Broadband Source Localization From an Eigenanalysis Perspective

Mehrez Souden, *Student Member, IEEE*, Jacob Benesty, and Sofiène Affes

***Abstract*—Broadband source localization has several applications ranging from automatic video camera steering to target signal tracking and enhancement through beamforming. Consequently, there has been a considerable amount of effort to develop reliable methods for accurate localization over the last few decades. Essentially, the localization process consists in finding the candidate source location that maximizes the *synchrony* between the properly time-shifted microphone outputs. In addition to using well known cross-correlation-based criteria such as the steered response power (SRP), minimum variance (MV), and multichannel cross-correlation (MCCC), this synchrony can also be measured using the averaged magnitude difference function (AMDF) and the averaged magnitude sum function (AMSF) whose calculations involve low computational cost. In earlier related works, the latter techniques have been used for time delay estimation (TDE) of a target source observed by only one pair of microphones. Their generalization to the multiple microphone case and application to source localization have not been studied yet. In this paper, we consider both categories, i.e., cross-correlation and AMDF (with AMSF)-based approaches, using an arbitrary number of microphones, and analyze their performance. Specifically, we first provide a unifying study of the most popular cross-correlation-based techniques, such as the SRP, MV, and MCCC. In this paper, we use the eigenanalysis of the parameterized spatial correlation matrix (PSCM) to classify these methods and gain some insight into their performance. We demonstrate, for instance, that the MV and SRP consist in searching the major eigenvalue of the PSCM, while the MCCC, essentially, combines its minor eigenvalues when scanning for the source location. Inspired by this analysis, we show, in the second part of this work, the efficiency of the AMDF and AMSF in localizing an acoustic source using multiple microphones. Indeed, we propose two new parameterized matrices named as the parameterized averaged magnitude difference matrix (PAMDM) and the parameterized averaged magnitude sum matrix (PAMSM). The eigenanalysis of these matrices also reveals new criteria for acoustic source localization. Simulation results are provided to illustrate the effectiveness of all the investigated and proposed methods.**

***Index Terms*—Acoustic source localization, averaged magnitude difference function (AMDF), averaged magnitude sum function (AMSF), beamforming, minimum variance (MV), multichannel cross-correlation, steered response power (SRP).**

I. Introduction

THE knowledge of the acoustic source location is of paramount importance in several applications including hands-free communication systems, teleconferencing, automatic camera steering, etc. In these applications, an array of microphones is commonly deployed to monitor the source location and perform either a required spatial filtering (focussing the beam-pattern of the array toward the source of interest and zeroing the array response toward the interference locations) [1] or automatically pointing the deployed cameras toward an active talker in a video-conferencing room for example [2]. This wide range of applications has boosted the development of several localization methods over the last few decades in order to achieve more robustness to the hostile nature of the acoustic environments where reverberation and noise are still major hindrances.

Essentially, the process of acoustic source localization consists in measuring the synchrony between properly delayed (noise-free) microphone outputs. Following this statement, we classify the source localization methods into three main categories. First, the most popular techniques are based on the second-order-statistics of the microphone array outputs. The SRP [3], MV [4], and MCCC [5] are considered as the benchmark methods in this category. Indeed, they have been applied in several earlier contributions to either time delay estimation (TDE) or localization. In [1], an overview of TDE techniques and discussions of the pros. and cons. of the different frequency-domain weighting functions were provided. In [6], [7], the SRP with the phase transform was used for multiple speakers localization. In [8], Domochowski *et al.* applied these well-known TDE methods to the localization by directly taking into account the relationship between the array geometry and the spatial location of the source. Therein, the authors have shown that these techniques are formulated using the PSCM that was first introduced in [4]. Herein, we further study these methods based on the eigenanalysis of the PSCM to justify their performance in noisy and reverberant conditions. The synchrony between the processed microphone outputs can also be measured from an information theoretic point of view. This corresponds to the second category of localization methods. In [9], for example, it was shown that the mutual information maximization (or joint entropy minimization) allows for accurate source localization. This approach leads to the (cross-correlation-based) MCCC for Gaussian signals. The speech distribution, however, is generally assumed to have a Laplacian shape [10] and more complicated calculations are required in the computation of the mutual information of the multivariate Laplacian-distributed microphone outputs. The final category consists in methods based on simpler criteria in the sense that neither second- nor higher-order statistics (or assumed distributions) are required. In this category, the synchrony between the outputs of each pair of microphones is measured using either the AMDF or the AMSF [14]. The AMDF was previously applied for pitch

Manuscript received May 05, 2009; revised October 18, 2009. First published December 15, 2009; current version published July 14, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Nakatani Tomohiro.

The authors are with the Université du Québec, INRS-EMT, Montreal, QC H5A 1K6, Canada (e-mail: souden@emt.inrs.ca).

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.2009.2038556

estimation [11], [12], and also TDE [13], [14] using a single pair of microphones. To the best of our knowledge, both methods have not been yet generalized to the multiple microphone case with application to source localization. In this paper, we will only focus on the first and third categories since they require much less computational load and can achieve good performance without any prior knowledge about the wideband source distribution.

Our contribution in this paper is twofold. First, we use the eigenanalysis as a powerful tool to analyze and classify the cross-correlation-based broadband source localization techniques. The underlying idea of the eigenanalysis framework investigated herein is that when the PSCM is steered toward the source location, two subspaces can be identified. The first one corresponds to the one-dimensional subspace spanned by the vector associated with the largest eigenvalue of the PSCM. Whereas the second subspace is spanned by the eigenvectors of the PSCM which are associated with the remaining eigenvalues of this matrix. Instead of using a predefined form of the steering vector as in narrowband high-resolution techniques [15], [16], we scan the potential source locations and observe the variations of the eigenvalues associated with these two subspaces. To gain a better understanding of the functioning of the MV and the SRP, we devise both criteria from a covariance fitting perspective [17], [18] and show that they are essentially equivalent to tracing the source location using the maximum eigenvalue. We also show that the MCCC, in contrast, takes advantage of the joint variations of the noise-subspace-associated eigenvalues. Alike the cross-correlation-based framework, our second contribution consists in demonstrating the efficiency of the AMDF and AMSF to localize acoustic sources when multiple microphones are deployed. Indeed, for a given pair of microphones the AMDF and the AMSF aim at maximizing the synchrony between properly time-shifted output signals by calculating the absolute difference or sum, respectively. By taking advantage of all microphone pairs as in the PSCM-based analysis, we propose two new parameterized matrices, namely, the PAMDM and the PAMSM that contain all the combinations of the AMDF and AMSF relating each pair of microphones. The eigenanalysis of both matrices reveals new efficient criteria for source localization.

This paper is organized as follows. Section II describes the data model and the assumptions required to obtain a tractable formulation of the problem. Section III reviews the cross-correlation-based methods and presents the eigenanalysis of the PSCM as a powerful tool for their understanding. In addition, new criteria are devised using the same framework. Section IV outlines the underlying idea of generalizing the AMDF and the AMSF to the multi-microphone case and their application to wideband source localization. The eigenanalysis of the new proposed matrices, i.e., the PAMDM and the PAMSM, reveals other criteria for source localization. Finally, Section V evaluates the performance of the proposed methods with comparisons to the existing ones.

II. Data Model

Let $s(t)$ denote a signal generated by a broadband source and captured by an array of N microphones. Let also $\mathbf{r}_s = [\theta_s \ \phi_s \ d_s]^T$ denote the 3–D space coordinates of the source, where $(\cdot)^T$ denotes the transpose operator. The three entries of $\mathbf{r}_s$ are the azimuth, elevation, and range in the spherical coordinate system. The output of the nth $(n = 1, \ldots, N)$ microphone is given by

$$x_n(t) = a_n(\mathbf{r}_s) s\left(t - \tau_n(\mathbf{r}_s)\right) + v_n(t) \tag{1}$$

where $a_n(\mathbf{r}_s)$ is the channel attenuation, $v_n(t)$ is an additive noise, and $\tau_n(\mathbf{r}_s)$ is the propagation time delay from the source to the nth microphone element. In the free-field case, the received signal magnitude is attenuated at a rate proportional to the inverse of the distance between the source and the microphone. Such an information can be exploited to develop energy-based source localization algorithms as it was shown in [21], for instance. However, such techniques generally exhibit high sensitivity to additive noise and more importantly reverberation. The time delay of arrival $\tau_n(\mathbf{r}_s)$ is also a function of the source location. Indeed,

$$\tau_n(\mathbf{r}_s) = \frac{d_n(\mathbf{r}_s)}{c} \tag{2}$$

where $d_n(\mathbf{r}_s)$ is the distance between the source and the nth microphone element and c is the sound velocity. Therefore, estimating $\tau_n(\mathbf{r}_s)$ amounts to estimating $d_n(\mathbf{r}_s)$ which defines a sphere with radius $d_n(\mathbf{r}_s)$ and the nth microphone as a center. Instead, the time difference of arrival (TDOA) between pairs of microphones are commonly used for localization in addition to some other simplifying assumptions [3], [8], [20]. The TDOA between microphones n and m $(n,\ m \in \{1, \ldots, N\})$ is simply given by

$$\mathcal{F}_{nm}(\mathbf{r}_s) = \tau_m(\mathbf{r}_s) - \tau_n(\mathbf{r}_s) = \frac{d_m(\mathbf{r}_s) - d_n(\mathbf{r}_s)}{c}. \tag{3}$$

By assuming that the source lies in the far-field, it can be shown that $\mathcal{F}_{nm}(\mathbf{r}_s)$ is, approximately, only dependent on θ_s and ϕ_s. To further simplify the problem, we can assume that the source and microphone array are located on the same plane. Consequently, the TDOA depends on θ_s only. In the sequel, we will make this assumption for the sake of simplicity. Finally, note that $\mathcal{F}_{nm}(\cdot)$ depends on the array geometry. Without loss of generality, we will consider a uniform circular array (UCA) whose advantages are discussed in [8]. In this case, we have

$$\mathcal{F}_{nm}(\theta_s) = \frac{r}{c}\left[\cos\left(\theta_s - \frac{2(n-1)\pi}{N}\right) - \cos\left(\theta_s - \frac{2(m-1)\pi}{N}\right)\right] \tag{4}$$

where r is the array radius.

Using a general terminology, we state that the process of locating an acoustic source consists in measuring the synchrony between the properly delayed microphone outputs. Although most of the source localization techniques take advantage of the microphone array outputs cross-correlation as a measure of synchrony, one can also expect the AMDF and AMSF to be of great help. Indeed, it was shown in [13] and [14] that these two synchrony measures allow for accurate estimation of the TDOA when a pair of microphones is used. Herein, we generalize both criteria to the multi-microphone case and show their effectiveness in locating acoustic sources. It is also worth mentioning that the mutual information is another potential synchrony measure. Its effectiveness in TDE was demonstrated in [9]. However, the estimation of the mutual information requires the prior knowledge of the microphone outputs distribution and leads to quite involved calculations [9]. The utilization of this criterion falls beyond the scope of this work due to space constraint and we would rather focus our attention in this contribution on the first two synchrony measures (i.e., the cross-correlation and the AMDF with AMSF).

The parameterized processing for acoustic source localization [5], [8], [14], [19] consists in applying a delay $\mathcal{F}_{1n}(\theta)$ to the observations seen by the nth microphone and optimize certain criteria over the parameter θ. Note that we take the first microphone as a reference, without loss of generality. The optimality is reached when $\theta = \theta_s$. For a given parameter θ, we define

$$\mathbf{x}(t,\theta) = [x_1(t)\; x_2\left(t + \mathcal{F}_{12}(\theta)\right)\; \cdots\; x_N\left(t + \mathcal{F}_{1N}(\theta)\right)]^T \quad (5)$$

where $x_n\left(t+\mathcal{F}_{1n}(\theta)\right) = a_n s(t - \tau_n(\mathbf{r}_s) + \mathcal{F}_{1n}(\theta)) + v_n(t + \mathcal{F}_{1n}(\theta))$. This parameterized vector will be directly involved in all the forthcoming processing.

Most of the cross-correlation-based techniques use the PSCM either explicitly or implicitly [8]. In the sequel, we revisit the definition of this matrix. Then, we propose a unifying framework for PSCM-based source localization techniques which takes advantage of the eigenanalysis of the PSCM. By analogy, we next propose the PAMDM and PAMSM that generalize the AMDF and AMSF to the multi-microphone case.

III. Cross-Correlation-Based Source Localization

The PSCM is simply defined as the correlation matrix of $\mathbf{x}(t,\theta)$ and is given by

$$\begin{aligned}\mathbf{R}_{xx}(\theta) &= E\left\{\mathbf{x}(t,\theta)\mathbf{x}^T(t,\theta)\right\}\\ &= \mathbf{A}\mathbf{R}_{ss}(\theta)\mathbf{A} + \mathbf{R}_{vv}(\theta)\end{aligned} \quad (6)$$

where $\mathbf{A} = \mathrm{diag}\left[\mathbf{a}\right]$, $\mathbf{a} = \left[a_1, a_2, \ldots, a_N\right]^T$, and $E\left\{\cdot\right\}$ denotes the mathematical expectation. Note that we avoid mentioning the dependence of the channel attenuation coefficient on the source location for notational convenience. $\mathbf{R}_{ss}(\theta)$ and $\mathbf{R}_{vv}(\theta)$ are the resulting parameterized covariance matrices of the source and noise, respectively. The (i,j)th entry of the PSCM is given by

$$\begin{aligned}\left[\mathbf{R}_{xx}(\theta)\right]_{i,j} &= r_{x_i x_j}(\mathcal{F}_{ij}(\theta))\\ &= a_i a_j r_{ss}(\mathcal{F}_{ij}(\theta) - \mathcal{F}_{ij}(\theta_s)) + r_{v_i v_j}(\mathcal{F}_{ij}(\theta))\end{aligned} \quad (7)$$

where $r_{x_i x_j}(\tau) = E\left\{x_i(t)x_j(t+\tau)\right\}$, $r_{v_i v_j}(\tau) = E\left\{v_i(t)v_j(t+\tau)\right\}$, and $r_{ss}(\tau) = E\left\{s(t)s(t+\tau)\right\}$, for a given time delay τ. The general procedure for parameterized processing that was proposed in [8] for source localization and that we consider in this paper consists in investigating the broadband spatial spectrum (a certain criteria which is function of the PSCM), denoted herein as $\mathcal{S}\left\{\mathbf{R}_{xx}(\theta)\right\}$, and identifying its peak that corresponds to the source location. Formally, the approach consists in estimating θ_s as [8]

$$\hat{\theta}_s = \arg\max_{\theta} \mathcal{S}\left\{\mathbf{R}_{xx}(\theta)\right\}. \quad (8)$$

So far, several criteria have been proposed for source localization such as the SRP [3], MV [4], and MCCC that stems from the linear spatial prediction [5], etc. Herein, we show that all these methods can be devised from a general eigenanalysis-based framework.

A. Eigenanalysis-Based Framework

When $\theta = \theta_s$, the PSCM becomes

$$\mathbf{R}_{xx}(\theta_s) = \sigma_s^2 \mathbf{a}\mathbf{a}^T + \mathbf{R}_{vv}(\theta_s) \quad (9)$$

where $\sigma_s^2 = E\{s^2(t)\}$. Clearly, $\mathbf{R}_{xx}(\theta_s)$ can be used to identify two subspaces: signal-plus-noise (major) and noise (minor) subspaces. The first one is of dimension 1 while the second is of dimension $N-1$. In narrowband high-resolution techniques such as MUSIC [15] and ESPRIT [16], the knowledge of the so-called steering vector allows one to determine the DOA. Indeed, the latter belongs to the signal subspace when $\theta = \theta_s$. In order to estimate the source direction of arrival, these methods take advantage of the orthogonality of the steering vector to the eigenvectors corresponding to the $N-1$ smallest eigenvalues of the observations' covariance matrix (spanning the noise subspace). In our case, however, no explicit expression for this steering vector is available because of the wideband nature of the acoustic signal and the convolution involved in the data model (1). In [22], a notable attempt was made to generalize the MUSIC algorithm to broadband signals. But the proposed approach was confronted to the choice of the steering vector. Transforming the problem to the frequency domain may be interesting in the sense that every frequency bin can be processed separately from the others using the MUSIC algorithm [15], for instance. Unfortunately, the combination of the location estimates is not straightforward since the speech signal is sparse in the frequency domain, thereby leading to poor location estimates in many frequency bins and affecting the overall performance when combining the estimates at all frequencies [1]. Fortunately, the eigenanalysis of $\mathbf{R}_{xx}(\theta)$ can still be of a great help. To proceed, let us first decompose $\mathbf{R}_{xx}(\theta)$ as

$$\mathbf{R}_{xx}(\theta) = \mathbf{U}^T(\theta)\mathbf{D}(\theta)\mathbf{U}(\theta) \quad (10)$$

where $\mathbf{U}(\theta)$ is a unitary matrix, $\mathbf{D}(\theta) = \mathrm{diag}\left[\lambda_1(\theta), \ldots, \lambda_N(\theta)\right]$, and $\lambda_1(\theta) \geq \lambda_2(\theta) \geq \cdots \geq \lambda_N(\theta) \geq 0$ denote the eigenvalues of $\mathbf{R}_{xx}(\theta)$ sorted in a decreasing order. This eigenanalysis will allow us to gain good insights into source localization techniques.

B. Justifying the Importance of the Eigenanalysis for Source Localization

Ignoring the noise component in (1), we see that when the PSCM is steered toward θ_s, we have $\mathbf{R}_{xx}(\theta_s) = \sigma_s^2\mathbf{a}\mathbf{a}^T$. Then, one can see that $\mathbf{R}_{xx}(\theta_s)\mathbf{a} = \sigma_s^2\left\|\mathbf{a}\right\|^2\mathbf{a}$, meaning that $\mathbf{a}$ is an eigenvector of $\mathbf{R}_{xx}(\theta_s)$ associated to the major eigenvalue (all other eigenvalues equal 0) of this matrix

$$\lambda_1(\theta_s) = \sigma_s^2\left\|\mathbf{a}\right\|^2 = \sigma_s^2\sum_{n=1}^{N} a_n^2. \quad (11)$$

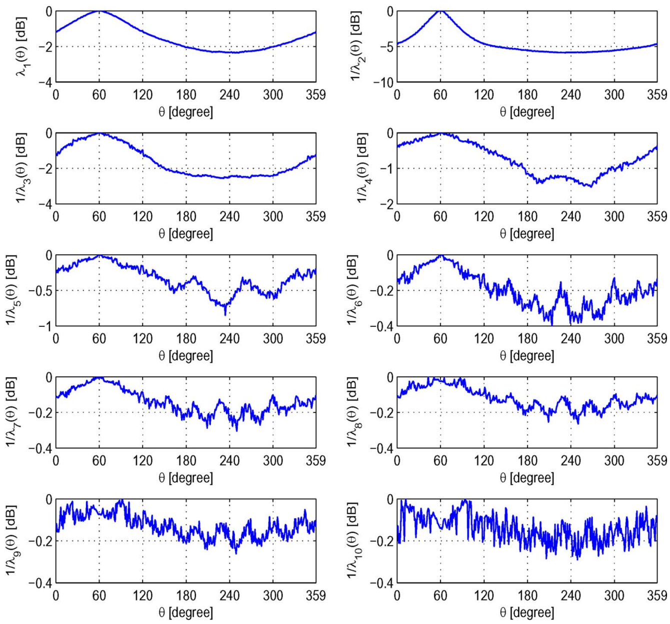


Fig. 1. PSCM eigenvalues versus azimuth; anechoic enclosure, SNR $= 0$ dB.

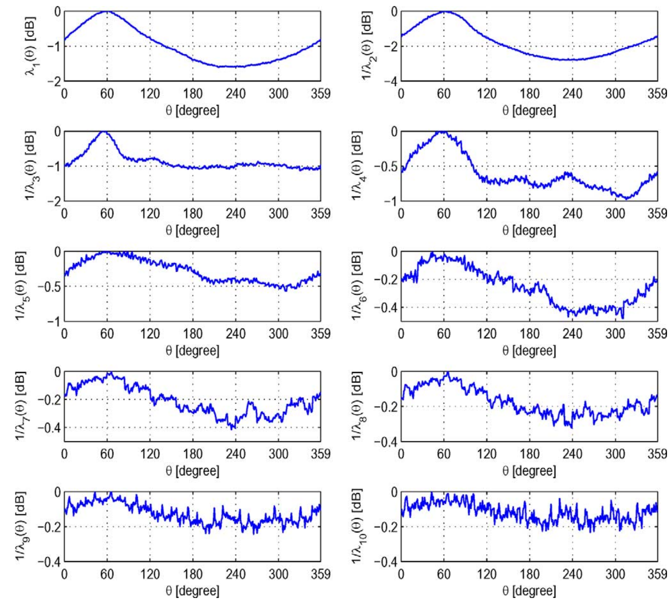


Fig. 2. PSCM eigenvalues versus azimuth; reverberant enclosure, SNR $=$ 0 dB.

It is apparent that $\lambda_1(\theta_s)$ captures the overall energy of the received signal (including the channel effect). Herein, we state that when the PSCM is not steered toward the effective source location, the received noise-free energy is smeared on more than one dimension. To understand this phenomenon, we consider the particular case where the desired source is temporally white with identically distributed components. In this case, when $\theta \neq \theta_s$,

$$\mathbf{R}_{xx}(\theta_s) = \sigma_s^2 \text{diag}\left[a_1^2, \ldots, a_N^2\right].$$

The nth eigenvalue of this matrix is obviously

$$\lambda_n(\theta) = \sigma_s^2 a_n^2 \tag{12}$$

and

$$\lambda_1(\theta_s) = \sum_{n=1}^{N} \lambda_n(\theta). \tag{13}$$

Hence, when the spatial correlation matrix is not steered toward the main direction, the source energy is spread over the N dimensions. The analysis in the case of a temporally correlated process such as speech is not straightforward, but one can expect a similar behavior. Figs. 1 and 2 depict the variations of all the eigenvalues of $\mathbf{R}_{xx}(\theta)$ with respect to the azimuth. These results were obtained in both anechoic and reverberant environments where a speech source is located at an azimuthal angle $\theta_s = 60^\circ$ and captured by a circular array of ten microphones as described in Section V. These figures support our expectations since the major eigenvalue is maximized while all other eigenvalues are minimized (see discussions in Section V-A). In the light of this example, we gained some insight into the effect of the choice of θ on the behavior of the desired signal energy distribution over the N dimensions: the energy is spread over multiple dimensions when $\theta \neq \theta_s$ and focused on a single one when $\theta = \theta_s$.

C. Analysis of the Major Eigenvalue

When $\theta = \theta_s$, there is only one dominant eigenvalue (supposing that the noise is weak enough or has independent and identically distributed (i.i.d.) components) that corresponds to the source (plus noise) energy. Physically, this can be explained by the fact that the overall signal energy is impinging on the microphone array from a single direction. When $\theta \neq \theta_s$, however, the rank of $\mathbf{R}_{xx}(\theta)$ is larger than 1, meaning that the energy of the source is spread over many dimensions. This intuition was exploited in [8] by observing the maximum eigenvalue and choosing this criterion

$$\mathcal{S}_{\text{MaxEig}}\left\{\mathbf{R}_{xx}(\theta)\right\} = \lambda_1(\theta). \tag{14}$$

In what follows, we take advantage of the covariance fitting approach [17], [18] to demonstrate that the SRP and MV aim in essence at analyzing the major eigenvalue. To this end, let us suppose that $\mathbf{a}$ has a unit norm[1]. First, we see from (9) that in the absence of noise, $\mathbf{a}$ is an eigenvector of $\mathbf{R}_{xx}(\theta_s)$ associated with $\lambda_1(\theta_s)$. If we further suppose the knowledge of $\mathbf{a}$, one can think about finding an estimate of $\lambda_1(\theta_s)$ (denoted as $\hat{\lambda}_1$) knowing a certain estimate of $\mathbf{R}_{xx}(\theta_s)$, say $\hat{\mathbf{R}}_{xx}$. From a covariance fitting perspective [17], [18], this can be easily formulated as

$$\hat{\lambda}_1 = \arg\min_{\lambda} \left\| \lambda \mathbf{a}\mathbf{a}^T - \hat{\mathbf{R}}_{xx} \right\|^2. \tag{15}$$

The straightforward solution to this optimization problem is

$$\hat{\lambda}_1 = \mathbf{a}^T \hat{\mathbf{R}}_{xx} \mathbf{a}. \tag{16}$$

If we further assume that there is no channel attenuation in the data model (1), i.e., $\mathbf{a} = \mathbf{1}$ which is an N–dimensional vector with all its entries being equal to 1, and replacing $\hat{\mathbf{R}}_{xx}$ by the

[1]The analysis is valid in the general case $\|\mathbf{a}\| \neq 1$, but with $\mathbf{a}$ independent of θ, which is the case here.

best available estimate of this matrix at a given direction θ, i.e., $\mathbf{R}_{xx}(\theta)$, we obtain the well known SRP criterion [3]

$$\hat{\lambda}_1(\theta) = \mathcal{S}_{\mathrm{SRP}}\left\{\mathbf{R}_{xx}(\theta)\right\} = \mathbf{1}^T\mathbf{R}_{xx}(\theta)\mathbf{1} \tag{17}$$

up to a constant scaling factor. Hence, in the absence of channel attenuations, maximizing $\mathcal{S}_{\mathrm{SRP}}\left\{\mathbf{R}_{xx}(\theta)\right\}$ is equivalent to maximizing the maximum eigenvalue of $\mathbf{R}_{xx}(\theta)$ over θ that leads to $\lambda_1(\theta_s)$.

A second way to find $\hat{\lambda}_1$ consists in solving the following covariance matching problem [17], [18]

$$\hat{\lambda}_1 = \arg\min_{\lambda}\left\|(\lambda\mathbf{a}\mathbf{a}^T)^{\#} - \hat{\mathbf{R}}_{xx}^{-1}\right\|^2 \tag{18}$$

where $^{\#}$ denotes the pseudo-inverse of a matrix, leading to[2]

$$\hat{\lambda}_1 = (\mathbf{a}^T\hat{\mathbf{R}}_{xx}^{-1}\mathbf{a})^{-1}. \tag{19}$$

Again, supposing that $\mathbf{a} = \mathbf{1}$, and replacing $\hat{\mathbf{R}}_{xx}$ by the best available estimate of this matrix at a given direction θ, we obtain the well known MV criterion [4], [8]

$$\hat{\lambda}_1(\theta) = \mathcal{S}_{\mathrm{MV}}\left\{\mathbf{R}_{xx}(\theta)\right\} = \left(\mathbf{1}^T\mathbf{R}_{xx}^{-1}(\theta)\mathbf{1}\right)^{-1} \tag{20}$$

up to a constant scaling factor. We conclude that the maximum eigenvalue, the SRP, and the MV have essentially the same objective which consists in maximizing the maximum eigenvalue of the PSCM over all potential candidate locations. Note, however, that the computation of the MV criterion involves the inversion of the PSCM which makes it sensitive to the ill conditioning of this matrix. This fact will be numerically illustrated in Section V.

D. Analysis of the Minor Eigenvalues

In the absence of noise, we have

$$\lambda_2(\theta_s) = \cdots = \lambda_N(\theta_s) = 0. \tag{21}$$

This is not the case in the presence of noise. However, one hopes that the energy of speech is high enough compared to the noise's so that $\lambda_1(\theta_s)$ is much higher than all other eigenvalues, thereby allowing to distinguish between the noise and speech contributions to the PSCM when steering it towards the source location.

The criteria that we propose herein are based on observing the energy of the minor subspace. This energy can be calculated by averaging it over $\lambda_2(\theta), \ldots, \lambda_N(\theta)$. Moreover, one can use either geometric or arithmetic averaging to obtain the two criteria given as follows:

$$\mathcal{S}_{\mathrm{NSA}}\left\{\mathbf{R}_{xx}(\theta)\right\} = \left(\frac{1}{N-1}\sum_{n=2}^{N}\lambda_n(\theta)\right)^{-1} \tag{22}$$

$$\mathcal{S}_{\mathrm{NSG}}\left\{\mathbf{R}_{xx}(\theta)\right\} = \left(\prod_{n=2}^{N}\lambda_n(\theta)\right)^{-1/(N-1)}. \tag{23}$$

The subscripts "NSA" and "NSG" stand for "noise subspace arithmetic averaging" and "noise subspace geometric averaging," respectively. We expect both criteria to reach their maximum at θ_s.

[2]It can be easily verified that $(\lambda\mathbf{a}\mathbf{a}^T)^{\#} = (1/\lambda)\mathbf{a}\mathbf{a}^T$.

Another notable acoustic source localization method was proposed in [5] where Benesty *et al.* took advantage of the linear spatial predictability of the noise-free microphone signals from each other (multiple redundancies) to develop the following criterion:

$$\mathcal{S}_{\mathrm{MCCC}}\left\{\mathbf{R}_{xx}(\theta)\right\} = 1 - \det\left\{\tilde{\mathbf{R}}_{xx}(\theta)\right\} \tag{24}$$

where $\tilde{\mathbf{R}}_{xx}(\theta) = \mathbf{\Lambda}^{-1/2}\mathbf{R}_{xx}(\theta)\mathbf{\Lambda}^{-1/2}$ and $\mathbf{\Lambda} = \mathrm{diag}\left[E\{x_1^2(t)\}, \ldots, E\left\{x_N^2\left(t + \mathcal{F}_{1N}(\theta)\right)\right\}\right]$ is a diagonal matrix containing all the diagonal terms of $\mathbf{R}_{xx}(\theta)$ which turn out to be scaling factors independent of θ due to the speech local stationarity.

In addition, maximizing $\mathcal{S}_{\mathrm{MCCC}}\left\{\mathbf{R}_{xx}(\theta)\right\}$ is equivalent to maximizing $1/\det\left\{\tilde{\mathbf{R}}_{xx}(\theta)\right\}$ or equivalently $1/\det\left\{\mathbf{R}_{xx}(\theta)\right\}$ since $\mathbf{\Lambda}$ is independent of θ. It is known that

$$\det\left\{\mathbf{R}_{xx}(\theta)\right\} = \prod_{n=1}^{N}\lambda_n(\theta). \tag{25}$$

Thus, we find, once again, that we are able to define another well known criterion for source localization by using the eigenvalues of the PSCM. Furthermore, we deduce that the MCCC is quite different from the SRP and MV since its objective is to look for the minor subspace and reach the optimality when $\det\left[\mathbf{R}_{xx}(\theta)\right]$ is minimal, i.e., when the effect of the minor eigenvalues is dominant (due to the geometric averaging).

E. Common Eigenanalysis Framework

So far, we have shown that the information about the source location can be traced using both types (i.e., major and minor) of eigenvalues. Herein, we propose the following general form that combines all the eigenvalues:

$$\mathcal{S}_{\mathrm{GSA}}\left\{\mathbf{R}_{xx}(\theta)\right\} = \sum_{n=1}^{N}\alpha_n\lambda_n^{\nu_n}(\theta) \tag{26}$$

$$\mathcal{S}_{\mathrm{GSG}}\left\{\mathbf{R}_{xx}(\theta)\right\} = \prod_{n=1}^{N}\lambda_n^{\beta_n}(\theta) \tag{27}$$

where α_n, ν_n, and β_n , $n \in \{1, \ldots, N\}$, are some multiplicative and exponential weighting factors that have to be chosen properly. The subscripts "GSA" and "GSG" stand for "generalized spectrum arithmetic" and "generalized spectrum geometric" averaging, respectively. A similar framework for spatial spectral estimation for *narrowband* sources has been recently proposed by Stoica *et al.* in [18]. In our proposal, we take advantage of this framework in the context of acoustic source localization by focusing on the eigenalysis of the PSCM. It is clear that all these localization techniques exploit the fact that when the spatial correlation matrix is steered toward θ_s, their averaged spectrum is minimal (or maximal) due to the dominance of the effect of the eigenvalues associated with either the minor or major subspace. However, different weights have to be attributed to the eigenvalues depending on whether they are associated with one of either subspaces. By doing so, one wishes to obtain better spectrum resolution and potentially improved localization. In the absence of any prior knowledge of the noise statistics, we

TABLE I
CLASSIFICATION OF THE PSCM-BASED EXISTING AND NEW SOURCE LOCALIZATION METHODS FROM AN EIGENANALYSIS PERSPECTIVE

Subspace	Major subspace			Minor subspace			Minor & major subspaces	
Localization method	SRP	MV	MaxEig	MCCC	NSA	NSG	CEig	CMCCC
α_1	$\left\{[\mathbf{U}(\theta)\mathbf{1}]_1\right\}^2$	$\left\{[\mathbf{U}(\theta)\mathbf{1}]_1\right\}^2$	1	–	0	–	1	–
$\alpha_n;\ n=2...N$	$\left\{[\mathbf{U}(\theta)\mathbf{1}]_n\right\}^2$	$\left\{[\mathbf{U}(\theta)\mathbf{1}]_n\right\}^2$	0	–	1	–	$-\frac{1}{N-1}$	–
ν_1	1	-1	1	–	–	–	1	–
$\nu_n;\ n=2...N$	1	-1	–	–	1	–	1	–
β_1	–	–	1	1	–	0	–	1
$\beta_n;\ n=2...N$	–	–	0	1	–	-1	–	-1

can simply assign the same weighting terms to the minor eigenvalues. Below, we give two examples to illustrate this strategy

$$\mathcal{S}_{\mathrm{CEig}}\left\{\mathbf{R}_{xx}(\theta)\right\}=\lambda_1(\theta)-\frac{1}{N-1}\sum_{n=2}^{N}\lambda_n(\theta) \quad (28)$$

$$\mathcal{S}_{\mathrm{CMCCC}}\left\{\mathbf{R}_{xx}(\theta)\right\}=\lambda_1(\theta)\prod_{n=2}^{N}\lambda_n^{-1}(\theta). \quad (29)$$

The above criteria are actually based on the contrast between both subspaces which reaches its maximum when the matrix is steered toward θ_s. Since we have no prior knowledge about the noise energy distribution, we averaged its eigenvalues (using either arithmetic or geometric averages). Also different weights are assigned to both minor and major subspace associated terms to properly combine the effects of both subspaces on the spatial spectrum. In Table I, we summarize and classify the PSCM-based source localization techniques from an eigenanalysis perspective[3].

As we stated above, the candidate azimuthal angle that maximizes the synchrony between properly time-shifted microphone outputs corresponds to the desired source location (angle of arrival). We have shown that by using the cross-correlation (through the PSCM) as a measure of synchrony, several criteria can be devised thanks to the new proposed eigenanalysis-based framework. Other alternatives to measure this synchrony consist in properly applying the AMDF and AMSF criteria when multiple microphone observations are available. In the following section, we take advantage of both criteria in addition to the eigenanalysis to propose new source localization methods.

IV. MULTI-MICROPHONE AVERAGED MAGNITUDE DIFFERENCE AND SUM-BASED SOURCE LOCALIZATION

The AMDF is a well-known criterion in pitch estimation literature [11], [12]. More recently, it has been applied to the estimation of the time delay of propagation between a pair of microphones [13], [14]. Essentially, this criterion consists in scanning for the proper parameter (time period in the context of pitch estimation and time delay in the context of TDE) that maximizes the synchrony between two properly time-shifted signals. This synchrony is measured using the magnitude of the difference between both signals of interest. This approach is known to offer good performance in favorable noise conditions and even in reverberant environments [13], [14]. Furthermore, the calculation of this criterion has low complexity since it involves no multiplications. In the light of this discussion, it is surprising to find that this criterion has neither been generalized to the multi-microphone case for TDE nor applied to source localization. Moreover, it was proven by Chen *et al.* that the AMSF is a promising simple and accurate synchrony measure for TDE [14]. Indeed, it is measured using the magnitude of the sum between both signals of interest. In this part, we take advantage of both criteria to localize acoustic sources using multiple microphones. Similar to the cross-correlation-based approaches, we process the parameterized vector $\mathbf{x}(t,\theta)$ by scanning all potential candidate locations and measuring the synchrony (using AMDF or AMSF-based criteria) between its entries.

When a block of data having a size $Q>1$, say for example $\mathbf{x}(t,\theta),\ \mathbf{x}(t-1,\theta),\ \ldots,\ \mathbf{x}(t-Q+1,\theta)$, is available, the AMDF of a given pair (i,j) of microphones, $i,j\in\{1,\ldots,N\}$, is expressed as

$$\tilde{J}_{ij,\mathrm{AMDF}}(\theta)=\frac{1}{Q}\sum_{t=1}^{Q}\left|x_i(t+\mathcal{F}_{1i}(\theta))-x_j(t+\mathcal{F}_{1j}(\theta))\right|. \quad (30)$$

To avoid rounding issues, it is better to use the TDOA $\mathcal{F}_{ij}(\theta)=\mathcal{F}_{1j}(\theta)-\mathcal{F}_{1i}(\theta)$ and obtain the following criterion, instead

$$J_{ij,\mathrm{AMDF}}(\theta)=\frac{1}{Q}\sum_{t=1}^{Q}\left|x_i(t)-x_j\left(t+\mathcal{F}_{ij}(\theta)\right)\right|. \quad (31)$$

In order to take full advantage of the multiple microphones, we define the PAMDM, $\mathbf{\Delta}(\theta)$, whose $(i,j)th$ entry is defined as

$$[\mathbf{\Delta}(\theta)]_{ij}=J_{ij,\mathrm{AMDF}}(\theta). \quad (32)$$

Note first that for equal channel attenuation coefficients and no additive noise in the data model (1), $\|\mathbf{\Delta}(\theta)\|^2\rightarrow 0$ when $\theta\rightarrow\theta_s$, where $\|\cdot\|$ denotes any matrix norm. In particular, we consider the norm $\|\mathbf{\Delta}(\theta)\|^2=\mathrm{tr}\left\{\mathbf{\Delta}(\theta)\mathbf{\Delta}^T(\theta)\right\}=\sum_{n=1}^{N}\delta_n^2(\theta)$, where $\delta_n(\theta), n=1,\ldots,N$ are the eigenvalues of the PAMDM sorted such that $|\delta_1(\theta)|\geq|\delta_2(\theta)|\geq\cdots\geq|\delta_N(\theta)|$. Hence, we deduce that $\|\mathbf{\Delta}(\theta)\|$ reaches its minimum at the source location and so do $|\delta_1(\theta)|,\ |\delta_2(\theta)|,\ \ldots,\ |\delta_N(\theta)|$. Figs. 3 and 4 depict the variations of the inverses of the magnitudes of the eigenvalues of $\mathbf{\Delta}(\theta)$ with respect to the azimuth. These results were obtained in the same conditions as in Figs. 1 and 2 discussed

[3]Note that for the MV and SRP criteria, we took advantage of the eigendecomposition in (10). Also, we considered minimizing $\mathcal{S}_{\mathrm{MV}}^{-1}\left\{\mathbf{R}_{xx}(\theta)\right\}$ instead of maximizing $\mathcal{S}_{\mathrm{MV}}\left\{\mathbf{R}_{xx}(\theta)\right\}$ for the MV.

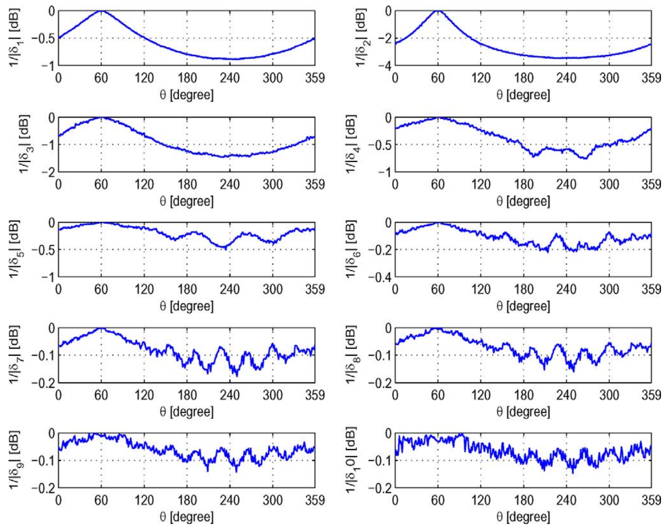


Fig. 3. PAMDM eigenvalues versus azimuth; anechoic enclosure, SNR $= 0$ dB.

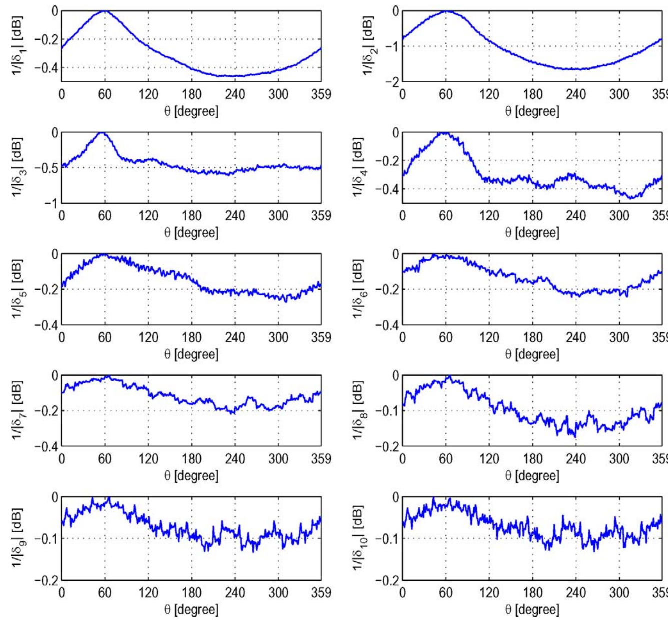


Fig. 4. PAMDM eigenvalues versus azimuth; reverberant enclosure, SNR $=$ 0 dB.

in Sections III-B and V-A. We deduce that most of these eigenvalues can be used for source localization by scanning all candidate source locations and retrieving the maximum of the spatial spectra that corresponds to θ_s. However, one can notice that the spectra of the smallest eigenvalues exhibit several spikes. It turns out that these spikes have a deleterious effect on source localization as we verified through several numerical examples. Large eigenvalues seem, in contrast, quite robust. We also noticed that in the presence of reverberation, the largest eigenvalue is the most reliable to be used as a criterion for source localization. Consequently, we propose the first new multiple microphone AMDF-based criterion

$$\mathcal{S}_{\mathrm{EigAMDF}}\left\{\mathbf{\Delta}(\theta)\right\} = \frac{1}{|\delta_1(\theta)|}. \tag{33}$$

Moreover, we propose the second new criterion which is inspired from the SRP and termed herein as *steered magnitude difference* (SMD)

$$\mathcal{S}_{\mathrm{SMD}}\left\{\mathbf{\Delta}(\theta)\right\} = \frac{1}{\mathbf{1}^T\mathbf{\Delta}(\theta)\mathbf{1}}. \tag{34}$$

This criterion is rather heuristic, yet efficient, as it will be proven by numerical examples. Furthermore, it reduces to the classical pairwise AMDF criterion for source localization when only a pair of microphones is used.

In [14], it was shown that the AMSF is also a promising criterion for TDE. Herein, we generalize this measure to the multichannel case in a similar fashion to the multi-microphone AMDF. Essentially, for a given pair of microphones, the AMSF is maximized when the signals are perfectly aligned. In other words, the synchrony between the outputs of a given pair of microphones (i,j), $i,j \in \{1, \ldots, N\}$, is maximized when the AMSF criterion

$$\tilde{J}_{ij,\mathrm{AMSF}}(\theta) = \frac{1}{Q}\sum_{t=1}^{Q}|x_i(t) + x_j\left(t + \mathcal{F}_{ij}(\theta)\right)| \tag{35}$$

is maximized. Following the same procedure that led to the new generalized multi-microphone AMDF-based criteria, we first define the PAMSM, $\mathbf{S}(\theta)$, such that its (i,j)th entry is given by

$$[\mathbf{S}(\theta)]_{ij} = J_{ij,\mathrm{AMSF}}(\theta). \tag{36}$$

In [14], it was demonstrated that the correlation coefficients between AMDF and AMSF are approximately zero, thereby meaning that both criteria contain supplementary information. This fact is also observed herein in the PAMSM and PAMDM. Indeed, since the maximum synchrony between all pairs is achieved when $\theta = \theta_s$, we expect $\|\mathbf{S}(\theta)\|^2 = \mathrm{tr}\left\{\mathbf{S}(\theta)\mathbf{S}^T(\theta)\right\} = \sum_{n=1}^{N}\gamma_n^2(\theta)$, where $\gamma_1(\theta),\ \gamma_2(\theta),\ \ldots,\ \gamma_N(\theta)$ are the eigenvalues of $\mathbf{S}(\theta)$ sorted such that $|\gamma_1(\theta)| \geq |\gamma_2(\theta)| \geq \cdots \geq |\gamma_N(\theta)|$, to reach its maximum when all pairs are aligned, in contrast to $\|\mathbf{\Delta}(\theta)\|$ which reaches its minimum. Figs. 5 and 6 depict the variations of the absolute value of the first eigenvalue and the inverses of the $N-1$ other eigenvalues $\mathbf{S}(\theta)$ with respect to the azimuth. These results were obtained in the same conditions as in Figs. 3 and 4 discussed in Sections III-B and V-A. We see that $|\gamma_1(\theta)|$ reaches its maximum at θ_s while all other eigenvalues are minimized. To explain this result, recall that in the ideal case (neither channel attenuation nor noise) $\mathbf{S}(\theta_s)$ is of rank one. Thus, similarly to the PSCM, the energy of the PAMSM is maximized and focussed on the maximum eigenvalue when it is steered toward the direction of arrival of the source. Otherwise, it is smeared toward other dimensions. Similar to the PAMDM eigenvalues, it is clear that most of these eigenvalues can be used for source localization. However, the spectra of the smallest eigenvalues exhibit several spikes that have a detrimental effect on source localization as we verified through several numerical examples. We empirically found that the largest eigenvalue is the most

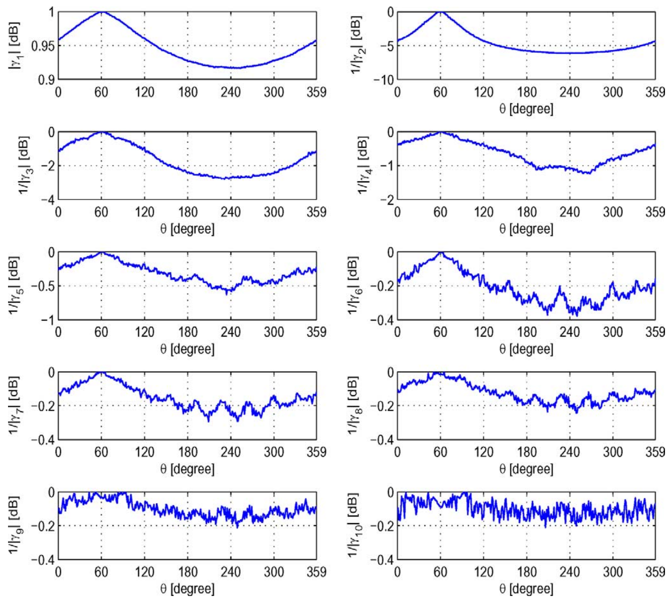


Fig. 5. PAMSM eigenvalues versus azimuth; anechoic enclosure, SNR = 0 dB.

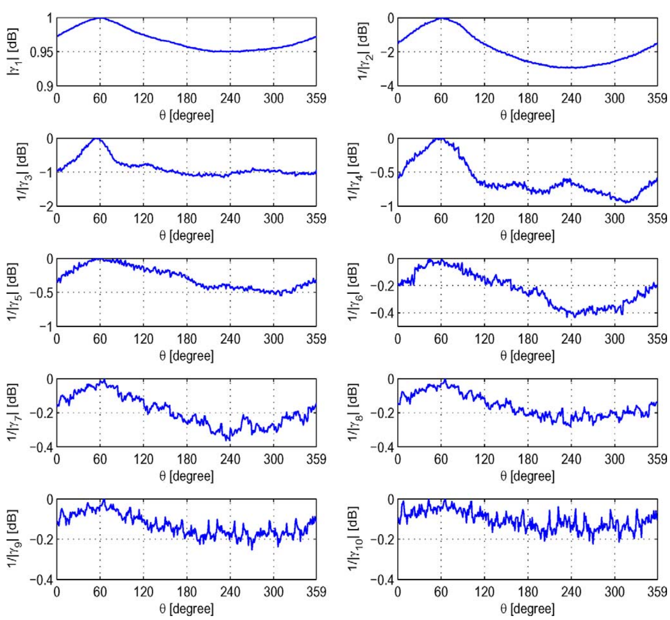


Fig. 6. PAMSM eigenvalues versus azimuth; reverberant enclosure, SNR = 0 dB.

reliable to be used for source localization. Consequently, we define the following new criterion:

$$\mathcal{S}_{\text{EigAMSF}}\left\{\mathbf{S}(\theta)\right\} = |\gamma_1(\theta)|. \tag{37}$$

The other eigenvalues can be used, but they have been empirically verified to provide poorer results due to their sensitivity to reverberation. Finally, we propose another ad-hoc yet simple and accurate multi-microphone AMSF-based criterion for source localization. This criterion is termed, herein, as *steered magnitude sum* (SMS) and is given by

$$\mathcal{S}_{\text{SMS}}\left\{\mathbf{S}(\theta)\right\} = \mathbf{1}^T\mathbf{S}(\theta)\mathbf{1}. \tag{38}$$

In addition to their ability to accurately localize the acoustic source, both criteria $\mathcal{S}_{\text{SMD}}$ and $\mathcal{S}_{\text{SMS}}$ have very low complexity because they simply consist in summing up all the entries of the PAMDM and PAMSM, respectively. These two matrices are, in their turn, simpler to compute than the PSCM since they involve no multiplications. However, the eigendecompositions of the PAMDM and PAMSM are required to, respectively, compute $\mathcal{S}_{\text{EigAMDF}}$ and $\mathcal{S}_{\text{SMS}}$, thereby requiring a computational complexity in the order of $O(N^3)$.

V. Numerical Examples

In this section, we analyze the performance of all the investigated and proposed source localization approaches. We proceed as in [8] and present our results in terms of percentage of anomalies (estimates that differ from the actual angle of arrival by more than 5 degrees), and root mean-square error of non-anomalous azimuth estimates. Moreover, we provide the spatial spectra of all the criteria.

In the investigated scenarios, the speaker is located in a reverberant room with dimensions: length $= 304.8$ cm, width $= 457.2$ cm, and height $= 381$ cm ($x \times y \times z$). The reverberant enclosure is simulated using the modified version of Allen and Berkley's image method [23], [24]. We consider a uniform circular array of $N = 10$ microphones. The center of the circular array is located at (152.4 cm, 228.6 cm, 101.6 cm) and its radius is chosen as to prevent the spatial aliasing for circular arrays (see [8] and references therein). Precisely, we choose $r = c/(4f_{\max}\sin(\pi/N))$. In our case, the highest frequency of the speech signal is $f_{\max} = 4$ kHz. Hence, we have $r \approx 6.9$ cm. The speaker generates a 2-minutes-long female speech. It is situated at a distance 200 cm from the center of the array and forms an azimuthal angle $\theta_s = 60$ degrees. Similar to [8], a white Gaussian noise was added to all sensors with SNR values of 0 and 10 dB. The speech signal is sampled at a rate 48 kHz to achieve a good angular resolution, and the frame length used to estimate the required criteria is 128 ms. To scan the whole plane, the spatial spectra are estimated at every degree over the range $[0, 359]$ degrees. The walls, ceiling, and floor reflection coefficients are set to 0 and 0.75 to respectively model an anechoic and reverberant room with reverberation time $T_{60} = 210$ ms measured using the backward integration method (see [25], Chapter 2 for more details). In what follows, we start by analyzing the eigenvalues of the three parameterized matrices: PSCM, PAMDM, and PAMSM. Then, we compare the performance of all the localization criteria considered in this paper in both anechoic and reverberant environments.

A. Eigenvalues of the Parameterized Matrices

In the classical high-resolution techniques for *narrowband* source localization, a general trend has been to scan all the potential source locations and observe the variations of a certain criterion (or spatial spectrum). For instance, this criterion is known to be the orthogonality between the noise subspace (defined by the lowest eigenvalues and their associated eigenvectors) and the steering vector in MUSIC [15]. Unfortunately, speech signals are wideband by nature and there is no particular form of the steering vector to be used as in narrowband approaches. However, we have shown that the eigenanalysis of

TABLE II
PERCENTAGE OF ANOMALIES OF ALL INVESTIGATED AND PROPOSED LOCALIZATION METHODS: SNR = 0 AND 10 dB, ANECHOIC AND REVERBERANT ROOMS

		$T_{60} = 0$ ms		$T_{60} = 210$ ms	
Category	Criterion	SNR = 0 dB	SNR = 10 dB	SNR = 0 dB	SNR = 10 dB
Crosscorrelation-based	$\mathcal{S}_{\mathrm{SRP}}$	3.71	0	28.43	23.54
	$\mathcal{S}_{\mathrm{MV}}$ (not reg.)	4.60	21.41	32.0	45.0
	$\mathcal{S}_{\mathrm{MV}}$ (reg.)	3.71	0	28.43	23.35
	$\mathcal{S}_{\mathrm{MaxEig}}$	3.53	0	28.24	22.97
	$\mathcal{S}_{\mathrm{MCCC}}$ (not reg.)	3.89	20.17	36.15	55.93
	$\mathcal{S}_{\mathrm{MCCC}}$ (reg.)	3.71	0	28.62	21.84
	$\mathcal{S}_{\mathrm{NSA}}$	3.53	0	28.24	22.97
	$\mathcal{S}_{\mathrm{NSG}}$	4.07	19.46	33.52	55.17
	$\mathcal{S}_{\mathrm{CEig}}$	3.53	0	28.24	22.97
	$\mathcal{S}_{\mathrm{CMCCC}}$	4.07	19.29	32.95	54.04
AMDF-based	$\mathcal{S}_{\mathrm{EigAMDF}}$	3.71	0	30.88	20.28
	$\mathcal{S}_{\mathrm{SMD}}$	3.71	0	30.69	21.71
AMSF-based	$\mathcal{S}_{\mathrm{EigAMSF}}$	6.54	0.17	33.89	25.98
	$\mathcal{S}_{\mathrm{SMS}}$	6.72	0.17	33.71	25.98

PSCM can be of great help in Section III. We further support this fact by analyzing the effect of the steering angle θ on all the eigenvalues of the PSCM matrix in Figs. 1 and 2. Similarly, the eigenvalues of the PAMDM and PAMSM also exhibit remarkable behaviors that can be exploited in acoustic source localization as shown in Figs. 3–6. Note that all the results are obtained by estimating the eigenvalues at the whole range of θ, averaging them over all the data frames, and normalizing them for the sake of clarity. First, we see that while the absolute values of the maximum eigenvalues of the PSCM, $\lambda_1(\theta)$, and PAMSM, $|\gamma_1(\theta)|$, reach their maximum at $\theta_s = 60^\circ$, almost all the other eigenvalues reach their minimum at this particular angle (note that the other spectra correspond to the inverses of these eigenvalues). This is justified by the fact that when both matrices are steered toward the source location, they become almost rank deficient (of rank one in the absence of channel attenuations and noise). The synchrony between the properly time-shifted microphone signals is maximized and the overall captured speech energy is focussed on a single direction that corresponds to the major subspace. The speech signal contribution to the energy contained in the other eigenvalues is minimized when $\theta = \theta_s$. Otherwise, some speech energy is spread over these dimensions, thereby explaining the particular peaks at the source location. Regarding the PAMDM, we have shown that when it is steered toward the source location, its norm is minimal (zero in the absence of channel attenuations and noise). Therefore, the absolute values of all its eigenvalues are reduced when $\theta = \theta_s$. Comparing the behaviors of all eigenvalues for a given parameterized matrix, we notice that the smallest eigenvalues, in contrast to the largest ones, have very irregular spectra with several spurious spikes, which is the case for $1/\lambda_9(\theta)$, $1/\lambda_{10}(\theta)$, $1/|\delta_9(\theta)|$, $1/|\delta_{10}(\theta)|$, $1/|\gamma_9(\theta)|$, and $1/|\gamma_{10}(\theta)|$, for instance. Thus, relying on them may lead to poor location estimates. A good solution would consist in masking the effect of these small eigenvalues as it will be better illustrated in the following subsection. The variations of the largest eigenvalues seem to be more regular and robust to the effects of the reverberation in spite of the remarkable flattening of all the spatial spectra.

B. Performance of the Proposed and Analyzed Localization Methods

We implemented all the PSCM-based methods described in Table I in addition to the criteria in (33), (34), (37), and (38) which are based on the PAMSM and PAMDM[4] and evaluate the effect of the SNR and reverberation on their performance. Note that we also added the regularized MCCC and regularized MV (where we added a positive diagonal loading factor to the PSCM). We notice in Tables II and III that in an anechoic environment and at SNR $= 0$ dB, all the PSCM-based localization methods exhibit very similar accuracy in terms of both percentage of anomalies and RMSE of the non-anomalous source location estimates. When the SNR is increased to 10 dB, we see that the SRP, MaxEig, NSA, CEig criteria yield almost exact source location estimate (0% anomalies and an RMSE of around 0.65°). The performance of the MCCC (without regularization), the NSG, the CMCCC, and the MV (without regularization) are deteriorated. To explain this fact, recall that we have previously shown that the MCCC, the NSG, CMCCC depend on the minor subspace eigenvalues. We have also shown that the smallest eigenvalues exhibit irregular variations with respect to the azimuthal angle. When the SNR is increased to 10 dB, the masking effect of the noise (spatially white) is reduced and the effect of the minor eigenvalues become significant. Regarding the MV, recall that its criterion requires the inversion of the PSCM which becomes problematic when this matrix is ill conditioned. These problems translate into a loss

[4]Note that other criteria can be devised from this eigenanalysis framework. However, the ones investigated herein are empirically found to be the most robust.

TABLE III
RMSE of Non-Anomalous Estimates of All Investigated and Proposed Localization Methods: SNR $= 0$ and 10 dB, Anechoic and Reverberant Rooms

		$T_{60} = 0$ ms		$T_{60} = 210$ ms	
Category	Criterion	SNR $= 0$ dB	SNR $= 10$ dB	SNR $= 0$ dB	SNR $= 10$ dB
Crosscorrelation-based	$\mathcal{S}_{\mathrm{SRP}}$	1.95	0.65	2.75	2.92
	$\mathcal{S}_{\mathrm{MV}}$ (not reg.)	1.99	2.62	2.74	2.90
	$\mathcal{S}_{\mathrm{MV}}$ (reg.)	1.95	0.65	2.75	2.92
	$\mathcal{S}_{\mathrm{MaxEig}}$	1.97	0.64	2.74	2.96
	$\mathcal{S}_{\mathrm{MCCC}}$ (not reg.)	1.98	0.86	3.0	3.92
	$\mathcal{S}_{\mathrm{MCCC}}$ (reg.)	1.96	0.64	2.84	3.0
	$\mathcal{S}_{\mathrm{NSA}}$	1.97	0.64	2.74	2.96
	$\mathcal{S}_{\mathrm{NSG}}$	1.99	0.86	2.99	3.87
	$\mathcal{S}_{\mathrm{CEig}}$	1.97	0.64	2.74	2.96
	$\mathcal{S}_{\mathrm{CMCCC}}$	1.98	0.86	2.92	3.76
AMDF-based	$\mathcal{S}_{\mathrm{EigAMDF}}$	2.06	0.65	2.75	3.0
	$\mathcal{S}_{\mathrm{SMD}}$	2.07	0.66	2.76	2.99
AMSF-based	$\mathcal{S}_{\mathrm{EigAMSF}}$	2.29	1.27	2.77	2.96
	$\mathcal{S}_{\mathrm{SMS}}$	2.29	1.27	2.78	2.95

of accuracy for the three minor-subspace-based methods in addition to the MV. The regularized MCCC and MV exhibit a much more robust behavior and their performance is as good as the SRP, MaxEig, and CEig methods since the regularization factor masks the effect of the smallest eigenvalues for the MCCC and improves the conditioning of the PSCM for the MV. In the reverberant scenario, similar behaviors are observed and the regularization of the PSCM seems important for both the MV and the MCCC to improve their accuracy especially for SNR $= 10$ dB. The new proposed NSA criterion provides comparable and even better accuracy than all other PSCM-based criteria. The geometric averaging of the minor-subspace-associated eigenvalues (NSG) and CMCCC are sensitive to the effect of the smallest eigenvalues and lead to high percentage of anomalies. Moreover, it is quite remarkable that PAMSM and especially the PAMDM-based criteria, namely the SMD, EigAMDF, SMS, and EigAMSF, are also good candidates for source localization in reverberant and anechoic environments. For instance, the two PAMDM-based criterion yield the lowest percentage of anomalies in the reverberant environment at SNR $= 10$ dB.

In Figs. 7–10, we show the spatial spectra of all the investigated methods. We notice that the minor subspace-based methods, especially the MCCC without regularization, the NSG, and the CMCCC, exhibit the best spatial resolutions. The arithmetic averaging over the minor-subspace eigenvalues leads to improved performance as compared to other methods from the same class in terms of percentage of anomalies, but, it results in less spectral resolution. Assigning different weights to the minor and major eigenvalues yields better spectral resolutions as it can be clearly seen when comparing the spectra of the CMCCC to the MCCC, and the CEig to the MaxEig. When the SNR is set to 10 dB, we see that the spectra of the MV, the MCCC, the NSG, and the CMCCC exhibit several spikes due to the inaccuracies caused by the lowest eigenvalues whose effect is dominant when a geometric averaging is used or when the matrix inversion is involved as in the MV criterion. By comparing Figs. 7 and 8 to Figs. 9 and 10, we observe the deleterious effect of the reverberation on the spatial spectra of the minor-subspace-based criteria as it increases the number of their spurious spikes, especially for SNR $= 10$ dB, and flattens the spectra. These two behaviors account for the increased number of anomalies seen with the minor-subspace-based methods. By diagonally loading the PSCM, the spurious spikes disappear at the price of a deteriorated resolution as it can be seen with the spatial spectra of the regularized MV and MCCC.

VI. Conclusion

In this paper, a new eigenanalysis-based framework for broadband source localization was proposed. First, we analyzed and classified cross-correlation-based source localization techniques using the eigenanalysis of the PSCM. This study was motivated by the fact that when the PSCM is steered toward the source location, two subspaces can be identified: the first subspace is spanned by the vector associated with the largest eigenvalue of the PSCM (major subspace) while the second corresponds to the minor subspace and is spanned by the eigenvectors of the PSCM which are associated with all the remaining eigenvalues. By scanning the potential source locations and observing the variations of both types of eigenvalues, we concluded that these eigenvalues bear very useful information on the source location. To gain a better understanding of the functioning of the MV and the SRP, we devised both criteria from a covariance fitting perspective and demonstrated that they essentially consist in tracing the source location in the maximum eigenvalue. Moreover, we showed that the MCCC takes advantage of the minor-subspace-associated eigenvalues. Other criteria combining the PSCM eigevalues were also proposed, namely the CEig, CMCCC, NSA, and NSG. In

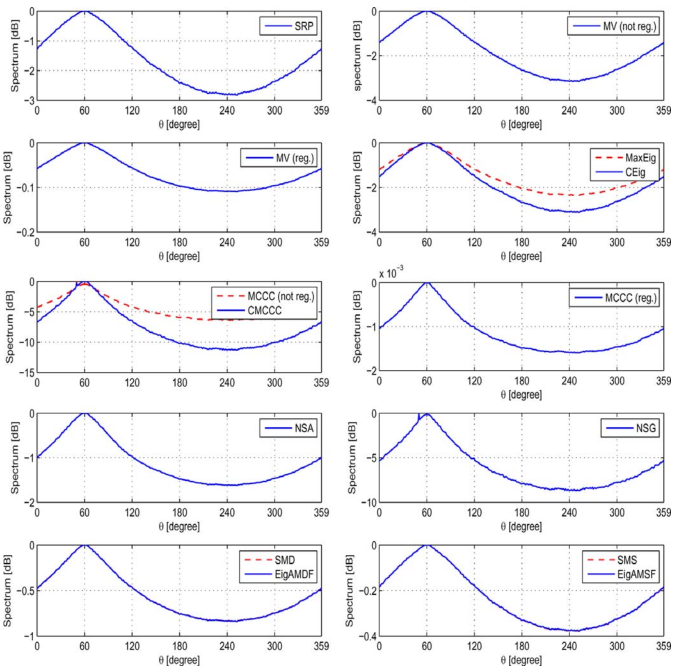


Fig. 7. All spatial spectra versus azimuth; anechoic enclosure, SNR $= 0$ dB.

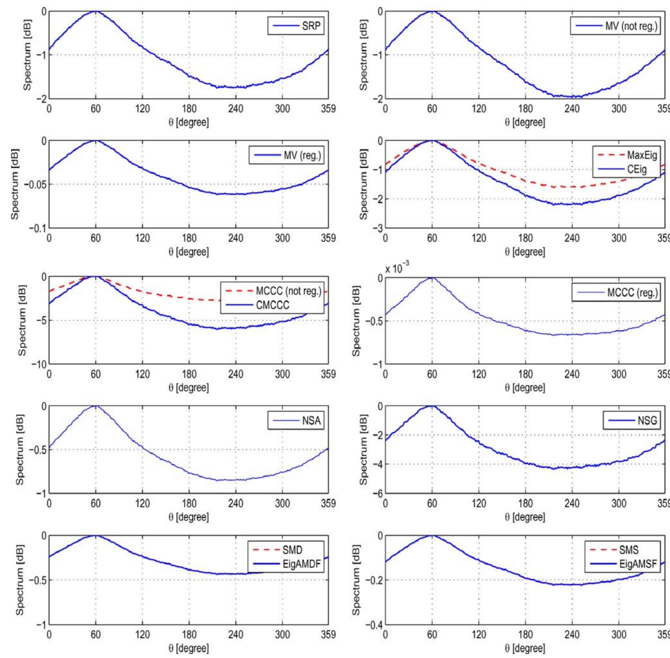


Fig. 9. All spatial spectra versus azimuth; reverberant enclosure, SNR $= 0$ dB.

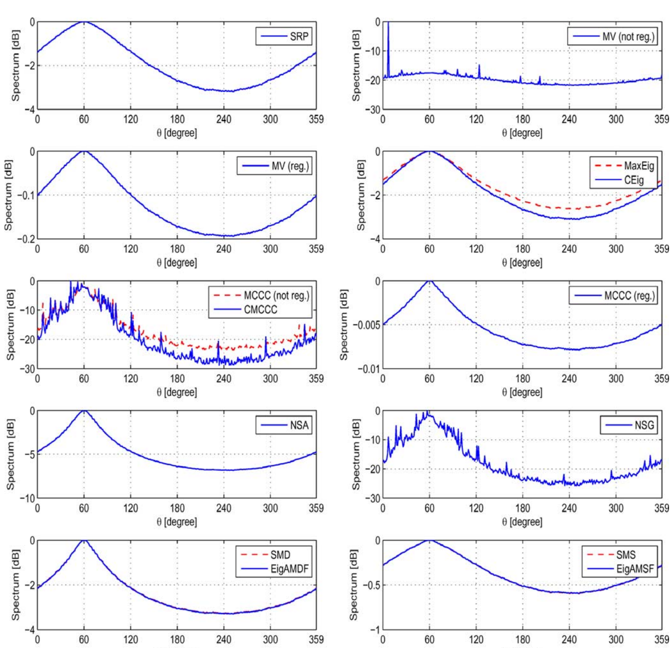


Fig. 8. All spatial spectra versus azimuth; anechoic enclosure, SNR $= 10$ dB.

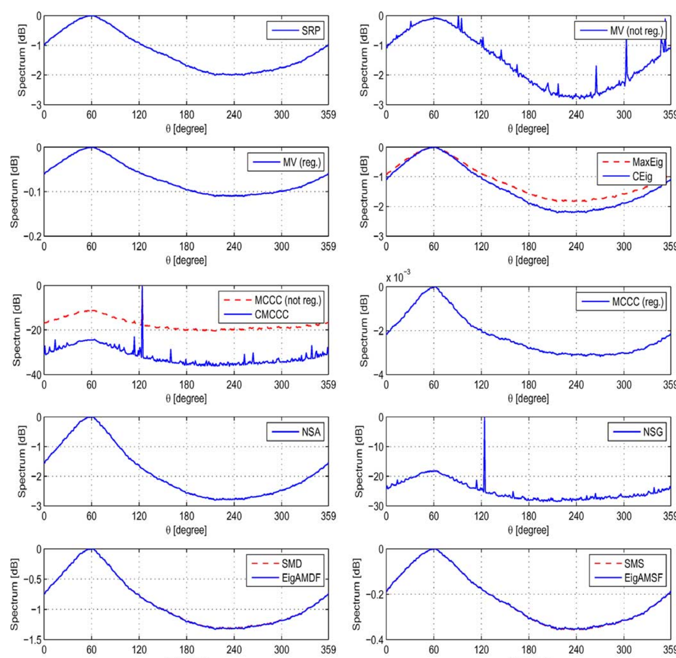


Fig. 10. All spatial spectra versus azimuth; reverberant enclosure, SNR $=$ 10 dB.

the second part of this work, we generalized the AMDF and AMSF to the multi-microphone case and applied both criteria to acoustic source localization. Both criteria aim at maximizing the synchrony between properly time-shifted microphone outputs by calculating the absolute difference or sum, respectively, and require less computational load than cross-correlation approaches. In an analogous fashion to the first part of this work, we proposed two new parameterized matrices, namely, PAMDM and the PAMSM that contain all the combinations of the AMDF and AMSF relating each pair of microphones. The eigenanalysis of both matrices revealed new efficient criteria for source localization. Our numerical evaluations showed that, both types of eigenvalues (minor and major) can be used for source localization. However, the lowest eigenvalues can have a deleterious effect on the accuracy of some algorithms such that the MCCC and the new proposed NSG and CMCCC criteria. Furthermore, the MV requires a matrix inversion. This fact makes it also sensitive to the ill conditioning of the PSCM. By diagonally loading this matrix in this case, an improved accuracy in terms of percentage anomalous estimates can be obtained at the price of decreased spatial resolution. The new proposed contrast-based methods assign different weights to the eigenvalues depending on which subspace they belong to, and exhibit improved spectrum resolutions. Finally, we showed that the new generalized AMSF and especially the AMDF-based methods are good candidates that allow for accurate source localization.

ACKNOWLEDGMENT

The authors would like to thank Dr. J. Dmochowski and Dr. J. Chen for their valuable help with the simulations.

REFERENCES

[1] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Berlin, Germany: Springer-Verlag, 2008.
[2] H. Wang and P. Chu, "Voice source localization for automatic camera pointing system in videoconferencing," in *Proc. IEEE ICASSP*, 1997, pp. 187–190.
[3] 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, 2001.
[4] 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.
[5] 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.
[6] G. Lathoud and M. Magimai-Doss, "A sector-based, frequency-domain approach to detection and localization of multiple speakers," in *Proc. IEEE ICASSP*, 2005, pp. 265–268.
[7] H. Do and H. F. Silverman, "A method for locating multiple sources using a frame of a large-aperture microphone array data without tracking," in *Proc. IEEE ICASSP*, 2008, pp. 301–304.
[8] 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.
[9] J. Benesty, Y. Huang, and J. Chen, "Time delay estimation via minimum entropy," *IEEE Signal Process. Lett.*, vol. 14, no. 3, pp. 157–160, Mar. 2007.
[10] S. Gazor and W. Zhang, "Speech probability distribution," *IEEE Signal Process. Lett.*, vol. 10, no. 7, pp. 204–207, Jul. 2003.
[11] M. Ross, H. Shaffer, A. Cohen, R. Freudberg, and H. Manley, "Average magnitude difference function pitch extractor," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-22, no. 5, pp. 353–362, Oct. 1974.
[12] L. R. Rabiner, M. J. Cheng, A. E. Rosenberg, and C. A. Mc-Gonegal, "A comparative performance study of several pitch detection algorithms," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-24, no. 5, pp. 399–418, Oct. 1976.
[13] G. Jacovetti and G. Scarano, "Discrete time techniques for time delay estimation," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 525–533, Feb. 1993.
[14] J. Chen, J. Benesty, and Y. Huang, "Performance of GCC and AMDF based time delay estimation in practical reverberant environments," *EURASIP J. Appl. Signal Process.*, vol. 2005, pp. 25–36, Jan. 2005.
[15] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagat.*, vol. AP-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] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing applications," *Signal Process.*, vol. 8, pp. 185–210, Apr. 1998.
[18] P. Stoica, J. Li, and X. Tan, "On spatial power spectrum and signal estimation using the Pisarenko framework," *IEEE Trans. Signal Process.*, vol. 56, pp. 5109–5119, Oct. 2008.
[19] 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.
[20] M. S. Brandstein, J. E. Adcock, J. H. DiBiase, and H. F. Silverman, "A closed-form method for finding source locations from microphone-array time-delay estimates," in *Proc. IEEE ICASSP*, 1995, pp. 3019–3022.
[21] K. C. Ho and M. Sun, "An accurate algebraic closed-form solution for energy-based source localization," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. 15, no. 8, pp. 2542–2550, Nov. 2007.
[22] J. P. Dmochowski, J. Benesty, and S. Affes, "Broadband MUSIC: opportunities and challenges for multiple source localization," in *Proc. IEEE WASPAA*, 2007, pp. 18–21.
[23] 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.
[24] P. Peterson, "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," *J. Acoust. Soc. Amer.*, vol. 80, pp. 1527–1529, Nov. 1986.
[25] Y. Huang, J. Benesty, and J. Chen, *Acoustic MIMO Signal Processing*. Berlin, Germany: Springer-Verlag, 2006.

Mehrez Souden (S'07) was born in 1980. He received the Diplôme d'Ingénieur degree in signals and systems (with honors) from the École Polytechnique de Tunisie, Tunis, Tunisia, in 2004 and the M.Sc. degree in telecommunications from the Institut National de la Recherche Scientifique-Énergie, Matériaux, et Télécommunications (INRS-EMT), University of Quebec, Montreal, QC, Canada, in 2006, respectively. He is currently pursuing the Ph.D. degree in telecommunications engineering at the INRS-EMT.

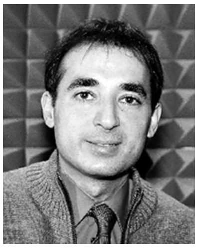
His current research focuses on microphone array processing with an emphasis noise reduction and source localization.

Mr. Souden is the recipient of the Alexander-Graham-Bell Canada graduate scholarship from the National Sciences and Engineering Research Council (2008–2010) and the National grant from the Tunisian Government at the Master and Doctoral Levels.

Jacob Benesty 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 April 1991.

During the Ph.D. degree (from November 1989 to April 1991), he worked on adaptive filters and fast algorithms at the Centre National d'Etudes des Telecommunications (CNET), Paris, France. From January 1994 to July 1995, he was with 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 INRS-EMT, University of Quebec, Montreal, QC, Canada, as a Professor. His research interests are in signal processing, acoustic signal processing, and multimedia communications. He coauthored the books *Noise Reduction in Speech Processing* (Springer-Verlag, 2009), *Microphone Array Signal Processing* (Springer-Verlag, 2008), *Acoustic MIMO Signal Processing* (Springer-Verlag, 2006), and *Advances in Network and Acoustic Echo Cancellation* (Springer-Verlag, 2001). He is the editor-in-chief of the reference *Springer Handbook of Speech Processing* (Springer-Verlag, 2007). He is also a coeditor/coauthor of the books *Speech Enhancement* (Springer-Verlag, 2005), *Audio Signal Processing for Next Generation Multimedia communication Systems* (Kluwer, 2004), *Adaptive Signal Processing: Applications to Real-World Problems* (Springer-Verlag, 2003), and *Acoustic Signal Processing for Telecommunication* (Kluwer, 2000).

Dr. Benesty received the 2001 and 2008 Best Paper Awards from the IEEE Signal Processing Society. He was a member of the editorial board of the *EURASIP Journal on Applied Signal Processing*, a member of the IEEE Audio and Electroacoustics Technical Committee, and the cochair of the 1999 International Workshop on Acoustic Echo and Noise Control (IWAENC). He is the general cochair of the 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA).



Sofiène Affes (S'94–M95–SM04) received the Diplôme d'Ingénieur in electrical engineering in 1992, and the Ph.D. degree with honors in signal processing in 1995, both from theÉcole Nationale Supérieure des Télécommunications (ENST), Paris, France.

He has been since with INRS-EMT, University of Quebec, Montreal, QC, Canada, as a Research Associate from 1995–1997, then as an Assistant Professor until 2000. Currently, he is an Associate Professor in the Wireless 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-EMT, Montreal. Since 2004, he has been actively involved in major projects in wireless of PROMPT (Partnerships for Research on Microelectronics, Photonics and Telecommunications).

Prof. Affes was the corecipient of the 2002 Prize for Research Excellence of INRS. He currently holds a Canada Research Chair in Wireless Communications and a Discovery Accelerator Supplement Award from NSERC (Natural Sciences and Engineering Research Council of Canada). In 2006, Professor Affes served as a General Co-Chair of the IEEE VTC'2006-Fall Conference, Montreal. In 2008, he received from the IEEE Vehicular Technology Society the IEEE VTC Chair Recognition Award for exemplary contributions to the success of IEEE VTC. He currently acts as a member of the Editorial Board of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and of the *Wiley Journal on Wireless Communications and Mobile Computing*.