



Joint DOA and multi-pitch estimation based on subspace techniques

Zhang, Johan Xi; Christensen, Mads Græsbøll; Jensen, Søren Holdt; Moonen, Marc

Published in:

Eurasip Journal on Advances in Signal Processing

DOI (link to publication from Publisher):

[10.1186/1687-6180-2012-1](https://doi.org/10.1186/1687-6180-2012-1)

Publication date:

2012

Document Version

Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

Citation for published version (APA):

Zhang, J. X., Christensen, M. G., Jensen, S. H., & Moonen, M. (2012). Joint DOA and multi-pitch estimation based on subspace techniques. *Eurasip Journal on Advances in Signal Processing*, 2012(1), 1-11.

<https://doi.org/10.1186/1687-6180-2012-1>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

RESEARCH

Open Access

Joint DOA and multi-pitch estimation based on subspace techniques

Johan Xi Zhang^{1*}, Mads Græsbøll Christensen², Søren Holdt Jensen¹ and Marc Moonen³

Abstract

In this article, we present a novel method for high-resolution joint direction-of-arrivals (DOA) and multi-pitch estimation based on subspaces decomposed from a spatio-temporal data model. The resulting estimator is termed multi-channel harmonic MUSIC (MC-HMUSIC). It is capable of resolving sources under adverse conditions, unlike traditional methods, for example when multiple sources are impinging on the array from approximately the same angle or similar pitches. The effectiveness of the method is demonstrated on a simulated an-echoic array recordings with source signals from real recorded speech and clarinet. Furthermore, statistical evaluation with synthetic signals shows the increased robustness in DOA and fundamental frequency estimation, as compared with a state-of-the-art reference method.

Keywords: multi-pitch estimation, direction-of-arrival estimation, subspace orthogonality, array processing

1. Introduction

The problem of estimating the fundamental frequency, or pitch, of a period waveform has been of interest to the signal processing community for many years. Fundamental frequency estimators are important for many practical applications such as automatic note transcription in music, audio and speech coding, classification of music, and speech analysis. Numerous algorithms have been proposed for both the single- and multi-pitch scenarios [1-5]. The problem for single-pitch scenarios is considered as well-posed. However, in real-world signals, the multi-pitch scenario occurs quite frequently [2,6]. The multi-pitch estimation algorithms are often based on, i.e., various modification of the auto-correlation function [1,7], maximum likelihood, optimal filtering, and subspace techniques [2,3,8]. In real-life recordings, problems such as frequency overlap of sources, reverberation, and colored noise will strongly limit the performance of multi-pitch estimator and estimator designed for single channel recordings often use simplified signal models. One widely used signal simplification in multi-pitch estimators, for example, is the sparseness of the signal, where the frequency spectrum

of sources are assumed to not overlap [2]. This assumption may be appropriate when sources consist of mixture of several speech signals having different pitches [9]. However, for audio signals it is less likely to be true. This is especially so in western music, where instruments are most often played in accord, something that causes the harmonics to overlap or even coincide. With only single-channel recording it is, therefore, hard, or perhaps even impossible, to estimate pitches with overlapping harmonics, unless additional information, such as a temporal or spectral model, is included.

Recently, multi-channel approaches have attracted considerable attention both in single- and multi-pitch scenarios. By exploring the spatial information of the sources, more robust pitch estimators have been proposed [10-14]. Most of those multi-channel methods are still mainly based on auto-correlation function-related approaches, however, although a few exceptions can be found in [15-18]. In direction-of-arrival (DOA) estimators, audio and speech signals are often modeled as broadband signal, and standard subspace methods such as MUSIC and ESPRIT are only defined for narrow-band signal model, which then fail to directly operate on broadband signals [19]. One often used concept is band-pass filtering of broadband signals into subbands, where narrow-band estimators can be applied to each subband [20]. In the narrow-band case, a delay in the

* Correspondence: jxz@es.aau.dk

¹Department of Electronic Systems (ES-MISP), Aalborg University, Aalborg, Denmark

Full list of author information is available at the end of the article

signal is equivalent to a phase shifts according to the frequencies of complex exponentials. An alternative method is, however, as follows: since harmonic signals consist of sinusoidal components, we can model each source as multiple narrow-band signal with distinct frequencies arriving at the same DOA.

In this article, we propose a parametric method for solving the problem of joint fundamental frequency and DOA estimation based on subspace techniques where the quantities of interest are jointly estimated using a MUSIC-like approach. We term the proposed estimator Multi-channel multi-pitch Harmonic MUSIC (MC-HMUSIC). The spatio-temporal data model used in MC-HMUSIC is based on the JAFE data model [21,22]. Originally, the JAFE data model was used for estimating joint unconstrained frequencies and DOAs estimates of complex exponential using ESPRIT, which is referred as joint angle-frequency estimation (JAFE) algorithm. Other-related work with joint frequency-DOA methods includes [23-25]. In this article, we have parametrized the harmonic structure of periodic signals in the signal model to model the fundamental frequency and the DOA of individual sources. An estimator is constructed for jointly estimating the parameters of interest. Incorporating the DOA parameter in finding the fundamental frequency may give better robustness against a signal with overlapping harmonics. Similarly, it can be expected that the DOA can be found more accurately when the nature of the signal of interest is taken into account.

The remainder of this article is comprised four sections: Section 2, in which we will introduce some notation, the spatio-temporal signal model, for which we also derive the associated Cramér-Rao lower bound, along with the JAFE data mode; Section 3, where we then present the proposed method; Section 4, in which we present the experimental results obtained using the proposed method; and, finally, Section 5, where we conclude on our work.

2. Fundamentals

2.1. Spatio-temporal signal model

Next, the signal model employed throughout the article will be presented. Without multi-path propagation of sources, it is given as follows: the signal x_i received by microphone element i arranged in a uniform linear array (ULA) configuration, $i = 1, \dots, M$, is given by

$$x_i(n) = \sum_{k=1}^K \sum_{l=1}^{L_k} \beta_{l,k} e^{j(\omega_k n + \phi_k l(i-1))} + e_i(n), \quad (1)$$

$$\beta_{l,k} = A_{l,k} e^{j\gamma_{l,k}},$$

for sample index $n = 0, \dots, N - 1$, where subscript k denotes the k th source and l the l th harmonic.

Moreover, $A_{l,k}$ is the real-valued positive amplitude of the complex exponential, L_k is the number of harmonics, K is number of sources, $\gamma_{l,k}$ is the phase of the individual harmonics, ϕ_k is the phase shift caused by the DOA, and $e_i(n)$ is complex symmetric white Gaussian noise. The phase shift between array elements is given as $\phi_k = \omega_k f_s \frac{d}{c} \sin(\theta_k)$, where d is the spacing between the elements measured in wavelengths, c is the speed of propagation in unit [m/s], θ_k is the DOA defined for $\theta_k \in [-90^\circ, 90^\circ]$, f_s is the signal sampling frequency. The problem of interest is to estimate ω_k and θ_k . We in the following assume that the number of sources K is known and the number of harmonics L_k of individual sources is known or found in some other, possibly joint, way. We note that a number of ways of doing this has been proposed in the past [26-28,2].

2.2. Cramér-Rao lower bound

We will now proceed to derive the exact Cramér-Rao lower bound (CRLB) for the problem of estimating the parameters of interest. First, we define the $M \times 1$ deterministic signal model vector $\mathbf{s}(n, \boldsymbol{\mu})$ with column element as

$$s_i(n, \boldsymbol{\mu}) = \sum_{k=1}^K \sum_{l=1}^{L_k} \beta_{l,k} e^{j(\omega_k n + \phi_k l(i-1))}, \quad \beta_{l,k} = A_{l,k} e^{j\gamma_{l,k}}, \quad (2)$$

where $\mathbf{s}(n, \boldsymbol{\mu}) = [s_1(n, \boldsymbol{\mu}) \dots s_M(n, \boldsymbol{\mu})]^T$. Furthermore, the parameter vector $\boldsymbol{\mu}$ is given by

$$\boldsymbol{\mu} = [\omega_1 \quad \dots \quad \omega_K \quad \theta_1 \quad \dots \quad \theta_K \quad A_{1,1} \quad \gamma_{1,1} \quad \dots \quad A_{L_k,K} \quad \gamma_{L_k,K}]. \quad (3)$$

Recall that the observed signal vector with additive white noise is given by

$$\mathbf{x}(n) = \mathbf{s}(n, \boldsymbol{\mu}) + \mathbf{e}(n) = \begin{bmatrix} s_1(n, \boldsymbol{\mu}) \\ \vdots \\ s_M(n, \boldsymbol{\mu}) \end{bmatrix} + \mathbf{e}(n), \quad (4)$$

with $\mathbf{e}(n)$ being the noise column vector. The CRLB is defined as the variance of an unbiased estimate of the p th element of $\boldsymbol{\mu}$, which is lower bounded as

$$\text{var}(\boldsymbol{\mu}_p) \geq [\mathbf{C}^{-1}]_{pp}, \quad (5)$$

where \mathbf{C} is the so-called Fisher information matrix given by

$$\mathbf{C} = \frac{2}{\sigma^2} \text{Re} \left(\sum_{n=0}^{N-1} \frac{\partial \mathbf{s}(n, \boldsymbol{\mu})^H}{\partial \boldsymbol{\mu}} \frac{\partial \mathbf{s}(n, \boldsymbol{\mu})}{\partial \boldsymbol{\mu}^T} \right). \quad (6)$$

principle used in MUSIC [19,32,27,26,4]. According to the signal noise subspace orthogonality principle, the following relationship holds:

$$\mathbf{J}_1 \bar{\mathbf{A}}_t \mathbf{G} = \mathbf{0}, \quad (29)$$

where we, for notational simplicity, have introduced $\mathbf{J}_1 \bar{\mathbf{A}}_t = \mathbf{A}_{ts}$. The matrix \mathbf{A}_{ts} is comprised Vandermonde matrices for sources $k = 1, \dots, K$. The matrix for each individual source is given by

$$\mathbf{A}_{ts,k} = \begin{bmatrix} 1 & \dots & 1 \\ e^{j\phi_k} & & e^{j\phi_k L_k} \\ \vdots & & \vdots \\ e^{j\phi_k S} & \dots & e^{j\phi_k L_k S} \\ \vdots & & \vdots \\ e^{j\omega_k(t-1)} & \dots & e^{j\omega_k L_k(t-1)} \\ e^{j\phi_k} e^{j\omega_k(t-1)} & & e^{j\phi_k L_k} e^{j\omega_k L_k(t-1)} \\ \vdots & & \vdots \\ e^{j\phi_k S} e^{j\omega_k(t-1)} & \dots & e^{j\phi_k L_k S} e^{j\omega_k L_k(t-1)} \end{bmatrix}. \quad (30)$$

The cost function of the proposed joint DOA and multi-pitch estimator is then

$$J(\omega_k, \theta_k) = \|\mathbf{A}_{ts,k}^H \mathbf{G}\|_F^2, \quad (31)$$

where $\|\cdot\|_F$ is the Frobenius Norm. Note that this measure is closely related to the angles between the subspaces as explained in [33] and can hence be used as a measure of the extent to which (29) holds for a candidate fundamental frequency and DOA. The pair of fundamental frequency and DOA can, therefore, be found as the combination that is the closest to being orthogonal to \mathbf{G} , i.e.,

$$\{\omega_k, \theta_k\}_{k=1}^K = \arg \min_{\{\theta_k\}_1^K, \{\omega_k\}_1^K} \|\mathbf{A}_{ts,k}^H \mathbf{G}\|_F^2. \quad (32)$$

The multi-channel estimators will have a cost function which is more well-behaved compared to those of single channel multi-pitch estimators (see, e.g., [26,32,28] for some examples of such).

3.2. Refined estimates

For many applications, only a coarse estimate of involved fundamental frequencies and DOAs are needed, in which case the cost function in (32) is evaluated on pre-defined search region with some specified granularity. If, however, very accurate estimates are desired, a refined estimate can be found as described next. For a rough estimate of the parameter of interests, refined estimates are obtained by minimizing the cost function in (32) using a cyclic minimization approach.

The gradient of the cost function (32) for fundamental frequency and DOA are given as

$$\frac{\partial}{\partial \omega_k} J(\omega_k, \theta_k) = 2 \operatorname{Re} \left(\operatorname{Tr} \left\{ \mathbf{A}_{ts,k}^H \mathbf{G} \mathbf{G}^H \frac{\partial}{\partial \omega_k} \mathbf{A}_{ts,k} \right\} \right), \quad (33)$$

$$\frac{\partial}{\partial \theta_k} J(\omega_k, \theta_k) = 2 \operatorname{Re} \left(\operatorname{Tr} \left\{ \mathbf{A}_{ts,k}^H \mathbf{G} \mathbf{G}^H \frac{\partial}{\partial \theta_k} \mathbf{A}_{ts,k} \right\} \right), \quad (34)$$

with $\operatorname{Re}(\cdot)$ denoting the real value. The gradient can be used for finding refined estimate using standard methods.

Here, we iteratively find a refined estimate using a cyclic approach. During an iteration, ω_k is first estimated with

$$\hat{\omega}_k^{i+1} = \hat{\omega}_k^i - \delta \frac{\partial}{\partial \omega_k} J(\hat{\omega}_k^i, \hat{\theta}_k^i), \quad (35)$$

where i is the iteration index and δ is a small positive constant that is found using line search. The estimated $\hat{\omega}_k^{i+1}$ is then used to initialize the minimization function for DOA, which is then found as

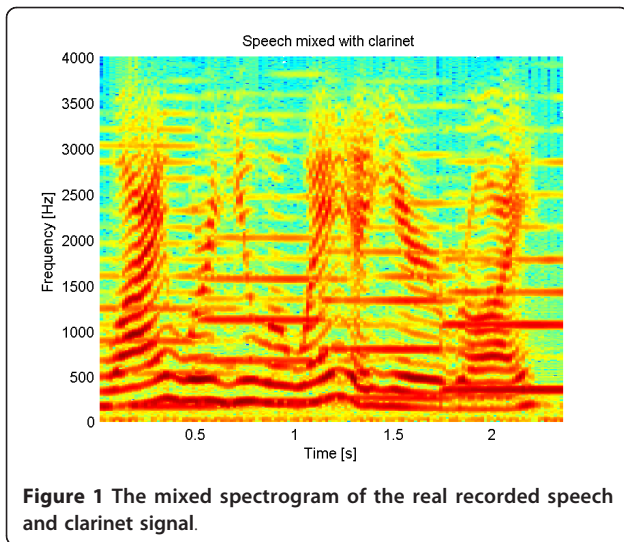
$$\hat{\theta}_k^{i+1} = \hat{\theta}_k^i - \delta \frac{\partial}{\partial \theta_k} J(\hat{\omega}_k^{i+1}, \hat{\theta}_k^i). \quad (36)$$

The method is initialized for $i = 0$ using the coarse estimates obtained from (32).

4. Experimental results

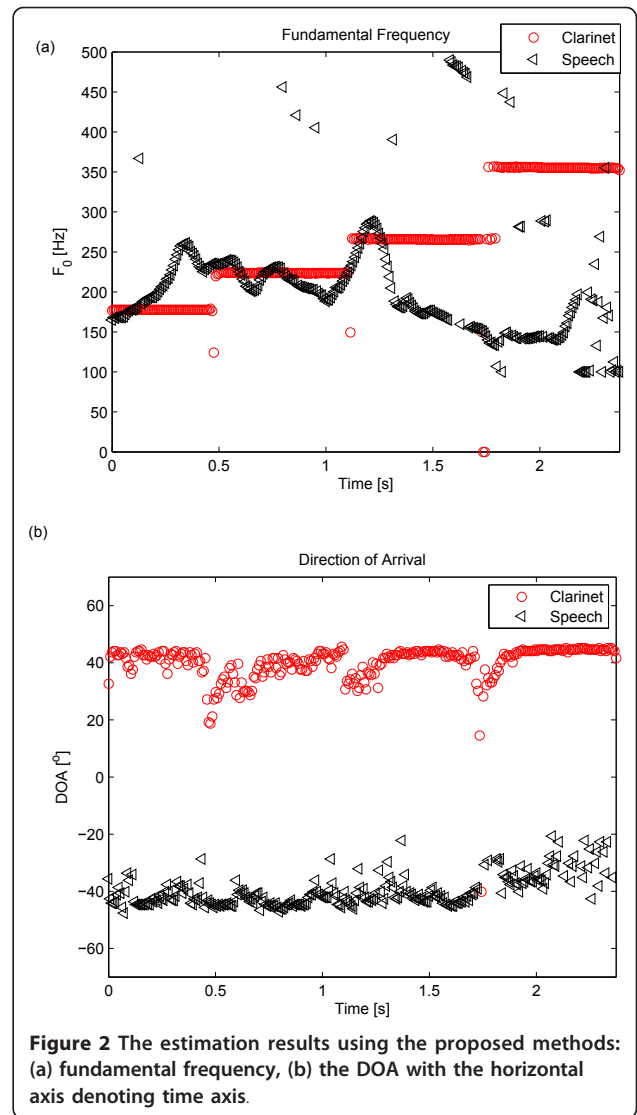
4.1. Signal examples

We start the experimental part of this article by illustrating the application of the proposed method to analyzing a mixed signal consisting of speech and clarinet signals, sampled at $f_s = 8000$ Hz. The single-channel signals are converted into a multi-channel signal by introducing different delays according to two pre-determined DOA to simulate a microphone array with $M = 8$ channels. The simulated DOAs of the speech and the clarinet signals are, respectively, $\theta_1 = -45^\circ$ and $\theta_2 = 45^\circ$. The spectrogram of the mixed signal of the first channel is illustrated in Figure 1. To avoid spatial ambiguities, the distance between two sensor is half the wavelength of the highest frequency in the observed signal, here $d = 0.0425$ m. The mixed signal is segmented into 50% overlapped signal segments with $N = 128$. The user parameter selected in this experiment is $t = \lfloor \frac{2N}{3} \rfloor$ and $s = \lfloor \frac{M}{2} \rfloor$. The cost function is evaluated with a Vandermonde matrix with $L = 5$ complex exponentials, and the noise subspace is formed from an overestimated signal subspace with assumption of signal subspace containing $N/2 = 64$ complex exponentials. The signal subspace



overestimation technique is usually used when the true order of the signal subspace is unknown, the signal subspace is assumed to be larger than the true one which can minimize the signal subspace components in the noise subspace. An added benefit of posing the problem as a joint estimation problem is that the multi-pitch estimation problem can be seen as several single-pitch problems for a distinct set of DOAs, one per source. Therefore, it is less important to select an exact signal model order than single-channel multi-pitch estimators would need [28]. The cost function is evaluated for frequencies from 100 to 500 with granularity of 0.52 Hz. The evaluated results are illustrated in Figure 2 where the upper panel contains the fundamental frequency estimates and lower panel the DOA estimates. It can be seen that the proposed algorithm can track the fundamental frequency and the DOA of the speech signal well, with only a few observed errors on regions with low signal energy. The clarinet signal's DOA and fundamental frequencies have also been estimated well for all segments.

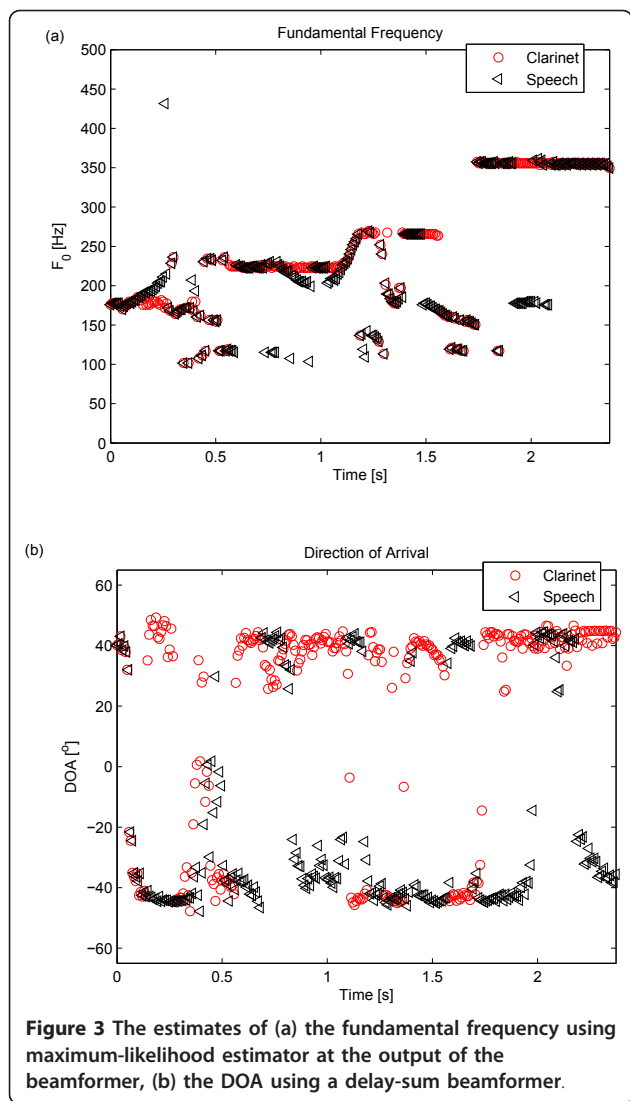
For the purpose of further comparison, the same signal will be analyzed using a standard time delay-and-sum beamformer [34] for DOA estimates and a single-channel maximum-likelihood based pitch estimator applied on the beamformed output signals [2]. The results are shown in Figure 3. The figure clearly shows that the delay-sum beamformer cannot satisfactorily resolve the DOAs with $M = 8$ array elements which will further affect the performance of the single-channel pitch estimator, as shown in the upper panel. In this example, the proposed algorithm shown in Figure 2 is superior compared to reference method shown in Figure 3. The low resolution performance of the reference method will make the statistical



evaluation of this method uninteresting, and we, therefore, will not be using it any further in the experiments to follow.

4.2. Statistical evaluation

Next, we use Monte Carlo simulations evaluated on synthetic signals embedded in noise in assessing the statistical properties of the proposed method and compare it with the exact CRLB. As a reference method for pitch and DOA estimation, we use the JAFE algorithm proposed in [22] for jointly estimating unconstrained frequencies and DOAs. Next, the unconstrained frequencies are grouped according to their corresponding DOAs where closely related directions are grouped together. A fundamental frequency is formed from these grouped frequencies in a weighted way as proposed in [35]. We refer this as the WLS estimator. In order to



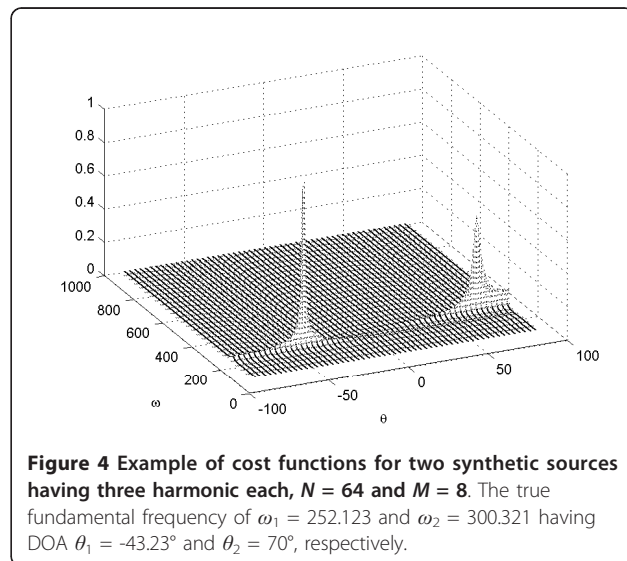
remove the errors due to the erroneous estimate of amplitudes, we assume WLS having the exact signal amplitude given. The WLS estimator is a computationally efficient pitch estimation method with good statistical properties. The reference DOA estimate is easily obtained in a similar way from the mean value of these grouped DOAs according to [22].

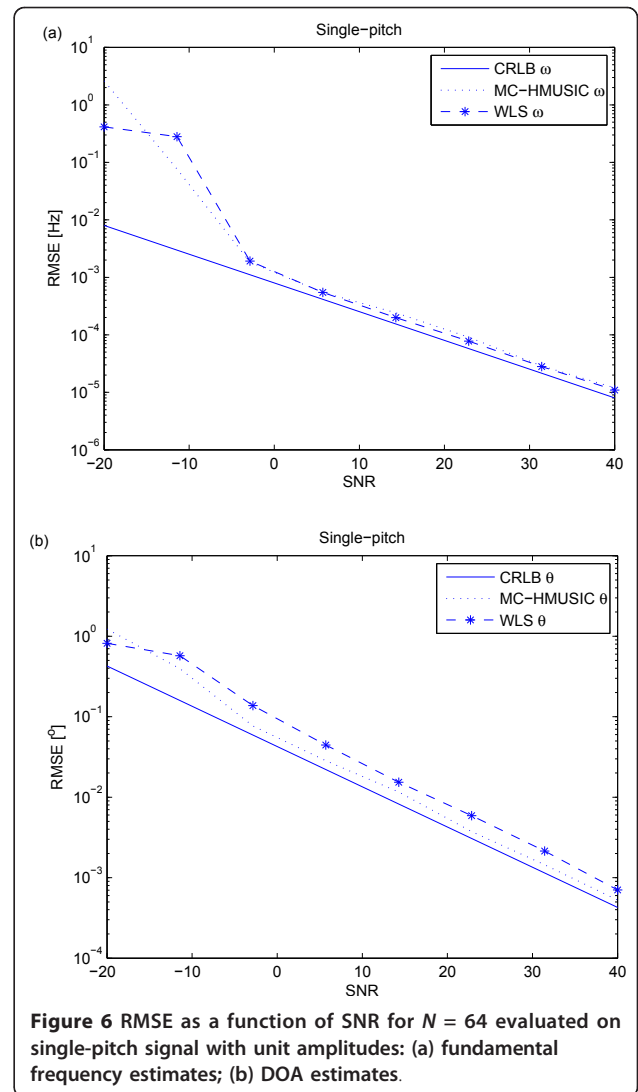
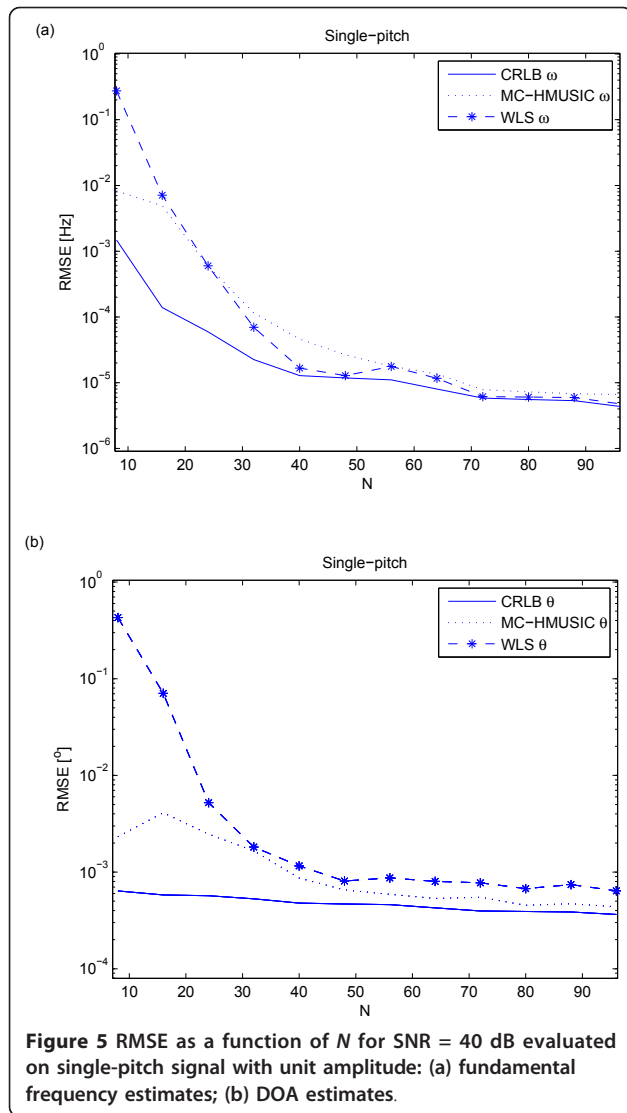
Here, we consider a $M = 8$ element ULA with sensor distance $d = 0.0425$ with a sampling frequency of $f_s = 8000$. The estimators are evaluated for two signal setups, first with two sources having $\omega_1 = 252.123$ and $\omega_2 = 300.321$ with $L_{1,2} = 3$, and second with one harmonic source of $\omega_1 = 252.123$ and $L_1 = 3$. All amplitudes on individual harmonics are set to unity $A_{k,l} = 1$ for tractability. Both sources are assumed to be far-field sources impinging on the array with DOAs at $\theta_1 = -43.23^\circ$ and $\theta_2 = 70^\circ$, respectively, and for one source having a DOA

of $\theta_1 = -43.23^\circ$. All simulation results are based on 100 Monte Carlo runs. The performance is measured using the root mean squared estimation error (RMSE) as defined in [28,32,26,27]. The user parameter for JAFE data model is selected to the optimal values as proposed in [22] with temporal and spatial smoothness parameters, $t = \lfloor \frac{2N}{3} \rfloor$ and $s = \lfloor \frac{M}{2} \rfloor$, respectively. We note that in practical applications, the computational complexity has to also be considered in selecting the appropriate parameters t and s . An example of the 2-dimensional (2D) cost function of our proposed method evaluated on two mixed signal is illustrated in Figure 4, where a coarser estimate of the DOA and fundamental estimates can be identified from the two peaks in the 2D cost function.

In the first simulation, we evaluate the proposed method's statistical properties in a single source scenario for varying sample lengths and SNRs. The RMSEs on signal with varying N are shown in Figure 5, and with varying SNR in Figure 6. It can be seen from these figures that both estimators perform well for all SNR above 0 dB with WLS being slightly better for fundamental frequency estimation while the proposed estimator is better in DOA estimation. Both methods are also able to follow CRLB closely for around sample length $N > 60$. The better DOA estimation capabilities of the proposed method can be explained by the joint estimation of the fundamental frequency and DOA, which leads to increased robustness under adverse conditions. Both estimators can be considered as consistent in the single-pitch scenario.

Next, we evaluate our method for the multi-pitch scenario. The so-obtained RMSEs for varying N and

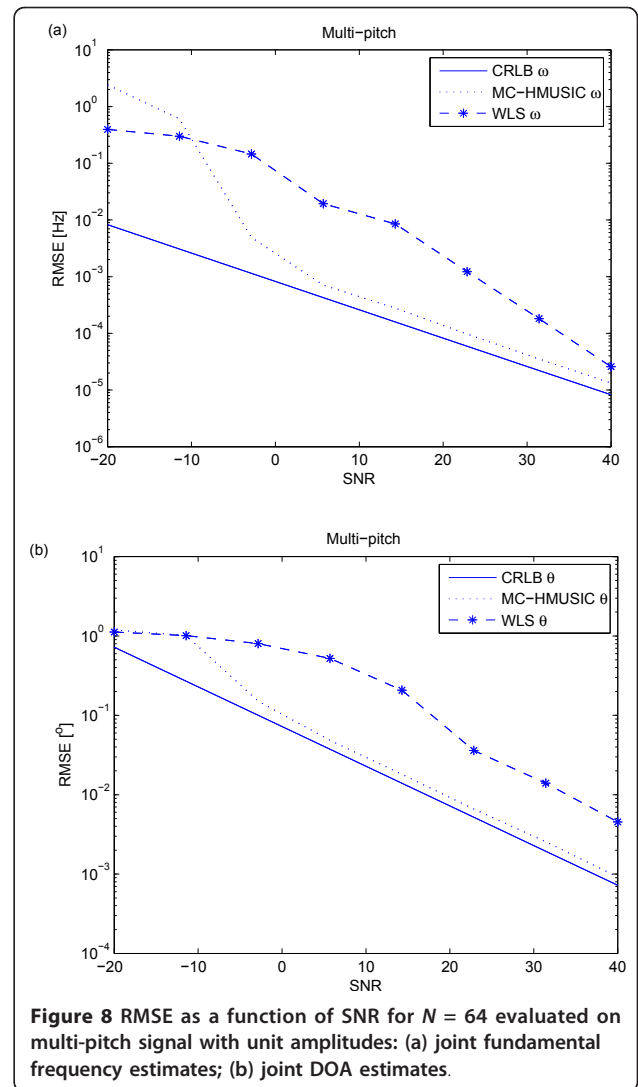
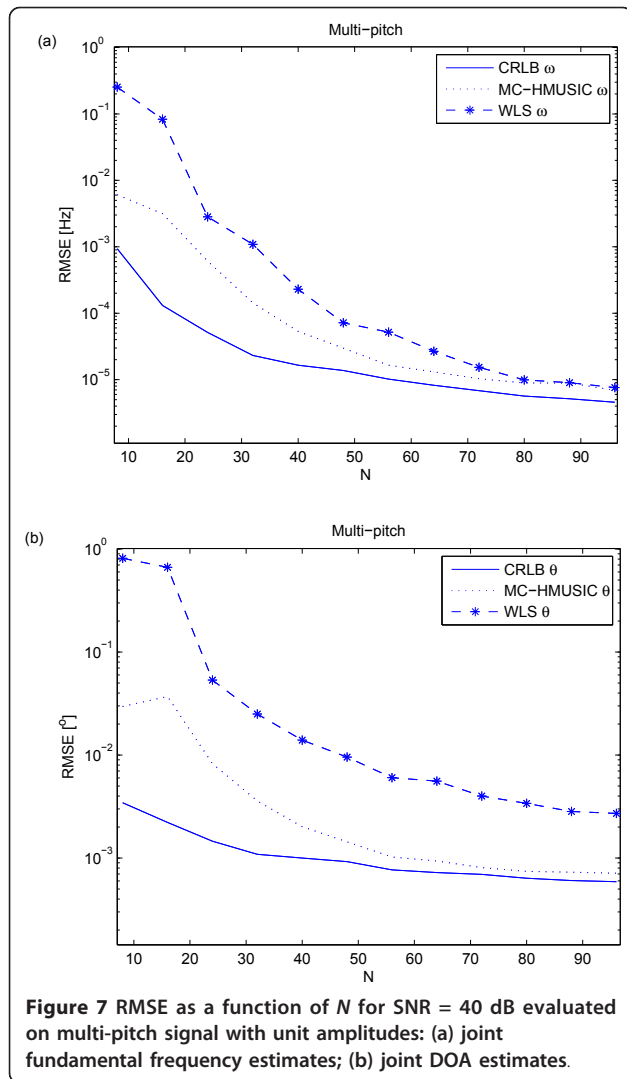




SNR are depicted in Figures 7 and 8. In Figure 7, it clearly shows that the proposed method is better than the WLS estimator for short sample lengths. The WLS estimator is not following CRLB until $N > 80$ samples while the proposed estimator is for $N > 64$. The remaining gap between CRLB and both evaluated estimators for $N > 80$ are due to the mutual interference between the harmonic sources. The slowly converging performance of WLS is mainly due to the bad estimate of the unconstrained frequency estimate using the JAFE method. With our selected simulation setup, the JAFE estimator is not giving consistent estimates for all harmonic components, which, in turn, results in poor performance in the WLS estimates. In general, the WLS estimator is sensitive to spurious estimate of the unconstrained frequencies. Moreover, the proposed

estimator, which is jointly estimating both the DOA and the fundamental frequency, yields better estimates for smaller sample length N . The results in terms of RMSEs for varying SNRs are shown in Figure 8. This figure shows that the proposed estimator is again more robust than the WLS estimator for both DOA and fundamental frequency estimation.

In next two experiments, we will study the performance as a function of the difference in fundamental frequencies and DOAs for multiple sources. We start with studying the RMSE as a function of the difference between the fundamental frequencies of two harmonic sources, i.e., $\Delta\omega = |\omega_1 - \omega_2|$, with $\theta_1 = -43.321^\circ$ and $\theta_2 = 70^\circ$. Here, we use an SNR set to 40 dB, and a sample length $N = 64$ with $M = 8$ array elements. The obtained RMSEs are shown in Figure 9. The figure clearly shows



that both methods can successfully estimate the fundamental frequencies and DOAs. Once again the proposed estimator gives more robust estimates, close to the CRLB. Additionally, it should be noted that both methods are correctly estimating the DOA even when the both fundamental frequencies are identical $\omega_1 = \omega_2$, something that would not be possible with only a single channel. MC-HMUSIC has the ability to estimate the fundamental frequencies when both harmonics are identical provided that the DOAs are distinct and vice versa. Estimation of the parameters of signals with overlapping harmonics is a crucial limitation in multi-pitch estimation using only single-channel recordings. In the final experiment, the RMSE as a function of the difference between the DOAs of two harmonic sources $\Delta\theta = |\theta_1 - \theta_2|$ is analyzed for an SNR set to 40 dB and a sample length of $N = 64$ with $M = 8$ array elements. The fundamental frequencies are $\omega_1 = 252.123$ and $\omega_2 = 300.321$,

respectively. The observations and conclusions are basically the same as before, with the proposed method outperforming the reference method so far.

5. Conclusion

In this article, we have generalized the single-channel multi-pitch problem into a multi-channel multi-pitch estimation problem. To solve this new problem, we propose an estimator for joint estimation of fundamental frequencies and DOAs of multiple sources. The proposed estimator is based on subspace analysis using a time-space data model. The method is shown to have potential in applications to real signals with simulated anechoic array recording, and a statistical evaluation demonstrates its robustness in DOA and fundamental frequency estimation as compared to a state-of-the-art reference method. Furthermore, the proposed method is shown to have good statistical performance under

30. T Shu, XZ Liu, Robust and computationally efficient signal-dependent method for joint DOA and frequency estimation. *EURASIP J Adv Signal Process*. **2008** (2008). Article ID 10.1155/2008/134853
31. H Krim, M Viberg, Two decades of array processing research-the parametric approach. *IEEE SP Mag* (1996)
32. MG Christensen, A Jakobsson, SH Jensen, Multi-pitch estimation using Harmonic MUSIC. in *Record of the Asilomar Conference on Signals, Systems, and Computers* 521–525 (2006)
33. MG Christensen, A Jakobsson, SH Jensen, Sinusoidal order estimation using angles between subspaces. *EURASIP J Adv Signal Process* 1–11 (2009). Article ID 948756
34. BDV Veen, KM Buckley, Beamforming: a versatile approach to spatial filtering. *IEEE ASSP Mag*. **5**, 4–24 (1988)
35. H Li, P Stoica, J Li, Computationally efficient parameter estimation for harmonic sinusoidal signals. *Elsevier Signal Process* 1937–1944 (2000)

doi:10.1186/1687-6180-2012-1

Cite this article as: Zhang et al.: Joint DOA and multi-pitch estimation based on subspace techniques. *EURASIP Journal on Advances in Signal Processing* 2012 **2012**:1.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com
