

DIRECTION OF ARRIVAL ESTIMATION USING EIGENANALYSIS OF THE PARAMETERIZED SPATIAL CORRELATION MATRIX

Jacek Dmochowski, Jacob Benesty, and Sofiène Affes

Université du Québec, INRS-EMT
800 rue de la Gauchetière Ouest
Montréal, Québec, Canada, H5A 1K6
{dmochow,benesty,affes}@emt.inrs.ca

ABSTRACT

The estimation of the direction-of-arrival (DOA) of one or more acoustic sources is an area that has generated much interest in recent years, with applications like automatic video camera steering and multi-party stereophonic teleconferencing entering the market. Time-difference-of-arrival (TDOA) based methods compute each relative delay using only two microphones, even though additional microphones are usually available, and thus suffer from the effects of background noise and reverberation. This paper deals with DOA estimation based on spatial spectral estimation, and proposes a novel DOA estimator based on the eigenvalues of the parameterized spatial correlation matrix. Simulation results confirm the ability of the proposed method to provide reliable estimates even in heavily reverberant environments.

Index Terms— Microphone arrays, DOA estimation, source localization, circular array, spatial correlation matrix.

1. INTRODUCTION

The location of a signal source is of much interest in many applications, and there exists a large and increasing need to locate and track sound sources. For example, a speech enhancing beamformer [1], [2] must continuously monitor the position of the speaker in order to provide the desired directivity and interference suppression.

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

The TDOA approach is based on the relationship between DOA and relative delays across the array. The problem of estimating these relative delays is termed “time delay estimation” [3]. The generalized cross-correlation (GCC) approach of [4], [5] is the most popular time delay estimation technique. Each computed time delay is derived from only two microphones, and often contains significant levels of corrupting noise and interference. It is thus well-known that current TDOA-based DOA estimation algorithms are plagued by the effects of noise and especially reverberation.

This paper focuses on DOA estimation based on spatial spectral estimation, proposing a DOA estimator based on the eigenanalysis of the parameterized spatial correlation matrix, which jointly utilizes

This work is supported by the National Sciences and Engineering Research Council of Canada (NSERC).

all microphones. Its relation to minimum variance spectral estimation is also examined, and simulation results show that the proposed methods yield a substantially lower anomaly rate in reverberant environments than that of TDOA-based methods. It is also shown that adding more microphones helps to combat the effects of reverberation.

2. SIGNAL MODEL

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

$$x_l[n] = \alpha_l s[n - t - f_l(\theta)] + v_l[n], \quad (1)$$

where α_l , $l = 0, 1, 2, \dots, L$, are the attenuation factors due to channel effects, t is the propagation time, in samples, from the unknown source $s[n]$ to microphone 0, $v_l[n]$ is the additive noise signal at the l th microphone and includes reverberation, and $f_l(\theta)$, $l = 0, 1, \dots, L$, is the relative delay between microphones 0 and l . The function f_l relates the angle of arrival to the relative delays between microphone elements 0 and l , and is derived for the case of an equispaced circular array in the following manner. When operating in the far-field, the time delay between microphone l and the centre of the array is given by [6]

$$g_l(\theta) = rc^{-1} \cos(\theta - \psi_l), \quad (2)$$

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

$$f_l(\theta) = g_0(\theta) - g_l(\theta) = rc^{-1} \left[\cos \theta - \cos \left(\theta - \frac{2\pi l}{L+1} \right) \right]. \quad (3)$$

3. PROPOSED APPROACH

3.1. Steered Response Power and the Parameterized Spatial Correlation Matrix

Using the model of Section 2, the output of a conventional delay-and-sum beamformer (DSB) steered to an angle of arrival of ϕ is given as

$$z_\phi[n] = \sum_{l=0}^L w_{\phi,l} x[n + f_l(\phi)]. \quad (4)$$

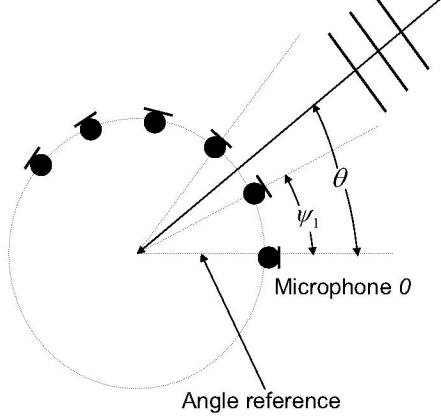


Fig. 1. Circular array geometry.

The delays $f_l(\phi)$ steer the beamformer to the desired DOA, while the angle-dependent beamformer weights $w_{\phi,l}$ help shape the beam accordingly. The estimate of the spatial spectral power at azimuth angle ϕ is given by the power of the beamformer output when steered to azimuth ϕ [7]:

$$S(\phi) = E\{z_\phi^2[n]\}, \quad (5)$$

which using (4) may be written in matrix notation as

$$S(\phi) = \mathbf{w}_\phi^T \mathbf{R}_\phi \mathbf{w}_\phi, \quad (6)$$

where

$$\mathbf{w}_\phi = [w_{\phi,0} \quad w_{\phi,1} \quad \cdots \quad w_{\phi,L}]^T, \quad (7)$$

and

$$\mathbf{R}_\phi = E\{\mathbf{x}_\phi[n]\mathbf{x}_\phi^T[n]\}. \quad (8)$$

is the parameterized spatial correlation matrix (proposed in a time delay estimation context in [8], [9]), which should not be confused with the array observation matrix that is commonly used in narrowband beamforming models. Each off-diagonal entry of \mathbf{R}_ϕ is a single cross-correlation term and a function of the azimuth angle ϕ . Notice that the various microphone pairs are combined in a *joint* fashion, in that altering the steering angle ϕ affects all off-diagonal entries of \mathbf{R}_ϕ .

The DOA estimate is thus given by

$$\hat{\theta} = \arg \max_{\phi} \mathbf{w}_\phi^T \mathbf{R}_\phi \mathbf{w}_\phi. \quad (9)$$

The question of how to choose the weights \mathbf{w}_ϕ remains. Notice that from (6), this is an effectively “narrowband” weight selection, in that the pre-aligning of the microphones requires only the selection of a single weight per channel. Note, however, that this weight selection must be performed for all angles ϕ . The simplest choice is the fixed weighting $\mathbf{w}_\phi = \mathbf{1}$ for all ϕ [10], [11], where $\mathbf{1}$ is a vector of $L+1$ ones. In this paper, two adaptive schemes are proposed, as presented in the next two subsections.

3.2. Eigenanalysis of the Parameterized Spatial Correlation Matrix

Using the signal model of Section 2, notice that when the steered azimuth ϕ matches the actual azimuth θ , the parameterized spatial correlation matrix may be decomposed into signal and noise components in the following manner:

$$\mathbf{R}_\theta = \sigma_s^2 \boldsymbol{\alpha} \boldsymbol{\alpha}^T + E\{\mathbf{v}_\theta \mathbf{v}_\theta^T\}, \quad (10)$$

where σ_s^2 is the signal power,

$$\boldsymbol{\alpha} = [\alpha_0 \quad \alpha_1 \quad \cdots \quad \alpha_L]^T, \quad (11)$$

and

$$\mathbf{v}_\theta = [v_0[n] \quad v_1[n + f_1(\theta)] \quad \cdots \quad v_L[n + f_L(\theta)]]^T. \quad (12)$$

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

To that end, consider an adaptive weight selection method, which follows from the ideas of narrowband beamforming [12]. This weight selection attempts to non-trivially maximize the output energy of the steered-beamformer for a given azimuth ϕ :

$$\mathbf{e}_{\max,\phi} = \arg \max_{\mathbf{w}_\phi} \mathbf{w}_\phi^T \mathbf{R}_\phi \mathbf{w}_\phi \quad (13)$$

subject to

$$\mathbf{w}_\phi^T \mathbf{w}_\phi = 1. \quad (14)$$

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

$$S^{\text{EIG}}(\phi) = \mathbf{e}_{\max,\phi}^T \mathbf{R}_\phi \mathbf{e}_{\max,\phi} = \lambda_{\max,\phi}, \quad (15)$$

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

$$\hat{\theta}_{\text{EIG}} = \arg \max_{\phi} \lambda_{\max,\phi}. \quad (16)$$

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

$$\hat{\boldsymbol{\alpha}} = \mathbf{e}_{\max,\hat{\theta}_{\text{EIG}}}. \quad (17)$$

3.3. Minimum Variance Estimation

The application of Capon’s MVDR weight-selection method [13] to spatial spectral estimation has been previously studied in [14]. The treatment therein is somewhat incomplete in that channel attenuation effects are not considered. With the estimate of $\boldsymbol{\alpha}$ now available, it is possible to derive the optimal MVDR weights for the steered response power approach of (9).

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

$$\sum_{l=0}^L w_{\phi,l} \alpha_l s[n-t-f_l(\phi)+f_l(\phi)] = s[n-t], \quad (18)$$

which may be simplified and written in vector notation as

$$\mathbf{w}_\phi^T \boldsymbol{\alpha} = 1. \quad (19)$$

Using the method of Lagrange multipliers, the optimal weights become

$$\mathbf{w}_{\text{mv}',\phi} = \frac{\mathbf{R}_\phi^{-1} \boldsymbol{\alpha}}{\boldsymbol{\alpha}^T \mathbf{R}_\phi^{-1} \boldsymbol{\alpha}}. \quad (20)$$

Substituting (20) into (6):

$$S^{\text{mv}'}(\phi) = \mathbf{w}_{\text{mv}',\phi}^T \mathbf{R}_\phi \mathbf{w}_{\text{mv}',\phi} = \left(\boldsymbol{\alpha}^T \mathbf{R}_\phi^{-1} \boldsymbol{\alpha} \right)^{-1}. \quad (21)$$

The proposed broadband minimum variance DOA estimator is thus given by:

$$\hat{\theta}_{\text{mv}'} = \arg \max_{\phi} \left(\boldsymbol{\alpha}^T \mathbf{R}_\phi^{-1} \boldsymbol{\alpha} \right)^{-1}. \quad (22)$$

Interestingly, from (20) and the definition of an eigenvector,

$$\mathbf{e}_{\text{max},\theta} \approx \mathbf{w}_{\text{mv}',\theta}. \quad (23)$$

and thus the EIG spectrum closely resembles the MV' spectrum at angles near the actual azimuth angle of arrival. It should also be mentioned that the classical minimum variance approach is a specific case of (21), where $\boldsymbol{\alpha} = \mathbf{1}$ [14].

4. SIMULATION EVALUATION

4.1. Simulation Environment

The proposed estimators are evaluated in a computer simulation. An equispaced circular array of 3-10 omnidirectional microphones is employed as the spatial aperture. The array radius is made as large as possible without suffering from spatial aliasing [6]:

$$r = \frac{c}{4 \sin \frac{\pi}{L+1} f_{\text{max}}}, \quad (24)$$

where f_{max} denotes the highest frequency of interest, and is chosen to be 4 kHz in the simulations. The signal sources are omnidirectional point sources.

A reverberant acoustic environment is simulated using the image model method [15]. The simulated room is rectangular with plane reflective boundaries (walls, ceiling and floor). Each boundary is characterized by a frequency- and angle-independent uniform reflection coefficient. The room dimensions in centimeters are (304.8, 457.2, 381). The centre of the array sits at (152.4, 228.6, 101.6). Two distinct scenarios are simulated, as described below.

The speaker is immobile and situated at (254, 406.4, 101.6) and (254, 406.4, 152.4) in the first and second simulation scenarios, respectively. The correct azimuth angle of arrival is 60 degrees. The SNR at the microphone elements is 0 dB. Here, SNR refers to spatially white sensor noise in the first scenario, and spherically isotropic (diffuse) noise in the second scenario (the noise is temporally white in both cases). Three reverberation levels are simulated for each scenario: anechoic, moderately reverberant ($T_{60} = 300$ ms), and highly reverberant ($T_{60} = 600$ ms). The reverberation times are measured using the reverse-time integrated impulse response method of [16]. In the first simulation scenario, the microphones are calibrated with unity gains. In the second simulation scenario, the presence of uncalibrated microphones is simulated, by setting α_l , $l = 0, 1, \dots, L$ to a uniformly distributed random number over the range (0.2, 1). The microphone outputs are filtered to the 300 - 4000 Hz range prior to processing. Two signal types are examined for each scenario: white Gaussian noise and female English speech. The DOA estimates are computed once per 128 ms frame. To achieve good angular resolution, the sampling rate is chosen to be 48 kHz, resulting in frames of $N = 6144$ samples each. A simulation run consists of $K = 890$ frames.

The estimated spatial spectra are plotted to observe mainlobe width and background values. For each frame, the spectrum is normalized such that the peak is 0 dB, and the spectra are then averaged

over the K frames. For each simulation, the algorithms are also evaluated from a DOA estimation standpoint using the percentage of anomalies ($\%_a$) – anomalies are classified as estimates that differ from the actual angle of arrival by more than 5 degrees – and root-mean-square (RMS) error measure for the nonanomalous estimates:

$$e_{\text{RMS}} = \sqrt{\frac{1}{K_{\text{na}}} \sum_{k \in \chi_{\text{na}}} (\hat{\theta}_k - \theta)^2}, \quad (25)$$

where χ_{na} is the set of all nonanomalous estimates, and K_{na} is the number of elements in χ_{na} .

4.2. TDOA Comparison Algorithm

The DOA estimation performance of the proposed estimators is compared to that of a standard two-step TDOA algorithm. The algorithm computes the time delay between microphone 0 and microphone l for $l = 1, \dots, L$ as:

$$\tau_{0l} = \arg \max_{\tau} E\{x_0[n]x_l[n + \tau]\}, \quad (26)$$

and then translates these relative delays to the azimuth angle of arrival using the least-squares criterion:

$$\hat{\theta}_{\text{TDOA}} = \arg \max_{\phi} \sum_{l=1}^L \left| \frac{c \cdot \tau_{0l}}{r} - \cos \phi + \cos \left(\phi - \frac{2\pi l}{L+1} \right) \right|^2. \quad (27)$$

4.3. Results

Figure 2 depicts the broadband spatial spectra of the proposed estimators in simulation scenario 1. Panels (a) and (b) are computed with moderate reverberation and a speech signal, and show the effect of adding more microphones on the resulting spectra. Panels (c) and (d) show the effect of increased reverberation on the spectral estimates and are computed with 10 mics and a speech signal. It is evident that adding extra microphones improves the resolution of the spatial spectra. The main lobe is narrowed, and the background level, which corresponds to the power of the reverberant field, is lowered. A lower reverberant field level decreases the probability of anomalies. As the reverberation level increases, the background level is increased, making anomalies more likely. Note that even with a T_{60} of 600 ms, the spectra clearly discriminate the source from the background.

Tables I and II pertain to the DOA estimation accuracy of the proposed and TDOA methods. It is obvious from the tables that the TDOA-based method provides very poor performance in reverberant conditions, as the proposed estimators greatly outperform the TDOA-based approach in all but the anechoic white signal case. This lends credence to the notion that jointly utilizing multiple microphone pairs combats reverberation, not just background noise. In the TDOA two-step method, a “hard-decision” is made in the computation of each τ_{0l} , and thus if this decision is incorrect, the error is propagated to the least-squares stage. On the other hand, spatial spectral estimators do not make such hard-decisions. Instead, the decision is deferred until after the contribution of all microphone pairs. As the eigenvalue and minimum variance spectra are equivalent at $\phi \approx \theta$ from (23), the DOA estimation performance of the two methods is identical. Notice that the introduction of variable microphone gains in scenario 2 seriously degrades the TDOA method's performance, but causes only a slight degradation in performance of the proposed methods, pointing to the efficacy of the estimate of $\boldsymbol{\alpha}$. Similarly, the elevation of the source does not pose problems to the proposed estimators.

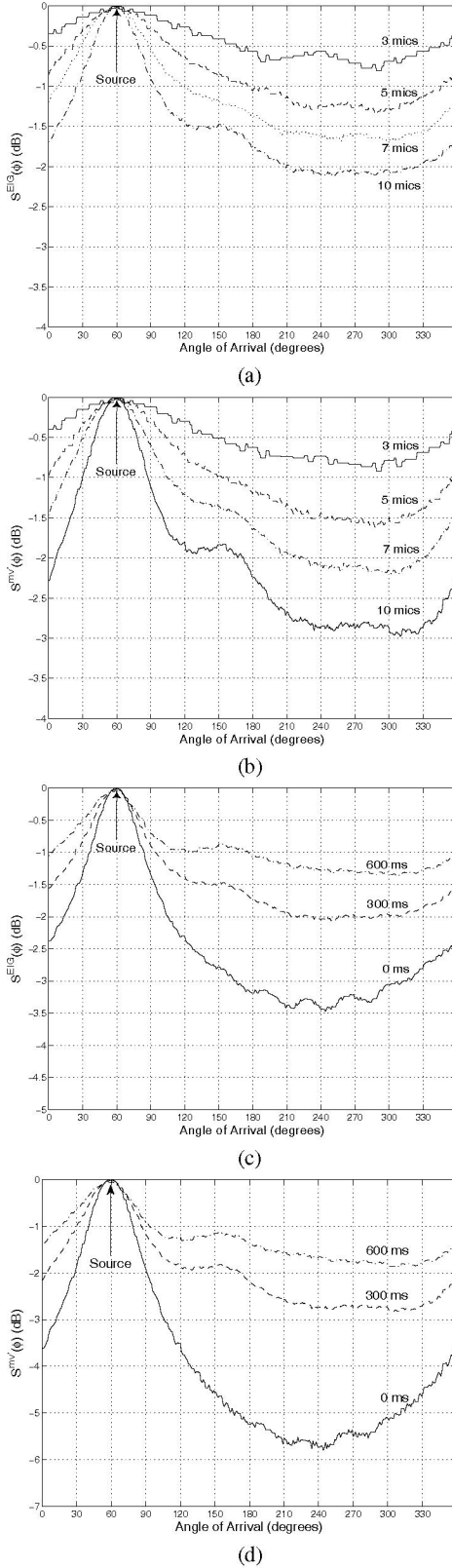


Fig. 2. Spatial spectral estimates: (a) $S^{\text{EIG}}(\phi)$ as a function of number of mics, (b) $S^{\text{mv}'}(\phi)$ as a function of number of mics, (c) $S^{\text{EIG}}(\phi)$ as a function of T_{60} , and (d) $S^{\text{mv}'}(\phi)$ as a function of T_{60} .

Table 1. DOA estimation results for simulation scenario 1.

Parameters	TDOA		EIG		MV'	
	% _a	e _{RMS}	% _a	e _{RMS}	% _a	e _{RMS}
$T_{60} = 0\text{ms, white signal}$	0	0.4	0	0.7	0	0.7
$T_{60} = 300\text{ms, white signal}$	11	2.5	0	1.4	0	1.4
$T_{60} = 600\text{ms, white signal}$	74	2.8	0	2.0	0	2.0
$T_{60} = 0\text{ms, speech signal}$	5	2.1	0	1.2	0	1.2
$T_{60} = 300\text{ms, speech signal}$	53	2.8	9	2.6	9	2.6
$T_{60} = 600\text{ms, speech signal}$	82	3.0	29	2.8	29	2.8

Table 2. DOA estimation results for simulation scenario 2.

Parameters	TDOA		EIG		MV'	
	% _a	e _{RMS}	% _a	e _{RMS}	% _a	e _{RMS}
$T_{60} = 0\text{ms, white signal}$	4	2.5	0	2.0	0	2.0
$T_{60} = 300\text{ms, white signal}$	29	3.1	0	3.5	0	3.5
$T_{60} = 600\text{ms, white signal}$	66	2.8	8	4.5	9	4.5
$T_{60} = 0\text{ms, speech signal}$	24	2.6	1	3.0	1	3.0
$T_{60} = 300\text{ms, speech signal}$	77	3.1	10	4.4	10	4.4
$T_{60} = 600\text{ms, speech signal}$	91	2.9	27	4.7	27	4.7

5. CONCLUSIONS

This paper has presented a novel approach to broadband DOA estimation based on the eigenanalysis of the parameterized spatial correlation matrix. It was shown that this estimate is equivalent to the MVDR spatial spectral estimate that takes into account channel attenuation effects. The addition of extra microphones increases the resolution of the proposed spectral estimate and helps shield the DOA estimate against the effects of reverberation. Simulation results showed that the proposed method provides reliable estimates in reverberant environments in which TDOA-based methods fail. This superior performance stems from the joint utilization of all microphones via the parameterized spatial correlation matrix.

6. REFERENCES

- [1] B. D. Van Veen and K. M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, pp. 4–24, Apr. 1988.
- [2] D. E. Dudgeon and D. H. Johnson, *Array Signal Processing*, Prentice-Hall, NJ, 1993.
- [3] J. Chen, J. Benesty, and Y. Huang, "Time delay estimation in room acoustic environments: an overview," *EURASIP Journal on Applied Signal Processing*, vol. 2006, article ID 26503, 19 pages, 2006.
- [4] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 24, pp. 320–327, Aug. 1976.
- [5] M. Omologo and P. Svaizer, "Use of the crosspower-spectrum phase in acoustic event location," *IEEE Trans. Speech and Audio Processing*, vol. 5, pp. 288–292, May 1997.
- [6] R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays*, SciTech Publishing, Raleigh, NC, USA, 2004.
- [7] D. H. Johnson, "The application of spectral estimation methods to bearing estimation problems," *Proc. IEEE*, vol. 70, pp. 1018–1028, Sept. 1982.
- [8] J. Chen, J. Benesty, and Y. Huang, "Robust time delay estimation exploiting redundancy among multiple microphones," *IEEE Trans. Speech and Audio Processing*, vol. 11, pp. 549–557, Nov. 2003.
- [9] J. Benesty, J. Chen, and Y. Huang, "Time-delay estimation via linear interpolation and cross-correlation," *IEEE Trans. Speech and Audio Processing*, vol. 12, pp. 509–519, Sept. 2004.
- [10] J. Dibiase, H. F. Silverman, and M. S. Brandstein, "Robust localization in reverberant rooms," in *Microphone Arrays: Signal Processing Techniques and Applications* (M. S. Brandstein and D. B. Ward, eds.), pp. 157–180, Springer-Verlag, Berlin, 2001.
- [11] D. N. Zofkin and R. Duraiswami, "Accelerated speech source localization via a hierarchical search of steered response power," *IEEE Trans. Speech and Audio Processing*, vol. 12, pp. 499–508, Sept. 2004.
- [12] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Mag.*, pp. 67–94, July 1996.
- [13] J. Capon, "High-resolution frequency-wavenumber analysis," *Proc. IEEE*, vol. 57, pp. 1408–1418, Aug. 1969.
- [14] N. Owsley, "Sonar array processing", in *Array Signal Processing*, S. Haykin, ed., Englewood Cliffs, NJ: Prentice Hall, 1984.
- [15] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, pp. 943–950, Apr. 1979.
- [16] M. R. Schroeder, "New method for measuring reverberation time," *J. Acoust. Soc. Am.*, vol. 37, pp. 409–412, 1965.