Broadband DOA Estimation Using Sensor Arrays on Complex-Shaped Rigid Bodies

Dumidu S. Talagala, *Student Member, IEEE*, Wen Zhang, *Member, IEEE*, and Thushara D. Abhayapala, *Senior Member, IEEE*

***Abstract*—Sensor arrays mounted on complex-shaped rigid bodies are a common feature in many practical broadband direction of arrival (DOA) estimation applications. The scattering and reflections caused by these rigid bodies introduce complexity and diversity in the frequency domain of the channel transfer function, which presents several challenges to existing broadband DOA estimators. This paper presents a novel high resolution broadband DOA estimation technique based on signal subspace decomposition. We describe how broadband signals can be decomposed into narrow subband components, and combined such that the frequency domain diversity is retained. The DOA estimation performance is compared with existing techniques using a uniform circular array and a sensor array on a hypothetical rigid body. An improvement in closely spaced source resolution of up to 6 dB is observed for the sensor array on the hypothetical rigid body, in comparison to the uniform circular array. The results suggest that frequency domain diversity, introduced by complex-shaped rigid bodies, can provide higher resolution and clearer separation of closely spaced broadband sound sources.**

***Index Terms*—Arbitrary array, array signal processing, direction of arrival (DOA), head related transfer function (HRTF), MUSIC, source localization.**

I. Introduction

DIRECTION of arrival (DOA) estimation of multiple sources using sensor arrays has been an active problem in signal processing for decades. Array geometry plays a major role in determining the source separation capabilities in sonar, radar, robotics and communications applications where location accuracy is important. Traditionally, various linear, circular and spherical sensor arrays used for DOA estimation assume free field propagation conditions between the source and sensors. However, this may not be true in certain applications, due to the scattering and reflections caused by the rigid body used as a sensor mount. These effects are exacerbated as the shape and structure of the mounting object becomes more complex; but they may also bring a source of diversity in the frequency domain of the channel transfer function. This paper introduces a DOA estimation method based on signal subspace techniques that exploits the additional diversity afforded by sensor arrays mounted on complex-shaped rigid bodies (e.g., a sensor array mounted on a submarine or a robotic platform).

The human auditory system is an example of a two-sensor array on a complex-shaped rigid body. The head and torso act as scattering objects, while the structures within the pinna produce reflections that act as multipath signals [1]. This results in direction specific changes to the phase and amplitude of a signal, collectively called the head related transfer function (HRTF). Perceptual studies in the past have found that the localization cues present in the HRTF provide the necessary diversity for 3-D DOA estimation [2]–[7]. These results suggest that frequency domain diversity can be used to reduce the resource requirements of the 3-D DOA estimation problem, while improving resolution between adjacent source locations. However, works in the area are limited, and especially focus on empirical modeling for binaural DOA estimation [8]–[11].

Broadband DOA estimation techniques can be broadly categorized into those based on cross-correlation analysis, high resolution signal subspace techniques or a combination of the two. Cross-correlation based techniques are typically described using simple array geometries in free field, yet are equally applicable to sensor arrays mounted on rigid objects. The time difference of arrival (TDOA) at the sensors is a function of source direction. The TDOA at the peaks of the cross-correlation function can therefore be used to identify the source directions of arrival. The Steered Response Power (SRP) [12] method is one such implementation, known as a multi-sensor variant of the Generalized Cross Correlation (GCC) method [13]. Although more successful multi-sensor techniques have been introduced [14], they are still fundamentally TDOA estimators that map time delay to a source location. Hence, any diversity present in the frequency domain remains unseen and unutilized by these DOA estimators.

Signal subspace techniques such as MUltiple Signal Classification (MUSIC) [15] and Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT) [16] are inherently narrowband methods for DOA estimation. The Coherent Signal Subspace (CSS) [17] was proposed in order to transform the broadband DOA estimation problem into a narrowband problem. This was achieved by focussing each subband of the broadband signal into a known narrowband frequency. The concept of the CSS has since been developed into a number of broadband DOA estimation techniques based on beamforming [18]–[20], modal decomposition [21], [22] and unitary focussing matrices [17], [23], [24]. Although uniform linear arrays (ULAs) are typically required for the DOA estimation algorithms based on ESPRIT, narrowband transformations that map arbitrary array geometries to ULAs [25]–[27] have been demonstrated. Hence, the broadband DOA estimators based

Manuscript received August 22, 2012; revised December 20, 2012, February 26, 2013; accepted March 11, 2013. Date of publication March 27, 2013; date of current version April 25, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Søren Holdt Jensen.

The authors are with the Applied Signal Processing Group, Research School of Engineering, College of Engineering and Computer Science, Australian National University, Canberra ACT 0200, Australia (e-mail: dumidu.talagala@anu.edu.au; wen.zhang@anu.edu.au; thushara.abhayapala@anu.edu.au).

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/TASL.2013.2255282

on MUSIC and ESPRIT can theoretically be adapted to any array geometry, including sensor arrays on complex-shaped rigid bodies. However, the complicated behavior of the channel transfer functions in the frequency domain make them far more susceptible to imperfections in the frequency focussing process, which can result in a reduction of the DOA estimation accuracy. Consider wideband MUSIC [17], a MUSIC broadband DOA estimator that implements the CSS concept. The broadband signals are segmented into multiple frequencies, multiplied by frequency focussing transformations, and the resulting focussed correlation matrices are then aggregated. The complicated frequency response of the channel transfer function (due to the sensors being mounted on a rigid body) results in imperfections in the numerical calculation of the focussing transformations; hence the signal spaces of the focussed correlation matrices are not fully aligned with each other. Aggregating these matrices results in a gradual increase of the rank of the coherent signal subspace as more frequency segments are included. This misalignment could eventually lead to the disappearance of the noise subspace and introduces additional complexity to the broadband DOA estimation process.

In this paper, we use biological inspiration to propose a broadband DOA estimation technique for sensor arrays on complex-shaped rigid bodies. In Section II, we present the background theory on the subband representation of broadband signals, and show that each subband carries both directional and source information. This includes a frequency dependent carrier term, which must be removed if multiple subband signals are to be combined. Section III presents the signal model used for DOA estimation. We introduce the subband signal extraction and focussing processes, and describe how the subband signals can be combined, such that the frequency domain diversity is retained. Next, the channel transformation matrix is defined, and used to derive the requirements for the existence of a noise subspace. Section IV describes the broadband DOA estimators for several DOA estimation scenarios; an ideal scenario where sources are uncorrelated between subbands, the real-world equivalent of the ideal scenario and the DOA estimation of known sources. The performance of our algorithm is compared with a correlation based DOA estimation technique, SRP-PHAT, and a signal subspace technique, wideband MUSIC. Section V briefly describes these algorithms, the performance measure for comparing the different algorithms and the simulation setup. Simulation results are discussed in Section VI, followed by concluding remarks.

II. System Model and Signal Representation

Consider a complex-shaped rigid body with M sensors located at distinct spatial positions as illustrated in Fig. 1. Let $h_m(\Theta_q, t)$ be the acoustic impulse response from the qth ($q = 1, \ldots, Q$) source in the direction $\Theta_q \equiv (\theta_q, \phi_q)$ to the mth ($m = 1, \ldots, M$) sensor. Note that sources are assumed to be in the far field (r_q the distance to the source satisfies the conditions $2\pi f r_q / c \gg 1$ and $r_q \gg r$, for an operating frequency f, speed of sound c and an array radius r [28]), where they exhibit plane wave behavior in the local region of the sensor array. The received signal at the mth sensor is given by

$$y_m(t) = \sum_{q=1}^{Q} h_m(\Theta_q, t) * s_q(t) + \tilde{n}_m(t), \tag{1}$$

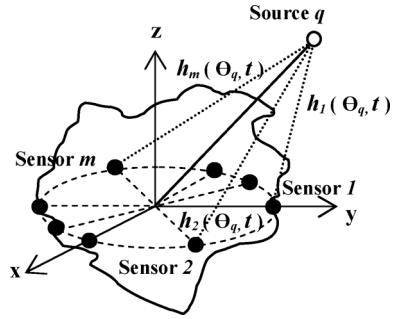


Fig. 1. Source-sensor channel impulse response of a sensor array on a complex-shaped rigid body.

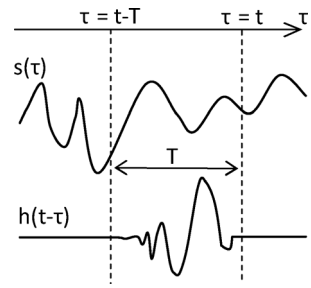


Fig. 2. Convolution of a source signal $s(\tau)$ and the source-sensor channel impulse response $h(\tau)$.

where $s_q(t)$ is the qth source signal, $\tilde{n}_m(t)$ is the diffuse ambient noise[1] at the mth sensor and '$*$' denotes the convolution operation. A Fourier series representation can now be used to model the received signal in (1) as a collection of subband signals.

A. Subband Expansion of Audio Signals

Consider the operation of audio compression [29], [30] and speech coding techniques [31], where a broadband signal is passed through a band-pass filter bank, down-converted, quantized and coded. An underlying assumption is that the source information can be represented by a small number of samples. This concept can be used to characterize the broadband source as a collection of subbands signals.

Suppose there exists a broadband signal $s(\tau)$ as shown in Fig. 2. A symmetric Fourier series representation can be used to decompose $s(\tau)$ into a collection of equally spaced, non-overlapping subband signals, within the interval $t - T \leq \tau \leq t$ (T is a window length determined by the length of the channel impulse response and desired frequency resolution) [32]. Thus,

$$s(\tau) = \frac{1}{\sqrt{T}} \sum_{k=-\infty}^{\infty} S(k, t) e^{jk\omega_0 \tau}, \tag{2}$$

where $k\omega_0$ represents the mid-band frequency of the kth subband and $\omega_0 = 2\pi/T$ is the frequency spacing between subbands.

$$S(k, t) = \frac{1}{\sqrt{T}} \int_{t-T}^{t} s(\tau) e^{-jk\omega_0 \tau} \mathrm{d}\tau$$

are the Fourier series coefficients that describe the time varying behavior of $s(\tau)$ in the kth subband, and are analogous to the

[1]In most practical systems, measured noise consists of noise originating at distinct spatial locations and ambient noise that lacks any directional attributes. Hence, the Q identifiable sources in the sound field may consist of both legitimate targets and noise sources. $\tilde{n}_m(t)$ represents the effects of the diffuse noise field that forms the ambient noise floor of the system.

short-time Fourier transform coefficients of $s(\tau)$. Hence, the kth subband signal $s(k\omega_0, t)$ is given by

$$s(k\omega_0, t) = \frac{1}{\sqrt{T}} S(k,t) e^{jk\omega_0 t}. \quad (3)$$

This implies that the information contained in each subband is completely described by a slowly time varying Fourier coefficient $S(k,t)$, and the corresponding carrier term $e^{jk\omega_0 t}$.

B. Direction Encoding of Source Signals

The channel impulse response $h_m(\Theta_q, t)$ is naturally a time limited function. Hence, $h_m(\Theta_q, t)$ can be described using the Fourier series representation

$$h_m(\Theta_q, t) = \frac{1}{\sqrt{T}} \sum_{k'=-\infty}^{\infty} H_{mq}(k') e^{jk'\omega_0 t} \quad 0 \le t \le T, \quad (4)$$

where

$$H_{mq}(k') = \frac{1}{\sqrt{T}} \int_0^T h_m(\Theta_q, \tau) e^{-jk'\omega_0 \tau} \mathrm{d}\tau$$

are time invariant Fourier coefficients that characterize the acoustic propagation channel between the (m,q)th source-sensor pair. This representation can now be used to simplify (1).

By substituting (2) and (4) into (1), the convolution of the (m,q)th source-sensor pair is given by

$$h_m(\Theta_q, t) * s_q(t) = \frac{1}{T} \int_{t-T}^{t} \left(\sum_{k=-\infty}^{\infty} S_q(k,t) e^{jk\omega_0 \tau} \right) \times \left(\sum_{k'=-\infty}^{\infty} H_{mq}(-k') e^{jk'\omega_0 t} e^{jk'\omega_0 \tau} \right) \mathrm{d}\tau. \quad (5)$$

Applying the following identity,

$$\int_{t-T}^{t} e^{jk'\omega_0 \tau} e^{jk\omega_0 \tau} \mathrm{d}\tau = \begin{cases} T & \text{if } k = -k' \\ 0 & \text{otherwise,} \end{cases}$$

then leads to

$$h_m(\Theta_q, t) * s_q(t) = \sum_{k=-\infty}^{\infty} H_{mq}(k) S_q(k,t) e^{-jk\omega_0 t}. \quad (6)$$

Thus, the received signal at the mth sensor can be expressed as the summation of subband signals,

$$y_m(t) = \sum_{q=1}^{Q} \sum_{k=-\infty}^{\infty} H_{mq}(k) S_q(k,t) e^{-jk\omega_0 t} + \tilde{n}_m(t). \quad (7)$$

Each subband signal now consists of two components; a narrowband source-direction information term $H_{mq}(k)S_q(k,t)$ and a carrier term $e^{-jk\omega_0 t}$. Eliminating this carrier term will lead to a set of focussed subband signals that can be used for DOA estimation.

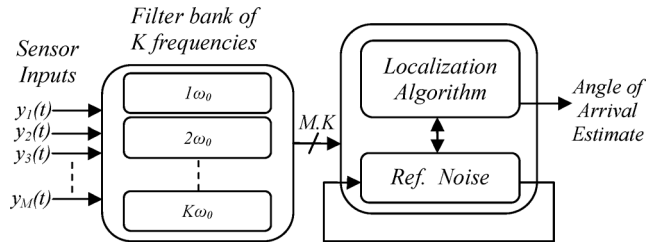


Fig. 3. The system model above consists of signal preprocessing and DOA estimation stages. Conceptually, the preprocessor separates each sensor signal into K subbands by passing it through a series of band-pass filters, before down-conversion and down-sampling. The localization algorithm estimates the source directions of arrival using the MK subband signals.

III. Signal Subspace Decomposition

In the previous section we observed that the subband signals of a direction encoded broadband source consist of a source-direction information term and a carrier term. The following section describes the process of decomposing the M broadband sensor signals into MK subbands, as shown in the system model in Fig. 3. We then describe a framework for combining the focussed subband signals for signal subspace decomposition.

A. Subband Signal Extraction and Focussing

In (7), we see that each subband signal is an amplitude modulation of the subband carrier frequency. This carrier term lacks any directional information, and is the source of spatial aliasing [33] on the DOA estimates at high frequencies. Demodulating each subband signal can help eliminate this problem, as different subband signals are simultaneously focussed into a set of low frequency signals of similar bandwidth. In practice, the functionality of the filter bank in Fig. 3 can be implemented as a series of mixing and low pass filtering operations of (7), as shown below.

When mixed with the complex exponential $e^{jk_0\omega_0 t}$, (7) can be expressed as

$$y_m(t) e^{jk_0\omega_0 t} = \sum_{k=-\infty}^{\infty} \left\{ \sum_{q=1}^{Q} H_{mq}(k) S_q(k,t) \right\} \times e^{-j(k-k_0)\omega_0 t} + \tilde{n}_m(t) e^{jk_0\omega_0 t}. \quad (8)$$

Passing this through a low pass filter with a filter cutoff bandwidth of $\omega_c \le \omega_0/2$, (8) becomes

$$\hat{y}_m(k_0, t) \triangleq \mathrm{LPF}\left\{ y_m(t) e^{jk_0\omega_0 t} \right\} = \sum_{q=1}^{Q} H_{mq}(k_0) S_q(k_0, t) + n_m(k_0, t), \quad (9)$$

where $\mathrm{LPF}\{\cdot\}$ denotes the low pass filter operation and $n_m(k_0, t)$ is the noise resulting from the mixing and low pass filtering of $\tilde{n}_m(t)$. Hence, the broadband signal can be separated into K subbands, where $H_{mq}(k)S_q(k,t)$ are of similar bandwidth for all $k = 1 \ldots K$. The signals in (9) now form a set of focussed subband signals that can be collated for signal subspace decomposition and DOA estimation.

B. Matrix Equation for Received Signals

Suppose the received broadband signal of the mth sensor is decomposed into K subband signals, where the subband components $H_{mq}(k)S_q(k,t)$ exist for all $k = 1 \dots K$. Let

$$\hat{y}_m(k,t) = \sum_{q=1}^{Q} \tilde{y}_{mq}(k,t) + n_m(k,t) \tag{10}$$

with

$$\tilde{y}_{mq}(k,t) \triangleq H_{mq}(k)S_q(k,t). \tag{11}$$

Equation (11) can now be extended to a matrix form

$$\tilde{\mathbf{y}}_q = \mathbf{D}_q \mathbf{s}_q, \tag{12}$$

where

$$\begin{aligned} \tilde{\mathbf{y}}_q &= [\, \tilde{y}_{1q}(1,t) \quad \tilde{y}_{2q}(1,t) \quad \cdots \quad \tilde{y}_{Mq}(K,t)\,]^T_{(1\times MK)}, \\ \mathbf{s}_q &= [\, S_q(1,t) \quad S_q(2,t) \quad \cdots \quad S_q(K,t)\,]^T_{(1\times K)} \end{aligned}$$

and

$$\mathbf{D}_q = \begin{bmatrix} H_{1q}(1) & 0 & \cdots & 0 \\ H_{2q}(1) & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ H_{Mq}(1) & 0 & \cdots & 0 \\ 0 & H_{1q}(2) & \cdots & 0 \\ 0 & H_{2q}(2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & H_{Mq}(2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & H_{1q}(K) \\ 0 & 0 & \cdots & H_{2q}(K) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & H_{Mq}(K) \end{bmatrix}_{(MK\times K)}$$

is the channel transformation matrix.

By employing a vector notation, the signals received at an array of M sensors can be compactly denoted as,

$$\hat{\mathbf{y}} = \sum_{q=1}^{Q} \tilde{\mathbf{y}}_q + \mathbf{n} = \sum_{q=1}^{Q} \mathbf{D}_q \mathbf{s}_q + \mathbf{n} \tag{13}$$

where

$$\begin{aligned} \hat{\mathbf{y}} &= [\, \hat{y}_1(1,t) \quad \hat{y}_2(1,t) \quad \cdots \quad \hat{y}_M(K,t)\,]^T_{(1\times MK)} \\ &\qquad\qquad \text{and} \\ \mathbf{n} &= [\, n_1(1,t) \quad n_2(1,t) \quad \cdots \quad n_M(K,t)\,]^T_{(1\times MK)}. \end{aligned}$$

The direction information of the channel transfer functions in the direction Θ_q is now represented by the channel transformation matrix $\mathbf{D}_q$, while its columns represent the directional vectors of each subband signal $S_q(k,t)$ for $k = 1 \dots K$.

C. Eigenstructure of the Received Signal Correlation Matrix

Signal subspace techniques such as MUSIC exploit the presence of orthogonal signal and noise subspaces for DOA estimation. The upper bound on the number of simultaneously active sources is determined by the existence of the noise subspace, and is related to the number of sources and sensors in the system. This condition can be evaluated by the eigenvalue decomposition of the received signal correlation matrix $\mathbf{R} \triangleq E\{\hat{\mathbf{y}}\hat{\mathbf{y}}^H\}$, where $E\{\cdot\}$ is the expectation operator.

Equation (13) can be reformulated as

$$\hat{\mathbf{y}} = \mathbf{D}\mathbf{s} + \mathbf{n}, \tag{14}$$

where

$$\mathbf{D} = [\, \mathbf{D}_1 \quad \mathbf{D}_2 \quad \cdots \quad \mathbf{D}_Q\,]_{(MK\times KQ)}$$

and

$$\mathbf{s} = [\, \mathbf{s}_1^T \quad \mathbf{s}_2^T \quad \cdots \quad \mathbf{s}_Q^T\,]^T_{(1\times KQ)}.$$

Thus,

$$\begin{aligned} \mathbf{R} &= \mathbf{D}E\{\mathbf{s}\mathbf{s}^H\}\mathbf{D}^H + E\{\mathbf{n}\mathbf{n}^H\} \\ &= \mathbf{D}\mathbf{R}_\mathbf{s}\mathbf{D}^H + \sigma_n^2\mathbf{I}_{(MK\times MK)}, \end{aligned} \tag{15}$$

where $\mathbf{R}_\mathbf{s} = E\{\mathbf{s}\mathbf{s}^H\}$ is the source correlation matrix. Assuming spatial whiteness of the noise, σ_n^2 represents the noise power and $\mathbf{I}$ is the identity matrix. For a full rank source correlation matrix $\mathbf{R}_\mathbf{s}$ (i.e., $S_q(k,t)$ are uncorrelated for all k and q), the eigenvalue decomposition of $\mathbf{R}$ becomes

$$\mathbf{R} = [\, \hat{\mathbf{D}}_\mathbf{S} \quad \hat{\mathbf{D}}_\mathbf{N}\,] \begin{bmatrix} \mathbf{\Lambda_s} & \mathbf{0} \\ \mathbf{0} & \sigma_n^2\mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{D}}_\mathbf{S}^H \\ \hat{\mathbf{D}}_\mathbf{N}^H \end{bmatrix}, \tag{16}$$

where $\mathbf{\Lambda_s}$ is a $KQ \times KQ$ diagonal matrix that contains the noise-perturbed eigenvalues of $\mathbf{D}\mathbf{R}_\mathbf{s}\mathbf{D}^H$ and $\sigma_n^2\mathbf{I}$ forms a $K(M-Q) \times K(M-Q)$ diagonal matrix that contains the noise power. $\hat{\mathbf{D}}_\mathbf{S}$ and $\hat{\mathbf{D}}_\mathbf{N}$ are matrices whose columns represent the signal and noise eigenvectors respectively. Given that the signal eigenvalues (corresponding to the subband signal powers $|S_q(k,t)|^2$) are greater than the noise power σ_n^2, the subspaces spanned by $\hat{\mathbf{D}}_\mathbf{S}$ and $\hat{\mathbf{D}}_\mathbf{N}$ can be identified and used for DOA estimation.

In general, the existence of a noise subspace depends on the existence of $\sigma_n^2\mathbf{I}$ in (16), and is related to the dimensions of $\mathbf{R}$ and $\mathbf{\Lambda_s}$. Since the dimension of $\mathbf{\Lambda_s}$ is the same as the rank of $\mathbf{R}_\mathbf{s}$, the noise subspace only exists when

$$\text{rank}\,(\mathbf{R}) > \text{rank}(\mathbf{R}_\mathbf{s}). \tag{17}$$

The rank of $\mathbf{R}_\mathbf{s}$ is greatest when $S_q(k,t)$ are uncorrelated, hence the worst case condition for the existence of the noise subspace is given by $MK > KQ$. This is analogous to the $M > Q$ condition in subspace techniques such as MUSIC, but (17) is relaxed when $\mathbf{R}_\mathbf{s}$ is rank deficient (i.e., subbands signals of the same source are correlated). This leads to a more general condition for the existence of a noise subspace, $MK > \text{rank}(\mathbf{R}_\mathbf{s})$. Further, it raises the possibility of source localization in under-determined systems ($M < Q$), where the correlation between subband signals is known.

IV. Direction of Arrival Estimation

In the previous section we described the signal and noise subspaces created by the focussed subband signals. The noise subspace exists in systems that satisfies (17), and this condition

hints at several potential DOA estimation scenarios. This section describes the direction of arrival estimation process for the following scenarios;

1) Unknown subband uncorrelated sources: Each source is uncorrelated across the different subbands of itself and other sources. Knowledge of the source is unavailable.
2) Unknown subband correlated sources: Sources may be correlated between subbands of the same source, but are uncorrelated between sources (speech signals are good examples of subband correlated sources). The correlation between subbands is unknown.
3) Known subband correlated sources: Similar to the previous scenario; sources may exhibit correlation between subbands, but some knowledge of this correlation is available.

A. DOA Estimation of Unknown, Subband Uncorrelated Sources

Subband signals that are uncorrelated between each other is the worst-case scenario for the existence of a noise subspace. Given that (17) is satisfied, the noise subspace exists, and the channel transformation matrix $\mathbf{D}_q$ (for all $q = 1 \ldots Q$) spans a subspace of the column space of $\hat{\mathbf{D}}_\mathbf{s}$. This implies that

$$\text{span}(\hat{\mathbf{D}}_\mathbf{s}) = \text{span}([\,\mathbf{D}_1 \quad \mathbf{D}_2 \quad \cdots \quad \mathbf{D}_Q\,]),$$

and

$$\text{span}(\mathbf{D}_q) \perp \text{span}(\hat{\mathbf{D}}_\mathbf{N}). \tag{18}$$

Since $\mathbf{D}_q$ is a matrix, a columnwise test of orthogonality leads to the general MUSIC broadband DOA estimate

$$\hat{P}(\theta_q, \phi_q) = \left\{ \sum_{k=1}^{K} \frac{|\mathbf{d}_q(k)^H \mathbf{P}_\mathbf{N} \mathbf{d}_q(k)|}{|\mathbf{d}_q(k)^H \mathbf{d}_q(k)|} \right\}^{-1}, \tag{19}$$

where

$$\mathbf{d}_q(k) = [\cdots \quad 0 \quad H_{1q}(k) \quad H_{2q}(k) \quad \cdots \quad H_{Mq}(k) \quad 0 \quad \cdots]^T$$

is the kth column of $\mathbf{D}_q$ and $\mathbf{P}_\mathbf{N} = \hat{\mathbf{D}}_\mathbf{N} \hat{\mathbf{D}}_\mathbf{N}^H$ represents the measured noise space. The summation in (19) tends to zero in the directions that sources exist, hence the peaks of $\hat{P}$ indicate the source directions of arrival. We have assumed that the direction information in each subband is equally important, but if necessary, a normalized weighting factor can be introduced for the selective weighting of different subbands.

The improvements in the accuracy of the DOA estimates is due to the increased dimensionality of the received signal correlation matrix, which retains the diversity contained in the frequency domain. This contrasts with traditional techniques, where the focussing process and the fixed dimension of the received signal correlation matrix results in a loss of diversity information encoded in frequency.

B. DOA Estimation of Unknown, Subband Correlated Sources

In the previous subsection we described a DOA estimator for uncorrelated subband signals (uncorrelated $S_q(k,t)$ for all k, q). For real-world sources, the different subband signals may be correlated due to the properties of the broadband source. Human speech sources are excellent examples of such sources, where a word or phrase is a result of a signal that is modulated and resonates at multiple frequencies [31]. Since the correlation between subband signals affects the signal subspace in (16), the DOA estimator in (19) is no longer applicable. However, a subset of the received signal correlation matrix $\mathbf{R}$ can still be used for DOA estimation.

An over-determined system ($M > Q$) naturally satisfies (17), and the received signal correlation matrix of each subband will contain its own noise subspace. This property can be used for DOA estimation as follows. Expressing (15) as

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_1 & \times & \cdots & \times \\ \times & \mathbf{R}_2 & \cdots & \times \\ \vdots & \vdots & \ddots & \vdots \\ \times & \times & \cdots & \mathbf{R}_K \end{bmatrix}, \tag{20}$$

it can be seen that the received signal correlation matrix has a block diagonal structure, where the kth subband forms a subband received signal correlation matrix $\mathbf{R}_k$ and $\times$ denotes terms of no relevance. Thus,

$$\begin{aligned} \mathbf{R}_k &= \mathbf{D}(k) E\left\{\mathbf{s}_k \mathbf{s}_k^H\right\} \mathbf{D}(k)^H + E\left\{\mathbf{n}_k \mathbf{n}_k^H\right\} \\ &= \mathbf{D}(k) \mathbf{R}_\mathbf{s}(k) \mathbf{D}(k)^H + \sigma_n^2(k) \mathbf{I}_{(M \times M)}, \end{aligned} \tag{21}$$

where $\mathbf{R}_\mathbf{s}(k)$ is the subband signal correlation matrix $E\{\mathbf{s}_k \mathbf{s}_k^H\}$, $\sigma_n^2(k)$ is the noise power in the kth subband, and

$$\begin{aligned} \mathbf{s}_k &= [\,S_1(k,t) \quad S_2(k,t) \quad \cdots \quad S_Q(k,t)\,]^T_{(1 \times Q)} \\ \mathbf{n}_k &= [\,n_1(k,t) \quad n_2(k,t) \quad \cdots \quad n_M(k,t)\,]^T_{(1 \times M)} \\ \mathbf{D}(k) &= \begin{bmatrix} H_{11}(k) & H_{12}(k) & \cdots & H_{1Q}(k) \\ H_{21}(k) & H_{22}(k) & \cdots & H_{2Q}(k) \\ \vdots & \vdots & \ddots & \vdots \\ H_{M1}(k) & H_{M2}(k) & \cdots & H_{MQ}(k) \end{bmatrix} \\ &= [\,\hat{\mathbf{d}}_1(k) \quad \cdots \quad \hat{\mathbf{d}}_Q(k)\,]_{(M \times Q)}. \end{aligned}$$

The eigenvalue decomposition of $\mathbf{R}_k$ is given by

$$\mathbf{R}_k = [\,\hat{\mathbf{D}}_\mathbf{S}(k) \quad \hat{\mathbf{D}}_\mathbf{N}(k)\,] \begin{bmatrix} \mathbf{\Lambda}_\mathbf{s}(k) & \mathbf{0} \\ \mathbf{0} & \sigma_n^2(k)\mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{D}}_\mathbf{S}(k)^H \\ \hat{\mathbf{D}}_\mathbf{N}(k)^H \end{bmatrix}, \tag{22}$$

where $\mathbf{\Lambda}_\mathbf{s}(k)$ is a $Q \times Q$ diagonal matrix that contains the noise-perturbed eigenvalues of $\mathbf{D}(k)\mathbf{R}_\mathbf{s}(k)\mathbf{D}(k)^H$, and $\hat{\mathbf{D}}_\mathbf{S}(k), \hat{\mathbf{D}}_\mathbf{N}(k)$ are matrices whose columns represent the eigenvectors of the signal and noise subspaces respectively.

Since $\mathbf{D}(k)$ spans the signal subspace of $\hat{\mathbf{D}}_\mathbf{S}(k)$, this implies that

$$\text{span}(\hat{\mathbf{D}}_\mathbf{N}(k)) \perp \text{span}([\,\hat{\mathbf{d}}_1(k) \quad \hat{\mathbf{d}}_2(k) \quad \cdots \quad \hat{\mathbf{d}}_Q(k)\,]).$$

The DOA estimates of different subbands can now be combined to form the MUSIC broadband DOA estimate

$$\hat{P}(\theta_q, \phi_q) = \left\{ \sum_{k=1}^{K} \frac{|\hat{\mathbf{d}}_q(k)^H \mathbf{P}_\mathbf{N}(k) \hat{\mathbf{d}}_q(k)|}{|\hat{\mathbf{d}}_q(k)^H \hat{\mathbf{d}}_q(k)|} \right\}^{-1}, \tag{23}$$

where $\mathbf{P}_\mathbf{N}(k) = \hat{\mathbf{D}}_\mathbf{N}(k) \hat{\mathbf{D}}_\mathbf{N}(k)^H$ represents the measured noise subspace of the kth subband.

C. DOA Estimation of Known, Subband Correlated Sources

In the previous DOA estimation scenario we assumed that the correlation between subband signals was unknown. However, this resulted in much of the diversity in $\mathbf{R}$ being discarded. Consider the following practical DOA estimation scenario; locating a known individual in a sound field of multiple speakers, with the knowledge of a known spoken phrase. In the case of speech sources (in short time intervals), the different subbands are correlated across frequency [31], while remaining uncorrelated between sources. Next, we observe how any knowledge of this correlation can be incorporated into the DOA estimator.

From (15) recall that

$$\mathbf{R} = \sum_{q=1}^{Q} \mathbf{D}_q E\left\{\mathbf{s}_q \mathbf{s}_q^H\right\} \mathbf{D}_q^H + E\{\mathbf{n}\mathbf{n}^H\}. \tag{24}$$

This can be simplified further using the eigenvalue decomposition of the source correlation matrix $E\{\mathbf{s}_q \mathbf{s}_q^H\}$. Let

$$E\left\{\mathbf{s}_q \mathbf{s}_q^H\right\} = \mathbf{U}_q \mathbf{\Lambda}_q \mathbf{U}_q^H, \tag{25}$$

where $\mathbf{\Lambda}_q$ is a diagonal matrix that contains the eigenvalues of the qth source correlation matrix and the columns of $\mathbf{U}_q$ contain the corresponding eigenvectors. $\mathbf{U}_q$ now describes the relationship between the subband signals. Thus, the knowledge of the source (for a known individual and phrase) implies that $\mathbf{U}_q$ is known. Equation (24) can now be expressed as

$$\mathbf{R} = \sum_{q=1}^{Q} [\mathbf{D}_q \mathbf{U}_q] \mathbf{\Lambda}_q [\mathbf{D}_q \mathbf{U}_q]^H + E\{\mathbf{n}\mathbf{n}^H\}, \tag{26}$$

where $\mathbf{D}_q \mathbf{U}_q$ contains the direction of arrival information of the qth sound source.

Given the existence of a noise space spanned by some $\hat{\mathbf{D}}_{\mathbf{N}}$, this implies that

$$\operatorname{span}(\mathbf{D}_q \mathbf{U}_q) \perp \operatorname{span}(\hat{\mathbf{D}}_{\mathbf{N}}). \tag{27}$$

Comparing the above with (18), the DOA estimator is clearly influenced by the correlation between the subband signals of each source. Thus, any knowledge of the source will be beneficial for identifying the direction of arrival of a specific sound source. For sources that are uncorrelated

$$\operatorname{span}(\mathbf{U}_q) \perp \operatorname{span}(\mathbf{U}_{q'}) \Longrightarrow \operatorname{span}(\mathbf{D}_q \mathbf{U}_q) \perp \operatorname{span}(\mathbf{D}_q \mathbf{U}_{q'}), \tag{28}$$

where $q \neq q'$ and $\mathbf{D}_q^H \mathbf{D}_q = \mathbf{I}$. Hence, the direction of arrival of a known source q' can be uniquely identified using the MUSIC broadband DOA estimate

$$\hat{P}(\theta_q, \phi_q, q') = \left\{ \sum_{l=1}^{L} \frac{|\tilde{\mathbf{d}}_q(l)^H \mathbf{P}_{\mathbf{N}} \tilde{\mathbf{d}}_q(l)|}{|\tilde{\mathbf{d}}_q(l)^H \tilde{\mathbf{d}}_q(l)|} \right\}^{-1}, \tag{29}$$

where $\tilde{\mathbf{d}}_q(l)$ is the lth column of $\mathbf{D}_q \mathbf{U}_{q'}$ and L is the number of most significant eigenvalues of $E\{\mathbf{s}_{q'} \mathbf{s}_{q'}^H\}$.

The correlation between subband signals results in a rank reduction of $E\{\mathbf{s}_q \mathbf{s}_q^H\}$, which suggests that (17) could be satisfied by some under-determined systems. This implies that the knowledge of the correlation between subband signals may be the crucial element needed for DOA estimation in under-determined systems.

V. Simulation Setup

A. Methods of Comparison

In order to evaluate the performance of the proposed technique, we compare it with two existing techniques; a wideband high resolution method based on narrowband MUSIC [15] and a multi-sensor variant of generalized cross correlation [13]. These techniques are briefly described below.

1) Wideband MUSIC: A broadband extension of the narrowband MUSIC [17], wideband MUSIC combines the narrowband spatial covariance matrices to form a coherent signal subspace of the broadband source. At any frequency $k\omega_0$ the narrowband covariance matrix can be expressed as

$$\mathbf{P}_{\mathbf{x}}(k) = \mathbf{A}(k, \Theta) \mathbf{P}_{\mathbf{s}}(k) \mathbf{A}^{\mathbf{H}}(k, \Theta) + \sigma_n^2 \mathbf{P}_{\mathbf{n}}(k), \tag{30}$$

where $\mathbf{A}(k, \Theta)$ is the array manifold matrix, and $\mathbf{P}_{\mathbf{x}}, \mathbf{P}_{\mathbf{s}}$ and $\mathbf{P}_{\mathbf{n}}$ represent the covariance matrices of the measured signals, source and noise respectively. Next, a transformation $\mathbf{T}(k)$ [30], where

$$\mathbf{T}(k) \mathbf{A}(k, \Theta) = \mathbf{A}(k_0, \Theta),$$

is used to transform and focus each $\mathbf{A}(k, \Theta)$ into a single frequency $k_0\omega_0$. Finally, the covariance matrices are combined to form the broadband spatial covariance matrix

$$\begin{aligned} \tilde{\mathbf{P}}_{\mathbf{x}} &= \sum_{k=1}^{K} \mathbf{T}(k) \mathbf{P}_{\mathbf{x}}(k) \mathbf{T}^{\mathbf{H}}(k) \\ &= \mathbf{A}(k_0, \Theta) \tilde{\mathbf{P}}_{\mathbf{s}} \mathbf{A}^{\mathbf{H}}(k_0, \Theta) + \tilde{\mathbf{P}}_{\mathbf{n}}, \end{aligned} \tag{31}$$

where $\tilde{\mathbf{P}}_{\mathbf{x}}, \tilde{\mathbf{P}}_{\mathbf{s}}$ and $\tilde{\mathbf{P}}_{\mathbf{n}}$ represent the broadband measured signal, source and noise respectively. The formulation in (31) is similar to that in narrowband MUSIC, and the same algorithm can be applied to estimate the source directions of arrival.

2) Steered Response Power—Phase Transform: A time difference of arrival estimator, the SRP-PHAT algorithm [12] is a combination of steered beamforming and the generalized cross correlation methods. The PHAT weighted cross correlation function of the (i, j)th sensor pair is given by the inverse Fourier transform

$$R_{ij}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{Y_i(\omega) Y_j^*(\omega)}{|Y_i(\omega) Y_j^*(\omega)|} e^{j\omega\tau} \mathrm{d}\omega, \tag{32}$$

where each source location is identified by a peak in $R_{ij}(\tau)$. Since each Θ_q corresponds to a specific TDOA $\tau_{ij}(\Theta_q)$ at the (i, j)th sensor pair, the power responses obtained from the individual cross correlation functions can be combined to form the SRP-PHAT estimate

$$S(\Theta_q) = \sum_{i=1}^{M} \sum_{j=1}^{M} R_{ij}(\tau_{ij}(\Theta_q)). \tag{33}$$

The SRP-PHAT spectrum is evaluated for all Θ, and the peaks in the spectrum now correspond to the directions of arrival of the broadband sources.

B. Localization Performance Measure

In this work we compare the performance of the proposed technique with wideband MUSIC and SRP-PHAT. Since the

DOA estimation spectra of the different techniques are not directly comparable, we define a new comparative performance measure, the 'Normalized Localization Confidence.' This can be written as

$$\mathrm{NLC}(\Theta_q) = 10\log_{10}\frac{\hat{P}(\Theta_q) - \min\{\hat{P}(\Theta)\}}{\max\{\hat{P}(\Theta)\} - \min\{\hat{P}(\Theta)\}}, \quad (34)$$

where $\hat{P}(\Theta)$ represents the DOA spectrum of each technique. Since the spectral peaks represent the locations with a higher probability of a source being present, this measure effectively scales the original DOA spectrum, with a probability of 1 and 0 assigned to the maximum and minimum probable source locations respectively. The scaling process normalizes the DOA spectra of the different techniques and enables a more meaningful comparison of the performance.

C. Transfer Functions of Sensor Arrays

In order to compare the DOA estimation performance, each algorithm is applied on a uniform circular array and a sensor array placed on a complex-shaped rigid body. This subsection describes the array configuration and the channel transfer function coefficients of the two sensor arrays.

1) Uniform Circular Array: Consider a plane wave propagating from the qth source in the direction (θ_q, ϕ_q) to the sth sensor at (θ_s, ϕ_s). At a frequency $k\omega_0$, the channel transfer function of a wave field incident on the sensor array at a radial distance r can be expressed [28] as

$$\begin{aligned} H_{sq}(k) &= e^{i\vec{\mathbf{k}}_q\cdot\vec{\mathbf{r}}_s} \\ &= 4\pi\sum_{n=0}^{\infty} i^n j_n\left(\frac{k\omega_0}{c}r\right) \\ &\quad\times\left[\sum_{m=-n}^{n} Y_n^m(\theta_s,\phi_s)Y_n^m(\theta_q,\phi_q)^*\right], \end{aligned} \quad (35)$$

where $\vec{\mathbf{k}}_q$ is the wave number vector in the direction of the source, $\vec{\mathbf{r}}_s$ is the direction vector to the sensor, c is the speed of sound in the medium, $j_n(\cdot)$ is the spherical Bessel function, $Y_n^m(\cdot)$ represents the spherical harmonic function and $(\cdot)^*$ denotes the complex conjugate operation.

For sources and sensors located on the same horizontal plane, (35) can be simplified further as

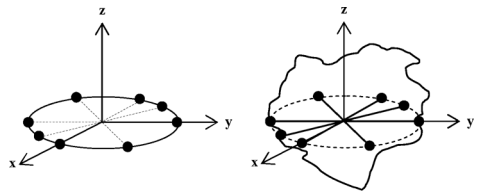
$$H_{sq}(k) = \sum_{m=-\infty}^{\infty} A_m\left(\frac{k\omega_0}{c}r\right)e^{jm\phi_s}e^{-jm\phi_q}, \quad (36)$$

where

$$A_m(k) = \sum_{n=|m|}^{\infty} j_n\left(\frac{k\omega_0}{c}r\right)i^n(2n+1)\frac{(n-|m|)!}{(n+|m|)!}\left[P_n^{|m|}(0)\right]^2$$

and $P_n^{|m|}(\cdot)$ represents the associated Legendre function. An eight sensor uniform circular array of 9 cm radius (shown in Fig. 4(a)) is used in our evaluations, and the transfer function coefficients calculated from (36) are used for $\mathbf{D}_q$ in (12).

2) Sensor Array on a Complex-Shaped Rigid Body: In order to evaluate the effect of frequency domain diversity on the DOA estimation performance, we consider a sensor array mounted on a complex-shaped rigid body as shown in Fig. 4(b).



(a) (b)

Fig. 4. Array geometry of (a) an eight sensor uniform circular array and (b) an eight sensor array on a complex-shaped rigid body.

We construct this hypothetical object using the HRTF information (known to contain diversity in the frequency domain) of four subjects in the CIPIC HRTF database [34]. The right and left ears of CIPIC subjects are treated as sensors located at $(\pi/2, \pi/2)$ and $(\pi/2, 3\pi/2)$ respectively. We distribute the sensors in the horizontal plane by introducing a rotation ϕ_{rot} around the vertical axis of each subjects. For example, by letting $\phi_{\mathrm{rot}} = [-\pi/2 \quad -\pi/4 \quad 0 \quad \pi/4]$ (each element of ϕ_{rot} corresponds to the rotation applied to a specific subject), a group of eight sensors will be uniformly distributed on the horizontal plane, 45° apart from each other. The resulting hypothetical object can be visualized as eight pinnae uniformly distributed on the horizontal plane of an approximately spherical object.

Using the HRTF measurements of the CIPIC subjects sampled at 5° intervals[2], the kth column of the channel transformation matrix $\mathbf{D}_q$ can be written as

$$\mathbf{d}_q(k) = [\cdots \quad 0 \quad H_{1q}(k) \quad H_{2q}(k) \quad \cdots \quad H_{8q}(k) \quad 0 \quad \cdots]^T, \quad (37)$$

where $H_{sq}(k)$ represents the HRTF in the direction Θ_q, and $s = 1, \ldots, 8$ indicates the different sensors on the object. The channel transfer functions now behave in the complicated fashion associated with the complex-shaped scatterer, and the resulting channel transformation matrix can be used for DOA estimation as described in Section IV.

VI. Evaluation and Discussion

In this section we compare the DOA estimation performance of the proposed MUSIC technique, wideband MUSIC and SRP-PHAT. The three DOA estimation scenarios in Section IV involve two types of sources; subband uncorrelated (ideal) sources and subband correlated (real-world) sources. We reproduce these conditions using the following signals.

- Ideal sources: Simulated by a collection of Gaussian pulses, modulated by a sinusoidal carrier with a random phase. The Gaussian pulses are time-shifted to remain uncorrelated between subbands and sources.
- Real-world sources: Real-world sources such as human speech exhibit some correlation between frequencies due to the natural processes that create these sources. Recordings of pure speech sources and of speech sources

[2]A continuous model of the channel transfer function can be obtained through the efficient sampling of the channel impulse response at a set of discrete locations. The sampling requirements for modeling the HRTF of a KEMAR manikin has been investigated by Zhang *et al.* in [35], where it was found that sampling at 5° was sufficient to recreate the HRTF up to 10 kHz.

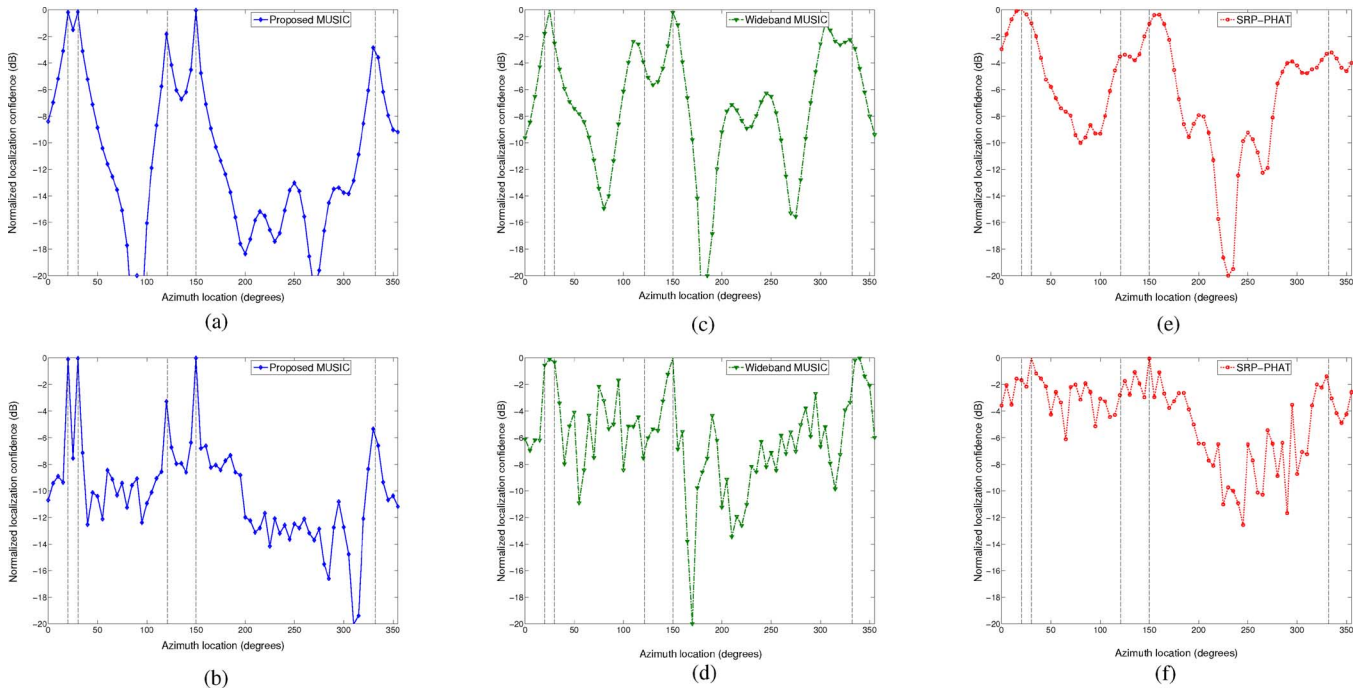


Fig. 5. DOA estimates of the Proposed MUSIC, Wideband MUSIC and SRP-PHAT techniques for subband uncorrelated sources at 10 dB SNR and 4 kHz audio bandwidth. Subfigures (a), (c), (e) are the DOA estimates of the uniform circular array and (b), (d), (f) are those of the sensor array on the complex-shaped rigid body. The five sources are located on the azimuth plane at 20°, 30°, 121°, 150° and 332°, indicated by the dashed vertical lines.

including musical content are used to simulate the subband correlated sources.

The DOA estimation scenarios are evaluated assuming free field propagation conditions between the sources and the sensor arrays described in Section V.C. We consider the case where sources and sensors are on the same horizontal plane, and the average signal to noise ratio (SNR) is 10 dB. SNR is defined as the ratio of the received source power to the noise power at the sensor, averaged across the frequency bandwidth used for DOA estimation. The proposed technique is implemented using a subband bandwidth of 50 Hz, where the subband central frequencies are at multiples of 50 Hz in the [0.3, 4] kHz or [0.3, 8] kHz frequency range, as applicable. All sound sources are two seconds in length and are sampled at 44.1 kHz. The channel transfer function coefficient $H_{mg}(k)$ is calculated at each frequency bin using an 882 point Discrete Fourier Transform (DFT) of the relevant impulse response data. A similar approach is used to implement wideband MUSIC, where the narrowband covariance matrices are calculated from an 882 point moving window DFT of the filtered broadband signal. A focussing frequency of 2.5 kHz is used to avoid spatial aliasing [33] and to maximize the spatial resolution. The results presented in this paper use the mean 'normalized localization confidence' obtained from 50 trial runs, where different sound and noise sources are considered.

A. DOA Estimation Scenario: Unknown, Subband Uncorrelated Sources

Consider the DOA estimation of five subband uncorrelated sources (ideal sources), located on the horizontal plane in the azimuth directions of 20°, 30°, 121°, 150° and 332°. Fig. 5 illustrates the DOA estimation performance of each technique, evaluated at every 5° in the azimuth plane using an audio bandwidth of 4 kHz.

Figs. 5(a), (c) and (e) illustrate the localization performance using the uniform circular array. The proposed MUSIC and wideband MUSIC techniques produce DOA estimation spectra of similar profile, where the proposed technique displays a higher estimation accuracy. In this scenario, the closely spaced sources at 20° and 30° are not separated by wideband MUSIC, whereas the proposed MUSIC technique is on the verge of identifying the two sources. The performance of the TDOA technique SRP-PHAT is comparatively less conclusive, where the four primary source regions are identified at a lower source location accuracy. The closely spaced sources are unresolvable as they are within the resolution limit of the circular array.

The DOA estimation performance using the sensor array on the complex-shaped rigid body is shown in Figs. 5(b), (d) and (f). The proposed MUSIC technique clearly identifies the five source locations, with an approximately 6 dB improvement in closely spaced source resolution (at the 20° and 30° locations) compared to the circular array in Fig. 5(a). The floor of the DOA spectrum has risen, but a 7 dB minimum difference between the source and adjacent locations is achieved. The DOA estimation performance of wideband MUSIC and SRP-PHAT are severely degraded, to the point where a reliable estimate of the source location is not possible. This can be attributed to the imperfect focussing matrices used by wideband MUSIC, while the failure of SRP-PHAT is related to the complicated TDOA behavior of the sensor array on the complex-shaped rigid body.

The DOA estimation performance of the proposed method with SNR is illustrated in Fig. 6 for 5 dB and -5 dB SNR at the sensor array on the complex-shaped rigid body. Simulation results suggest that a minimum SNR of -5 dB to 0 dB is

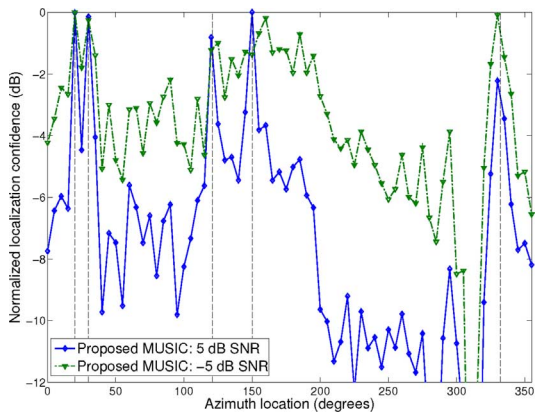


Fig. 6. Performance of the Proposed MUSIC estimator with SNR for uncorrelated sources using the sensor array on the complex-shaped rigid body at 4 kHz audio bandwidth. The five sources are located in the azimuth plane at 20°, 30°, 121°, 150° and 332°, indicated by the dashed vertical lines.

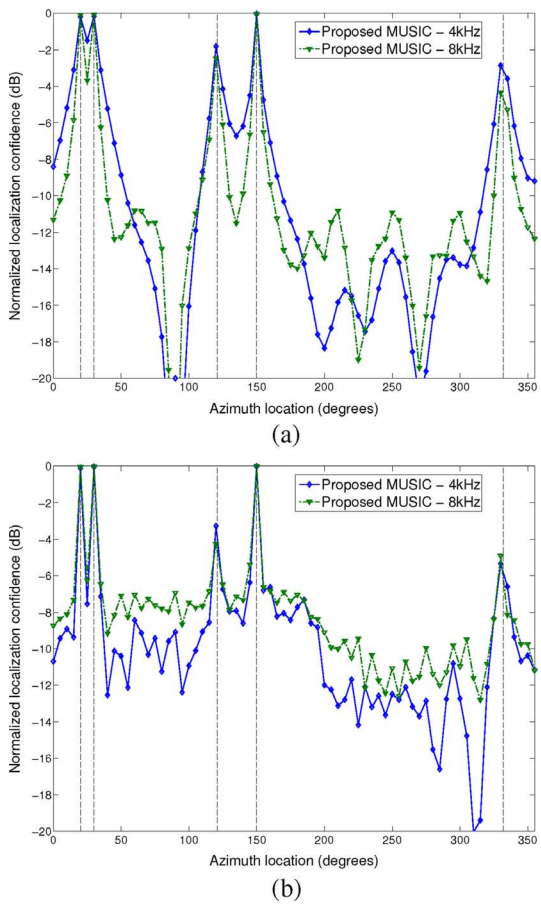


Fig. 7. DOA estimates of the Proposed MUSIC technique for subband uncorrelated sources at 10 dB SNR for 4 kHz (dotted line) and 8 kHz (dot-dash line) audio bandwidths. Subfigures (a) and (b) are the DOA estimates of the uniform circular array and the sensor array on the complex-shaped rigid body respectively. The five sources are located in the azimuth plane at 20°, 30°, 121°, 150° and 332°, indicated by the dashed vertical lines.

required to accurately resolve source locations, which loosely corresponds with the SNR requirements of wideband MUSIC. Fig. 7 illustrates the DOA estimation performance of the proposed technique with increasing audio bandwidth. In Fig. 7(a) we observe that the doubling of the audio bandwidth from 4 kHz to 8 kHz improves the closely spaced source resolution by up to 3 dB. However, the resolution improvements for the sensor array on the complex-shaped rigid body in Fig. 7(b) are marginal. This suggests that the number of subbands (K) used by the DOA estimator is a key factor that determines the resolution of the proposed technique.

This phenomena can be explained as follows. Consider the circular array, where diversity is encoded in the TDOA between subband signals. Intuitively, increasing the number of subbands can be considered a process of averaging the TDOA estimates over multiple subbands. Thus, the source location accuracy is expected to improve with increasing K. This basic relationship is applicable to the sensor array on the complex-shaped rigid body, but it is just one factor that affects the DOA estimation accuracy. From (19), it is seen that the proposed DOA estimator employs a sum of the orthogonality tests of each subband. This implies that increasing K will improve resolution by exploiting the frequency diversity of the subbands, although the marginal contribution by each additional subband is decreasing. Hence, the improvement in source resolution reaches a limit, beyond which increasing K becomes ineffective. This limit corresponds to a bandwidth of approximately 4 kHz for the hypothetical object used in our simulations. Overall, the number of subbands K and the subband bandwidth ω_0 can be described as design parameters related to the rigid body.

B. DOA Estimation Scenario: Unknown, Subband Correlated Sources

Consider the DOA estimation of five unknown subband correlated sources (real-world sources that are independent of each other), located on the horizontal plane in the azimuth directions of 20°, 30°, 121°, 150° and 332°. Fig. 8 illustrates the DOA estimation performance of each technique, evaluated at 5° intervals using an audio bandwidth of 4 kHz.

Typically, the extracted subband signals of a real-world speech source are correlated, due to the physiological processes that create speech. At low frequencies, each subband is essentially a scaled version of the modulating signal (the spoken word or phrase) of the speaker. Since the relationship between the different subbands is unknown, the additional diversity information in $\mathbf{R}$ must be discarded to produce the proposed MUSIC DOA estimate in (23). Figs. 8(a), (c) and (e) illustrate the DOA estimation performance of the three techniques using the uniform circular array. As expected, the estimation accuracy of wideband MUSIC and SRP-PHAT are similar to the previous scenario, and the closely spaced sources cannot be resolved. The performance of the proposed MUSIC technique is superior, and is just beginning to resolve the closely spaced sources at 20° and 30°.

The DOA estimation performance using the sensor array on the complex-shaped rigid body is illustrated in Figs. 8(b), (d) and (f). As in the previous scenario, wideband MUSIC and SRP-PHAT are unable to produce a clear estimate of the source directions of arrival, whereas the proposed MUSIC technique accurately identifies the source directions for SNRs above 0 dB. A 6 dB minimum separation between the source and adjacent

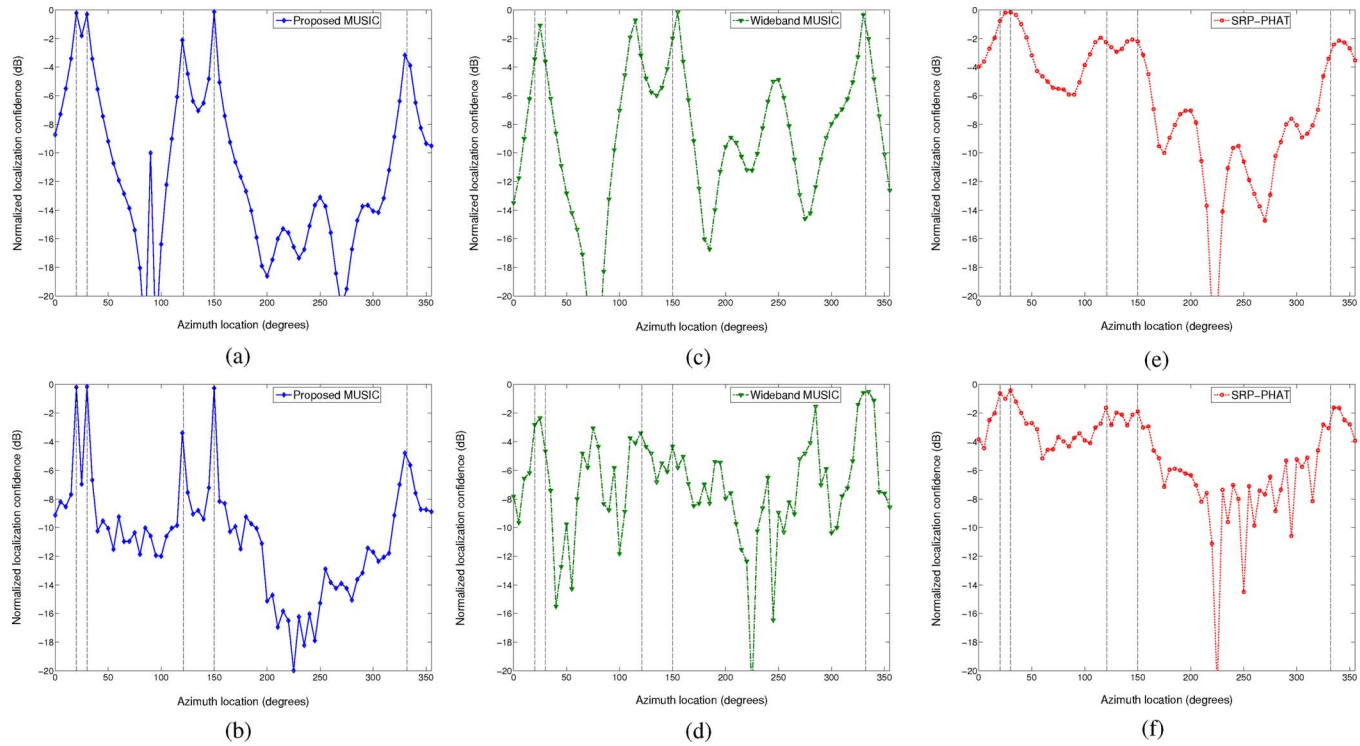


Fig. 8. DOA estimates of the Proposed MUSIC, Wideband MUSIC and SRP-PHAT techniques for subband correlated sources at 10 dB SNR and 4 kHz audio bandwidth. Subfigures (a), (c), (e) are the DOA estimates of the uniform circular array and (b), (d), (f) are those of the sensor array on the complex-shaped rigid body. The five sources are located in the azimuth plane at 20°, 30°, 121°, 150° and 332°, indicated by the dashed vertical lines.

locations is achieved at a SNR of 10 dB; a 1 dB reduction in comparison with the previous estimation scenario.

These results imply that real-world sources can be localized without any source knowledge, and that the lack of inter-subband information in (15) has a negligible impact. Hence, the proposed technique effectively utilizes the frequency domain diversity of the complicated channel transfer functions, while reducing the computational complexity for the real-world DOA estimation scenario.

C. DOA Estimation Scenario: Known, Subband Correlated Sources

Consider the scenario where a particular speaker is to be located in a multi-source sound field using a spoken word or phrase; the classic cocktail party scenario. Using the knowledge of the speaker and what is being said, the relationship between the subband signals can be established. Thus, it is possible to obtain a DOA estimate of a particular source as described in Section IV.C. Figs. 9 and 10 illustrate the DOA estimation performance of the proposed MUSIC technique, where the sources are located on the horizontal plane in the directions 120°, 150° and 330°.

Fig. 9 shows the DOA estimates of ideal sources (partially correlated across subbands), where the two plots are the DOA estimates of the sources at 120° and 330°, respectively. We have assumed that our knowledge of the source is incomplete, hence L in (29) is selected to include the eigenvalues greater than 50% of the maximum eigenvalue of each source. Fig. 9 suggests that the specified source can be localized with similar performance as the multi-source DOA estimation scenarios discussed previously. As expected, the localization performance using the sensor array on the complex-shaped rigid body is superior to the uniform circular array, and a sharper separation of adjacent locations is observed.

Fig. 10 illustrates the DOA estimation performance of real-world sources, i.e., the speech and speech + music signals described previously. As in the previous case, L is selected to include the eigenvalues greater than 50% of the maximum eigenvalue, while the source correlation matrices are calculated in 25 ms intervals; the time period is selected to ensure that the statistics of the speech signal remains stationary [31], [36], [37]. In practice, the speaker's words and tone of voice can be used to derive the relationship between the subband signals. The DOA spectrum is averaged across multiple time intervals (syllables) to obtain the DOA estimates as shown in Fig. 10. Both sensor arrays identify the actual source locations, although the performance gained by using the sensor array on the complex-shaped rigid body is reduced in Fig. 10(b). Comparing Figs. 9 and 10, it can be seen that the imperfect knowledge of the source statistics affects the DOA estimation performance. However, any knowledge gained can now be applied to other subspace reduction techniques [38] for iterative DOA estimation in under-determined systems.

D. Computational Complexity

The computational complexity of the proposed DOA estimation scenarios vary. Scenario A (DOA estimation of unknown, uncorrelated sources) and Scenario B (DOA estimation of unknown, correlated sources) represent the most and least computationally complex scenarios respectively. Table I compares the computational complexity of these scenarios in the big-$\mathcal{O}$

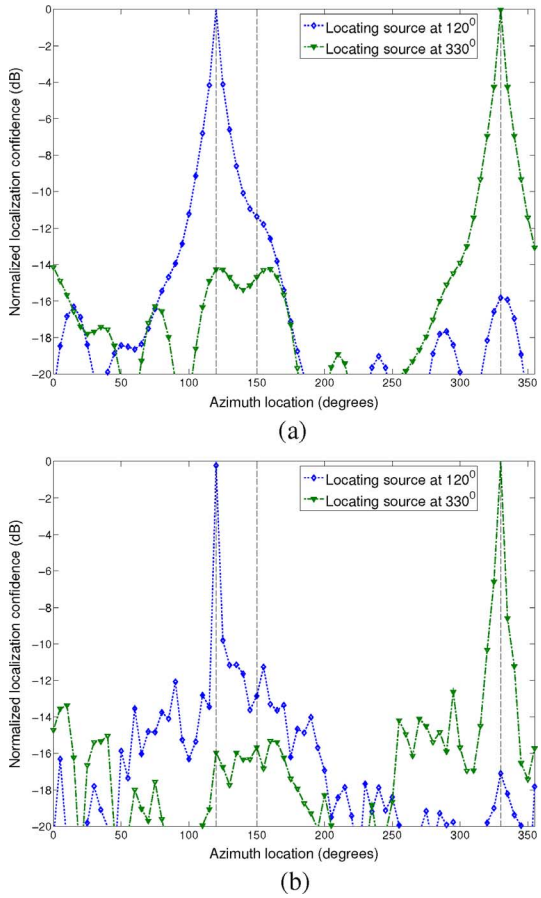


Fig. 9. DOA estimates of two known sources (imperfect source knowledge) in a sound field of three sources, using the Proposed MUSIC technique at 10 dB SNR and 4 kHz audio bandwidth. The simulated sources are correlated between subbands, but uncorrelated between each other. Subfigures (a) and (b) are the DOA estimates of the uniform circular array and the sensor array on the complex-shaped rigid body respectively. The sources are located in the azimuth plane at 120°, 150° and 330°, indicated by dashed vertical lines.

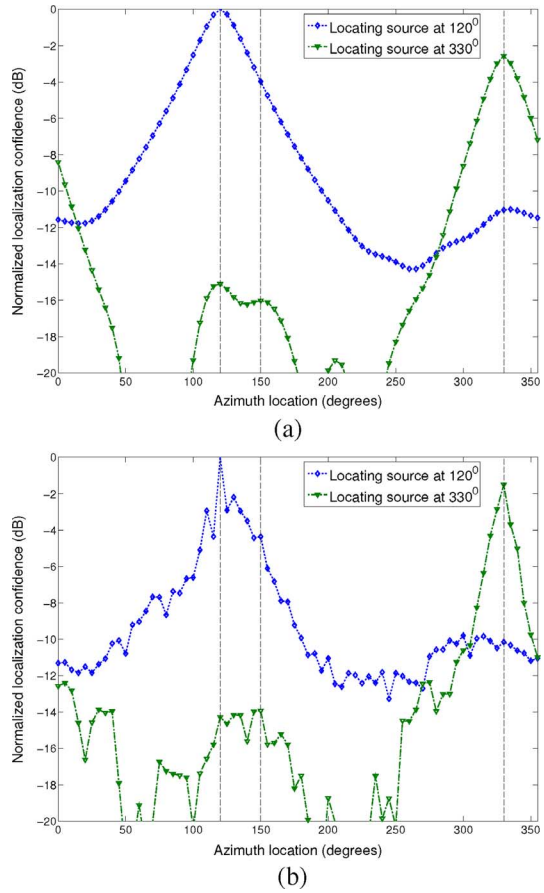


Fig. 10. DOA estimates of two known sources (imperfect source knowledge) in a sound field of three sources, using the Proposed MUSIC technique at 10 dB SNR and 4 kHz audio bandwidth. Three real-world speech and speech + music sources are simulated. Subfigures (a) and (b) are the DOA estimates of the uniform circular array and the sensor array on the complex-shaped rigid body respectively. The sources are located in the azimuth plane at 120°, 150° and 330°, indicated by the dashed vertical lines.

notation using the number of multiplication operations as an estimate of the complexity. Each algorithm can be separated into four main stages; subband extraction, correlation matrix calculation, eigenvalue decomposition and direction of arrival estimation. The symbols used in the big-$\mathcal{O}$ notation, M, K, N, L and T represent the number of sensors, number of subbands, FFT window length, number of filter taps and the number of samples used in the calculations respectively.

We note that the proposed technique is generally more complex than wideband MUSIC, and that the increased complexity can be attributed to the latter two stages. In the most complex case (Scenario A), the increased complexity is primarily due to the eigenvalue decomposition, where the required computations increase as the cube of K. The complexity of the correlation calculation and DOA estimation stages increase as the square of K. Thus, the most complex scenario represents a cubic increase in complexity. However, for the least complex case (Scenario B), the computational complexity increases linearly with K. Since this scenario corresponds to the DOA estimation of unknown real-world sources, for many practical applications, the resolution gained by using a sensor array on a rigid body will represent a linear increase in complexity with the number of subbands.

TABLE I
COMPUTATIONAL COMPLEXITY OF THE PROPOSED TECHNIQUE AND WIDEBAND MUSIC

	Proposed method	Wideband MUSIC
Subband decomposition / FFT	$K \cdot \mathcal{O}(L)$	$\mathcal{O}(N \log N)$
Correlation matrix computation	$KT \cdot \mathcal{O}(M^2)$ to $T \cdot \mathcal{O}(M^2K^2)$	$KT \cdot \mathcal{O}(M^2)$
Eigenvalue decomposition	$K \cdot \mathcal{O}(M^3)$ to $\mathcal{O}(M^3K^3)$	$\mathcal{O}(M^3)$
DOA estimation	$K \cdot \mathcal{O}(M^2)$ to $\mathcal{O}(M^2K^2)$	$\mathcal{O}(M^2)$

VII. CONCLUSION

In this paper, we propose a high resolution broadband DOA estimation technique for sensor arrays mounted on complex-shaped rigid bodies. We describe the broadband source as a collection of modulated subband signals and introduce a frequency focussing method to extract these signals. The subband signals can then be combined to create a higher dimensional signal correlation matrix, where the frequency domain diversity is retained. We find that DOA estimation using the signal subspace decomposition of the new correlation matrix provides higher resolution in various DOA estimation scenarios. In conclusion, we demonstrate that the complexity introduced in the frequency domain is not a hindrance, but is a source of diversity that can be exploited to achieve higher resolution and clearer separation

of closely spaced sources. We should state that mathematically modeling the diversity information for the sensor array on the rigid body is not straightforward due to the large audio bandwidth and subband design of the estimator, which complicates the process of deriving the Cramér-Rao Bound (CRB). However, the CRB is an important benchmark for the analysis of the efficiency of the proposed estimator, and will be addressed in the future.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions that helped to improve the clarity and quality of this manuscript.

REFERENCES

[1] T. Gold and R. J. Pumphrey, "Hearing. I. the cochlea as a frequency analyzer," *Proc. R. Soc. Lond. B Biol. Sci.*, vol. 135, no. 881, pp. 462–491, Dec. 1948.

[2] P. M. Hofman, J. G. Van Riswick, and A. J. Van Opstal, "Relearning sound localization with new ears," *Nat. Neurosci.*, vol. 1, no. 5, pp. 417–421, Sept. 1998.

[3] V. Best, S. Carlile, C. Jin, and A. van Schaik, "The role of high frequencies in speech localization," *J. Acoust. Soc. Amer.*, vol. 118, no. 1, pp. 353–363, Jul. 2005.

[4] M. Morimoto, K. Iida, and M. Itoh, "Upper hemisphere sound localization using head-related transfer functions in the median plane and interaural differences," *Acoust. Sci. Technol.*, vol. 24, no. 5, pp. 267–275, Sep. 2003.

[5] B. Rakerd, W. M. Hartmann, and T. L. McCaskey, "Identification and localization of sound sources in the median sagittal plane," *J. Acoust. Soc. Amer.*, vol. 106, no. 5, pp. 2812–2820, Nov. 1999.

[6] B. G. Shinn-Cunningham, S. Santarelli, and N. Kopco, "Tori of confusion: Binaural localization cues for sources within reach of a listener," *J. Acoust. Soc. Amer.*, vol. 107, no. 3, pp. 1627–1636, Mar. 2000.

[7] M. Aytekin, E. Grassi, M. Sahota, and C. F. Moss, "The bat head-related transfer function reveals binaural cues for sound localization in azimuth and elevation," *J. Acoust. Soc. Amer.*, vol. 116, no. 6, pp. 3594–3605, Dec. 2004.

[8] J. Woodruff and D. Wang, "Binaural localization of multiple sources in reverberant and noisy environments," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 5, pp. 1503–1512, Jul. 2012.

[9] K. Iida, "Estimation of sound source elevation by extracting the vertical localization cues from binaural signals," in *Proc. Meetings Acoust., 155th Meeting Acoust. Soc. Amer.*, Paris, France, Jul. 2008, vol. 4, p. 050002.

[10] M. Raspaud, H. Viste, and G. Evangelista, "Binaural source localization by joint estimation of ILD and ITD," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 1, pp. 68–77, Jan. 2010.

[11] P. Zakarauskas and M. S. Cynader, "A computational theory of spectral cue localization," *J. Acoust. Soc. Amer.*, vol. 94, no. 3, pp. 1323–1331, Sept. 1993.

[12] J. H. DiBiase, "A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays," Ph.D. dissertation, Brown Univ., Providence, RI, USA, 2000.

[13] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-24, no. 4, pp. 320–327, Aug. 1976.

[14] J. Dmochowski, J. Benesty, and S. Affes, "Direction of arrival estimation using the parameterized spatial correlation matrix," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 4, pp. 1327–1339, May 2007.

[15] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. ASSP-34, no. 3, pp. 276–280, Mar. 1986.

[16] 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.

[17] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 4, pp. 823–831, Aug. 1985.

[18] N. Wang, P. Agathoklis, and A. Antoniou, "A new DOA estimation technique based on subarray beamforming," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3279–3290, Sep. 2006.

[19] D. Ward, Z. Ding, and R. Kennedy, "Broadband DOA estimation using frequency invariant beamforming," *IEEE Trans. Signal Process.*, vol. 46, no. 5, pp. 1463–1469, May 1998.

[20] T.-S. Lee, "Efficient wideband source localization using beamforming invariance technique," *IEEE Trans. Signal Process.*, vol. 42, no. 6, pp. 1376–1387, Jun. 1994.

[21] T. D. Abhayapala and H. Bhatta, "Coherent broadband source localization by modal space processing," in *Proc. 10th Int. Conf. Telecomm., ICT '03*, Papeete, Tahiti, French Polynesia, Feb. 2003, vol. 2, pp. 1617–1623.

[22] H. Teutsch and W. Kellermann, "Acoustic source detection and localization based on wavefield decomposition using circular microphone arrays," *J. Acoust. Soc. Amer.*, vol. 120, no. 5, pp. 2724–2736, 2006.

[23] H. Hung and M. Kaveh, "Focussing matrices for coherent signal-subspace processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 8, pp. 1272–1281, Aug. 1988.

[24] H. Hung and M. Kaveh, "Coherent wide-band ESPRIT method for directions-of-arrival estimation of multiple wide-band sources," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 2, pp. 354–356, Feb. 1990.

[25] F. Belloni, A. Richter, and V. Koivunen, "DoA estimation via manifold separation for arbitrary array structures," *IEEE Trans. Signal Process.*, vol. 55, no. 10, pp. 4800–4810, Oct. 2007.

[26] M. Costa, A. Richter, and V. Koivunen, "DoA and polarization estimation for arbitrary array configurations," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2330–2343, May 2012.

[27] Q. Yuan, Q. Chen, and K. Sawaya, "Accurate DOA estimation using array antenna with arbitrary geometry," *IEEE Trans. Antennas Propag.*, vol. 53, no. 4, pp. 1352–1357, Apr. 2005.

[28] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. New York, NY, USA: Academic, 1999.

[29] K. Brandenburg and G. Stoll, "ISO/MPEG-1 audio: A generic standard for coding of high-quality digital audio," *J. Audio Eng. Soc.*, vol. 42, no. 10, pp. 780–792, Oct. 1994.

[30] D. Pan, "A tutorial on MPEG/audio compression," *IEEE Multimedia*, vol. 2, no. 2, pp. 60–74, Jun. 1995.

[31] L. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[32] A. V. Oppenheim, A. S. Willsky, and S. Hamid, *Signals and Systems*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1996.

[33] R. Monzingo and T. Miller, *Introduction to Adaptive Arrays*. New York, NY, USA: Wiley, 1980.

[34] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoustics, WASPAA '01*, New Paltz, NY, USA, Oct. 2001, pp. 99–102.

[35] W. Zhang, M. Zhang, R. A. Kennedy, and T. D. Abhayapala, "On high-resolution head-related transfer function measurements: An efficient sampling scheme," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 2, pp. 575–584, Feb. 2012.

[36] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-28, no. 4, pp. 357–366, Aug. 1980.

[37] R. McAulay and T. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 4, pp. 744–754, Aug. 1986.

[38] G. Bouleux and R. Boyer, "Zero-forcing based sequential MUSIC algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., ICASSP '07*, Honolulu, HI, USA, Apr. 2007, vol. 3, pp. 1017–1020.

Dumidu S. Talagala (S'11) received the B.Sc. Eng. (Hons) in electronic and telecommunication engineering from the University of Moratuwa, Sri Lanka, in 2007. From 2007 to 2009, he was an Engineer at Dialog Axiata PLC, Sri Lanka. He is currently pursuing the Ph.D. degree within the Applied Signal Processing Group, College of Engineering and Computer Science, at the Australian National University, Canberra.

His research interests are in the areas of sound source localization, spatial soundfield reproduction, active noise control and array signal processing. He is a graduate student member of the IEEE.

Wen Zhang (S'06–M'09) received the B.E. degree in telecommunication engineering from Xidian University, Xi'an, China, in 2003 and the M.E. degree in electrical engineering (with first class honors) and the Ph.D. degree from the Australian National University, Canberra, in 2005 and 2010, respectively.

From 2010 to 2012, she was an OCE Postdoctoral Fellow at CSIRO Process Science and Engineering, Sydney, Australia. She is currently a research fellow in the Research School of Engineering at the Australian National University. Her primary research interests are in the field of spatial sound-field recording and reconstruction, source separation and localization, and active noise cancellation. She is currently an editorial board member for Science Journal of Circuits, Systems and Signal Processing.

Thushara D. Abhayapala (M'00–SM'08) received the B.E. degree (with honors) in interdisciplinary systems engineering and the Ph.D. degree in telecommunications engineering, from the Australian National University (ANU), Canberra, in 1994 and 1999, respectively. From 1995 to 1997, he was a Research Engineer with the Arthur C. Clarke Center for Modern Technologies, Sri Lanka. Since December 1999, he has been a faculty member at ANU.

He was the Leader of the Wireless Signal Processing (WSP) Program at the National ICT Australia (NICTA) from November 2005 to June 2007. Currently, he is the Director of the Research School of Engineering at the Australian National University.

His research interests are in the areas of spatial audio and acoustic signal processing, space-time signal processing for wireless communication systems, and array signal processing. Professor Abhayapala has supervised 28 research students and coauthored over 180 peer-reviewed papers. He is an Associate Editor for the *EURASIP Journal on Wireless Communications and Networking*. Professor Abhayapala is a Member of the Audio and Acoustic Signal Processing Technical Committee (2011–2013) of the IEEE Signal Processing Society.