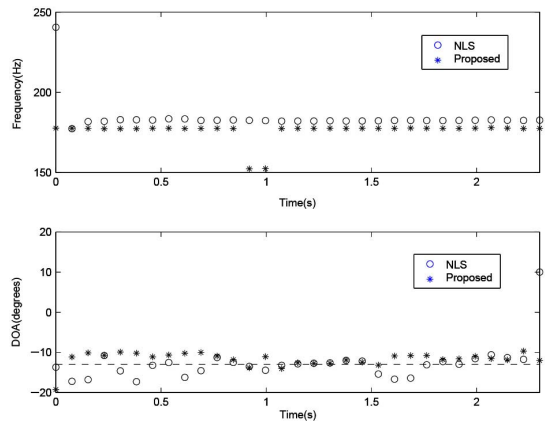

Fig. 9. Comparison of the estimated pitch and DOA parameters for real recorded data.

when the results of the pitch parameter were only compared to the corresponding spectrogram shown in Fig. 8. However, the estimated DOAs using the two methods were relatively close to the true DOA.

VII. CONCLUSION

In this paper, the ESPRIT algorithm was used for joint pitch and DOA estimation of multiple channel multi-pitch with a ULA. Based on a constructed spatio-temporal signal model, the invariance property in both the time-domain and the spatial domain was exploited to estimate the pitch and DOA parameters using the ESPRIT method. We also demonstrated that the spatio-temporal smoothing technique can be used to solve the single-pitch multi-path propagation case. Compared to the available state-of-the-art algorithms, the proposed method based on ESPRIT without 2-D searching is computationally more efficient. The asymptotic performance of the estimated pitch and DOA for multi-pitch case was also analyzed. Finally, numerous simulations using synthetic data were conducted to compare the performance of the proposed method with both the NLS-based method and the MC-HMUSIC method for the single pitch case and the MC-HUMUSIC method for the multi-pitch case. The results show that the performance of the proposed estimator is not only comparable to the other methods for both pitch and DOA estimation in terms of the root mean square error but is also close to the corresponding CRLB. Furthermore, the experimental test on real-life data demonstrates that the

proposed method can be applied to real-life signals and yields a similar estimation performance to the NLS-based method. In future work, the problem of joint DOA and pitch estimation for multi-pitch and multi-path scenarios will be considered.

APPENDIX

Following the steps in [41], the derivation of the variances of the estimated pitch and DOA is given as follows:

(1) Variance of pitch estimation: For the pitch estimation, the parameter $\hat{\omega}_p$ is related to the eigenvalue $\mu_l^{(p)}$ of $\mathbf{\Phi}$ by

$$\mu_l^{(p)} = e^{jl\omega_p}, l = 1, \cdots, L_p. \tag{65}$$

Using the first-order Taylor series expansion, yields

$$\delta\mu_l^{(p)} = j\mu_l^{(p)} l\delta\omega_p, l = 1, \cdots, L_p. \tag{66}$$

Thus, we have

$$\delta\omega_p = \frac{1}{j\mu_l^{(p)} l}\delta\mu_l^{(p)}. \tag{67}$$

For $l = 1, \cdots, L_p$, we take the average of all the $\delta\omega_p$,

$$\delta\omega_p = \frac{1}{L_p}\sum_{l=1}^{L_p}\frac{1}{j\mu_l^{(p)} l}\delta\mu_l^{(p)}. \tag{68}$$

Since the $\delta\omega_p$ is a real quantity, we get

$$\begin{aligned}\delta\omega_p &= \frac{1}{2L_p}\sum_{l=1}^{L_p} real\{\frac{1}{j\mu_l^{(p)} l}\delta\mu_l^{(p)} + \frac{1}{-j\mu_l^{*(p)} l}\delta\mu_l^{*(p)}\} \\ &= \frac{1}{L_p}\sum_{l=1}^{L_p}\frac{real\{-j\mu_l^{*(p)}\delta\mu_l^{(p)}\}}{l}.\end{aligned} \tag{69}$$

The variance of the estimated pitch parameter is

$$E\{(\delta\omega_p)^2\} = \frac{1}{L_p^2}E\{[\sum_{l=1}^{L_p}\frac{real\{j\mu_l^{*(p)}\delta\mu_l^{(p)}\}}{l}]^2\}. \tag{70}$$

For simplicity, here we consider two harmonic components with $L_p = 2$ for $E\{(\omega_P)^2\}$ and $E\{(\theta_P)^2\}$, respectively.

When $L_p = 2$, we have

$$\delta\omega_p = \frac{1}{2}(\frac{1}{j\mu_1^{(p)}}\delta\mu_1^{(p)} + \frac{1}{2j\mu_2^{(p)}}\delta\mu_2^{(p)}). \tag{71}$$

Since the $\delta\omega_p$ is a real quantity, we get

$$\delta\omega_p = \frac{1}{2}(\delta\omega_p + \delta\omega_p'). \tag{72}$$

After complicated algebra, the variance of the estimated pitch parameter is

$$\begin{aligned}E\{(\delta\omega_p)^2\} &= \frac{1}{4}(\delta\omega_p + \delta\omega_p^*)(\delta\omega_p + \delta\omega_p^*) \\ &= \frac{1}{8}|\mu_1^{(p)}|^2 E\{|\delta\mu_1^{(p)}|^2\} \\ &\quad + \frac{1}{32}|\mu_2^{(p)}|^2 E\{|\delta\mu_2^{(p)}|^2\} \\ &\quad - \frac{1}{8}real\Big[(\mu_1^{*(p)})^2 E\{(\delta\mu_1^{(p)})^2\} \\ &\quad - \frac{1}{32}(\mu_2^{*(p)})^2 E\{(\delta\mu_2^{(p)})^2\} \\ &\quad - \frac{1}{8}\mu_1^{*(p)}\mu_2^{*(p)} E\{\delta\mu_1^{(p)}\delta\mu_2^{(p)}\} \\ &\quad + \frac{1}{8}\mu_1^{(p)}\mu_2^{*(p)} E\{\delta\mu_1^{*(p)}\delta\mu_2^{(p)}\}\Big].\end{aligned} \tag{73}$$

(2) DOA: For DOA estimation, the parameter θ_p is dependent on ω_p, which is related to $\nu_l^{(p)}$ by

$$\nu_l^{(p)} = e^{jl\phi_p} = e^{jl\omega_p f_s\frac{d}{c}sin(\theta_p)}, l = 1, \cdots, L_p. \tag{74}$$

Using the first-order Taylor series expansion and (40), after some algebra, yields

$$\begin{aligned}\delta\theta_p &= \frac{\delta\nu_l^{(p)} - j\nu_l^{(p)} l f_s\frac{d}{c}\sin\theta_p\delta\omega_p}{j\nu_l^{(p)} l\omega_p f_s\frac{d}{c}\cos\theta_p} \\ &= \frac{\delta\nu_l^{(p)} - \nu_l^{(p)}\mu_l^{*(p)} f_s\frac{d}{c}\sin\theta_p\delta\mu_l^{(p)}}{j\nu_l^{(p)} l\omega_p f_s\frac{d}{c}\cos\theta_p}.\end{aligned} \tag{75}$$

For $l = 1, \cdots, L_p$, we take the average of all the $\delta\theta_p$, and get

$$\begin{aligned}\delta\theta_p &= \frac{1}{L_p}\sum_{l=1}^{L_p}\frac{\delta\nu_l^{(p)} - \nu_l^{(p)}\mu_l^{*(p)} f_s\frac{d}{c}\sin\theta_p\delta\mu_l^{(p)}}{j\nu_l^{(p)} l\omega_p f_s\frac{d}{c}\cos\theta_p} \\ &= \frac{1}{L_p\omega_p f_s\frac{d}{c}\cos\theta_p}\sum_{l=1}^{L_p}\frac{1}{l}(-j\nu_l^{*(p)}\delta\nu_l^{(p)} \\ &\quad + j\mu_l^{*(p)} f_s\frac{d}{c}\sin\theta_p\delta\mu_l^{(p)}).\end{aligned} \tag{76}$$

Similarly, we take

$$\begin{aligned}\delta\theta_p &= \frac{1}{L_p\omega_p f_s\frac{d}{c}\cos\theta_p} real\{\sum_{l=1}^{L_p}\frac{1}{l}(-j\nu_l^{*(p)}\delta\nu_l^{(p)} \\ &\quad + j\mu_l^{*(p)} f_s\frac{d}{c}\sin\theta_p\delta\mu_l^{(p)})\},\end{aligned} \tag{77}$$

and, hence, the variance of the estimated DOA parameter is computed as

$$\begin{aligned}E\{(\delta\theta_p)^2\} &= \frac{1}{(L_p\omega_p f_s\frac{d}{c}\cos\theta_p)^2}E\{[\sum_{l=1}^{L_p}\frac{1}{l}real\{ \\ &\quad - j\nu_l^{*(p)}\delta\nu_l^{(p)} + j\mu_l^{*(p)} f_s\frac{d}{c}\sin\theta_p\delta\mu_l^{(p)}\}]^2\}.\end{aligned} \tag{78}$$

Similarly, for $L_p = 2$, we have

$$\begin{aligned}\delta\theta_p &= \frac{1}{2\omega_p f_s\frac{d}{c}\cos\theta_p}\sum_{l=1}^{2}\frac{1}{l}(-j\nu_l^{*(p)}\delta\nu_l^{(p)} \\ &\quad + j\mu_l^{*(p)} f_s\frac{d}{c}\sin\theta_p\delta\mu_l^{(p)}) \\ &= \frac{1}{2jA}[\nu_1^{*(p)}\delta\nu_1^{(p)} - B\mu_1^{*(p)}\delta\mu_1^{(p)} + \frac{1}{2}\nu_2^{*(p)}\delta\nu_2^{(p)} \\ &\quad - \frac{1}{2}B\mu_2^{*(p)}\delta\mu_2^{(p)}],\end{aligned} \tag{79}$$

where $A = \omega_p f_s\frac{d}{c}\cos\theta_p$ and $B = f_s\frac{d}{c}\sin\theta_p$.

Let $\delta\theta_p = 12(\delta\theta_p + \delta\theta_p^*)$, and then we have the equation shown at the bottom of the next page.

Using a similar derivation similar to the one in [41], the quantities $E\{|\delta\mu_l^{(p)}|^2\}$, $E\{(\delta\mu_l^{(p)})^2\}$, $E\{|\delta\nu_l^{(p)}|^2\}$, $E\{(\delta\nu_l^{(p)})^2\}$, $E\{\delta\mu_l^{(p)}\delta\nu_l^{(p)}\}$ and $E\{\delta\mu_l^{(p)*}\delta\nu_l^{(p)}\}$ are the covariances of eigenvalues, which are derived one by one.

$$\delta\mu_l^{(p)} = \mathbf{u}_l^{(p)}\mathbf{U}_1^{\#}(\mathbf{W}_2 - \mu_l^{(p)}\mathbf{W}_1)\delta\mathbf{U}_s\mathbf{q}_l^{(p)}, \tag{81}$$

where $\mathbf{u}_l^{(p)}, \mathbf{q}_l^{(p)}$ are the corresponding left and right eigenvectors of $\mathbf{\Phi}$, and $\delta\mathbf{U}_s = \hat{\mathbf{U}}_s - \mathbf{U}_s$, which is given as follows (see [42] for details),

$$\delta\mathbf{U}_s = \mathbf{U}_n\mathbf{U}_n^H\delta\mathbf{Y}\mathbf{V}_s\Sigma_s^{-1}, \tag{82}$$

$$\delta\mathbf{Y} = \hat{\mathbf{Y}} - \mathbf{Y}. \tag{83}$$

Similarly, $\mathbf{v}_l^{(p)}, \mathbf{h}_l^{(p)}$ are the corresponding left and right eigenvectors of the matrix $\mathbf{\Phi}_s$, and the error of $\delta\nu_i$ is then given by

$$\delta\nu_l^{(p)} = \mathbf{v}_l^{(p)}\mathbf{U}_3^{\#}(\mathbf{W}_4 - \nu_l^{(p)}\mathbf{W}_3)\delta\mathbf{U}_s\mathbf{h}_l^{(p)}. \tag{84}$$

From (81) and (84), the variances of $\hat{\mu}_l^{(p)}$ and $\hat{\nu}_l^{(p)}$ are thus computed by the following expressions

$$\begin{aligned}
E\{|\delta\mu_l^{(p)}|^2\} &= \mathbf{u}_l^{(p)}\mathbf{U}_1^{\#}(\mathbf{W}_2 - \mu_l^{(p)}\mathbf{W}_1)E\{\delta\mathbf{U}_s\mathbf{q}_l^{(p)} \\
&\quad \mathbf{q}_l^{H(p)}\delta\mathbf{U}_s^H\}(\mathbf{W}_2 - \mu_l^{(p)}\mathbf{W}_1)^H(\mathbf{u}_l^{(p)}\mathbf{U}_1^{\#})^H \\
&= \mathbf{u}_l^{(p)}\mathbf{U}_1^{\#}(\mathbf{W}_2 - \mu_l^{(p)}\mathbf{W}_1)\mathbf{U}_n\mathbf{U}_n^H E\{\delta\mathbf{Y}\mathbf{V}_s \\
&\quad \Sigma_s^{-1}\mathbf{q}_l^{(p)}\mathbf{q}_l^{H(p)}\Sigma_s^{-1*}\mathbf{V}_s^H\delta\mathbf{Y}^H\} \\
&\quad \mathbf{U}_n\mathbf{U}_n^H(\mathbf{W}_2 - \mu_l^{(p)}\mathbf{W}_1)^H(\mathbf{u}_l^{(p)}\mathbf{U}_1^{\#})^H
\end{aligned} \tag{85}$$

$$\begin{aligned}
E\{(\delta\mu_l^{(p)})^2\} &= \mathbf{u}_l^{(p)}\mathbf{U}_1^{\#}(\mathbf{W}_2 - \mu_l^{(p)}\mathbf{W}_1)\mathbf{U}_n\mathbf{U}_n^H E\{\delta\mathbf{Y}\mathbf{V}_s\Sigma_s^{-1} \\
&\quad \mathbf{q}_l^{(p)}\mathbf{q}_l^{T(p)}\Sigma_s^{-1}\mathbf{V}_s^T\delta\mathbf{Y}^T\}\mathbf{U}_n\mathbf{U}_n^T(\mathbf{W}_2 - \mu_l^{(p)}\mathbf{W}_1)^T(\mathbf{u}_l^{(p)}\mathbf{U}_1^{\#})^T
\end{aligned}$$

$$\begin{aligned}
E\{|\delta\nu_l^{(p)}|^2\} &= \mathbf{v}_l^{(p)}\mathbf{U}_3^{\#}(\mathbf{W}_4 - \nu_l^{(p)}\mathbf{W}_3)\mathbf{U}_n\mathbf{U}_n^H \\
&\quad E\{\delta\mathbf{Y}\mathbf{V}_s\Sigma_s^{-1}\mathbf{h}_l^{(p)}\mathbf{h}_l^{H(p)}\Sigma_s^{-1*}\mathbf{V}_s^H\delta\mathbf{Y}^H\} \\
&\quad \mathbf{U}_n\mathbf{U}_n^H(\mathbf{W}_4 - \nu_l^{(p)}\mathbf{W}_3)^H(\mathbf{v}_l^{(p)}\mathbf{U}_3^{\#})^H
\end{aligned} \tag{86}$$

$$\begin{aligned}
E\{(\delta\nu_l^{(p)})^2\} &= \mathbf{v}_l^{(p)}\mathbf{U}_3^{\#}(\mathbf{W}_4 - \nu_l^{(p)}\mathbf{W}_3)\mathbf{U}_n\mathbf{U}_n^H \\
&\quad E\{\delta\mathbf{Y}\mathbf{V}_s\Sigma_s^{-1}\mathbf{h}_l^{(p)}\mathbf{h}_l^{T(p)}\Sigma_s^{-1}\mathbf{V}_s^T\delta\mathbf{Y}^T\} \\
&\quad \mathbf{U}_n\mathbf{U}_n^T(\mathbf{W}_4 - \nu_l^{(p)}\mathbf{W}_3)^T(\mathbf{v}_l^{(p)}\mathbf{U}_3^{\#})^T
\end{aligned} \tag{87}$$

$$\begin{aligned}
E\{\delta\mu_l^{(p)}\delta\nu_l^{*(p)}\} &= \mathbf{u}_l^{(p)}\mathbf{U}_1^{\#}(\mathbf{W}_2 - \mu_l^{(p)}\mathbf{W}_1)\mathbf{U}_n\mathbf{U}_n^H \\
&\quad E\{\delta\mathbf{Y}\mathbf{V}_s\Sigma_s^{-1}\mathbf{q}_l^{(p)}\mathbf{h}_l^{H(p)}\Sigma_s^{-1*}\mathbf{V}_s^H\delta\mathbf{Y}^H\} \\
&\quad \mathbf{U}_n\mathbf{U}_n^H(\mathbf{W}_4 - \nu_l^{(p)}\mathbf{W}_3)^H(\mathbf{v}_l^{(p)}\mathbf{U}_3^{\#})^H
\end{aligned} \tag{88}$$

$$\begin{aligned}
E\{\delta\mu_l^{(p)}\delta\nu_l^{(p)}\} &= \mathbf{u}_l^{(p)}\mathbf{U}_1^{\#}(\mathbf{W}_2 - \mu_l^{(p)}\mathbf{W}_1)\mathbf{U}_n\mathbf{U}_n^H \\
&\quad E\{\delta\mathbf{Y}\mathbf{V}_s\Sigma_s^{-1}\mathbf{q}_l^{(p)}\mathbf{h}_l^{T(p)}\Sigma_s^{-1}\mathbf{V}_s^T\delta\mathbf{Y}^T\} \\
&\quad \mathbf{U}_n\mathbf{U}_n^T(\mathbf{W}_4 - \nu_l^{(p)}\mathbf{W}_3)^T(\mathbf{v}_l^{(p)}\mathbf{U}_3^{\#})^T
\end{aligned} \tag{89}$$

In the equation (85), let $\mathbf{f}_l^{(p)} = \mathbf{V}_s\Sigma_s^{-1}\mathbf{q}_l^{(p)}$ and $\mathbf{g}_l^{(p)} = \mathbf{V}_s\Sigma_s^{-1}\mathbf{h}_l^{(p)}$, and then we get

$$\begin{aligned}
&E\{\delta\mathbf{Y}\mathbf{V}_s\Sigma_s^{-1}\mathbf{q}_l^{(p)}\mathbf{q}_l^{H(p)}\Sigma_s^{-1*}\mathbf{V}_s^H\delta\mathbf{Y}^H\} \\
&= E\{\delta\mathbf{Y}\mathbf{f}_l^{(p)}\mathbf{f}_l^{H(p)}\delta\mathbf{Y}^H\} \\
&= \sum_{i=1}^{N-M+1}\sum_{j=1}^{N-M+1} f_{l,i}^{(p)}f_{l,j}^{*(p)}E\{\delta\mathbf{y}_i\delta\mathbf{y}_j^H\},
\end{aligned} \tag{90}$$

$$\begin{aligned}
&E\{\delta\mathbf{Y}\mathbf{V}_s\Sigma_s^{-1}\mathbf{h}_l^{(p)}\mathbf{h}_l^{H(p)}\Sigma_s^{-1*}\mathbf{V}_s^H\delta\mathbf{Y}^H\} \\
&= E\{\delta\mathbf{Y}\mathbf{g}_l^{(p)}\mathbf{g}_l^{H(p)}\delta\mathbf{Y}^H\} \\
&= \sum_{i=1}^{N-M+1}\sum_{j=1}^{N-M+1} g_{l,i}^{(p)}g_{l,j}^{*(p)}E\{\delta\mathbf{y}_i\delta\mathbf{y}_j^H\}.
\end{aligned} \tag{91}$$

where $f_{l,i}^{(p)}$ is the i-th element of the vector $\mathbf{f}_l^{(p)}$, $g_{l,j}^{(p)}$ is the j-th element of the vector $\mathbf{g}_l^{(p)}$, and $\delta\mathbf{y}_i$ is the i-th column of the

---

$$\begin{aligned}
\delta\theta_p &= \frac{1}{4jA}[\nu_1^{*(p)}\delta\nu_1^{(p)} - B\mu_1^{*(p)}\delta\mu_1^{(p)} + \frac{1}{2}\nu_2^{*(p)}\delta\nu_2^{(p)} \\
&\quad - \frac{1}{2}B\mu_2^{*(p)}\delta\mu_2^{(p)} - \nu_1^{(p)}\delta\nu_1^{*(p)} + B\mu_1^{(p)}\delta\mu_1^{*(p)} \\
&\quad - \frac{1}{2}\nu_2^{(p)}\delta\nu_2^{*(p)} + 12B\mu_2^{(p)}\delta\mu_2^{*(p)}],
\end{aligned} \tag{80}$$

$$\begin{aligned}
E\{(\delta\theta_p)^2\} &= \frac{1}{16A^2}\{2|\nu_1^{(p)}|^2E\{|\delta\nu_1^{(p)}|^2\} + 2B^2|\mu_1^{(p)}|^2E\{|\delta\mu_1^{(p)}|^2\} \\
&\quad + \frac{1}{2}B^2|\mu_2^{(p)}|^2E\{|\delta\mu_2^{(p)}|^2\} + \frac{1}{2}|\nu_2^{(p)}|^2E\{|\delta\nu_2^{(p)}|^2\} \\
&\quad - 2real[(\nu_1^{*(p)})^2E\{(\delta\nu_1^{(p)})^2\} + B^2(\mu_1^{*(p)})^2E\{(\delta\mu_1^{(p)})^2\} \\
&\quad + \frac{1}{4}(\nu_2^{*(p)})^2E\{(\delta\nu_2^{(p)})^2\} + \frac{1}{4}B^2(\mu_2^{*(p)})^2E\{(\delta\mu_2^{(p)})^2\}] \\
&\quad - 2real[-2B\mu_1^{*(p)}\nu_1^{*(p)}E\{\delta\nu_1^{(p)}\delta\mu_1^{(p)}\} + \nu_1^{*(p)}\nu_2^{*(p)}E\{\delta\nu_1^{(p)}\delta\nu_2^{(p)}\} \\
&\quad + B^2\mu_1^{*(p)}\mu_2^{*(p)}E\{\delta\mu_1^{(p)}\delta\mu_2^{(p)}\} - B\mu_2^{*(p)}\nu_1^{*(p)}E\{\delta\nu_1^{(p)}\delta\mu_2^{(p)}\} \\
&\quad + B\mu_2^{*(p)}\nu_1^{(p)}E\{\delta\mu_2^{(p)}\delta\nu_1'\} - \nu_1'\nu_2E\{\delta\nu_1^{(p)}\delta\nu_2^{*(p)}\} \\
&\quad - B\mu_1^{*(p)}\nu_2^{*(p)}E\{\delta\nu_2^{(p)}\delta\mu_1^{(p)}\} + B\mu_1^{*(p)}\nu_2^{(p)}E\{\delta\nu_2^{*(p)}\delta\mu_1^{(p)}\} \\
&\quad + 2B\mu_1^{*(p)}\nu_1^{(p)}E\{\delta\nu_1^{*(p)}\delta\mu_1^{(p)}\} - B^2\mu_1^{*(p)}\mu_2^{(p)}E\{\delta\mu_1^{(p)}\delta\mu_2^{*(p)}\} \\
&\quad - \frac{1}{2}B\mu_2^{*(p)}\nu_2^{*(p)}E\{\delta\nu_2^{(p)}\delta\mu_2^{(p)}\} + \frac{1}{2}B\mu_2^{*(p)}\nu_2^{(p)}E\{\delta\nu_2^{*(p)}\delta\mu_2^{(p)}\}]\}.
\end{aligned}$$

matrix $\delta\mathbf{Y}$. The simplified formulas in Equations (86)-(89) can be derived similarly and are omitted here.

REFERENCES

[1] D. Wang and G. J. Brown, *Computational Auditory Scene Analysis Principles, Algorithm, and Applications*. New York, NY, USA: Wiley, 2006.

[2] S. Makino, T. W. Lee, and H. Sawada, *Blind Speech Separation*. New York, NY, USA: Springer, 2007.

[3] M. G. Christensen and A. Jakobsson, "Multi-pitch estimation," *Synth. Lect. Speech Audio Process*, vol. 5, no. 1, pp. 1–160, 2009.

[4] Y. Medan, E. Yair, and D. Chazan, "Super resolution pitch determination of speech signals," *IEEE Trans. Signal Process.*, vol. 39, no. 1, pp. 40–48, Jan. 1991.

[5] M. G. Christensen, A. Jakobsson, and S. H. Jensen, "Fundamental frequency estimation using the shift-invariance property," in *Proc. 41st Asilomar Conf. Signals, Systems, Comput. (ACSSC'07)*, 2007, pp. 631–635.

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

[7] V. Emiya, B. David, and R. Badeau, "A parametric method for pitch estimation of piano tones," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 1, pp. 249–52, Apr. 2007.

[8] M. G. Christensen, P. Stoica, A. Jakobsson, and S. H. Jensen, "Multipitch estimation," *Elsevier Signal Process.*, vol. 88, no. 4, pp. 972–983, Apr. 2008.

[9] M. G. Christensen, "Multi-channel maximum likelihood pitch estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2012, pp. 409–412.

[10] H. Krim and M. Viberg, "Two decades of array processing research-the parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, Jul. 1996.

[11] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ, USA: Pearson Education, 2005.

[12] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, *Exact and large sample maximum likelihood techniques for parameter estimation and detection in array processing*. New York, NY, USA: Springer, 1993.

[13] N. L. Owsley, *Array Signal Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1985, pp. 115–193.

[14] R. Levanda and A. Leshem, "Synthetic aperture radio telescopes," *IEEE Signal Process. Mag.*, vol. 21, no. 1, pp. 14–29, Jan. 2010.

[15] P. Stoica and J. Li, "Lecture notes-source localization from range-difference measurements," *IEEE Signal Process. Mag.*, vol. 23, no. 6, pp. 63–66, Nov. 2006.

[16] J. Chen, S. Gannot, and J. Benesty, "Part H: Speech enhancement, in springer handbook of speech processing," in , J. Benesty, Y. A. Huang, and M. MohanSondhi, Eds. New York, NY, USA: Springer, Nov. 2007.

[17] M. Wohlmayr and M. Kepesi, "Joint position-pitch extraction from multichannel audio," in *Proc. Interspeech*, Aug. 2007, pp. 1629–1632.

[18] M. Kepesi, L. Ottowitz, and T. Habib, "Joint position-pitch estimation for multiple speaker scenarios," in *Proc. Hands-Free Speech Commun. Microphone Arrays*, May 2008, pp. 85–88.

[19] R. Roy and T. Kailath, "ESPRIT- Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process*, vol. 37, no. 7, pp. 984–995, Jul. 1989.

[20] M. Viberg and P. Stoica, "A computationally efficient method for joint direction finding and frequency estimation in colored noise," *Rec. Asilomar Conf. Signals, Syst., Comput*, vol. 2, pp. 1547–1551, Nov. 1998.

[21] T. Shu and X. Liu, "Robust and computationally efficient signal-dependent method for joint DOA and frequency estimation," *EURASIP J. Adv. Signal Process.*, vol. 1, no. 4, pp. 1–16, Apr. 2008.

[22] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.

[23] J. Capon, "Maximum-likelihood spectral estimation," in *Nonlinear Methods of Spectral Analysis*. New York, NY, USA: Springer-Verlag, 1983.

[24] A. Jakobsson, S. L. Marple, Jr, and P. Stoica, "Computationally efficient two-dimensional capon spectrum analysis," *IEEE Trans. Signal Process.*, vol. 48, no. 9, pp. 2651–2661, Sep. 2000.

[25] A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multi-channel wiener filtering for noise reduction," *Elsevier Signal Process.*, vol. 84, no. 12, pp. 2367–2387, Dec. 2004.

[26] M. S. Brandstein and H. F. Silverman, "A robust method for speech signal time-delay estimation in reverberant rooms," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 1, pp. 375–378, Apr. 1997.

[27] M. Jian, A. C. Kot, and M. H. Er, "DOA estimation of speech source with microphone arrays," *Proc. IEEE Int. Symp. Circuits Syst.*, vol. 5, pp. 293–296, May 1998.

[28] X. Qian and R. Kumaresan, "Joint estimation of time delay and pitch of voiced speech signals," *Rec. Asilomar Conf. Signals, Syst., Comput.*, vol. 1, pp. 735–739, Oct. 1995.

[29] Y. Wu, H. C. So, and Y. Tan, "Joint time-delay and frequency estimation using parallel factor analysis," *Elsevier Signal Process.*, vol. 89, pp. 1667–1670, 2009.

[30] L. Y. Ngan, Y. Wu, H. C. So, P. C. Ching, and S. W. Lee, "Joint time delay and pitch estimation for speaker localization," *Proc. IEEE Int. Symp. Circuits Syst.*, vol. 3, pp. 722–725, May 2003.

[31] Y. Wu, H. C. So, and P. C. Ching, "Joint time delay and frequency estimation via state-space realization," *IEEE Signal Process. Lett.*, vol. 10, no. 11, pp. 339–342, Nov. 2003.

[32] J. Dmochowski, J. Benesty, and S. Affes, "Linearly constrained minimum variance source localization and spectral estimation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 8, pp. 1490–1502, Nov. 2008.

[33] J. R. Jensen, M. G. Christensen, and S. H. Jensen, "Joint DOA and fundamental frequency estimation methods based on 2-D filtering," in *Proc. Eur. Signal Process. Conf.*, Aug. 2010, pp. 2091–2095.

[34] J. X. Zhang, M. G. Christensen, S. H. Jensen, and M. Moonen, "Joint DOA and multi-pitch estimation based on subspace techniques," *EURASIP J. Adv. Signal Process.*, no. 1, pp. 1–11, Jan. 2012.

[35] Z. H. Zhou, M. G. Christensen, and H. C. So, "Two stage DOA and fundamental frequency estimation based on subspace techniques," in *Proc. IEEE Int. Conf. Signal Process. (ICSP'12)*, Oct. 2012, vol. 1, pp. 210–214.

[36] J. R. Jensen, M. G. Christensen, and S. H. Jensen, "Nonlinear least squares methods for joint DOA and pitch estimation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 5, pp. 924–933, May. 2013.

[37] C. Duofang, C. Baixiao, and Q. Guodong, "Angle estimation using ESPRIT in MIMO radar," *Electron. Lett.*, vol. 44, no. 12, pp. 770–771, Jun. 2008.

[38] T. H. Liu and J. M. Mendel, "Azimuth and elevation direction finding using arbitrary array geometries," *IEEE Trans. Signal Process.*, vol. 46, no. 7, pp. 2061–2065, Jul. 1998.

[39] T. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 4, pp. 806–811, Aug. 1985.

[40] A. N. Lemma, A. J. Van der Veen, and E. F. Deprettere, "Analysis of joint angle-frequency estimation using ESPRIT," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1264–1283, May 2003.

[41] N. Yuen and B. Friedlander, "Asymptotic performance analysis of ESPRIT, higher-order ESPRIT, and virtual ESPRIT algorithms," *IEEE Trans. Signal Process.*, vol. 44, no. 10, pp. 2537–2550, Oct. 1996.

[42] J. Liu, X.-Q. Liu, and X.-L. Ma, "First-order perturbation analysis of singular value decomposition," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3044–3049, Jul. 2008.

[43] L. Armani and M. Omologo, "Weighted autocorrelation-based f0 estimation for distant-talking interaction with a distributed microphone network," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2004, vol. 1, pp. 113–116.

Yuntao Wu (M'12) received the Ph.D. degree in information and communication engineering from the National Key Lab. for Radar Signal Processing at Xidian University in December 2003 From April 2004 to September 2006, he was a Post-Doctoral Researcher at Institute of Acoustics, Chinese Academy of Sciences, Beijing, China. From Oct. 2006 to Feb. 2008, he was a Senior Research Fellow at the City University of Hong Kong. From April 2013 to March 2014, he was a Visiting Researcher in the Faculty of Engineering, Bar-Ilan University, Israel. Currently, he is a Full-Time Professor at the Wuhan Institute of Technology. At the same time, he is also a Chutian Scholar Project in Hubei Province Distinguish Professor at the same university. He has published over 60 journal and conference papers. He is a senior member of the Chinese Institute of Electronic Engineers (CIE), and an Associate Editor of Multidimensional Systems and Signal Processing (an international journal of Springer). His research interests include signal detection, parameter estimation in array signal processing and source localization for wireless sensor networks, biomedicine signal analysis, etc.

Amir Leshem (SM'06) received the B.Sc. (cum laude) in mathematics and physics, the M.Sc. (cum laude) in mathematics, and the Ph.D. in mathematics all from the Hebrew University, Jerusalem, Israel, in 1986, 1990, and 1998, respectively. He is a Professor and one of the founders of the faculty of engineering at Bar-Ilan University where he heads the signal processing track. From 1998 to 2000, he was with the Faculty of Information Technology and Systems, Delft University of Technology, The Netherlands, as a postdoctoral fellow. From 2000 to 2003, he was director of advanced technologies with Metalink Broadband where he was responsible for research and development of new DSL and wireless MIMO modem technologies and served as a member of ITU-T SG15, ETSI TM06, NIPP-NAI, IEEE 802.3, and 802.11. From 2000 to 2002, he was also a Visiting Researcher at the Delft University of Technology. From 2003 to 2005, he was also the Technical Manager of the U-BROAD consortium developing technologies to provide 100 Mbps and beyond over DSL lines. He was the leading guest editor a special issue of IEEE JOURNAL ON SELECTED TOPICS IN SIGNAL PROCESSING, dedicated to signal processing for space research and for a special issue of the SIGNAL PROCESSING MAGAZINE, dedicated to signal processing in astronomy and cosmology. In 2008-2011, he was an associate editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING.

His main research interests include multichannel wireless and wireline communication, applications of game theory to dynamic and adaptive spectrum management of communication and sensor networks, array and statistical signal processing with applications to multiple element sensor arrays and networks in radio-astronomy, brain research, wireless communications and radio-astronomical imaging, set theory, logic, and foundations of mathematics.

Jesper Rindom Jensen (S'09–M'12) was born in Ringkøbing, Denmark, in August 1984. He received the M.Sc. degree (cum laude) for completing the elite candidate education in 2009 from Aalborg University in Denmark. In 2012, he received the Ph.D. degree from Aalborg University.

Currently, he is a Postdoctoral Researcher at the Department of Architecture, Design, and Media Technology at Aalborg University in Denmark, where he is also a member of the Audio Analysis Lab. He has been a Visiting Researcher at the University of Quebec, INRS-EMT, in Montreal, Quebec, Canada, and at the Friedrich-Alexander Universität Erlangen-Nürnberg in Erlangen, Germany. He has published more than 30 papers in peer-reviewed conference proceedings and journals. Among others, his research interests are digital signal processing and microphone array signal processing theory and methods with application to speech and audio signals. In particular, he is interested in parametric analysis, modeling and extraction of such signals. Dr. Jensen has received an individual postdoc grant from the Danish Independent Research Council as well as several travel grants from private foundations.

Guisheng Liao (M'96) was born in Guiling, China. He received the B.S. degree from Guangxi University, Guangxi, China, in 1985 and the M.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 1990 and 1992, respectively. He is currently a Professor with Xidian University, where he is also the Dean of the School of Electronic Engineering. He has been a Senior Visiting Scholar with the Chinese University of Hong Kong, Shatin, Hong Kong. His research interests include synthetic aperture radar (SAR), space-time adaptive processing, SAR ground moving target indication, and distributed small satellite SAR system design. Prof. Liao is a member of the National Outstanding Person and the Cheung Kong Scholars in China.