EIGENANALYSIS-BASED BROADBAND SOURCE LOCALIZATION

Mehrez Souden, Jacob Benesty, and Sofiène Affes
INRS-ÉMT, 800, de la Gauchetière Ouest, Suite 6900, Montréal, H5A 1K6, QC, Canada.
{souden,benesty,affes}@emt.inrs.ca

ABSTRACT

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 crosscorrelation-based criteria such as the steered response power (SRP), minimum variance (MV), and multichannel crosscorrelation (MCCC), this synchrony can be measured using the averaged magnitude difference function (AMDF) and the averaged magnitude sum function (AMSF) whose calculations involve low computational cost. In this paper, we study the crosscorrelation and AMDF (with AMSF) based approaches using an arbitrary number of microphones. Specifically, we use the eigenanalysis of the parameterized spatial correlation matrix (PSCM) to first provide a unifying study of the most popular crosscorrelation-based techniques. Then, we show the efficiency of the AMDF and AMSF in localizing an acoustic source using multiple microphones by proposing 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 two matrices reveals new criteria.

***Index Terms*—** Source localization, microphone array, eigenanalysis.

1. INTRODUCTION

The process of acoustic source localization consists in measuring the synchrony between properly delayed (noise-free) microphone outputs. Consequently, current acoustic source localization methods can be classified into three main categories. First, the most popular techniques are based on the second-order-statistics of the microphone outputs such as the SRP [1], MV [2], and MCCC [3]. The second category consists of methods that measure the synchrony between the processed microphone outputs from an information theoretic point of view [4]. Nevertheless, the speech is generally assumed to be Laplacian-distributed [5] and intricate calculations are required to estimate the mutual information of the multivariate Laplacian-distributed microphone outputs. The final category consists of methods based on simpler criteria since neither second-order-statistics nor 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. Both criteria have been previously applied for time delay estimation (TDE) [6, 7] using a single pair of microphones. But they have not been generalized to the multiple microphone case with application to source localization yet. This contribution is concerned with the first and third categories.

In this paper, we first analyze and classify the crosscorrelation-based broadband source localization techniques from an eigenanalysis perspective. The underlying idea of the proposed approach 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 associated with the largest eigenvalue of the PSCM, while the second is defined by its remaining eigenvalues. We take advantage of the variations of both types of eigenvalues to gain a better understanding of the functioning of existing methods and propose new ones. Our second contribution consists in generalizing the AMDF and AMSF to the localization of acoustic sources using multiple microphones. For a given pair of microphones, both criteria aim at maximizing the synchrony between time-shifted output signals by calculating the absolute difference and sum, respectively. In an analogous fashion to the crosscorrelation-based framework, we propose two new parameterized matrices, namely, the PAMDM and PAMSM that contain all the combinations of the AMDF and AMSF relating each pair. The eigenanalysis of both matrices reveals new efficient criteria for source localization.

2. PROBLEM STATEMENT

Let $s(t)$ denote a signal generated by a broadband source and captured by an array of N microphones. We are interested in estimating the azimuthal angle of arrival, θ_s, of this source which is assumed to lie in the far-field (for simplicity, we also assume that the source and microphone array are located on the same plane). The output of the nth ($n = 1, ..., N$) microphone is given by

$$x_n(t) \quad = \quad a_n s\left[t - \tau_n(\theta_s)\right] + v_n(t), \tag{1}$$

where a_n is the channel attenuation, $v_n(t)$ is an additive noise, and $\tau_n(\theta_s)$ is the propagation time delay from the source to the nth microphone element which is a function of the source location and array geometry. The time difference of arrival (TDOA) between pairs of microphones is commonly used for localization [1, 8] and is denoted as $\mathcal{F}_{nm}(\cdot)$ for a pair (n, m). For a given parameter θ, we define

$$\mathbf{x}(t,\theta) = \left[x_1[t]\; x_2[t + \mathcal{F}_{12}(\theta)] \;\cdots\; x_N[t + \mathcal{F}_{1N}(\theta)]\right]^T, \tag{2}$$

where $x_n[t+\mathcal{F}_{1n}(\theta)] = a_n s[t-\tau_n(\theta_s)+\mathcal{F}_{1n}(\theta)]+v_n[t+\mathcal{F}_{1n}(\theta)]$.

3. CROSS-CORRELATION-BASED METHODS

The PSCM is defined as

$$\mathbf{R}_{xx}(\theta) = E\left\{\mathbf{x}(t,\theta)\mathbf{x}^T(t,\theta)\right\} = \mathbf{A}\mathbf{R}_{ss}(\theta)\mathbf{A} + \mathbf{R}_{vv}(\theta), \tag{3}$$

where $\mathbf{A} = \mathrm{diag}\left[\mathbf{a}\right]$ and $\mathbf{a} = \left[a_1, a_2, ..., a_N\right]^T$. $\mathbf{R}_{ss}(\theta)$ 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 $\left[\mathbf{R}_{xx}(\theta)\right]_{i,j} = r_{x_i x_j}[\mathcal{F}_{ij}(\theta)] = a_i a_j r_{ss}[\mathcal{F}_{ij}(\theta) - \mathcal{F}_{ij}(\theta_s)] + r_{v_i v_j}[\mathcal{F}_{ij}(\theta)]$, where $r_{x_i x_j}(\tau) = E\left\{x_i(t)x_j(t-\tau)\right\}$ for a given time delay τ. The general procedure for parameterized processing that we consider in this paper consists in investigating the broadband spatial spectrum, denoted herein as $\mathcal{S}\left[\mathbf{R}_{xx}(\theta)\right]$, and identifying its peak that corresponds to the source location. Formally, this approach consists in estimating θ_s as [8]

$$\hat{\theta}_s = \arg\max_{\theta} \mathcal{S}\left[\mathbf{R}_{xx}(\theta)\right]. \tag{4}$$

To date, several criteria have been proposed such as the SRP [1], MV [2], and MCCC [3]. In the following, we show that all these methods can be devised from a general eigenanalysis-based framework.

Justifying the Eigenanalysis-Based Framework: In narrowband high-resolution techniques such as MUSIC [9], the knowledge of the so-called steering vector (that lies in the signal subspace when $\theta = \theta_s$) allows one to determine the DOA. 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 (1). Fortunately, the eigenanalysis of $\mathbf{R}_{xx}(\theta)$ can be of great help. Indeed, when $\theta = \theta_s$, we have

$$\mathbf{R}_{xx}(\theta_s) = \sigma_s^2 \mathbf{a}\mathbf{a}^T + \mathbf{R}_{vv}(\theta_s), \tag{5}$$

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 with dimensions 1 and $N-1$, respectively. Consequently, for a given θ, we decompose $\mathbf{R}_{xx}(\theta)$ as $\mathbf{R}_{xx}(\theta) = \mathbf{U}^T(\theta)\mathbf{D}(\theta)\mathbf{U}(\theta)$, where $\mathbf{U}(\theta)$ is a unitary matrix, $\mathbf{D}(\theta) = \mathrm{diag}\left[\lambda_1(\theta), ..., \lambda_N(\theta)\right]$, and $\lambda_1(\theta), ..., \lambda_N(\theta)$ are the eigenvalues of $\mathbf{R}_{xx}(\theta)$ sorted in a decreasing order. Now, by ignoring the noise component in (1), we see that when the PSCM is steered toward θ_s, the major eigenvalue of $\mathbf{R}_{xx}(\theta_s)$ is $\lambda_1(\theta_s) = \sigma_s^2 \sum_{n=1}^{N} a_n^2$. Hence, $\lambda_1(\theta_s)$ captures the overall energy of the received signal (including the channel effect). Let us further consider the particular case where the desired source is temporally white with identically distributed components. In this case, when $\theta \neq \theta_s$, $\mathbf{R}_{xx}(\theta_s) = \sigma_s^2 \mathrm{diag}\left(a_1^2, ..., a_N^2\right)$. Obviously, $\lambda_n(\theta) = \sigma_s^2 a_n^2$ and $\lambda_1(\theta_s) = \sum_{n=1}^{N} \lambda_n(\theta)$. Hence, when $\theta \neq \theta_s$, 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. 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 focussed on a single one when $\theta = \theta_s$.

Analysis of the Major Eigenvalue: When $\theta = \theta_s$, there is only one dominant eigenvalue (supposing that the noise is weak enough or has 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 has been partially exploited in [8] by observing the maximum eigenvalue and choosing this criterion

$$\mathcal{S}_{\mathrm{MaxEig}}\left[\mathbf{R}_{xx}(\theta)\right] = \lambda_1(\theta). \tag{6}$$

Now, we demonstrate that the SRP and MV aim in essence at analyzing the major eigenvalue. To this end, we further suppose the knowledge of $\mathbf{a}$ (also assumed to have a unit norm without loss of generality) and try to find an estimate of $\lambda_1(\theta_s)$ knowing a certain estimate of $\mathbf{R}_{xx}(\theta_s)$, say $\hat{\mathbf{R}}_{xx}$. From a covariance fitting perspective [10], this can be easily formulated as

$$\hat{\lambda}_1 \quad = \quad \arg\min_{\lambda} \left\| \lambda \mathbf{a}\mathbf{a}^T - \hat{\mathbf{R}}_{xx} \right\|^2, \tag{7}$$

or

$$\hat{\lambda}_1' \quad = \quad \arg\min_{\lambda} \left\| (\lambda \mathbf{a}\mathbf{a}^T)^{\#} - \hat{\mathbf{R}}_{xx}^{-1} \right\|^2, \tag{8}$$

where $^{\#}$ denotes the pseudo-inverse of a matrix. The straightforward solutions to these two optimization problems are respectively

$$\hat{\lambda}_1 = \mathbf{a}^T \hat{\mathbf{R}}_{xx} \mathbf{a}, \tag{9}$$

and

$$\hat{\lambda}_1' = \left(\mathbf{a}^T \hat{\mathbf{R}}_{xx}^{-1} \mathbf{a}\right)^{-1}. \tag{10}$$

If we assume that there is no channel attenuation in the data model (1), i.e., $\mathbf{a} = \mathbf{1} = [1\ ...\ 1]^T$, and replace $\hat{\mathbf{R}}_{xx}$ by the best available estimate of this matrix at a given direction θ, i.e., $\mathbf{R}_{xx}(\theta)$, we obtain the well known SRP and MV criteria [1, 2]

$$\hat{\lambda}_1(\theta) \quad = \quad \mathcal{S}_{\mathrm{SRP}}\left[\mathbf{R}_{xx}(\theta)\right] = \mathbf{1}^T \mathbf{R}_{xx}(\theta)\mathbf{1}, \tag{11}$$

$$\hat{\lambda}_1'(\theta) \quad = \quad \mathcal{S}_{\mathrm{MV}}\left[\mathbf{R}_{xx}(\theta)\right] = \left[\mathbf{1}^T \mathbf{R}_{xx}^{-1}(\theta)\mathbf{1}\right]^{-1}, \tag{12}$$

up to constant scaling factors, respectively.

Analysis of the Minor Eigenvalues: In the absence of noise, we have $\lambda_2(\theta_s) = \cdots = \lambda_N(\theta_s) = 0$. But, when the noise is present, 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 calculated using $\lambda_2(\theta), ..., \lambda_N(\theta)$ by either geometric or arithmetic averaging

$$\mathcal{S}_{\mathrm{NSA}}\left[\mathbf{R}_{xx}(\theta)\right] \quad = \quad \left[\frac{1}{N-1}\sum_{n=2}^{N} \lambda_n(\theta)\right]^{-1}, \tag{13}$$

$$\mathcal{S}_{\mathrm{NSG}}\left[\mathbf{R}_{xx}(\theta)\right] \quad = \quad \left[\prod_{n=2}^{N} \lambda_n(\theta)\right]^{-1/(N-1)}. \tag{14}$$

The subscripts NSA and NSG stand for "noise subspace arithmetic" and "noise subspace geometric" averaging, respectively. Another notable acoustic source localization method was proposed in [3] where Benesty et al. took advantage of the linear spatial predictability of the noise-free microphone signals from each other (multiple redundancies) to develop a new criterion that essentially minimizes the determinant of the PSCM. It is known that

$$\det\left[\mathbf{R}_{xx}(\theta)\right] = \prod_{n=1}^{N} \lambda_n(\theta). \tag{15}$$

Again, we find that the eigenanalysis allows for defining another well known criterion for source localization. Also, 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).

Common Eigenalysis Framework: 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

$$\mathcal{S}_{\mathrm{GSA}}\left[\mathbf{R}_{xx}(\theta)\right] \quad = \quad \sum_{n=1}^{N} \alpha_n \lambda_n^{\nu_n}(\theta), \tag{16}$$

$$\mathcal{S}_{\mathrm{GSG}}\left[\mathbf{R}_{xx}(\theta)\right] \quad = \quad \prod_{n=1}^{N} \lambda_n^{\beta_n}(\theta), \tag{17}$$

where GSA and GSG stand for "generalized spectrum arithmetic" and "generalized spectrum geometric" averaging, respectively. α_n, ν_n, and β_n , $n \in \{1, ..., N\}$, are some multiplicative and exponential weighting factors that have to be chosen properly to define the MV, SRP, MCCC, MaxEig, NSA, and NSG criteria. 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 can simply assign the same weights to the minor eigenvalues as given below

$$\mathcal{S}_{\text{CEig}}\left[\mathbf{R}_{xx}(\theta)\right] = \lambda_1(\theta) - \frac{1}{N-1}\sum_{n=2}^{N}\lambda_n(\theta) \quad (18)$$

$$\mathcal{S}_{\text{CMCCC}}\left[\mathbf{R}_{xx}(\theta)\right] = \lambda_1(\theta)\prod_{n=2}^{N}\lambda_n^{-1}(\theta). \quad (19)$$

The above criteria are based on the contrast between both subspaces which reaches its maximum when the matrix is steered toward θ_s.

4. AMDF AND AMSF-BASED METHODS

The AMDF is a well-known criterion in pitch estimation literature [11, 12] and has been recently shown to offer good performance with low complexity when applied to TDE for a pair of microphones [6, 7]. In [7], Chen et al. showed that the AMSF is another promising simple and accurate synchrony measure for TDE for two microphones. Here, we take advantage of both criteria to localize acoustic sources using multiple microphones. Similar to the crosscorrelation-based approaches, we process the parameterized vector $\mathbf{x}(t,\theta)$ using AMDF or AMSF-based criteria.

The AMDF criterion is less complex than the cross-correlation-based criteria since it involves no multiplications. Specifically, it is defined for a given pair (i,j) of microphones, $i,j \in \{1,...,N\}$, as

$$J_{ij,\text{AMDF}}(\theta) = E\left\{|x_i[t] - x_j[t+\mathcal{F}_{ij}(\theta)]|\right\}. \quad (20)$$

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 $[\mathbf{\Delta}(\theta)]_{ij} = J_{ij,\text{AMDF}}(\theta)$. For equal channel attenuation coefficients and no additive noise in the data model (1), $\|\mathbf{\Delta}(\theta)\|^2 \to 0$ when $\theta \to \theta_s$ where $\|\cdot\|$ denotes any matrix norm. In particular, we consider the norm $\|\mathbf{\Delta}(\theta)\|^2 = \text{tr}\left[\mathbf{\Delta}(\theta)\mathbf{\Delta}^T(\theta)\right] = \sum_{n=1}^{N}\delta_n^2(\theta)$, where $\delta_n(\theta), n = 1,...,N$, 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 $|\delta_1(\theta)|,\ |\delta_2(\theta)|,\ ...,\ |\delta_N(\theta)|$. It can be empirically verified that most of these eigenvalues have regular variations that can be used for localization. However, we found that the largest eigenvalue is the most reliable to be used as a criterion. Consequently, we propose the first new multiple microphone AMDF-based criterion

$$\mathcal{S}_{\text{EigAMDF}}(\theta) = \frac{1}{|\delta_1(\theta)|}. \quad (21)$$

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

$$\mathcal{S}_{\text{SMD}}(\theta) = \frac{1}{\mathbf{1}^T\mathbf{\Delta}(\theta)\mathbf{1}}. \quad (22)$$

In [7], it was shown that the AMSF is maximized when the signals are perfectly aligned. In our case, the synchrony between the outputs of a given pair of microphones (i,j), $i,j \in \{1,\ ...,\ N\}$, is maximized when the AMSF criterion

$$J_{ij,\text{AMSF}}(\theta) = E\left\{|x_i[t] + x_j[t+\mathcal{F}_{ij}(\theta)]|\right\} \quad (23)$$

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 $[\mathbf{S}(\theta)]_{ij} = J_{ij,\text{AMSF}}(\theta)$. In [7], 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 $\|\mathbf{S}(\theta)\|^2 = \text{tr}\left[\mathbf{S}(\theta)\mathbf{S}^T(\theta)\right] = \sum_{n=1}^{N}\gamma_n^2(\theta)$, where $\gamma_1(\theta),\ \gamma_2(\theta),\ ...,\ \gamma_N(\theta)$ 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. It can be empirically verified that most of these eigenvalues have regular variations that can be used for localization. However, we found that that the largest eigenvalue is the most reliable to be used for source localization. Consequently, we define the following new criterion

$$\mathcal{S}_{\text{EigAMSF}}(\theta) = |\gamma_1(\theta)|. \quad (24)$$

The other eigenvalues can be used. But they have been empirically verified to provide poorer results due to their sensitivity to reverberation and noise. 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

$$\mathcal{S}_{\text{SMS}}(\theta) = \mathbf{1}^T\mathbf{S}(\theta)\mathbf{1}. \quad (25)$$

5. NUMERICAL EXAMPLES

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 Allen and Berkley's image method [13, 14]. We consider a uniform circular array of $N = 10$ microphones whose center is located at $(152.4, 228.6, 101.6)$ cm and its radius is chosen as $r = 6.9$ cm. The speaker is a 2 minutes-long female speech and situated at a distance 200 cm from the center of the array and forms an azimuthal angle $\theta_s = 60$ degrees. 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. 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 (with reverberation time $T_{60} = 210$ ms) environments. The results are presented 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.

In Fig. 1, we show the variations of the first and last two eigenvalues of the three matrices, e.g., PSCM, PAMDM, and PAMSM. We see that $\lambda_1(\theta)$ and $|\gamma_1(\theta)|$ reach their maximum while $|\delta_1(\theta)|$ is minimized at θ_s. To explain this result, recall that in the ideal case (neither channel attenuation nor noise) $\mathbf{R}_{xx}(\theta_s)$ and $\mathbf{S}(\theta_s)$ are each of rank one. Thus, the energy of the PSCM (or 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. In contrast, the norm of the PAMDM is minimized at θ_s. This explains the special peak of $1/|\delta_1(\theta)|$ at θ_s. It is also clear that almost all other eigenvalues reach their minimum at the source location and can consequently be used for source localization. However, the spectra of the smallest eigenvalues exhibit several spikes that have a detrimental effect on source localization. The results of Tables 1 and 2 show that when 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 estimates. The performance of the MCCC, NSG, CMCCC, and MV are deteriorated. To explain this fact, recall that we have previously shown that the MCCC, NSG, and CMCCC depend on the minor subspace eigenvalues which may exhibit irregular variations. 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. Also, recall that the MV requires the inversion of the PSCM which becomes problematic when this matrix is ill conditioned. Both problems translate into inaccuracies for the three minor-subspace-based methods in addition to the MV. The regularized (the regularization consists in diagonally loading the PSCM) MCCC and MV exhibit a more robust behavior and perform as well as the SRP, MaxEig, and CEig since the regularization factor masks the effect of the smallest eigenvalues for the MCCC and improves the conditioning of the PSCM for the MV. The proposed NSA criterion provides comparable and even better accuracy than all other PSCM-based criteria. The NSG and CMCCC are sensitive to the effect of the smallest eigenvalues and lead to high percentage of anomalies for high SNR. Finally, it is quite remarkable that 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 $\mathrm{SNR} = 10$ dB.

6. CONCLUSION

We proposed a new eigenanalysis-based framework for broadband source localization. First, we analyzed and classified crosscorrelation-based source localization techniques using the eigenanalysis of the PSCM. By observing the variations of minor and major PSCM eigenvalues, we concluded that they bear very useful information about the source location. Then, we generalized the AMDF and AMSF to the multi-microphone case and applied both criteria to source localization. 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 localization criteria.

7. REFERENCES

[1] J.H. Dibiase, *A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays.* Ph.D. thesis, Brown university, source DAI-B 61/09, Mar. 2001.

[2] J. Krolik and D. Swingler, "Multiple broad-band source location using steered covariance matrices," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 37, pp. 1481–1494, Oct. 1989.

[3] J. Benesty, J. Chen, and Y. Huang, "Time delay estimation via linear interpolation and crosscorrelation," *IEEE Trans. Speech Audio Processing,* vol. 12, pp. 509–519, Sept. 2004.

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

[5] S. Gazor and W. Zhang, "Speech probability distribution," *IEEE Signal Process. Lett.,* vol. 10, pp. 204–207, Jul. 2003.

[6] G. Jacovetti and G. Scarano, "Discrete time techniques for time delay estimation," *IEEE Trans. Signal Process.*, vol. 41, pp. 525–533, Feb. 1993.

[7] J. Chen, J. Benesty, and Y. Huang, "Performance of GCC and AMDF based time delay estimation in practical reverberant environments," *EURASIP J. Applied Signal Process.*, vol. 2005, pp. 25–36, Jan. 2005.

[8] J. Dmochowski, J. Benesty, and S. Affes, "Direction of arrival estimation using the parameterized spatial correlation matrix," *IEEE Trans. Audio, Speech, Language Processing,* vol. 15, pp. 1327–1339, May 2007.

[9] R. O. Schmidt " Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagat.*, vol. AP-34, pp. 276–280, Mar. 1986.

[10] 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.

[11] M. Ross, H. Shaffer, A. Cohen, R. Freudberg, and H. Manley, "Average magnitude difference function pitch extractor," *IEEE Trans. Acoustics Speech Signal Process.*, vol. 22, 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. Acoustics, Speech, and Signal Process.*, vol. 24, pp. 399–418, Oct. 1976.

[13] 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.

[14] 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.

	$T_{60} = 0$ ms		$T_{60} = 210$ ms	
SNR [dB]	0	10	0	10
$\mathcal{S}_{\mathrm{SRP}}$	3.71	0	28.43	23.54
$\mathcal{S}_{\mathrm{MV}}$	4.60	21.41	32.0	45.0
$\mathcal{S}_{\mathrm{MV}}$(regularized)	3.71	0	28.43	23.35
$\mathcal{S}_{\mathrm{MaxEig}}$	3.53	0	28.24	22.97
$\mathcal{S}_{\mathrm{MCCC}}$	3.89	20.17	36.15	55.93
$\mathcal{S}_{\mathrm{MCCC}}$(regularized)	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
$\mathcal{S}_{\mathrm{EigAMDF}}$	3.71	0	30.88	20.28
$\mathcal{S}_{\mathrm{SMD}}$	3.71	0	30.69	21.71
$\mathcal{S}_{\mathrm{EigAMSF}}$	6.54	0.17	33.89	25.98
$\mathcal{S}_{\mathrm{SMS}}$	6.72	0.17	33.71	25.98

Table 1. Percentage of anomalies.

	$T_{60} = 0$ ms		$T_{60} = 210$ ms	
SNR [dB]	0	10	0	10
$\mathcal{S}_{\mathrm{SRP}}$	1.95	0.65	2.75	2.92
$\mathcal{S}_{\mathrm{MV}}$	1.99	2.62	2.74	2.90
$\mathcal{S}_{\mathrm{MV}}$(regularized)	1.95	0.65	2.75	2.92
$\mathcal{S}_{\mathrm{MaxEig}}$	1.97	0.64	2.74	2.96
$\mathcal{S}_{\mathrm{MCCC}}$	1.98	0.86	3.0	3.92
$\mathcal{S}_{\mathrm{MCCC}}$(regularized)	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
$\mathcal{S}_{\mathrm{EigAMDF}}$	2.06	0.65	2.75	3.0
$\mathcal{S}_{\mathrm{SMD}}$	2.07	0.66	2.76	2.99
$\mathcal{S}_{\mathrm{EigAMSF}}$	2.29	1.27	2.77	2.96
$\mathcal{S}_{\mathrm{SMS}}$	2.29	1.27	2.78	2.95

Table 2. RMSE of non-anomalous estimates.

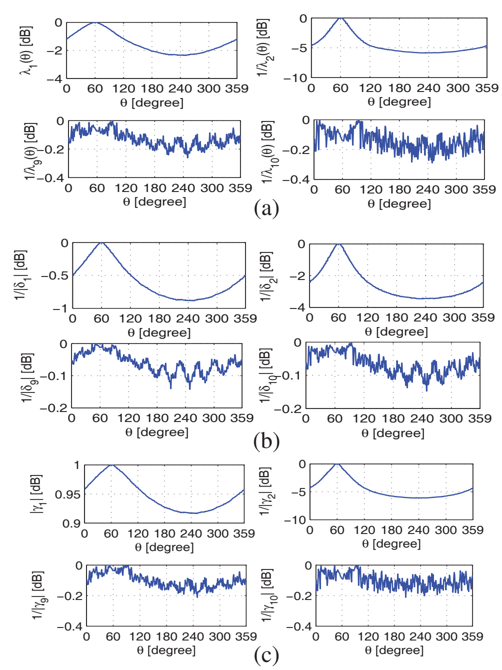

Fig. 1. Eigenvalues vs. θ: (a) PSCM, (b) PAMDM, and (c) PAMSM.