Direction of Arrival Estimation Using the Parameterized Spatial Correlation Matrix

Jacek Dmochowski, Jacob Benesty, *Senior Member, IEEE*, and Sofiène Affes, *Senior Member, IEEE*

***Abstract*—The estimation of the direction-of-arrival (DOA) of one or more acoustic sources is an area that has generated much interest in recent years, with applications like automatic video camera steering and multiparty stereophonic teleconferencing entering the market. DOA estimation algorithms are hindered by the effects of background noise and reverberation. Methods based on the time-differences-of-arrival (TDOA) are commonly used to determine the azimuth angle of arrival of an acoustic source. TDOA-based methods compute each relative delay using only two microphones, even though additional microphones are usually available. This paper deals with DOA estimation based on spatial spectral estimation, and establishes the parameterized spatial correlation matrix as the framework for this class of DOA estimators. This matrix jointly takes into account all pairs of microphones, and is at the heart of several broadband spatial spectral estimators, including steered-response power (SRP) algorithms. This paper reviews and evaluates these broadband spatial spectral estimators, comparing their performance to TDOA-based locators. In addition, an eigenanalysis of the parameterized spatial correlation matrix is performed and reveals that such analysis allows one to estimate the channel attenuation from factors such as uncalibrated microphones. This estimate generalizes the broadband minimum variance spatial spectral estimator to more general signal models. A DOA estimator based on the multichannel cross correlation coefficient (MCCC) is also proposed. The performance of all proposed algorithms is included in the evaluation. It is shown that adding extra microphones helps combat the effects of background noise and reverberation. Furthermore, the link between accurate spatial spectral estimation and corresponding DOA estimation is investigated. The application of the minimum variance and MCCC methods to the spatial spectral estimation problem leads to better resolution than that of the commonly used fixed-weighted SRP spectrum. However, this increased spatial spectral resolution does not always translate to more accurate DOA estimation.**

***Index Terms*—Circular arrays, delay-and-sum beamforming (DSB), direction-of-arrival (DOA) estimation, linear spatial prediction, microphone arrays, multichannel cross correlation coefficient (MCCC), spatial correlation matrix, time delay estimation.**

I. Introduction

PROPAGATING signals contain much information about the sources that emit them. Indeed, the location of a signal source is of much interest in many applications, and there exists a large and increasing need to locate and track sound sources. For example, a signal-enhancing beamformer [1], [2] must continuously monitor the position of the desired signal source in order to provide the desired directivity and interference suppression. This paper is concerned with estimating the direction-of-arrival (DOA) of acoustic sources in the presence of significant levels of both noise and reverberation.

The two major classes of broadband DOA estimation techniques are those based on the time-differences-of-arrival (TDOA) and spatial spectral estimators. The latter terminology arises from the fact that spatial frequency corresponds to the wavenumber vector, whose direction is that of the propagating signal. Therefore, by looking for peaks in the spatial spectrum, one is determining the DOAs of the dominant signal sources.

The TDOA approach is based on the relationship between DOA and relative delays across the array. The problem of estimating these relative delays is termed "time delay estimation" [3]. The generalized cross-correlation (GCC) approach of [4], [5] is the most popular time delay estimation technique. Alternative methods of estimating the TDOA include phase regression [6] and linear prediction preprocessing [7]. The resulting relative delays are then mapped to the DOA by an appropriate inverse function that takes into account array geometry.

Even though multiple-microphone arrays are commonplace in time delay estimation algorithms, there has not emerged a clearly preferred way of combining the various measurements from multiple microphones. Notice that in the TDOA approach, the time delays are estimated using only *two* microphones at a time, even though one usually has several more sensor outputs at one's disposal. The averaging of measurements from independent pairs of microphones is not an optimal way of combining the measurements, as each computed time delay is derived from only two microphones, and thus often contains significant levels of corrupting noise and interference. It is thus well known that current TDOA-based DOA estimation algorithms are plagued by the effects of both noise and especially reverberation.

To that end, Griebel and Brandstein [8] map all "realizable" combinations of microphone-pair delays to the corresponding source locations, and maximize simultaneously the *sum* (across various microphone pairs) of cross-correlations across all possible locations. This approach is notable, as it jointly maximizes the results of the cross-correlations between the various microphone pairs.

The spatial spectral estimation problem is well defined in the narrowband signal community. There are three major methods: the steered conventional beamformer approach (also termed the "Bartlett" estimate), the minimum variance estimator (also termed the "Capon" or maximum-likelihood estimator), and the linear spatial predictive spectral estimator. Reference [9]

Manuscript received September 6, 2006; revised November 8, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Hiroshi Sawada.

The authors are with the Institut National de la Recherche Scientifique-Énergie, Matériaux, et Télécommunications (INRS-EMT), Université du Québec, Montréal, QC H5A 1K6, Canada (e-mail: dmochow@emt.inrs.ca).

Digital Object Identifier 10.1109/TASL.2006.889795

provides an excellent overview of these approaches. These three approaches are unified in their use of the narrowband spatial correlation matrix, as outlined in the next section.

The situation is more scattered in the broadband signal case. Various spectral estimators have been proposed, but there does not exist any common framework for organizing these approaches. The steered conventional beamformer approach applies to broadband signals. The delay-and-sum beamformer (DSB) is steered to all possible DOAs to determine the DOA which emits the most energy. An alternative formulation of this approach is termed the "steered-response power" (SRP) method, which exploits the fact that the DSB output power may be written as a sum of cross-correlations. The computational requirements of the SRP method are a hindrance to practical implementation [8]. A detailed treatment of steered-beamformer approaches to source localization is given in [10], and the statistical optimality of the approach is shown in [11]–[13]. Krolik and Swingler develop a broadband minimum variance estimator based on the steered conventional beamformer [14], which may be viewed as an adaptive weighted SRP algorithm. There have also been approaches that generalize narrowband localization algorithms (i.e., MUSIC [15]) to broadband signals through subband processing and subsequent combining (see [16], for example). A broadband linear spatial predictive approach to *time delay estimation* is outlined in [17] and [18]. This approach, which is limited to linear array geometries, makes use of all the channels in a joint fashion via the time delay parameterized spatial correlation matrix.

This paper attempts to unify broadband spatial spectral estimators into a single framework and compares their performance from a DOA estimation standpoint to TDOA-based algorithms. This unified framework is the azimuth parameterized spatial correlation matrix, which is at the heart of all broadband spatial spectral estimators.

In addition, several new ideas are presented. First, due to the parametrization, well-known narrowband array processing notions [19] are applied to the DOA estimation problem, generalizing these ideas to the broadband case. A DOA estimator based on the eigenanalysis of the parameterized spatial correlation matrix ensues. More importantly, it is shown that this eigenanalysis allows one to estimate the channel attenuation from factors such as uncalibrated microphones. The existing minimum variance approach to broadband spatial spectral estimation is reformulated in the context of a more general signal model which accounts for such attenuation factors. Furthermore, the ideas of [17] and [18] are extended to more general array geometries (i.e., circular) via the azimuth parameterized spatial correlation matrix, resulting in a minimum entropy DOA estimator.

Circular arrays (see [20]–[22], for example) offer some advantages over their linear counterparts. A circular array provides spatial discrimination over the entire 360° azimuth range, which is particularly important for applications that require front-to-back signal enhancement, such as teleconferencing. Furthermore, a circular array geometry allows for more compact designs. While the contents of this paper apply generally to planar array geometries, the circular geometry is used throughout the simulation portion.

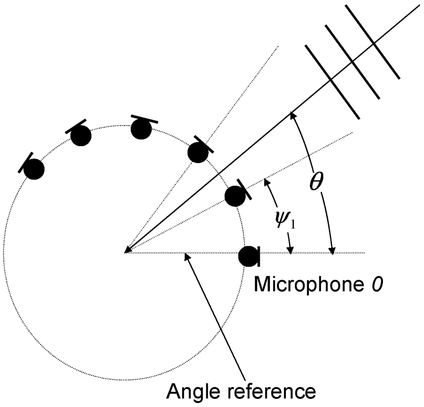

Fig. 1. Circular array geometry.

Section II presents the signal propagation model in planar (i.e., circular) arrays and serves as the foundation for the remainder of the paper. Section III reviews the role of the traditional, nonparameterized spatial correlation matrix in narrowband DOA estimation, and shows how the parameterized version of the spatial correlation matrix allows for generalization to broadband signals. Section IV describes the existing and proposed broadband spatial spectral estimators in terms of the parameterized spatial correlation matrix. Section V outlines the simulation model employed throughout this paper and evaluates the performance of all spatial spectral estimators and TDOA-based methods in both reverberation- and noise-limited environments. Concluding statements are given in Section VI.

The spatial spectral estimation approach to DOA estimation has limitations in certain reverberant environments. If an interfering signal or reflection arrives at the array with a higher energy than the direct-path signal, the DOA estimate will be false, even though the spatial spectral estimate is accurate. Such situations arise when the source is oriented towards a reflective barrier and away from the array. This problem is beyond the scope of this paper and is not addressed herein. Rather, the focus of this paper is on the evaluation of spatial spectral estimators in noisy and reverberant environments and on their application to DOA estimation.

II. Signal Model

Assume a planar array of $L+1$ elements in a 2-D geometry, shown in Fig. 1 (i.e., circular geometry), whose outputs are denoted by $x_l[n], l = 0, 1, \ldots, L$, where n is the time index. Denoting the azimuth angle of arrival by θ, propagation of the signal from a far-field source to microphone l is modeled as:

$$x_l[n] = \alpha_l s\left[n - t - f_l(\theta)\right] + v_l[n] \tag{1}$$

where α_l, $l = 0, 1, 2, \ldots, L$, are the attenuation factors due to channel effects, t is the propagation time, in samples, from the unknown source $s[n]$ to microphone 0, $v_l[n]$ is an additive noise signal at the lth microphone, and $f_l(\theta)$, $l = 0, 1, \ldots, L$, is the relative delay between microphones 0 and l. In matrix form, the array signal model becomes:

$$\begin{bmatrix} x_0[n] \\ x_1[n] \\ x_2[n] \\ \vdots \\ x_L[n] \end{bmatrix} = \begin{bmatrix} \alpha_0 & 0 & 0 & \cdots & 0 \\ 0 & \alpha_1 & 0 & \cdots & 0 \\ 0 & 0 & \alpha_2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \alpha_L \end{bmatrix} \begin{bmatrix} s[n-t] \\ s[n-t-f_1(\theta)] \\ s[n-t-f_2(\theta)] \\ \vdots \\ s[n-t-f_L(\theta)] \end{bmatrix} + \begin{bmatrix} v_0[n] \\ v_1[n] \\ v_2[n] \\ \vdots \\ v_L[n] \end{bmatrix}. \quad (2)$$

The function f_l relates the angle of arrival to the relative delays between microphone elements 0 and l, and is derived for the case of an equispaced circular array in the following manner. When operating in the far-field, the time delay between microphone l and the center of the array is given by [23]

$$g_l(\theta) = rc^{-1}\cos(\theta - \psi_l) \quad (3)$$

where the azimuth angle (relative to the selected angle reference) of the lth microphone is denoted by $\psi_l = 2\pi l/(L+1)$, $\psi_0 = 0$, r denotes the array radius, and c is the speed of signal propagation. It easily follows that

$$f_l(\theta) = g_0(\theta) - g_l(\theta) = rc^{-1}\left[\cos\theta - \cos\left(\theta - \frac{2\pi l}{L+1}\right)\right]. \quad (4)$$

It is also worth mentioning that the additive noise $v_l[n]$ may be temporally correlated with the desired signal $s[n]$. In that case, a reverberant environment is modeled. The anechoic environment is modeled by making the additive noise temporally uncorrelated with the source signal. In either case, the additive noise may be spatially correlated across the sensors.

It should also be stated that the signal model presented above makes use of the far-field assumption, in that the incoming wave is assumed to be planar, such that all sensors perceive the same DOA. An error is incurred if the signal source is actually located in the near-field; in that case, the relative delays are also a function of the range. In the most general case (i.e., a source in the near-field of a 3-D geometry), the function f_l takes three parameters: the azimuth, range, and elevation. This paper focuses on a specific subset of this general model: a source located in the far-field with only a slight elevation, such that a single parameter suffices. This is commonly the case in a teleconferencing environment. Nevertheless, the concepts of this paper, although presented in far-field planar context, easily generalize to the near-field spherical case by including the range and elevation in the forthcoming parametrization.

III. Parameterized Spatial Correlation Matrix

In narrowband signal applications, a common space-time statistic is that of the spatial correlation matrix [19], which is given by

$$\mathbf{R} = E\{\mathbf{x}[n]\mathbf{x}^H[n]\} \quad (5)$$

where

$$\mathbf{x}[n] = [x_0[n] \quad x_1[n] \quad \cdots \quad x_L[n]]^T \quad (6)$$

the superscript ^{H} denotes conjugate transpose, as complex signals are commonly used in narrowband applications, and ^{T} denotes the transpose of a matrix or vector. To steer these array outputs to a particular DOA, one applies a complex weight to each sensor output, whose phase performs the steering, and then sums the sensor outputs to form the output beam. Now, if the input signal is no longer narrowband, each frequency requires its own complex weight to appropriately phase-shift the signal at that frequency. In the context of broadband spatial spectral estimation, the spatial correlation matrix may be computed at each temporal frequency, and the resulting spatial spectrum is now a function of the temporal frequency. For broadband applications, these narrowband estimates may be assimilated into a time-domain statistic, a procedure termed "focusing," which is described in [24]. The resulting structure is termed a "focused covariance matrix."

In this paper, broadband spatial spectral estimation is addressed in another manner. Instead of implementing the steering delays in the complex weighting at each sensor, the delays are actually implemented as a time-delay in the spatial correlation matrix, which is now parameterized. Thus, each microphone output is appropriately delayed before computing this parameterized spatial correlation matrix:

$$\mathbf{x}_\phi[n] = [x_0[n] \quad x_1[n+f_1(\phi)] \quad \cdots \quad x_L[n+f_L(\phi)]]^T \quad (7)$$

and real signals are assumed from this point on. The delays are a function of the assumed azimuth DOA, which becomes the parameter. The parameterized spatial correlation matrix is formally written as shown by (8) and (9) at the bottom of the page. The matrix $\mathbf{R}_\phi$ is not simply the array observation matrix, as is commonly used in narrowband beamforming models. Instead, it is a parameterized correlation matrix that represents the signal

$$\begin{aligned}\mathbf{R}_\phi &= E\{\mathbf{x}_\phi[n]\mathbf{x}_\phi^T[n]\} \quad (8) \\ &= \begin{bmatrix} E\{x_0^2[n]\} & E\{x_0[n]x_1[n+f_1(\phi)]\} & \cdots & E\{x_0[n]x_L[n+f_L(\phi)]\} \\ E\{x_1[n+f_1(\phi)]x_0[n]\} & E\{x_1^2[n]\} & \cdots & E\{x_1[n+f_1(\phi)]x_L[n+f_L(\phi)]\} \\ \vdots & \vdots & \ddots & \vdots \\ E\{x_L[n+f_L(\phi)]x_0[n]\} & E\{x_L[n+f_L(\phi)]x_1[n+f_1(\phi)] & \cdots & E\{x_L^2[n]\} \end{bmatrix} \quad (9)\end{aligned}$$

powers across the array emanating from azimuth ϕ. Each off-diagonal entry in the matrix $\mathbf{R}_\phi$ is a single cross-correlation term and a function of the azimuth angle ϕ. Notice that the various microphone pairs are combined in a *joint* fashion, in that altering the steering angle ϕ affects all off-diagonal entries of $\mathbf{R}_\phi$. This property allows for the more prudent combining of microphone measurements as compared to the *ad hoc* method of averaging independent pairs of cross-correlation results.

This paper relates broadband spatial spectral estimators in terms of the parameterized spatial correlation matrix

$$S(\phi) = f(\mathbf{R}_\phi) \tag{10}$$

where f is some estimation function, ϕ is the steered azimuth angle, and $S(\phi)$ is the estimate of the broadband spatial spectrum at azimuth angle ϕ.

The DOA estimate follows directly from the spatial spectrum, in that peaks in the spectrum correspond to assumed source locations. For the case of a single source, which is the case throughout this paper, the estimate of the source's DOA is given by

$$\hat{\theta} = \arg\max_\phi S(\phi) = \arg\max_\phi f(\mathbf{R}_\phi) \tag{11}$$

where $\hat{\theta}$ is the DOA estimate.

Note that this broadband extension is not without caveats: care must be taken when spacing the microphones to ensure that spatial aliasing [2] does not result.

It is also important to point out that the GCC method is quite compatible with DOA estimation based on the parameterized spatial correlation matrix—the cross-correlation estimates that comprise the matrix may be computed in the frequency-domain using a GCC variant such as the phase transform (PHAT) [4]. This paper focuses on how to extract the DOA estimate from the parameterized spatial correlation matrix; the ideas presented are general in that they do not hinge on any particular method for computing the actual cross correlations.

IV. Broadband Spatial Spectral Estimators

The following subsections detail the existing and proposed broadband spatial spectral estimation methods, relating each to the parameterized spatial correlation matrix.

A. *Steered Conventional Beamforming and the SRP Algorithm*

The aim of a DSB is to time-align the received signals in the array aperture, such that the desired signal is coherently summed, while signals from other directions are incoherently summed and thus attenuated. Using the model of Section II, the output of a DSB steered to an angle of arrival of ϕ is given as

$$z_\phi[n] = \sum_{l=0}^{L} w_{\phi,l} x_l [n + f_l(\phi)]. \tag{12}$$

The delays $f_l(\phi)$ steer the beamformer to the desired DOA, while the beamformer weights $w_{\phi,l}$ help shape the beam accordingly. The weights here have been made dependent on the desired angle of arrival ϕ, for a reason that will become apparent in future subsections. In (12), the received signals are delayed [or advanced, depending on the sign of $f_l(\phi)$], by an amount that takes into account the array geometry, via the function f_l.

The estimate of the spatial spectral power at azimuth angle ϕ is given by the power of the beamformer output when steered to azimuth ϕ. Therefore, to form the entire spectrum, one needs to steer the beam and compute the output power across the entire azimuth space.

The steered-beamformer spectral estimate is given by

$$S^{\mathrm{DSB}}(\phi) = E\left\{z_\phi^2[n]\right\}. \tag{13}$$

Substitution of (12) into (13) leads to

$$S^{\mathrm{DSB}}(\phi) = \sum_{l_1=0}^{L}\sum_{l_2=0}^{L} w_{\phi,l_1} w_{\phi,l_2} E\left\{x_{l_1}[n + f_{l_1}(\phi)]\, x_{l_2}[n + f_{l_2}(\phi)]\right\}. \tag{14}$$

Expression (14) may be written more neatly in matrix notation as

$$S^{\mathrm{DSB}}(\phi) = \mathbf{w}_\phi^T \mathbf{R}_\phi \mathbf{w}_\phi \tag{15}$$

where

$$\mathbf{w}_\phi = \begin{bmatrix} w_{\phi,0} & w_{\phi,1} & \cdots & w_{\phi,L} \end{bmatrix}^T. \tag{16}$$

The DOA estimate is thus given by

$$\hat{\theta} = \arg\max_\phi \mathbf{w}_\phi^T \mathbf{R}_\phi \mathbf{w}_\phi. \tag{17}$$

The maximization of a steered beamformer output power is equivalent to maximizing a quadratic of the beamformer weight vector with respect to the angle of arrival. Altering the angle affects the parameter in the quadratic form, namely, the parameterized spatial correlation matrix.

The well-known SRP algorithm [10] follows directly from a special case of (17), where $\mathbf{w}_\phi = \mathbf{1}$ for all ϕ, and $\mathbf{1}$ is a vector of $L+1$ ones:

$$\hat{\theta}_{\mathrm{SRP}} = \arg\max_\phi \mathbf{1}^T \mathbf{R}_\phi \mathbf{1}. \tag{18}$$

For this special case of fixed unit weights, this means that the maximization of the power of a steered DSB is equivalent to the maximization of the sum of the entries of $\mathbf{R}_\phi$.

The SRP algorithm has garnered significant attention recently: see [10], [25], and [26]. In all of these implementations, the weighting of $\mathbf{w}_\phi = \mathbf{1}$ is used, which is fixed with respect to both the data and the steering angle. Given the well-known classical results on the advantages of adaptive beamforming over fixed beamforming, it is therefore surprising that adaptive weighting schemes have not been investigated more in the context of DOA estimation based on the parameterized spatial correlation matrix (A fixed weighting scheme is proposed in [27]). Notice that from (15), this is an effectively "narrowband" weight selection, in that the pre-aligning of the microphones requires only the selection of a single weight per channel. Note, however, that this weight selection must be performed for all angles ϕ. To that end, the following section presents one such adaptive weighting scheme, proposed by Krolik [14].

B. Minimum Variance

The minimum variance approach to spatial spectral estimation involves selecting weights that pass a signal [i.e., a broadband plane wave $s(n)$] propagating from azimuth ϕ with unity gain, while minimizing the total output power, given by $\mathbf{w}_\phi^T \mathbf{R}_\phi \mathbf{w}_\phi$. The application of the minimum variance method to broadband spatial spectral estimation is given in [14].

The unity gain constraint proposed by [14] is

$$\mathbf{w}_\phi^T \mathbf{1} = 1. \tag{19}$$

and the $\mathbf{1}$ vector follows from the fact that the signal is already time-aligned across the array before minimum variance processing. It is as if the signal is coming from the broadside of a linear array.

Using the method of Lagrange multipliers in conjunction with the cost function $\mathbf{w}_\phi^T \mathbf{R}_\phi \mathbf{w}_\phi$, the minimum variance weights become

$$\mathbf{w}_{\mathrm{mv},\phi} = \frac{\mathbf{R}_\phi^{-1}\mathbf{1}}{\mathbf{1}^T \mathbf{R}_\phi^{-1} \mathbf{1}}. \tag{20}$$

The resulting minimum variance spatial spectral estimate is found by substituting the weights of (20) into the cost function:

$$S^{\mathrm{mv}}(\phi) = \mathbf{w}_{\mathrm{mv},\phi}^T \mathbf{R}_\phi \mathbf{w}_{\mathrm{mv},\phi} = \left(\mathbf{1}^T \mathbf{R}_\phi^{-1} \mathbf{1}\right)^{-1}. \tag{21}$$

The broadband minimum variance DOA estimator is thus given by

$$\hat{\theta}_{\mathrm{mv}} = \arg\max_\phi \left(\mathbf{1}^T \mathbf{R}_\phi^{-1} \mathbf{1}\right)^{-1}. \tag{22}$$

The next section presents a new idea: the eigenanalysis of the parameterized spatial correlation matrix.

C. Eigenanalysis of the Parameterized Spatial Correlation Matrix

Using the signal model of Section II, notice that when the steered azimuth ϕ matches the actual azimuth θ, the parameterized spatial correlation matrix may be decomposed into signal and noise components in the following manner:

$$\mathbf{R}_\theta = \sigma_s^2 \boldsymbol{\alpha}\boldsymbol{\alpha}^T + E\{\, \mathbf{v}_\theta \mathbf{v}_\theta^T \}\tag{23}$$

where σ_s^2 is the signal power

$$\boldsymbol{\alpha} = [\alpha_0 \quad \alpha_1 \quad \cdots \quad \alpha_L]^T \tag{24}$$

and

$$\mathbf{v}_\theta = [v_0[n] \quad v_1[n + f_1(\theta)] \quad \cdots \quad v_L[n + f_L(\theta)]]^T. \tag{25}$$

Note that it has been implicitly assumed that the desired signal is wide-sense stationary, zero-mean, and temporally uncorrelated with the additive noise. Consider only the signal component of $\mathbf{R}_\theta$. It may be easily shown that this matrix has one nonzero eigenvector, that eigenvector being $\boldsymbol{\alpha}$, with the corresponding eigenvalue being $\sigma_s^2 \|\boldsymbol{\alpha}\|^2$. The vector of attenuation constants $\boldsymbol{\alpha}$ is generally unknown; however, from the above discussion, it is apparent that the vector may be estimated from the eigenanalysis of $\mathbf{R}_\theta$.

To that end, consider another adaptive weight selection method, which follows from the ideas of narrowband beamforming [19]. This weight selection attempts to nontrivially maximize the output energy of the steered-beamformer for a given azimuth ϕ

$$\mathbf{e}_{\max,\phi} = \arg\max_{\mathbf{w}_\phi} \mathbf{w}_\phi^T \mathbf{R}_\phi \mathbf{w}_\phi \tag{26}$$

subject to

$$\mathbf{w}_\phi^T \mathbf{w}_\phi = 1. \tag{27}$$

It is well known that the solution to the above constrained optimization is the vector that maximizes the Rayleigh quotient [2] $\mathbf{w}_\phi^T \mathbf{R}_\phi \mathbf{w}_\phi / \mathbf{w}_\phi^T \mathbf{w}_\phi$, which is in turn given by the eigenvector corresponding to the maximum eigenvalue of $\mathbf{R}_\phi$. The resulting spatial spectral estimate is given by

$$S^{\mathrm{EIG}}(\phi) = \mathbf{e}_{\max,\phi}^T \mathbf{R}_\phi \mathbf{e}_{\max,\phi} = \lambda_{\max,\phi} \tag{28}$$

where $\lambda_{\max,\phi}$ is the maximum eigenvalue of $\mathbf{R}_\phi$, and $\mathbf{e}_{\max,\phi}$ is the corresponding eigenvector. The DOA estimation involves searching for the angle that produces the largest maximum eigenvalue of $\mathbf{R}_\phi$:

$$\hat{\theta}_{\mathrm{EIG}} = \arg\max_\phi \lambda_{\max,\phi}. \tag{29}$$

In addition to producing another spatial spectrum estimate, the above eigenanalysis allows one to estimate: $\boldsymbol{\alpha}$

$$\hat{\boldsymbol{\alpha}} = \mathbf{e}_{\max,\hat{\theta}_{\mathrm{EIG}}}. \tag{30}$$

Now that an estimate of the attenuation vector $\boldsymbol{\alpha}$ is available, the minimum variance method of [14] may be improved to reflect the presence of channel attenuation factors, which were omitted in the developments of Section IV-B.

D. Improved Minimum Variance

The broadband minimum variance spatial spectral estimation proposed by [14] assumes that the attenuation vector $\boldsymbol{\alpha}$ is equal to $\mathbf{1}$, or a scaled version of $\mathbf{1}$. In practice, it is not uncommon for this assumption to be violated by factors such as uncalibrated microphones, for example. To that end, the unity gain constraint proposed by [14] is modified to reflect the more general signal model of Section II.

Taking into account the channel attenuation vector $\boldsymbol{\alpha}$, the proposed unity gain constraint is

$$\sum_{l=0}^{L} w_{\phi,l}\alpha_l s[n - t - f_l(\phi) + f_l(\phi)] = s[n - t] \tag{31}$$

which may be simplified and written in vector notation as

$$\mathbf{w}_\phi^T \boldsymbol{\alpha} = 1. \tag{32}$$

Therefore, the optimal minimum variance weights become

$$\mathbf{w}_{\mathrm{mv}',\phi} = \frac{\mathbf{R}_\phi^{-1}\boldsymbol{\alpha}}{\boldsymbol{\alpha}^T \mathbf{R}_\phi^{-1} \boldsymbol{\alpha}}. \tag{33}$$

The resulting proposed minimum variance spatial spectral estimate is found by substituting the weights of (33) into the cost function

$$S^{\mathrm{mv}'}(\phi) = \mathbf{w}_{\mathrm{mv}',\phi}^T \mathbf{R}_\phi \mathbf{w}_{\mathrm{mv}',\phi} = \left(\boldsymbol{\alpha}^T \mathbf{R}_\phi^{-1} \boldsymbol{\alpha}\right)^{-1}. \tag{34}$$

The proposed broadband minimum variance DOA estimator is thus given by

$$\hat{\theta} = \arg\max_{\phi} \left(\boldsymbol{\alpha}^T \mathbf{R}_\phi^{-1} \boldsymbol{\alpha}\right)^{-1}. \tag{35}$$

E. Linear Spatial Prediction and the Multichannel Cross-Correlation Coefficient

Spatial spectral estimation using linear prediction is well defined for the case of narrowband signals, as the narrowband assumption allows one to write one of the microphone outputs as a complex-weighted linear combination of the other microphone outputs [2]. To extend this idea to the broadband case, the same method as that of the previous sections is used, in that the time delay is applied prior to computing the predictive coefficients.

This concept was first presented in [17] and [18] in the context of time delay estimation; the approach was limited to linear array geometries, and yielded only a single relative delay. This section generalizes the idea to planar array geometries, transforming the problem from time delay estimation to DOA estimation.

The idea is to predict, using real predictive coefficients, the output of $x_0[n]$ using a linear combination of $x_l[n + f_l(\phi)]$, $l = 1, 2, \ldots, L$. Using a spatial autoregressive (AR) model, the linear predictive framework is given by

$$x_0[n] = \sum_{l=1}^{L} a_{\phi,l} x_l[n + f_l(\phi)] + e[n] \tag{36}$$

where $e[n]$ may be interpreted as either the spatially white noise that drives the AR model, or the prediction error. For each ϕ in the azimuth space, one finds the weight vector

$$\mathbf{a}_\phi = \begin{bmatrix} -a_{\phi,0} & a_{\phi,1} & \cdots & a_{\phi,L} \end{bmatrix}^T \tag{37}$$

which minimizes the criterion

$$J_\phi = E\left\{ \left| -\mathbf{a}_\phi^T \mathbf{x}_\phi[n] \right|^2 \right\} \tag{38}$$

subject to the constraint

$$\delta^T \mathbf{a}_\phi = -1 \tag{39}$$

where

$$\delta = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T. \tag{40}$$

Using the method of Lagrange multipliers, the optimal predictive weights are given by

$$\mathbf{a}_{\mathrm{opt},\phi} = -\frac{\mathbf{R}_\phi^{-1}\delta}{\delta^T \mathbf{R}_\phi^{-1} \delta} \tag{41}$$

and the resulting minimum mean-squared error (mmse) is

$$J_{\min,\phi} = \frac{1}{\delta^T \mathbf{R}_\phi^{-1} \delta}. \tag{42}$$

Note that both the optimal predictive coefficients and the mmse are a function of the steered angle ϕ.

The classical approach to spectral estimation using linear prediction is to map the optimal predictive coefficients to an AR transfer function. However, it is well known that this method is very sensitive to the presence of additive noise in the observations [2]. This is because the AR model breaks down when additive noise is present. To that end, a more robust implementation of linear spatial prediction is proposed in [17] and [18]. The idea is to not estimate an AR spectrum, but rather to find the parameter (i.e., the angle ϕ) that minimizes the prediction error.

In [17] and [18], the idea of linear spatial prediction was used to derive the (time delay parameterized) multichannel cross correlation coefficient (MCCC) in the context of linear array time delay estimation. These ideas are now extended to planar array geometries, and the azimuth angle-parameterized MCCC is presented as another broadband spatial spectral estimator.

The matrix $\mathbf{R}_\phi$ may be factorized as [17], [18]:

$$\mathbf{R}_\phi = \mathbf{D}\widetilde{\mathbf{R}}_\phi \mathbf{D} \tag{43}$$

where

$$\mathbf{D} = \begin{bmatrix} \sqrt{E\{x_0^2[n]\}} & 0 & \cdots & 0 \\ 0 & \sqrt{E\{x_1^2[n]\}} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \sqrt{E\{x_L^2[n]\}} \end{bmatrix} \tag{44}$$

is a diagonal matrix

$$\widetilde{\mathbf{R}}_\phi = \begin{bmatrix} 1 & \rho_{\phi,01} & \cdots & \rho_{\phi,0L} \\ \rho_{\phi,01} & 1 & \cdots & \rho_{\phi,1L} \\ \vdots & \ddots & \ddots & \vdots \\ \rho_{\phi,0L} & \cdots & \rho_{\phi,L-1L} & 1 \end{bmatrix} \tag{45}$$

is a symmetric matrix, and

$$\rho_{\phi,kl} = \frac{E\{x_k[n + f_k(\phi)] x_l[n + f_l(\phi)]\}}{\sqrt{E\{x_k^2[n]\} E\{x_l^2[n]\}}}, \quad k, l = 0, 1, \ldots, L \tag{46}$$

is the cross-correlation coefficient between $x_k[n + f_k(\phi)]$ and $x_l[n + f_l(\phi)]$.

The azimuth-angle dependent mmse may be written using (43) as

$$J_{\min,\phi} = E\left\{x_0^2[n]\right\} \frac{\det\left(\widetilde{\mathbf{R}}_\phi\right)}{\det\left(\widetilde{\mathbf{R}}_{\phi,1:L}\right)} \tag{47}$$

where $\widetilde{\mathbf{R}}_{\phi,1:L}$ is the submatrix formed by removing the first row and column from $\widetilde{\mathbf{R}}_{\phi}$, and det stands for "determinant." It is shown in [17] and [18] that

$$\det\left(\widetilde{\mathbf{R}}_{\phi}\right) \leq \det\left(\widetilde{\mathbf{R}}_{\phi,1:L}\right) \leq \cdots \leq 1 \tag{48}$$

and thus the following relationship is established:

$$\det\left(\widetilde{\mathbf{R}}_{\phi}\right) \leq \frac{J_{\phi,\min}}{E\left\{x_0^2[n]\right\}} \leq 1. \tag{49}$$

From this relationship, it is easily observed that minimizing the spatial prediction error corresponds to minimizing the quantity $\det\left(\widetilde{\mathbf{R}}_{\phi}\right)$. Notice that $\det\left(\widetilde{\mathbf{R}}_{\phi}\right) = 0$ when every entry of $\widetilde{\mathbf{R}}_{\phi}$ is equal to unity (i.e., perfectly correlated microphone signals). Conversely, in the case of mutually uncorrelated microphone outputs, $\det\left(\widetilde{\mathbf{R}}_{\phi}\right) = 1$. Putting all of this together, the azimuth angle parameterized MCCC is defined as

$$\rho_{\phi}^2 = 1 - \det\left(\widetilde{\mathbf{R}}_{\phi}\right) = 1 - \det\left(\mathbf{D}^{-1}\mathbf{R}_{\phi}\mathbf{D}^{-1}\right). \tag{50}$$

The MCCC broadband spatial spectral estimate is given by

$$S^{\mathrm{MCCC}}(\phi) = \rho_{\phi}^2 \tag{51}$$

from which the DOA estimation easily follows as

$$\hat{\theta}_{\mathrm{MCCC}} = \arg\max_{\phi} \rho_{\phi}^2. \tag{52}$$

It is interesting to note that even though the linear spatial predictive approach is used here to arrive at the azimuth parameterized MCCC estimator, maximizing the MCCC actually corresponds more closely to the minimization of the joint entropy of the received signals [28], assuming that the signals are jointly Gaussian distributed. This follows from the fact that for jointly Gaussian distributed $\mathbf{x}_{\phi}[n]$, the joint entropy of $\mathbf{x}_{\phi}[n]$ is directly proportional to $\det(\mathbf{R}_{\phi})$ [28].

V. Simulation Evaluation

A. *Simulation Environment*

The various broadband spatial spectral estimators are evaluated in a computer simulation. An equispaced circular array of three to ten omnidirectional microphones is employed as the spatial aperture. The radius of the array is chosen as the distance that fulfills the spatial aliasing equality for circular arrays. In other words, the array radius is made as large as possible without suffering from spatial aliasing [23]

$$r = \frac{\frac{c}{f_{\max}}}{4\sin\frac{\pi}{L+1}} \tag{53}$$

where $f_{\max}$ denotes the highest frequency of interest, and is chosen to be 4 kHz in the simulations. For a ten-element circular array, the array radius becomes 6.9 cm. The signal sources are omnidirectional point sources. This means that the direct-path component is stronger than any individual reflected component—as mentioned in the Introduction, it is beyond the scope of this paper to handle cases where due to source directivity and orientation, a reflected component contains more energy than the direct-path component.

A reverberant acoustic environment is simulated using the image model method [29]. The simulated room is rectangular with plane reflective boundaries (walls, ceiling, and floor). Each boundary is characterized by a frequency-independent uniform reflection coefficient which does not vary with the angle of incidence of the source signal.

The room dimensions in centimeters are (304.8, 457.2, 381). The circular array is located in the center of the room: the center of the array sits at (152.4, 228.6, 101.6). Two distinct scenarios are simulated, as described below.

The speaker is immobile and situated at (254, 406.4, 101.6) and (254, 406.4, 152.4) in the first and second simulation scenarios, respectively. The immobility of the source means that the evaluation does not consider frames during which the source exhibits movement. The correct azimuth angle of arrival is 60°. The distance from the center of the array to the source is 204.7 cm.

The SNR at the microphone elements is 0 dB. Here, SNR refers to spatially white sensor noise in the first scenario and spherically isotropic (diffuse) noise in the second scenario. The generation of spherically isotropic noise is performed by transforming a vector of uncorrelated Gaussian random variables into a vector of correlated (i.e., according to a given covariance matrix) Gaussian random variables by premultiplying the original (uncorrelated) vector with the Cholesky factorization [30] of the covariance matrix of a diffuse noise field [2]. The covariance matrix of the diffuse noise field is computed by averaging over the entire frequency range (300–4000 Hz). For the computation of the SNR, the signal component includes reverberation. In terms of reverberation, three levels are simulated for each scenario: anechoic, moderately reverberant, and highly reverberant. The reverberation times are measured using the reverse-time integrated impulse response method of [31]. The frequency-independent reflection coefficients of the walls and ceiling are adjusted to achieve the desired level of reverberation: a 60-dB reverberation decay time of 300 ms for the moderately reverberant case, and 600 ms for the highly reverberant case.

In the first simulation scenario, the microphones are all perfectly calibrated with unity gains. In the second simulation scenario, the presence of uncalibrated microphones is simulated, by setting $\alpha_l, l = 0, 1, \ldots, L$ to a uniformly distributed random number over the range (0.2, 1).

The source signal is convolved with the synthetic impulse responses. Appropriately scaled temporally white Gaussian noise is then added at the microphones to achieve the required SNR. The microphone outputs are filtered to the 300–4000-Hz range prior to processing.

Two signal types are examined for each scenario: stationary white Gaussian noise and nonstationary female English speech. The DOA estimates are computed once per 128-ms frame. To achieve good angular resolution, the sampling rate is chosen to be 48 kHz, resulting in frames of $N = 6144$ samples each. A simulation run consists of $K = 890$ frames.

Throughout the simulations, the parameterized spatial correlation matrix is estimated once per frame by time-averaging

$$\hat{\mathbf{R}}_\phi = \frac{1}{N}\sum_{n=0}^{N-1} \mathbf{x}_\phi[n]\,\mathbf{x}_\phi^T[n] \tag{54}$$

where $\hat{\mathbf{R}}_\phi$ is the time-averaged estimate of the correlation matrix $\mathbf{R}_\phi$. For all algorithms, the spatial spectrum is estimated every degree over the range (0, 359) degrees.

The planar signal model assumed in Section II results in elevation ambiguity in a three-dimensional environment. The simulations do not take elevation into account, even though the height of the source (i.e., 1.52 m) does not equal the height of the planar array (i.e., 1.02 m) in the second simulation scenario. Notice that in a teleconferencing environment, the sources are commonly only slightly elevated compared to the array.

B. Performance Criteria

The estimated spatial spectra are plotted to observe mainlobe width and background values. These spectra are averaged over the K frames and normalized such that the peak of the averaged spectrum is 0 dB.

For each simulation, the algorithms are also evaluated from a DOA estimation standpoint using the percentage of anomalies (estimates that differ from the actual angle of arrival by more than 5°), and root-mean-square (rms) error measure for the nonanomalous estimates

$$\mathrm{e_{rms}} = \sqrt{\frac{1}{K_{\mathrm{na}}}\sum_{k\in\chi_{\mathrm{na}}}\left[\left|\hat{\theta}_k - \theta\right|'\right]^2} \tag{55}$$

where χ_{na} is the set of all nonanomalous estimates, K_{na} is the number of elements in χ_{na}, and the prime operator is included to take into account the cyclicity of the angular space

$$\left|\hat{\theta}_k - \theta\right|' = \begin{cases} \left|\hat{\theta}_k - \theta\right|, & \text{if } \left|\hat{\theta}_k - \theta\right| \le \pi \\ 2\pi - \left|\left(\hat{\theta}_k - \theta\right)\right|, & \text{if } \left|\hat{\theta}_k - \theta\right| > \pi \end{cases}. \tag{56}$$

C. TDOA Comparison Algorithm

The DOA estimation performance of the broadband spatial spectral estimators is compared to that of a standard two-step TDOA algorithm which consists of computing unweighted cross correlations in the first-step, and least-squares DOA estimation (translation of relative delays to DOA) in the second step. Formally, the algorithm computes the time delay between microphone 0 and microphone l for $l = 1, \ldots, L$ as

$$\tau_{0l} = \arg\max_\tau E\{x_0[n]x_l[n+\tau]\} \tag{57}$$

and then translates these relative delays to the azimuth angle of arrival using the least-squares criterion [32], [33]

$$\hat{\theta}_{\mathrm{TDOA}} = \arg\min_\phi \sum_{l=1}^{L}\left|\frac{c\cdot\tau_{0l}}{r} - \cos\phi + \cos\left(\phi - \frac{2\pi l}{L+1}\right)\right|^2. \tag{58}$$

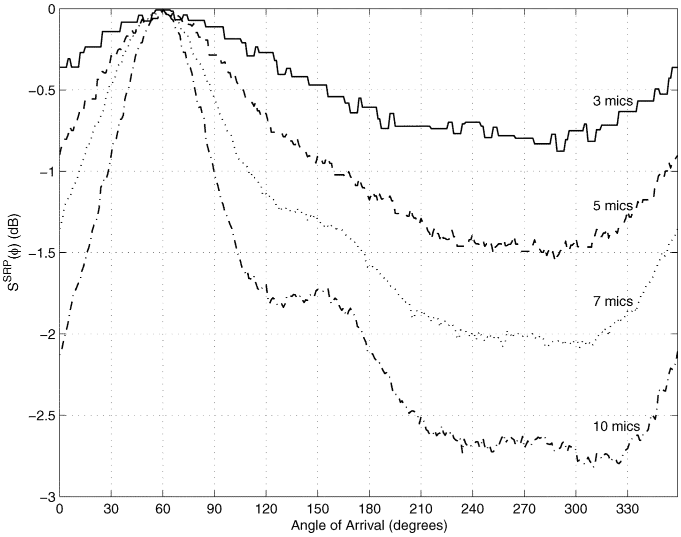

Fig. 2. SRP spatial spectral estimates for simulation scenario 1, with speech signal and moderate reverberation.

The computational requirements of TDOA-based algorithms are lower than that of the methods presented in this paper. Note that both approaches (TDOA-based and spectral estimation-based) use the cross-correlation measurements for each microphone pair and these cross correlations are computed over the range of physically realizable relative delays. The difference between the approaches lies in the manner in which these measurements are utilized. TDOA-based approaches simply compute the lag which produces the peak in cross correlation for each microphone pairing and use these optimal lags to arrive at an estimate of the DOA via an intersection or least-squares procedure. On the other hand, spatial spectral estimaton-based methods use *all* cross-correlation measurements (i.e., not just the peak values and their lag arguments) to form the parameterized spatial correlation matrix at each DOA, followed by a search procedure over the DOA space that identifies the peak in the spectrum. The computational requirements of this search process are a disadvantage of spatial spectral-based methods.

D. Discussion

Consider first Figs. 2–5. From these figures, it is evident that adding extra microphones improves the resolution of all broadband spatial spectra. The main lobe is narrowed, and the background level, which corresponds to the power of the reverberant field, is lowered. Thus, it is inferred that the addition of microphones combats the effects of reverberation as well as noise. A lower reverberant field level decreases the probability of anomalies.

From Figs. 6–8, it is evident that the MCCC spatial spectral estimate provides the narrowest main lobe and lowest background level. As the level of reverberation increases, however, the resolution gain of the MCCC method decreases. As the microphones are calibrated in scenario 1, the MV and MV' methods yield very similar spatial spectra, which are only slightly below that of the SRP spectrum—the difference is slight because in a spatially white noise field, the minimum

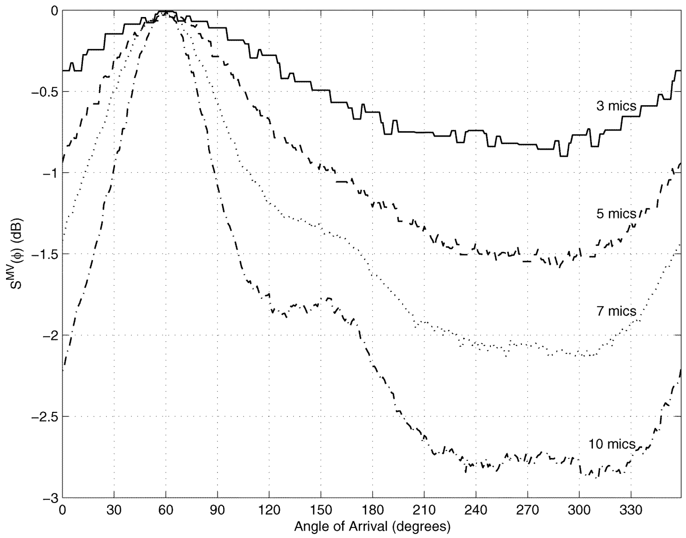

Fig. 3. Minimum variance (MV) spatial spectral estimates for simulation scenario 1, with speech signal and moderate reverberation.

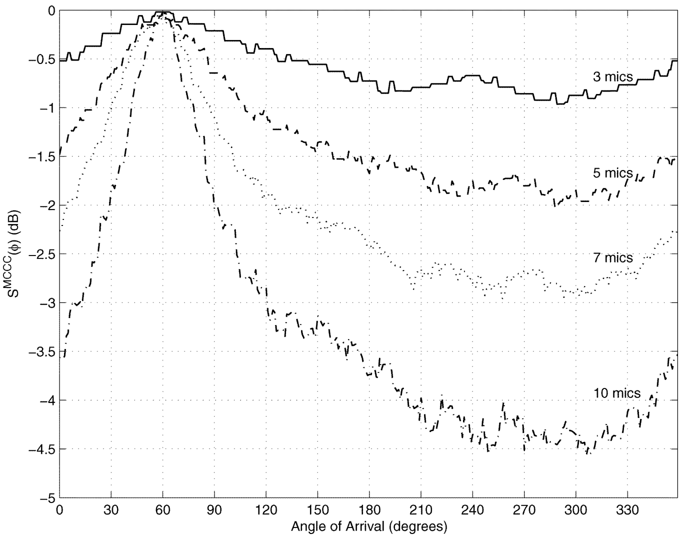

Fig. 5. MCCC spatial spectral estimates for simulation scenario 1, with speech signal and moderate reverberation.

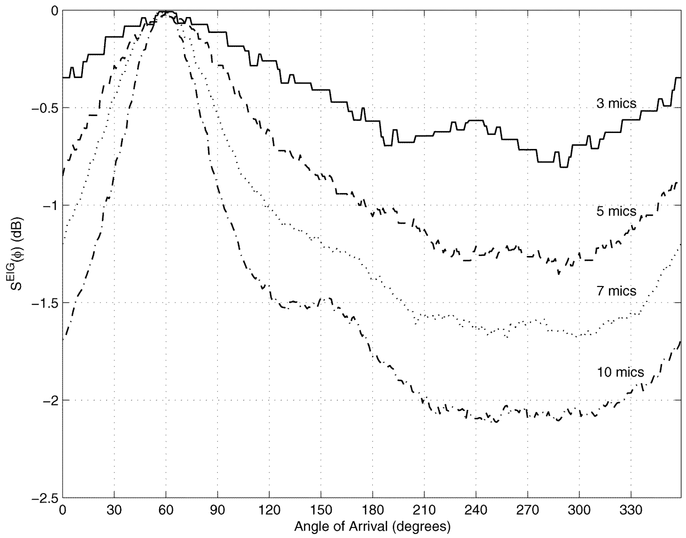

Fig. 4. Eigenvalue spatial spectral estimates for simulation scenario 1, with speech signal and moderate reverberation.

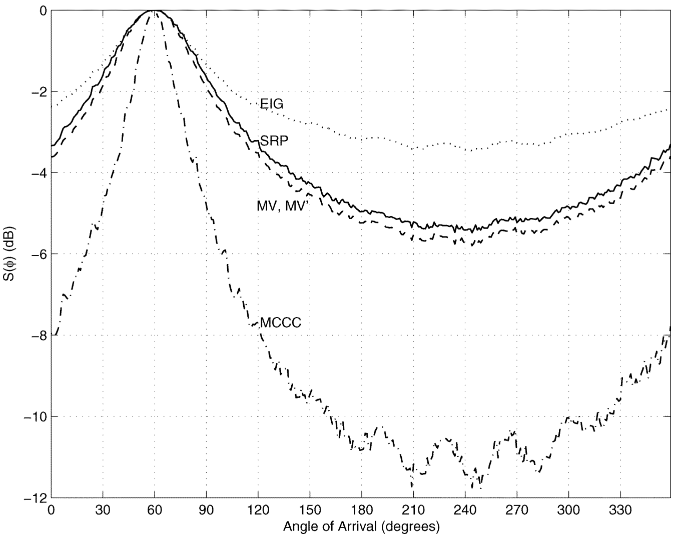

Fig. 6. Spatial spectral estimates for simulation scenario 1, with speech signal and no reverberation.

variance weights of (20) and (33) are simply uniform (as in the SRP spectrum). Notice also that for a spatially white noise field, parametrization of the matrix $\mathbf{R}_\phi$ does not alter the noise statistics: the time-shifting alters the off-diagonal elements only, which are all equal to zero with or without time-aligning. The eigenvalue spectrum offers the poorest resolution; however, the main idea behind the eigenanalysis of $\mathbf{R}_\phi$ is to provide an estimate of $\boldsymbol{\alpha}$, not to act as a spatial spectral estimator. Nevertheless, the eigenvalue spectrum may be thought of as an "intermediate" spectrum.

Consider now the spatial spectral estimates from the second simulation scenario, shown in Figs. 9–11. First of all, it is evident that Krolik's minimum variance method [14] fails in the presence of uncalibrated microphones—in fact, the "minimum variance" spectrum actually shows a higher background level than the SRP spectrum. The mismatch in the steering vector leads to serious spectral distortion. The proposed minimum variance method, which utilizes the eigenvalue method's estimate of $\boldsymbol{\alpha}$, rectifies this problem, as the improved minimum variance (MV') spectrum is clearly "minimum variance." However, it is rather disappointing to note that even in a diffuse noise field, the resolution gain of the MV' spectrum over the SRP spectrum is quite small. This may seem even more surprising given classical results from beamforming theory, which note the great advantage of minimum variance distortionless response (MVDR) beamformers over fixed beamformers. The reason for this small gain stems from the nature of the extension from narrowband to broadband via the parametrization of $\mathbf{R}_\phi$. Consider only the noise component of $\mathbf{R}_\phi$—it is very easy to show that after the time-alignment (i.e., parametrization), the off-diagonal elements of the diffuse noise matrix become decorrelated, and $\mathbf{R}_\phi$ approximates the identity matrix if the

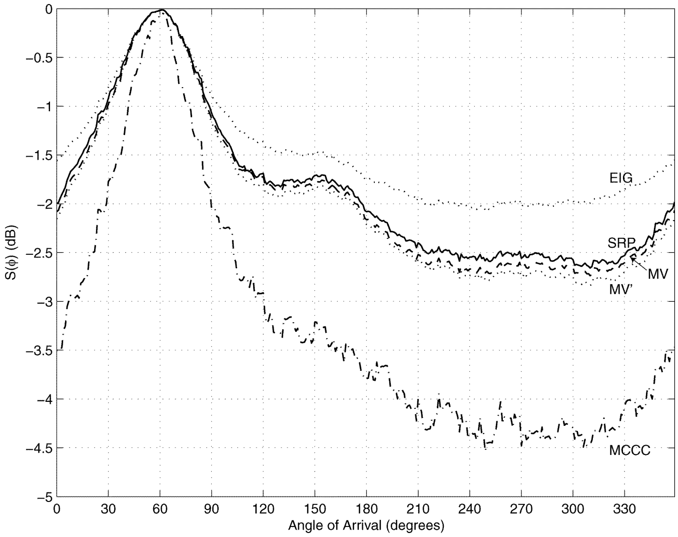

Fig. 7. Spatial spectral estimates for simulation scenario 1, with speech signal and moderate reverberation.

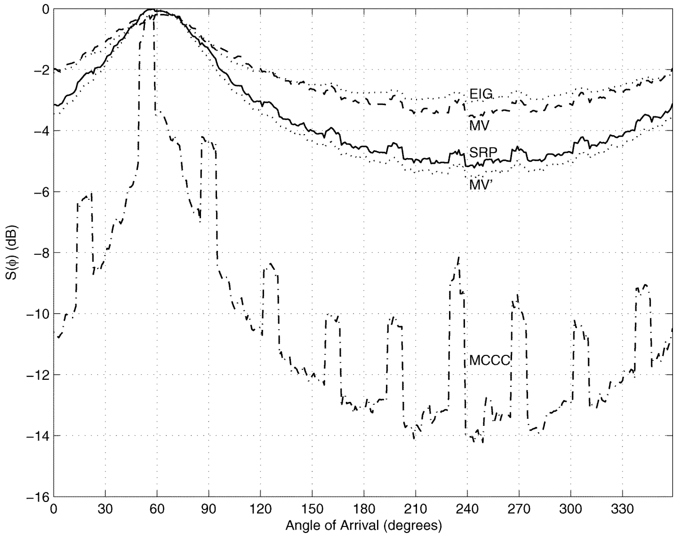

Fig. 9. Spatial spectral estimates for simulation scenario 2, with speech signal and no reverberation.

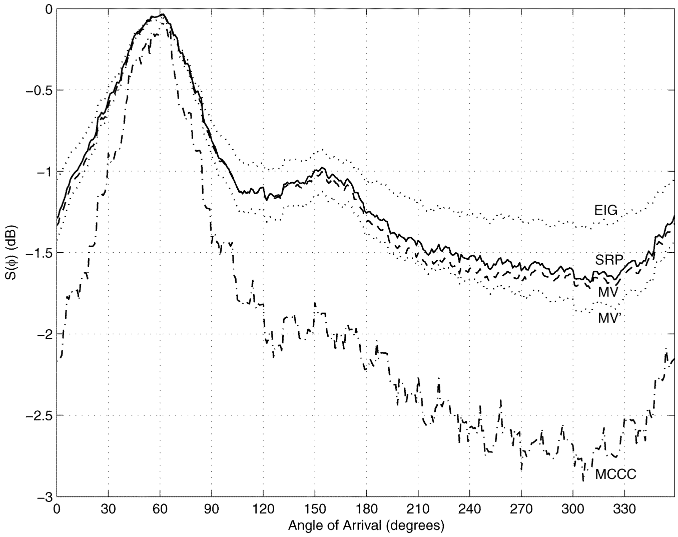

Fig. 8. Spatial spectral estimates for simulation scenario 1, with speech signal and heavy reverberation.

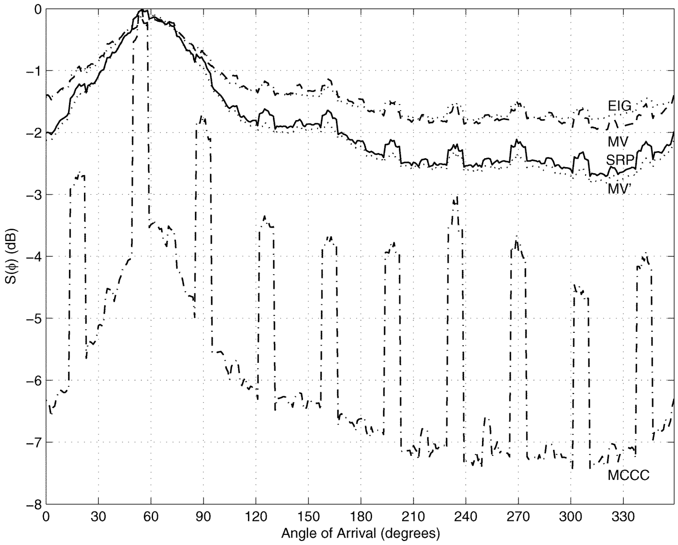

Fig. 10. Spatial spectral estimates for simulation scenario 2, with speech signal and moderate reverberation.

diffuse noise is temporally white (which is often the case with diffuse noise):

$$\mathbf{R}_\phi \,|_{\text{noise}} \approx \sigma_n^2 \mathbf{I}, \;\; \forall \phi \tag{59}$$

where σ_n^2 is the noise variance, and $\mathbf{I}$ is the identity matrix. The equation above holds strictly for circular arrays, but not linear arrays. When steering to the broadside of a linear array, there is no time-alignment necessary, and thus the spatial correlation of the diffuse noise remains. Due to the spatial decorrelation created by the parametrization, the weights of (33) are approximately equal to

$$\mathbf{w}_{\text{mv}',\phi} \approx \frac{\boldsymbol{\alpha}}{\|\boldsymbol{\alpha}\|^2}, \;\; \forall \phi. \tag{60}$$

The minimum variance coefficients vary only to the extent of the nonuniformity of the attenuation factors, and thus the MV' spectrum offers only a small benefit over the SRP spectrum. Moreover

$$\mathbf{e}_{\max,\theta} \approx \mathbf{w}_{\text{mv}',\theta}. \tag{61}$$

and thus the eigenvalue spectrum closely resembles the MV' spectrum at angles near the actual azimuth angle of arrival. Lastly, it is apparent that the MCCC spectrum is somewhat sensitive (i.e., the presence of deep ridges in spectrum) to factors such as variable microphone gains and source elevation.

It is important to analyze the relationship between spatial spectral estimation and DOA estimation—for example, it is interesting to investigate if the spatial spectral estimators that show a greater resolution (i.e., minimum variance, MCCC) lead to greater DOA estimation accuracy.

To that end, consider the findings of Tables I and II—these pertain to the DOA estimation accuracy of the various spatial

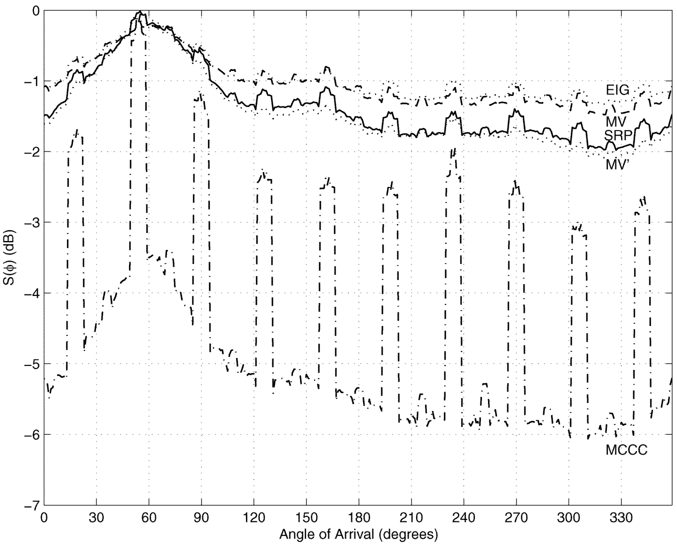

Fig. 11. Spatial spectral estimates for simulation scenario 2, with speech signal and heavy reverberation.

TABLE I
PERCENTAGE OF ANOMALIES FOR SIMULATION SCENARIO 1: SPATIALLY WHITE NOISE, NO SOURCE ELEVATION, UNIFORM MICROPHONE GAINS, TEN MICROPHONES

Parameters	**TDOA**	**SRP**	**MV**	**MV’**	**EIG**	**MCCC**
$T_{60} = 0\text{ms}, \text{SNR} = 0\text{dB}$, white signal	0	0	0	0	0	0
$T_{60} = 300\text{ms}, \text{SNR} = 0\text{dB}$, white signal	10.79	0	0	0	0	0
$T_{60} = 600\text{ms}, \text{SNR} = 0\text{dB}$, white signal	73.60	0	0	0.11	0.11	27.64
$T_{60} = 0\text{ms}, \text{SNR} = 0\text{dB}$, speech signal	4.94	0	0	0	0	0.11
$T_{60} = 300\text{ms}, \text{SNR} = 0\text{dB}$, speech signal	52.92	8.31	9.55	8.54	8.54	19.55
$T_{60} = 600\text{ms}, \text{SNR} = 0\text{dB}$, speech signal	82.13	26.52	27.75	28.88	28.88	52.02

TABLE II
RMS ERROR VALUES FOR SIMULATION SCENARIO 1: SPATIALLY WHITE NOISE, NO SOURCE ELEVATION, UNIFORM MICROPHONE GAINS, TEN MICROPHONES

Parameters	**TDOA**	**SRP**	**MV**	**MV’**	**EIG**	**MCCC**
$T_{60} = 0\text{ms}, \text{SNR} = 0\text{dB}$, white signal	0.40	0.71	0.76	0.71	0.71	0.75
$T_{60} = 300\text{ms}, \text{SNR} = 0\text{dB}$, white signal	2.54	1.42	1.37	1.45	1.45	1.44
$T_{60} = 600\text{ms}, \text{SNR} = 0\text{dB}$, white signal	2.75	1.86	1.70	2.01	2.01	2.14
$T_{60} = 0\text{ms}, \text{SNR} = 0\text{dB}$, speech signal	2.14	1.23	1.26	1.23	1.23	1.23
$T_{60} = 300\text{ms}, \text{SNR} = 0\text{dB}$, speech signal	2.80	2.57	2.60	2.57	2.57	2.41
$T_{60} = 600\text{ms}, \text{SNR} = 0\text{dB}$, speech signal	2.97	2.81	2.75	2.85	2.85	2.83

TABLE III
PERCENTAGE OF ANOMALIES FOR SIMULATION SCENARIO 2: SPHERICALLY ISOTROPIC NOISE, SOURCE ELEVATED, VARIABLE MICROPHONE GAINS, TEN MICROPHONES

Parameters	**TDOA**	**SRP**	**MV**	**MV’**	**EIG**	**MCCC**
$T_{60} = 0\text{ms}, \text{SNR} = 0\text{dB}$, white signal	4.04	0	24.94	0	0	1.01
$T_{60} = 300\text{ms}, \text{SNR} = 0\text{dB}$, white signal	29.33	0	4.94	0.34	0.34	26.07
$T_{60} = 600\text{ms}, \text{SNR} = 0\text{dB}$, white signal	65.62	0.34	14.38	7.87	7.87	58.88
$T_{60} = 0\text{ms}, \text{SNR} = 0\text{dB}$, speech signal	23.82	1.01	56.52	0.90	0.90	2.70
$T_{60} = 300\text{ms}, \text{SNR} = 0\text{dB}$, speech signal	76.52	7.64	51.69	10.34	10.34	24.38
$T_{60} = 600\text{ms}, \text{SNR} = 0\text{dB}$, speech signal	91.23	15.51	59.89	26.85	26.85	46.18

TABLE IV
RMS ERROR VALUES FOR SIMULATION SCENARIO 2: SPHERICALLY ISOTROPIC NOISE, SOURCE ELEVATED, VARIABLE MICROPHONE GAINS, TEN MICROPHONES.

Parameters	**TDOA**	**SRP**	**MV**	**MV’**	**EIG**	**MCCC**
$T_{60} = 0\text{ms}, \text{SNR} = 0\text{dB}$, white signal	2.46	2.00	2.76	1.95	1.95	2.02
$T_{60} = 300\text{ms}, \text{SNR} = 0\text{dB}$, white signal	3.13	3.91	3.69	3.46	3.46	4.91
$T_{60} = 600\text{ms}, \text{SNR} = 0\text{dB}$, white signal	2.75	4.94	4.53	4.46	4.46	5.00
$T_{60} = 0\text{ms}, \text{SNR} = 0\text{dB}$, speech signal	2.56	3.37	3.05	3.01	3.01	3.92
$T_{60} = 300\text{ms}, \text{SNR} = 0\text{dB}$, speech signal	3.08	4.66	3.59	4.38	4.38	4.93
$T_{60} = 600\text{ms}, \text{SNR} = 0\text{dB}$, speech signal	2.87	4.90	4.19	4.71	4.71	4.99

spectral estimation methods in the first simulation scenario. It is obvious from the tables that the TDOA-based method provides very poor performance in reverberant conditions, as the class of spatial spectral estimators greatly outperforms the TDOA-based approach in all but the anechoic white signal case. This lends credence to the notion that jointly utilizing multiple microphone pairs combats reverberation, not just background noise. In the TDOA two-step method, a “hard-decision” is made in the computation of each τ_{0l}, and thus if this decision is incorrect, the error is propagated to the least-squares stage. On the other hand, spatial spectral estimators do not make such hard decisions. Instead, the decision is deferred until after the contribution of all microphone pairs. As expected, the SRP, MV, and MV’ methods show virtually identical performance. The slight advantage of the SRP algorithm in the heavily reverberant speech case follows from the fact that the SRP algorithm does not suffer from the desired signal cancelation phenomenon. Furthermore, it is evident that the increased spatial resolution of the MCCC estimate does not translate to a greater DOA estimation accuracy. In fact, the MCCC spectrum exhibits anomalies in conditions where the SRP and minimum variance algorithms do not. The sensitivity of the MCCC spectrum to speech occurs due to the fact that speech is not well modeled by a Gaussian distribution, and thus, the determinant of the parameterized spatial correlation matrix no longer equals the joint entropy of the observations. This is well explained in [28].

Lastly, consider Tables III and IV. Krolik and Swingler’s minimum variance method performs very poorly from the DOA estimation standpoint with uncalibrated microphones, as expected. The improved minimum variance method corrects the shortcomings, but again, due to the spatial decorrelation of the parametrization, the minimum variance method does not lead to a greater DOA estimation accuracy than the SRP method. As the level of reverberation increases, the SRP actually outperforms the MV’ method—it is inferred that the SRP spectrum is more “robust” than the adaptively formed MV’ spectrum, in that it avoids the desired signal cancellation phenomenon that occurs when the noise is correlated with the target signal. The MCCC method performs better than the TDOA method, but significantly worse than the other spatial spectral estimators. The relationship between spatial spectral estimation and DOA estimation is also well illustrated by comparing the MV’ and eigenvalue methods: from Figs. 6–11, the background levels of the eigenvalue spectra are much higher than those of the MV’ spectra. However, because the spectra are equivalent at

$\phi \approx \theta$ from (61), the DOA estimation performance of the two methods is identical (from Tables I–IV).

Recently, experiments employing real data obtained from the IDIAP Smart Meeting Room database have been performed. In these experiments, the MCCC method was shown to yield the lowest anomaly rate among the various DOA estimators.

VI. Conclusion

The parametrization of the spatial correlation matrix generalizes narrowband spatial spectral estimation methods to the broadband environment. The parametrization spatially decorrelates noise fields, and if the noise is temporally white, the noise component of the parameterized spatial correlation matrix is simply the identity matrix. As a result, the application of minimum variance spectral estimation to the broadband spatial spectral estimation problem yields only marginal gains. On the other hand, the MCCC spatial spectrum produces a significantly higher resolution than SRP and minimum variance methods.

The addition of extra microphones increases the resolution of all spatial spectral estimation methods. The eigenanalysis of the parameterized spatial correlation matrix allows one to estimate the channel attenuation vector, and is useful for determining the microphone gains.

DOA estimation methods based on spatial spectral estimation provide a superior performance over TDOA-based methods due to the nature of the combining of the various microphone pairs. The application of higher resolution spatial spectral estimation such as the minimum variance and MCCC methods to the DOA estimation problem generally yields equivalent performance compared to the fixed-weighted SRP method in anechoic and moderately reverberant environments, and worse performance in heavily reverberant environments. In the case of the minimum variance method, the latter degradation stems from the desired signal cancellation that occurs due to reverberation; for the MCCC method, the performance degradation follows from the sensitivity of the MCCC to practical factors such as the non-Gaussian nature of speech.

Acknowledgment

The authors would like to thank J. Chen and Y. Huang for providing helpful suggestions and feedback regarding the simulation portion of this paper. They would also like to thank the anonymous reviewers for improving the clarity and quality of the manuscript.

References

[1] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE Acoust., Speech, Signal Process. Mag.*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[2] D. E. Dudgeon and D. H. Johnson*, Array Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[3] J. Chen, J. Benesty, and Y. Huang, "Time delay estimation in room acoustic environments: An overview," *EURASIP J. Appl. Signal Process.*, vol. 2006, p. 19, 2006.

[4] C. H. Knapp and G. C. 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.

[5] M. Omologo and P. Svaizer, "Use of the crosspower-spectrum phase in acoustic event location," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 3, pp. 288–292, May 1997.

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

[7] J. P. Ianiello, "High-resolution multipath time delay estimation for broad-band random signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 2, pp. 320–327, Mar. 1988.

[8] S. M. Griebel and M. S. Brandstein, "Microphone array source localization using realizable delay vectors," in *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust.*, 2001, pp. 71–74.

[9] D. H. Johnson, "The application of spectral estimation methods to bearing estimation problems," *Proc. IEEE*, vol. 70, no. 9, pp. 1018–1028, Sep. 1982.

[10] J. Dibiase, H. F. Silverman, and M. S. Brandstein, "Robust localization in reverberant rooms," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. S. Brstein and D. B. Ward, Eds. Berlin, Germany: Springer-Verlag, 2001, pp. 157–180.

[11] W. Bangs and P. Schultheis, "Space-time processing for optimal parameter estimation," in *Signal Processing*, J. Griffiths, P. Stocklin, and C. V. Schooneveld, Eds. New York: Academic, 1973, pp. 577–590.

[12] G. Carter, "Variance bounds for passively locating an acoustic source with a symmetric line array," *J. Acoust. Soc. Amer.*, vol. 62, pp. 922–926, Oct. 1977.

[13] W. Hahn and S. Tretter, "Optimum processing for delay-vector estimation in passive signal arrays," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 5, pp. 608–614, Sep. 1973.

[14] J. Krolik and D. Swingler, "Multiple broad-band source location using steered covariance matrices," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 10, pp. 1481–1494, Oct. 1989.

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

[16] F. Asano *et al.*, "Real-time sound source localization and and separation system and its application to automatic speech recognition," in *EUROSPEECH*, 2001, pp. 1013–1016.

[17] J. Chen, J. Benesty, and Y. Huang, "Robust time delay estimation exploiting redundancy among multiple microphones," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 549–557, Nov. 2003.

[18] J. Benesty, J. Chen, and Y. Huang, "Time-delay estimation via linear interpolation and cross-correlation," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 509–519, Sep. 2004.

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

[20] H. Teutsch and W. Kellermann, "EB-ESPRIT: 2D localization of multiple wideband acoustic sources using eigen-beams," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, 2005, vol. 3, pp. 89–92.

[21] I. Tashev and H. S. Malvar, "A new beamformer design for microphone arrays," in *Proc. IEEE Instrum. Meas. Technol. Conf.*, 2005, vol. 3, pp. 18–23.

[22] A. M. McKee and R. A. Goubran, "Sound localization in the human thorax," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, 2005, vol. 1, pp. 16–19.

[23] R. A. Monzingo and T. W. Miller*, Introduction to Adaptive Arrays*. Raleigh, NC: SciTech, 2004.

[24] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wideband sources," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 8, pp. 823–831, Aug. 1985.

[25] D. N. Zotkin and R. Duraiswami, "Accelerated speech source localization via a hierarchical search of steered response power," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 499–508, Sep. 2004.

[26] A. Johansson and S. Nordholm, "Robust acoustic direction of arrival estimation using Root-SRP-PHAT, A realtime implementation," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, 2005, vol. 4, pp. 933–936.

[27] B. Mungamuru and P. Aarabi, "Enhanced sound localization," *IEEE Trans. Syst., Man, Cybern. B: Cybern.*, vol. 34, no. 3, pp. 1526–1540, Jun. 2004.

[28] J. Benesty, Y. Huang, and J. Chen, "Time delay estimation via minimum entropy," *IEEE Signal Process. Lett.*, vol. 14, pp. 157–160, Mar. 2007.

[29] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, pp. 943–950, Apr. 1979.

[30] J. E. Gentle, "Cholesky factorization," in *Numerical Linear Algebra for Applications in Statistics*. Berlin, Germany: Springer-Verlag, 1998, pp. 93–95.

[31] M. R. Schroeder, "New method for measuring reverberation time," *J. Acoust. Soc. Amer.*, vol. 37, pp. 409–412, 1965.

[32] Y. Huang, J. Benesty, and J. Chen*, Acoustic MIMO Signal Processing*. Berlin, Germany: Springer-Verlag, 2006.

[33] S. L. Gay and J. Benesty*, Acoustic Signal Processing for Telecommunications*. Boston, MA: Kluwer.

Jacek Dmochowski was born in Gdansk, Poland, in December 1979. He received the B.Eng. degree (with high distinction) in communications engineering and the M.A.Sc. degree in electrical engineering from Carleton University, Ottawa, ON, Canada, in 2003 and 2005, respectively. He is currently pursuing the Ph.D. degree at the Institut National de la Recherche Scientifique-Énergie, Matériaux, et Télécommunications (INRS-EMT), Université du Québec, Montréal, QC, Canada.

His research interests are in the area of multichannel digital signal processing, and include microphone array beamforming and source localization.

Mr. Dmochowski is the recipient of the Ontario Graduate Scholarship (2004–2005), and the National Science and Engineering Research Council Post Graduate Scholarship at the Doctoral Level (2005–2008).

Jacob Benesty (M'92–SM'04) was born in 1963. He received the M.S. degree in microwaves from Pierre and Marie Curie University, Paris, France, in 1987, and the Ph.D. degree in control and signal processing from Orsay University, Orsay, France, in 1991.

During his Ph.D. program (from November 1989 to April 1991), he worked on adaptive filters and fast algorithms at the Centre National d'Etudes des Telecommunications (CNET), Paris. From January 1994 to July 1995, he worked at Telecom Paris University on multichannel adaptive filters and acoustic echo cancellation. From October 1995 to May 2003, he was first a Consultant and then a Member of the Technical Staff at Bell Laboratories, Murray Hill, NJ. In May 2003, he joined the Institut National de la Recherche Scientifique-Énergie, Matériaux, et Télécommunications (INRS-EMT), Université du Québec, Montréal, QC, Canada, as an Associate Professor. His research interests are in signal processing, acoustic signal processing, and multimedia communications. He coauthored the books *Acoustic MIMO Signal Processing* (Springer-Verlag, 2006) and *Advances in Network and Acoustic Echo Cancellation* (Springer-Verlag, 2001). He is also a coeditor/coauthor of the books *Speech Enhancement* (Spinger-Verlag, 2005), *Audio Signal Processing for Next Generation Multimedia Communications System* (Kluwer, 2004), *Adaptive Signal Processing: Applications to Real-World Problems* (Springer-Verlag, 2003), and *Acoustic Signal Processing for Telecommunications* (Kluwer, 2000).

Dr. Benesty received the 2001 Best Paper Award from the IEEE Signal Processing Society. He was a member of the editorial board of the EURASIP Journal on Applied Signal Processing and was the Co-Chair of the 1999 International Workshop on Acoustic Echo and Noise Control.

Sofiène Affes (M'94–SM'04) received the Diplôme d'Ingénieur in electrical engineering and the Ph.D. degree (with honors) in signal processing from the École Nationale Supérieure des Télécommunications (ENST), Paris, France, in 1992 and 1995, respectively.

He has been with Institut National de la Recherche Scientifique-Énergie, Matériaux, et Télécommunications (INRS-EMT), Université du Québec, Montréal, QC, Canada, as a Research Associate from 1995 until 1997, then as an Assistant Professor until 2000. Currently, he is an Associate Professor in the Personal Communications Group. His research interests are in wireless communications, statistical signal and array processing, adaptive space-time processing, and MIMO. From 1998 to 2002, he has been leading the radio design and signal processing activities of the Bell/Nortel/NSERC Industrial Research Chair in Personal Communications at INRS-ÉMT. Currently, he is actively involved in a major project in wireless of PROMPT-Québec (Partnerships for Research on Microelectronics, Photonics and Telecommunications).

Prof. Affes is the corecipient of the 2002 Prize for Research Excellence of INRS and currently holds a Canada Research Chair in High-Speed Wireless Communications. He served as a General Co-Chair of the IEEE VTC'2006-Fall conference, Montréal, and currently acts as a member of Editorial Board of the Wiley Journal on Wireless Communications and Mobile Computing.