Estimation of Acoustic Reflection Coefficients Through Pseudospectrum Matching

D. Marković, K. Kowalczyk, F. Antonacci, C. Hofmann, A. Sarti, and W. Kellermann

I. Introduction

KNOWING the acoustic properties of reflective surfaces a given environment is of crucial importance for a wide range of applications of space-time audio processing. Applications of acoustic rendering, for example, are critically dependent on the reverberation in the hosting environment. Failing to correctly account for reflections on the part of walls and/or obstacles could result in a severely impaired spatial impression of the acoustic scene. If the locations of walls and their reflective properties are known, reverberation can be canceled out up to a certain degree [1]. Early reflections can even be exploited to enhance the spatial impression [2], [3] or to improve the accuracy of source localization algorithms [4].

In this contribution we propose a methodology for the frequency-subband *in-situ* estimation of the reflection coefficients of planar surfaces. An accurate estimation of the reflection coefficients is a highly challenging task, which is typically conducted in an acoustically conditioned dedicated environment. For this purpose, several techniques are available in the literature, based on the use of impedance tubes [5], [6]. This category of techniques has been standardized in ISO 10534-1 (for narrowband estimation) [7] and ISO 10534-2 (for wideband estimation) [8]. Impedance tube measurements offer a good level of accuracy but they are costly and complex to carry out, as a portion of material under test needs to be removed. As an alternative to the impedance tube methods, reflective properties of wall material can also be measured in a reverberation chamber. The related measurement process is standardized in ISO 354:2003 [9]. It is worth noting that the use of a reverberation chamber requires the whole measurement procedure to be accomplished in acoustically treated rooms and, therefore, it turns out to be rather costly for many applications.

A different route is pursued for in-situ measurements. Relevant examples include [10]–[16]. The method in [10] is based on the analysis of the Acoustic Impulse Response (AIR) between a source and a microphone, which are aligned perpendicularly to the wall under analysis and in close proximity to it. In this configuration, the AIR consists of the direct-path signal between the source and the microphone, the echo related to the reflection coming from the material under test and all other reflections. Given a reference AIR between the source and the microphone, obtained in a separate free-field measurement, the direct-path signal is subtracted from the rest of the measured AIR. Under the assumption that the reflection from the material under test and other unwanted reflections do not overlap in time, they can be readily separated through temporal windowing. The reflection coefficient is then estimated by comparing the magnitude of direct and reflection paths. This type of measurement procedure has been standardized in ISO 13472-1 [17]. The method proposed in [11] follows a similar measurement procedure but estimates the acoustic impedance (and thus also the reflection coefficient) using the complex transfer function between the source and the receiver, obtained as a division of the spectrum of the reflected and direct-path signals. Authors in [12] propose to use space-time processing in order to avoid time-domain analysis of the Acoustic Impulse Response. The microphone array is placed between the

Manuscript received April 12, 2013; revised July 31, 2013; accepted September 15, 2013. Date of publication October 17, 2013; date of current version November 13, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Boaz Rafaely.

D. Marković, F. Antonacci, and A. Sarti are with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, 20133 Milan, Italy (e-mail: dmarkovic@elet.polimi.it; dejan.markovic@polimi.it; antonacc@elet.polimi.it; sarti@elet.polimi.it).

K. Kowalczyk was with the Chair of Multimedia Communication and Signal Processing (LMS), University of Erlangen-Nuremberg, 91058 Erlangen, Germany. He is now with Fraunhofer IIS, 91058 Erlangen, Germany (e-mail:kowalzcyk@lnt.de).

C. Hofmann and W. Kellermann are with the Chair of Multimedia Communications and Signal Processing, University of Erlangen-Nuremberg, 91058 Erlangen, Germany (e-mail: hofmann@LNT.de; wk@LNT.de).

source and the reflective surface. The magnitude of the direct and reflection paths are estimated by steering the maximum directivity of the spatial response towards the probing source or the material under test, respectively. The reflection coefficient is then estimated by comparing the magnitudes of the direct-path and the reflected signals, similarly to [10]. Spatial filtering is also used in [13] in order to separate direct and reflected paths. The amplitudes of these paths are determined by means of plane-wave decomposition and the complex ratio of the reflected and direct-path signal amplitudes determines the reflection coefficient value. Authors in [14] use a method based on the combined measurement of the instantaneous sound pressure and sound particle velocity. Provided that the particle velocity-pressure transducer is adequately calibrated, the proposed method gives reasonably accurate measurements of the field impedance close to the surface of interest. When the environment comprises walls of different materials, the methods in [10]–[14] require multiple measurements, one per wall, which results in a rather time-consuming procedure. The method proposed in [15] considers the in-situ estimation of the sound absorption of all walls in a room as a boundary inverse problem that is solved using a global optimization algorithm. In particular, the finite element method (FEM) or the finite difference method (FDM) is used to model the room and a evolutionary algorithm minimizes the difference between the modeled pressure values and the values observed at a few measurement points. In [16], for in-situ measurement of surface properties, a hemispherical microphone array is designed. Once the sound pressure distribution on the hemispherical shell is known, the data is transformed into the Spherical Harmonics (SH) domain. This data can then be used to derive angle-dependent reflection properties of surfaces.

The technique proposed in this manuscript aims at estimating the reflection coefficients of the walls through a single measurement in a frequency-subband fashion. A beamforming technique allows to estimate the angular distribution of the acoustic energy "observed" by the microphone array. This distribution is referred to as *angular pseudospectrum* [18] or *steered response power* [19]. The angular pseudospectrum is modeled as a linear combination of multiple contributions related to the most relevant reflection paths. If we assume to know the geometry; position and radiation pattern of the source; position and spatial response of the array; such contributions can be modeled up to a scale factor, which nonlinearly depends on the reflection coefficients. We propose a two-step iterative estimation technique based on the Expectation-Maximization (EM) algorithm. The first step estimates the scaling factors, while the latter determines the reflection coefficients from the scaling factors. Under the assumption of additive white Gaussian noise, the outlined methodology determines a Maximum Likelihood (ML) estimate.

In [20] the generative model of the pseudospectrum is introduced. However, the method is not based on Expectation-Maximization, resulting in less accurate estimations, especially at low frequencies and for highly absorptive materials. In this manuscript we aim at overcoming these limitations in two different ways. The first major novelty is in the use of Expectation-Maximization algorithm, which improves the accuracy of the estimation procedure. Secondly, we propose using a Minimum Variance Distortionless Response (MVDR) beamformer in order to measure pseudospectra with higher spatial selectivity.

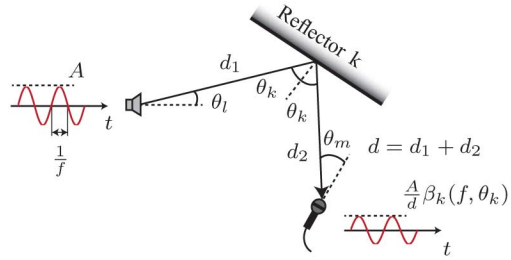

Fig. 1. Acoustic wave reflection from a wall.

Although in this paper we consider 3D wave propagation, we propose a solution based on 2D geometric analysis (i.e., we assume reflections from floor and ceiling to be negligible), which allows us to adopt an acquisition system based on a circular microphone array. Note that the extension of the presented method to the case of 3D geometry is straightforward. This choice, however, would render the acquisition system to be more expensive (spherical microphone array) and the acquisition and estimation process more time-consuming (more variables and unknowns), unless we were dealing with simple environments, e.g., where all walls of a shoe-box room were made of the same material type.

The rest of the manuscript is organized as follows: Section II formulates the problem of reflection coefficient estimation. Section III shortly reviews some beamforming techniques for pseudospectrum measurement. Section IV introduces the generative model of the pseudospectrum, which constitute the basis for the rest of the manuscript. In Section V, the algorithm for the estimation of the reflection coefficient is proposed. Section VI assesses the accuracy of the technique through experiments based on measured data. Finally, Section VII offers some concluding remarks.

II. Problem Formulation

As discussed in the Section I, in this manuscript we study a 2D floormap of the environment, which allows us to adopt a circular microphone array for the estimation of the reflection coefficients. The real-valued reflection coefficient of a material is defined as the amplitude of a reflected wave relative to an incident wave [21]. The reflection coefficient value $\beta_k(f,\theta_k)$ for the reflector k depends on the frequency f and the angle of incidence θ_k. We assume that the environment is an enclosure made of large - relative to the wavelength - planar surfaces, each with homogeneous properties. Incident waves are assumed to be reflected by the walls in a purely specular fashion (the angle of incidence equaling the angle of reflection). We also assume the size of the array to be small, so that the wavefronts that it senses can be considered as planar. As shown in Fig. 1, upon specular wave reflection, the amplitude of the reflected signal is reduced proportionally to the traveled distance d and it is scaled down by the reflection coefficient $\beta_k(f,\theta_k)$. It is known in the literature

(see [21]) that, assuming the boundary to be locally reacting, the angle-dependent reflection coefficient is given by

$$\beta_k(f,\theta_k) = \frac{\xi_k(f)\cos(\theta_k) - 1}{\xi_k(f)\cos(\theta_k) + 1}, \tag{1}$$

where $\xi_k(f)$ measures the ratio of the acoustic impendances of the kth wall and air at frequency f. If the wall is normal to the travel direction of the incident wave, (1) reduces to

$$\beta_k(f,0) = R_k(f) = \frac{\xi_k(f) - 1}{\xi_k(f) + 1}. \tag{2}$$

In this manuscript, the goal is to estimate $|R_k(f)|$ from acquisitions by a microphone array of a known source signal.

In addition to the traveled distance d and the reflection coefficient, the amplitude of the reflected signal acquired by a microphone also depends on the amplitude A of the emitted signal $s(t)$, the radiation pattern $l(f,\theta_l)$ of the loudspeaker (function of the direction θ_l), and the directivity pattern of the microphone $m(f,\theta_m)$ (function of the direction θ_m). As reflection coefficients are generally frequency-dependent, we estimate their values at different frequencies. However, in order to simplify the notation, we will omit specifying the frequency variable f (unless explicitly provided in Section III).

The *a priori* knowledge that is needed for the estimation procedure consists of the source signal $s(t)$; the environment geometry; and rough estimates of $l(\theta_l)$ and $m(\theta_m)$. In particular, the latter two can also be estimated or approximated using methods presented in [22], [23]. The parameters θ_k, θ_l, θ_m, d, and τ (i.e., delay corresponding to the travelled distance d) can be obtained from geometrical acoustics relationships (e.g. using the modeling engine in [24]). Thus the remaining unknowns are the signal amplitude A; the reflection coefficients at normal-incidence of all reflective surfaces in the environment, R_k; and the variance of the additive noise at the microphones.

In what follows, we denote the signal acquired by microphone j as $\bar{x}_j(t)$ and the M signals acquired by the microphone array as $\bar{\mathbf{X}} = [\bar{x}_1(t),\ldots,\bar{x}_M(t)]^T$. The signals $x_j(t)$, on the other hand, denote the modeled signals, obtained using the *a priori* information. These signals match $\bar{x}_j(t)$ except for the magnitude of the echoes (which depends on A and R_k) and the noise component (assumed to be Gaussian). From $\bar{\mathbf{X}}$ we obtain the measured pseudospectrum $\hat{P}(\theta)$. Using the *a priori* information, we obtain the modeled pseudospectrum $P(\theta)$, up to scalar coefficients, which depend on R_k, A and the noise variance. The aim of the proposed method is to estimate the reflection coefficients of all walls in the acoustic enclosure by matching these pseudospectra for all directions θ.

III. Measurement of the Pseudospectrum

When using a microphone array for sampling the sound field, the angular distribution of the acoustic energy can be estimated using beamforming techniques on the acquired microphone signals. For this purpose, the room is scanned using a beamformer and the output power for each look direction forms the so-called angular pseudospectrum [18]. Such a spatial power pseudospectrum $\hat{\mathbf{P}} = [\hat{P}(\theta_1),\hat{P}(\theta_2),\ldots]^T$ for sample look directions $\theta_1,\theta_2,\ldots$ can generally be expressed as

$$\hat{P}(\theta) = \mathbf{h}^H(\theta)\hat{\mathbf{R}}_{xx}\mathbf{h}(\theta), \tag{3}$$

where θ is the look direction, $\mathbf{h}(\theta)$ denotes the array weight vector, which depends on the specific beamforming technique, and $\hat{\mathbf{R}}_{xx}$ denotes the estimate of the autocorrelation matrix of the microphone signals

$$\hat{\mathbf{R}}_{xx} = \hat{E}[\overline{\mathbf{X}\mathbf{X}}^H], \tag{4}$$

$\hat{E}[.]$ denoting the sample estimate of the expected value. In principle, the pseudospectrum $\hat{\mathbf{P}}$ can be measured using (3) with $\mathbf{h}$ computed with any beamforming technique. In this manuscript, we consider both data-independent and statistically optimum beamformers, namely the Delay-And-Sum (DAS) and the Minimum Variance Distortionless Response (MVDR) beamformers.

A. Delay-and-Sum Beamformer

The pseudospectrum of a Delay-And-Sum (DAS) beamformer is obtained by substituting $\mathbf{h}(\theta) = \mathbf{a}(\theta)$ into (3), which yields

$$\hat{P}(\theta) = \mathbf{a}^H(\theta)\hat{\mathbf{R}}_{xx}\mathbf{a}(\theta), \tag{5}$$

where the steering vector of the array is given by

$$\begin{aligned}\mathbf{a}(\theta) &= [a_1(\theta), a_2(\theta), \ldots, a_M(\theta)]^T \\ &= \frac{1}{\sqrt{M}}\left[e^{j2\pi f u_1(\theta)}, e^{j2\pi f u_2(\theta)}, \ldots, e^{j2\pi f u_M(\theta)}\right]^T\end{aligned} \tag{6}$$

and f denotes the temporal frequency, θ is the angle of arrival, $u_j(\theta) = \rho\cos(\theta-\theta_j)/c$ for the circular array with radius ρ, θ_j is the angle of the jth microphone position with respect to the array center, and c is the speed of sound [18].

The DAS is a data-independent beamformer whose design maximizes the SNR for spatially uncorrelated noise. This means that it is robust against microphone self-noise and positioning errors in the microphone rig [25]. On the other hand, it offers limited spatial resolution at low frequencies.

B. MVDR Beamformer

In order to improve spatial selectivity in the mid/low frequency ranges, superdirective beamformers can be employed. One such beamformer is the MVDR, whose design is statistically optimum, as it minimizes the output variance (or power) subject to a distortionless constraint on its response in the look direction θ [26]. The MVDR beamformer solves the following minimization problem

$$\min_{\mathbf{h}} \mathbf{h}^H\hat{\mathbf{R}}_{xx}\mathbf{h}, \quad \text{subject to} \quad \mathbf{h}^H\mathbf{a}(\theta) = 1. \tag{7}$$

The closed-form solution of the weight vector of the beamformer can be derived, using the method of Lagrange multipliers [27], as

$$\mathbf{h}(\theta) = \frac{\hat{\mathbf{R}}_{xx}^{-1}\mathbf{a}(\theta)}{\mathbf{a}^H(\theta)\hat{\mathbf{R}}_{xx}^{-1}\mathbf{a}(\theta)}. \tag{8}$$

As superdirective beamformers are highly sensitive to the microphone self-noise and geometric errors in the array rig, controlling the robustness of the design is of utmost importance. With reference to (8), this can be achieved through diagonal loading (with frequency-dependent loading factors), obtained with an iterative design scheme [28]. In the optimization problem (7), improved robustness can be achieved using the additional White Noise Gain (WNG) constraint

$$\frac{1}{\mathbf{h}^H(f)\mathbf{h}(f)} \geq \zeta, \tag{9}$$

ζ being the array gain against white noise [25]. In addition, for the extraction of room reflection signals that have low energy and are strongly correlated with the direct-path signal (see model presented in Section IV), focusing matrices and frequency smoothing techniques [29], [30] can be used for alleviating the ill-conditioning problem associated to the autocorrelation matrix $\hat{\mathbf{R}}_{xx}$, thus increasing the robustness of this data-dependent beamformer for coherent sources (see [31], [32] for a detailed discussion). Further robustness can be gained by smoothing the autocorrelation matrix $\hat{\mathbf{R}}_{xx}$ over time.

The purpose of focusing matrices $\mathbf{T}(f_\upsilon, \theta)$ is to map the signal space at all frequency bins f_υ from the range $\{f_1, \ldots, f_K\}$ onto a common reference frequency $f_0 \in \{f_1, \ldots, f_K\}$, i.e.,

$$\mathbf{T}(f_\upsilon, \theta)\mathbf{A}(f_\upsilon, \theta) = \mathbf{A}(f_0, \theta), \tag{10}$$

where $\mathbf{A}(f_\upsilon, \theta) = [\mathbf{a}(f_\upsilon, \theta_1), \ldots, \mathbf{a}(f_\upsilon, \theta_D)]$ denotes the direction matrix for D sound sources. As these Directions Of Arrival (DOAs) are known in our scenario, the design of focusing matrices could in principle be performed by rearranging terms in (10). However, since it has been shown in [30] that unitary focusing matrices do not change the spatial correlation of noise, we apply the unitary focusing matrices, which can be computed by solving the following constrained problem

$$\min_{\mathbf{T}(f_\upsilon,\theta)} \|\mathbf{A}(f_0, \theta) - \mathbf{T}(f_\upsilon, \theta)\mathbf{A}(f_\upsilon, \theta)\|_F$$

subject to

$$\mathbf{T}^H(f_\upsilon, \theta)\mathbf{T}(f_\upsilon, \theta) = \mathbf{I}, \tag{11}$$

where $||.||_F$ denotes the Frobenius matrix norm and $\mathbf{I}$ is the identity matrix. Using (11), the focused and frequency-smoothed measured autocorrelation matrix is obtained as

$$\begin{aligned}\hat{\mathbf{R}}_{xx} &= \sum_{\upsilon=1}^{K} \mathbf{T}(f_\upsilon, \theta)\hat{\mathbf{R}}_{xx}(f_\upsilon)\mathbf{T}^H(f_\upsilon, \theta) \\ &= \mathbf{A}(f_0, \theta)\hat{\mathbf{R}}_{ss}\mathbf{A}^H(f_0, \theta) + \hat{\mathbf{R}}_{ww},\end{aligned} \tag{12}$$

where $\hat{\mathbf{R}}_{ww} = \sum_{\upsilon=1}^{K} \mathbf{T}(f_\upsilon, \theta)\hat{\mathbf{R}}_{ww}(f_\upsilon)\mathbf{T}^H(f_\upsilon, \theta)$ and $\hat{\mathbf{R}}_{ss} = \sum_{\upsilon=1}^{K} \hat{\mathbf{R}}_{ss}(f_\upsilon)$, and $\hat{\mathbf{R}}_{ss}(f_\upsilon)$ and $\hat{\mathbf{R}}_{ww}(f_\upsilon)$ denote the source and noise autocorrelation matrices, respectively. Finally, substituting (12) into (8) and (3), the measured MVDR pseudospectrum can be estimated.

The MVDR design offers high directivity even at low frequencies and automatic null placement to interferers, which is beneficial for automatic suppression of the direct path and strong reflection signals. However, as this is a signal-dependent superdirective beamformer, the robustness issues mentioned above, need to be addressed in order to deal with real microphone arrays and coherent reflection signals.

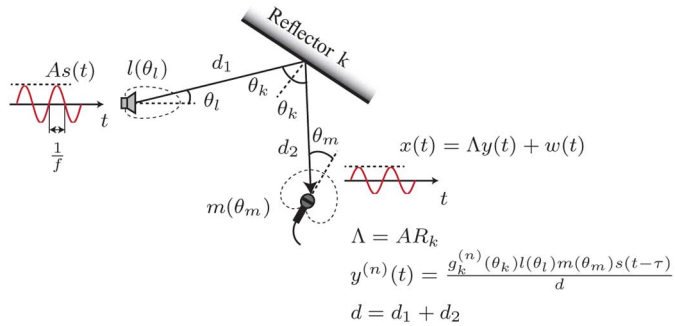

Fig. 2. Acoustic propagation and reflection model.

IV. Modeling the Pseudospectrum

In order to model the pseudospectrum for a given acoustic scenario, we begin with deriving the model $x_j(t)$ of the signal $\bar{x}_j(t)$ acquired by the microphone j. With reference to the propagation model of Fig. 2, a source signal $s(t)$ (emitted by a loudspeaker) of amplitude A is reflected by a wall in the acoustic enclosure. Assuming a point source (loudspeaker located in the far-field), the reflection signal acquired by a microphone can be written as [21]

$$z(t) = \frac{A\beta_k(\theta)l(\theta_l)m(\theta_m)}{d}s(t-\tau), \tag{13}$$

where $\beta(\theta_k)$ is the reflection coefficient of the kth reflector at angle θ; $l(\theta_l)$ is the loudspeaker gain in the direction θ_l; $m(\theta_m)$ is the microphone gain in the direction θ_m; d and τ are the traveled distance and the corresponding delay, respectively. As made explicit in Section II, the term $\beta_k(\theta)$ depends also on the frequency. The signal $z(t)$ should therefore be interpreted as the passband-filtered version of the signal acquired by the microphone, where the central frequency of the filter is f.

Eq. (13) shows that reflected paths have magnitudes that depend on the incidence angle, as already explained by (1). If we attempted to estimate the reflection coefficient for all possible incidence angles, we would soon find ourselves with more unknowns than measurements. We therefore need to reformulate the problem in such a way to reduce the number of unknowns to work on at the same time. This can be done through a process of "iterative factorization." We begin with rewriting the expression of the reflection coefficient in such a way to become the product of the normal-incidence reflection coefficient R_k (which is the term we aim to estimate) and a term $g(\theta_k)$. This second term depends on R_k as well, but this dependency can be removed by computing it from a value of R_k estimated in the previous iteration step. Through this iterative approximation, the unknown R_k is now shared by all of the incident acoustic paths on reflector k, which rebalances the number of unknowns with respect to the number of available measurements.

More specifically, we rewrite (1) as

$$\beta_k(\theta_k) = R_k g(\theta_k), \tag{14}$$

where

$$g(\theta_k) = \frac{[\xi_k \cos(\theta_k) - 1][\xi_k + 1]}{[\xi_k \cos(\theta_k) + 1][\xi_k - 1]}, \tag{15}$$

depends on R_k as well, because

$$\xi_k = \frac{1 + R_k}{1 - R_k}. \tag{16}$$

However, as our estimation algorithm is iterative, the estimate $\hat{\xi}_k^{(n-1)}$ of ξ_k at iteration $n-1$ can be obtained from (16) using the value of the reflection coefficient $R_k^{(n-1)}$ obtained in the previous iteration. More explicitly, we can write

$$\beta_k(\theta_k) \approx R_k g^{(n)}(\theta_k), \tag{17}$$

where

$$g^{(n)}(\theta_k) = \frac{[\hat{\xi}_k^{(n-1)} \cos(\theta_k) - 1][\hat{\xi}_k^{(n-1)} + 1]}{[\hat{\xi}_k^{(n-1)} \cos(\theta_k) + 1][\hat{\xi}_k^{(n-1)} - 1]}.$$

Using the iterative factorization (17) the term $g^{(n)}(\theta_k)$ is now known. In fact, $\hat{\xi}_k^{(n-1)}$ is available from the previous iteration and the incidence angle θ_k can be computed using a propagation modeling engine.

Under the assumptions expressed in Section II (prior knowledge of the signal $s(t)$, of the geometry of the environment, and rough knowledge of $l(\theta_l)$ and $m(\theta_m)$), we use the acoustic propagation modeling engine presented in [24] to obtain θ_k, θ_l, θ_m, d, and τ. As far as the initialization of the estimation algorithm is concerned, we set $g^{(0)}(\theta_k) = 1$. Note that the function $g^{(n)}(\theta_k)$ has two poles in $\xi_k = 1$ and $\xi_k = -1/\cos(\theta_k)$; and zeroes in $\xi_k = 1/\cos(\theta_k)$ and $\xi_k = -1$. As the normal-incidence reflection coefficient $|R_k|$ grows, $|\xi_k|$ in (16) grows as well and soon $g_k(\theta_k)$ tends to 1 (see eq. (15)). This means that with a sufficiently large $|R_k|$, the reflection coefficient (14) is nearly independent of the angle.

With the above approximation, the reflection signal of Fig. 2 becomes

$$z(t) \approx \frac{AR_k g_k^{(n)}(\theta_k) l(\theta_l) m(\theta_m)}{d} s(t - \tau).$$

The known portion of the observed signal $x(t)$ can be written as

$$y^{(n)}(t) = g_k^{(n)}(\theta_k) l(\theta_l) m(\theta_m) s(t - \tau)/d. \tag{18}$$

The signal amplitude A depends on the overall gain of the hardware used for acquiring the signal and, for practical reasons, it is assumed to be unknown. The unknowns are therefore the wall reflection coefficients and the signal amplitude of the source, which can be expressed as a product $\Lambda = AR_k$.

After a number of reflections, such as those shown in Fig. 3, the signal that is acquired by microphone j can be written as

$$\begin{aligned} x_j(t) &= z_j(t) + w_j(t) \\ &= \sum_{r=0}^{L} \sum_{i=1}^{N_r} \Lambda_{r,i} y_{j,r,i}^{(n)}(t) + w_j(t), \end{aligned} \tag{19}$$

where $y_{j,r,i}^{(n)}(t)$ is the known portion of the signal corresponding to the ith acoustic path of those that have undergone r reflections, as acquired by the jth microphone; $\Lambda_{r,i}$ is the unknown

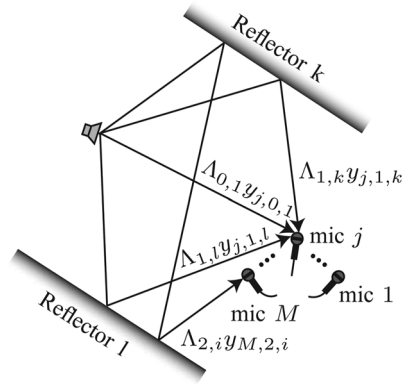

Fig. 3. Reflection model for an acoustic environment with two reflectors.

amplitude of the reflected signal; and $w_j(t)$ denotes the additive noise, related to the self-noise of the microphone and on the ADC circuitry. In particular, $r = 0$ is the direct path; $r = 1$ collects all the echoes that have bounced off one wall only, etc.; finally, N_r is the number of reflection paths of reflection order r. For the case of a rectangular 2D room, we have $N_0 = 1$ and $N_1 = 4$ and $\Lambda_{0,1} = A$ for the direct path, $\Lambda_{1,k} = AR_k$, $k = 1, \ldots, N_1$ for the first-order reflections, and $\Lambda_{2,i} = AR_kR_l$, $k \neq l$ for the second-order reflections, as shown in Fig. 3. We assume the noise component $w_j(t)$ to be Gaussian, statistically independent from $s(t)$, and spatially white. Note that in (19), reflections are modeled up to order L, whereas reflection paths with $r > L$ are assumed to have very low energy and to fall on the array from directions that are uniformly distributed, therefore they can be considered as part of the noise. Note from (19) that the signal-related component $y_{j,r,i}^{(n)}(t)$ of $x_j(t)$ contains the term $g^{(n)}(\theta_k)$, which is not known *a priori* but it is estimated iteratively. Consequently, an erroneous estimate of $g^{(n)}(\theta_k)$ could lead to an erroneous value of $y_{j,r,i}^{(n)}(t)$. However, note that all the paths that are incident on the same wall with index k share the same unknown $|R_k|$. An error on $g^{(n)}(\theta_k)$ for a specific path, therefore, does not necessarily leads to an erroneous estimate of $|R_k|$. To verify the convergence of the iterative estimation algorithm, simulations were conducted and the results are shown in Section VI.

The signals acquired by the M-element microphone array are written in matrix form as

$$\mathbf{X} = \mathbf{Z} + \mathbf{W} = \sum_{r=0}^{L} \sum_{i=1}^{N_r} \Lambda_{r,i} \mathbf{Y}_{r,i} + \mathbf{W}, \tag{20}$$

where

$$\begin{aligned} \mathbf{X} &= [x_1(t), \ldots, x_j(t), \ldots, x_M(t)]^T, \\ \mathbf{Z} &= [z_1(t), \ldots, z_j(t), \ldots, z_M(t)]^T, \\ \mathbf{Y}_{r,i} &= [y_{1,r,i}(t), \ldots, y_{j,r,i}(t), \ldots, y_{M,r,i}(t)]^T, \\ \mathbf{W} &= [w_1(t), \ldots, w_j(t), \ldots, w_M(t)]^T. \end{aligned}$$

In the following paragraphs, we apply the model of (20) as a model for the pseudospectrum. First, however, it is worth noticing that echoes corresponding to different propagation paths consist of delayed and attenuated replica of the same

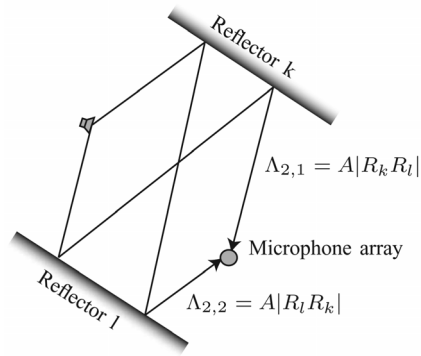

Fig. 4. Illustration of the amplitudes of two second-order reflections.

signal $s(t)$, and are therefore highly coherent. This fact has a relevant impact on the structure of the correlation matrix $\hat{\mathbf{R}}_{xx}$ in (3). We denote with

$$\hat{\mathbf{R}}_{r_1,i_1,r_2,i_2} = \hat{E}[\mathbf{Y}_{r_1,i_1}\mathbf{Y}_{r_2,i_2}^H] \tag{21}$$

the modeled correlation matrix between the generic echoes $\mathbf{Y}_{r_1,i_1}$ and $\mathbf{Y}_{r_2,i_2}$. Substituting (21) into (3), we obtain

$$P(\theta) = \mathbf{h}^H \left(\sum_{r_1=0}^{L} \sum_{i_1=1}^{N_{r_1}} \sum_{r_2=0}^{L} \sum_{i_2=1}^{N_{r_2}} \Lambda_{r_1,i_1}\Lambda_{r_2,i_2}\hat{\mathbf{R}}_{r_1,i_1,r_2,i_2} \right) \mathbf{h} + \sigma_w^2 \mathbf{h}^H\mathbf{h}, \tag{22}$$

where the noise $\mathbf{W}$ is assumed to be statistically independent of $\mathbf{Y}_{r,i}$ and spatially white with unknown variance σ_w^2, i.e.,

$$E[\mathbf{W}\mathbf{W}^H] = \sigma_w^2 \mathbf{I}_M. \tag{23}$$

Applying the DAS beamformer, $\mathbf{h}(\theta) = \mathbf{a}(\theta)$, yields obviously $\sigma_w^2 \mathbf{h}(\theta)^H\mathbf{h}(\theta) = \sigma_w^2$. For the MVDR beamformer, $\mathbf{h}(\theta)$ depends on the measured autocorrelation matrix, as in (8). Moreover, focusing and frequency smoothing techniques are applied to the modeled correlation matrices $\mathbf{R}_{r_1,i_1,r_2,i_2}$ and to the noise components, according to the equations that follow (12) for the measured counterparts. We denote with $\mathbf{P}_{r_1,i_1,r_2,i_2} = [P_{r_1,i_1,r_2,i_2}(\theta_1), \ldots, P_{r_1,i_1,r_2,i_2}(\theta_N)]^T$ the contribution to the pseudospectrum coming from the acoustic paths with indices r_1, i_1 and r_2, i_2, where the component $P_{r_1,i_1,r_2,i_2}(\theta)$ is defined as

$$P_{r_1,i_1,r_2,i_2}(\theta) = \mathbf{h}^H(\theta)\hat{\mathbf{R}}_{r_1,i_1,r_2,i_2}\mathbf{h}(\theta). \tag{24}$$

The model of the pseudospectrum $\mathbf{P} = [P(\theta_1), \ldots, P(\theta_N)]^T$ is obtained using (24) in (22), which yields

$$\mathbf{P} = \sum_{r_1=0}^{L} \sum_{i_1=1}^{N_{r_1}} \sum_{r_2=0}^{L} \sum_{i_2=1}^{N_{r_2}} \Lambda_{r_1,i_1}\Lambda_{r_2,i_2}\mathbf{P}_{r_1,i_1,r_2,i_2} + \sigma_w^2\mathbf{v}, \tag{25}$$

where $\mathbf{v} = [\mathbf{h}^H(\theta_1)\mathbf{h}(\theta_1), \ldots, \mathbf{h}^H(\theta_N)\mathbf{h}(\theta_N)]^T$.

In order to derive a simplified expression of $\mathbf{P}$, we note that in (25) there are sets of contributions $\mathbf{P}_{r_1,i_1,r_2,i_2}$ whose unknown scaling factors $\Lambda_{r_1,i_1}\Lambda_{r_2,i_2}$ correspond to the same combination of the reflection coefficients R_k and signal amplitude A. In order to clarify this fact, Fig. 4 shows two examples of second-order reflection paths that undergo reflections from the same walls but in a different order. The angular variation of reflection coefficients is included in the model $\mathbf{Y}_{r,i}$, and therefore these two paths have the same unknown amplitudes $\Lambda_{2,1} = \Lambda_{2,2} = AR_kR_l$. As a consequence, the autocorrelation in the pseudospectrum of the first path $\mathbf{P}_{2,1,2,1}$, the autocorrelation in the second path $\mathbf{P}_{2,2,2,2}$, and the crosscorrelations between these two paths (namely, $\mathbf{P}_{2,1,2,2}$ and $\mathbf{P}_{2,2,2,1}$), all share the same unknown scaling factor $A^2R_k^2R_l^2$ in (25). We therefore group together all pseudospectrum components that have the same scaling factors, which greatly simplifies the overall expression of $\mathbf{P}$. With reference to the simple scenario of Fig. 4, we can group $\mathbf{P}_{2,1,2,2}$ and $\mathbf{P}_{2,2,2,1}$ to obtain their weights and pseudospectrum components as $q_i = A^2R_k^2R_l^2$ and $\mathbf{P}_i = \mathbf{P}_{2,1,2,1} + \mathbf{P}_{2,2,2,2} + \mathbf{P}_{2,1,2,2} + \mathbf{P}_{2,2,2,1}$. This grouping operation can be performed using beam tracing [24], by performing a check on paths that undergo the same reflections. The modeled pseudospectrum finally takes the expression

$$\mathbf{P} = \sum_{i=1}^{N} \mathbf{P}_i q_i + \mathbf{v}\sigma_w^2 = \mathbf{M}\mathbf{q}, \tag{26}$$

where N is the number of pseudospectrum components after grouping, $\mathbf{M} = [\mathbf{P}_1, \ldots, \mathbf{P}_i, \ldots, \mathbf{P}_N, \mathbf{v}]$, $\mathbf{q} = [q_1, \ldots, q_i, \ldots, q_N, \sigma_w^2]^T$, and q_i indicates the scale factor of the pseudospectrum component $\mathbf{P}_i$. As an example, $q_i = A^2$ for the direct path component, $q_i = A^2R_k^2$ for the autocorrelation of the first-order reflection, $q_i = A^2|R_k|$ for the crosscorrelation between the direct path and the first-order reflection, and $q_i = A^2|R_kR_l|$ for the crosscorrelation between two first-order reflections.

Note that the scaling factors q_i are a product of a variable number of reflection coefficients $|R_k|$, the signal amplitude A, and the noise variance ($q_{N+1} = \sigma_w^2$). The nonlinear system of equations to be solved for calculating the reflection coefficients $|R_k|$ from the pseudospectrum scale factors $\mathbf{q}_i$ can be transformed into a linear system of equations by applying a logarithmic transform. For the example of $q_i = A^2|R_kR_l|$, the transform results in $\log(q_i) = 2\log(A) + \log|R_k| + \log|R_l|$. Let us now define $\mathbf{m} = [\log A, \log|R_1|, \ldots, \log|R_k|, \ldots, \log|R_l|, \ldots, \log\sigma_w^2]^T$ as a vector that contains the logarithms of unknown parameters, and $\mathbf{H}$ as a matrix that counts the number of occurrences of the elements in $\mathbf{m}$ for each component of $\log(\mathbf{q})$. As an example, the ith row of $\mathbf{H}$ corresponding to $\log(q_i) = \log(A^2|R_1R_2|)$ is given by $[2, 1, 1, 0, \ldots, 0]$. We can then write the relation between the component scale factors $\mathbf{q}$ and unknown parameters $\mathbf{m}$ as

$$\log(\mathbf{q}) = \mathbf{H}\mathbf{m}. \tag{27}$$

V. Reflection Coefficient Estimation

The procedure for the estimation of the reflection coefficients is based on the matching between the measured pseudospectrum (3) and the modeled pseudospectrum (26)

$$\hat{\mathbf{P}} \approx \mathbf{M}\mathbf{q}. \tag{28}$$

In (28), the pseudospectrum vector $\hat{\mathbf{P}}$ represents the measurements acquired by the microphone array and $\mathbf{M}$ contains the

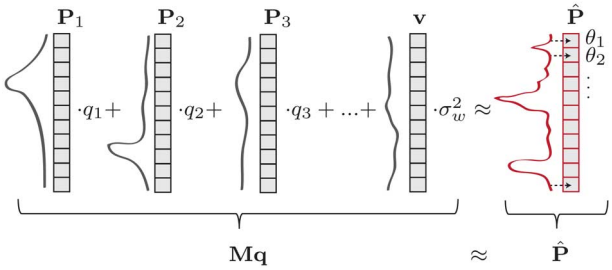

Fig. 5. Pseudospectrum generation model and matching with measurements.

modeled components of the pseudospectrum $\mathbf{P}$, as shown in Fig. 5.

In [20], we followed a two-step procedure:

1) Obtain the estimate $\hat{\mathbf{q}}$ of $\mathbf{q}$ from $\hat{\mathbf{P}}$ by minimizing the sum of squared differences between the observations and the model.
2) Obtain a least-squares solution $\hat{\mathbf{m}}$ from $\hat{\mathbf{q}}$ through the relationship (27). The final estimates $[\hat{A}, |\hat{R}_1|, |\hat{R}_2|, \ldots]^T$ are obtained as $\exp(\hat{\mathbf{m}})$.

This procedure showed promising results but was inaccurate at low frequencies and for absorptive materials. This can be explained by the fact that the number of variables in $\mathbf{q}$ is larger than the number of parameters to be estimated. Consequently, (28) can result in multiple solutions, among which we need to choose the least-squares one. However, the best solution in the least-squares sense for the first step does not necessarily guarantee the best solution for the second step. This is especially true when the measured pseudospectrum exhibits a smooth behavior, i.e., for low frequencies (a smaller resolution is possible) and for absorptive materials, which generate less pronounced peaks in the pseudospectrum. In order to alleviate this problem in [20], the matching was applied to multiple observations. In this manuscript, in order to overcome the above limitations, we pursue a different route.

We treat $\mathbf{q}$ as the unobserved latent variables linked to the observations $\hat{\mathbf{P}}$ through the non-invertible transformation $\mathbf{M}$. We then apply the Expectation-Maximization (EM) algorithm [33], [34], which consists of the following two steps:

1) **E-step**: Compute $E[\log p(\mathbf{q}|\mathbf{m})|\hat{\mathbf{P}}, \hat{\mathbf{m}}^{(n)}]$, where $\hat{\mathbf{m}}^{(n)}$ is the estimate of $\mathbf{m}$ at the current iteration n of the algorithm. Under the assumption that $p(\mathbf{q}|\mathbf{m})$ follows a Gaussian distribution, we obtain the Minimum Mean Square Error (MMSE) [35] estimate of $\mathbf{q}$.
2) **M-step**: Find $\hat{\mathbf{m}}^{(n+1)}$ that maximizes the log-likelihood $E[\log p(\mathbf{q}|\mathbf{m})]$ obtained in the previous step.

In order for the EM algorithm to provide the Maximum Likelihood (ML) estimate of $\mathbf{m}$, the distribution $p(\mathbf{q}|\mathbf{m})$ should be Gaussian. Before discussing the EM algorithm in detail, we show that $p(\mathbf{q}|\mathbf{m})$ can be approximated by a Gaussian distribution.

A. Considerations on $p(\mathbf{q}|\mathbf{m})$

In (3), the sample estimate $\hat{\mathbf{R}}_{xx}$ of the observation autocorrelation matrix $\mathbf{R}_{xx}$ is used, which can be written as a sum of signal and noise autocorrelation matrices, i.e., $\mathbf{R}_{xx} = \mathbf{R}_{zz} + \mathbf{R}_{ww}$, where $\mathbf{R}_{zz} = E[\mathbf{Z}\mathbf{Z}^H]$ and $\mathbf{R}_{ww} = E[\mathbf{W}\mathbf{W}^H]$ (source and noise signals are assumed to be uncorrelated). The same estimation procedure is performed in (21) in order to obtain the modeled components of the signal autocorrelation matrix $\hat{\mathbf{R}}_{r_1,i_1,r_2,i_2}$. The signal $s(t)$ is deterministic and known, therefore the sum of components $\hat{\mathbf{R}}_{r_1,i_1,r_2,i_2}$ with corresponding amplitudes in (22) should be equal to the signal autocorrelation matrix $\mathbf{R}_{zz}$ contained in $\hat{\mathbf{R}}_{xx}$. However, in (22) we use the true noise autocorrelation matrix $\mathbf{R}_{ww}$ (under the assumption of spatially white noise), while (3) contains an estimate. For Gaussian noise with covariance matrix $\mathbf{C}_{ww} = \mathbf{R}_{ww}$ (the noise has zero mean), $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{ww} = \sigma_w^2 \mathbf{I}_M)$, it can be shown that the sample autocorrelation matrix estimate $\hat{\mathbf{R}}_{xx}$ obtained using K samples has a Wishart distribution with $K-1$ degrees of freedom $W_M(\mathbf{R}_{xx}/K, K-1)$ [36]. The pseudospectrum $\hat{P}(\theta)$ in (3) is therefore distributed as $\hat{P}(\theta) \sim \frac{P(\theta)}{K}\chi^2_{K-1}$, where χ^2_{K-1} is a χ^2 distribution with $K-1$ degrees of freedom. The mean value of $\hat{P}(\theta)$ is $(K-1)\frac{P(\theta)}{K}$ and the variance amounts to $2(K-1)\frac{P(\theta)^2}{K^2}$. The χ^2 distribution with $K-1$ degrees of freedom is the sum of $K-1$ independent random variables with finite mean and variance, and thus for sufficiently large K, using the central limit theorem, it can be approximated as a Gaussian distribution, i.e., $P(\theta) \sim \mathcal{N}(P(\theta), 2P(\theta)^2/K)$. Here we assume that the number of samples is sufficiently high and the pseudospectrum noise has a Gaussian distribution. The model in (26) could suffer from some model inaccuracies (e.g., geometry) and measurement errors. However, the distribution of these errors is more difficult to determine. The geometric information (environment geometry, loudspeaker and microphone array positions) can be measured in a robust fashion and with good accuracy [37], [38]. However, small measurement errors are unavoidable, especially for the radiometric information ($l(\theta_l)$ and $m(\theta_m)$), which affects the accuracy of the estimate of $\mathbf{q}$. For sufficiently small additional errors of $g_k^{(n)}(\theta_k)$, $l(\theta_l)$, and $m(\theta_m)$, we ascertain that $\mathbf{q}$ has approximately a Gaussian distribution. In order to verify this statement, 1000 independent simulations of measurements within a 2D rectangular room were performed. For this test, we set $A = 2$ while the reflection coefficient is kept constant for all room walls, $R_k = 0.7$, $k = 1, \ldots, 4$. The angle-dependent gain of the loudspeaker $l(\theta_l)$ (modeled as cardioid) and the angle-dependent gain of the microphone $m(\theta_m)$ (modeled as omnidirectional) are affected by an additive Gaussian noise. We analyze the distribution of

$$\begin{cases} q_2 = A^2|R_1|^2 = 1.96 \\ q_3 = A^2|R_1R_2| = 1.96 \\ q_4 = A^2|R_1|^4 = 0.9604 \\ q_5 = A^2|R_1R_2R_3| = 1.372 \\ q_6 = A^2|R_1|^6 = 0.4706 \\ q_7 = A^2|R_1|^8 = 0.2306. \end{cases} \tag{29}$$

Fig. 6 shows the norm-plots [39] of $P(q_i|\mathbf{m})$, $i = 2, \ldots, 7$. In order for q_i to be Gaussian, its cumulative distribution should fit the dashed line. Note that the match turns out to be quite good, especially when the value of q_i is sufficiently large, which confirms the validity of the assumption behind the Expectation-Maximization algorithm.

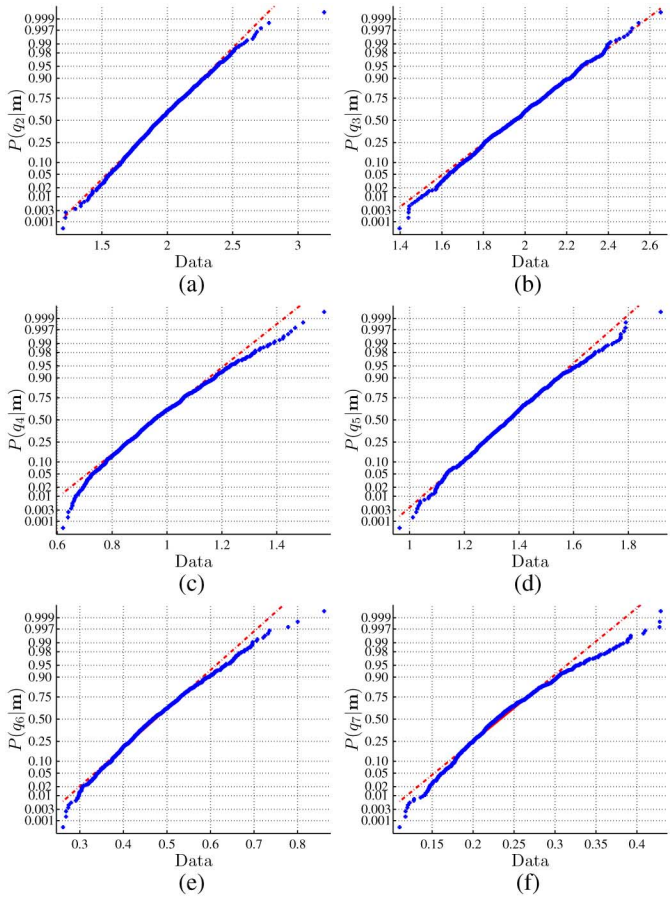

Fig. 6. Norm-plots of $p(q_i|\mathbf{m})$, $i = 2, \ldots, 7$. Dashed lines denote the norm-plot values for a Gaussian distribution.

B. Expectation-Maximization Algorithm

Assuming the pseudospectrum noise $\mathbf{n}$ and the component scale factors $\mathbf{q}$ to be normally distributed with covariance matrices $\mathbf{C}_{nn}$ and $\mathbf{C}_{qq}$, respectively, pseudospectra matching can be formulated as

$$\hat{\mathbf{P}} = \mathbf{M}\mathbf{q} + \mathbf{n}, \tag{30}$$

where

$$\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{nn}), \quad \mathbf{q} \sim \mathcal{N}(e^{\mathbf{Hm}}, \mathbf{C}_{qq}).$$

1) The E-step calculates the expected value of $\log p(\mathbf{q}|\mathbf{m})$ given the current estimate $\hat{\mathbf{m}}^{(n)}$, i.e.,

$$Q(\mathbf{m}|\hat{\mathbf{m}}^{(n)}) = E[\log p(\mathbf{q}|\mathbf{m})|\hat{\mathbf{P}}, \hat{\mathbf{m}}^{(n)}]. \tag{31}$$

For a Gaussian distribution of $\mathbf{q}$, the log-likelihood $\log p(\mathbf{q}|\mathbf{m})$ is given by

$$\begin{aligned}\log p(\mathbf{q}|\mathbf{m}) &= c - \frac{1}{2}(\mathbf{q} - e^{\mathbf{Hm}})^T \mathbf{C}_{qq}^{-1}(\mathbf{q} - e^{\mathbf{Hm}}) \\ &= \frac{1}{2}(e^{\mathbf{Hm}})^T \mathbf{C}_{qq}^{-1}\mathbf{q} + \frac{1}{2}\mathbf{q}^T \mathbf{C}_{qq}^{-1} e^{\mathbf{Hm}} \\ &\quad - \frac{1}{2}(e^{\mathbf{Hm}})^T \mathbf{C}_{qq}^{-1} e^{\mathbf{Hm}} + d, \end{aligned} \tag{32}$$

where $c = -\frac{1}{2}\log(\det(2\pi\mathbf{C}_{qq}))$ and $d = c - \frac{1}{2}\mathbf{q}^T\mathbf{C}_{qq}^{-1}\mathbf{q}$ indicate the terms that do not depend on $\mathbf{m}$. Substituting (32) into (31), we obtain

$$\begin{aligned} Q(\mathbf{m}|\hat{\mathbf{m}}^{(n)}) &= \frac{1}{2}(e^{\mathbf{Hm}})^T \mathbf{C}_{qq}^{-1}\hat{\mathbf{q}}^{(n)} + \frac{1}{2}(\hat{\mathbf{q}}^{(n)})^T \mathbf{C}_{qq}^{-1} e^{\mathbf{Hm}} \\ &\quad - \frac{1}{2}(e^{\mathbf{Hm}})^T \mathbf{C}_{qq}^{-1} e^{\mathbf{Hm}} + d \\ &= -\frac{1}{2}(\hat{\mathbf{q}}^{(n)} - e^{\mathbf{Hm}})^T \mathbf{C}_{qq}^{-1}(\hat{\mathbf{q}}^{(n)} - e^{\mathbf{Hm}}) + e, \end{aligned} \tag{33}$$

where $e = d + \frac{1}{2}(\hat{\mathbf{q}}^{(n)})^T \mathbf{C}_{qq}^{-1}\hat{\mathbf{q}}^{(n)}$ and $\hat{\mathbf{q}}^{(n)} = E[\mathbf{q}|\hat{\mathbf{P}}, \hat{\mathbf{m}}^{(n)}]$ is the MMSE estimate of $\mathbf{q}$ at the current iteration n [35]. For $\hat{\mathbf{P}}$ and $\mathbf{q}$ jointly Gaussian we have

$$\begin{aligned} p(\mathbf{q}|\hat{\mathbf{P}}, \hat{\mathbf{m}}^{(n)}) \sim \mathcal{N}(&\mu_q + \mathbf{C}_{q\hat{P}}\mathbf{C}_{\hat{P}\hat{P}}^{-1} \\ &\times (\hat{\mathbf{P}} - \mu_{\hat{P}}), \mathbf{C}_{qq} - \mathbf{C}_{q\hat{P}}\mathbf{C}_{\hat{P}\hat{P}}^{-1}\mathbf{C}_{\hat{P}q}), \end{aligned}$$

where the respective mean values and covariance matrices are given by

$$\begin{aligned} \mu_q &= E[\mathbf{q}] = e^{\mathbf{H}\hat{\mathbf{m}}^{(n)}}, \\ \mu_{\hat{P}} &= E[\hat{\mathbf{P}}] = \mathbf{M}\mu_q = \mathbf{M}e^{\mathbf{H}\hat{\mathbf{m}}^{(n)}}, \\ \mathbf{C}_{\hat{P}\hat{P}} &= E[(\hat{\mathbf{P}} - \mu_{\hat{P}})(\hat{\mathbf{P}} - \mu_{\hat{P}})^T] = \mathbf{M}\mathbf{C}_{qq}\mathbf{M}^T + \mathbf{C}_{nn}, \\ \mathbf{C}_{q\hat{P}} &= E[(\mathbf{q} - \mu_q)(\hat{\mathbf{P}} - \mu_{\hat{P}})^T] = \mathbf{C}_{qq}\mathbf{M}^T. \end{aligned}$$

Therefore the MMSE estimate $\hat{\mathbf{q}}^{(n)}$ is the mean value of the *a posteriori* probability $p(\mathbf{q}|\hat{\mathbf{P}}, \hat{\mathbf{m}}^{(n)})$, i.e.,

$$\begin{aligned} \hat{\mathbf{q}}^{(n)} = E[\mathbf{q}|\hat{\mathbf{P}}, \hat{\mathbf{m}}^{(n)}] &= \mu_q + \mathbf{C}_{q\hat{P}}\mathbf{C}_{\hat{P}\hat{P}}^{-1}(\hat{\mathbf{P}} - \mu_{\hat{P}}) = e^{\mathbf{H}\hat{\mathbf{m}}^{(n)}} \\ &\quad + \mathbf{C}_{qq}\mathbf{M}^T(\mathbf{M}\mathbf{C}_{qq}\mathbf{M}^T + \mathbf{C}_{nn})^{-1}(\hat{\mathbf{P}} - \mathbf{M}e^{\mathbf{H}\hat{\mathbf{m}}^{(n)}}). \end{aligned} \tag{34}$$

2) The M-step finds the new estimate $\hat{\mathbf{m}}^{(n+1)}$ that maximizes the expected value of $Q(\mathbf{m}|\hat{\mathbf{m}}^{(n)})$ computed using (33) and (34), i.e.,

$$\begin{aligned} \hat{\mathbf{m}}^{(n+1)} &= \arg\max_{\mathbf{m}} Q(\mathbf{m}|\hat{\mathbf{m}}^{(n)}) \\ &= \arg\min_{\mathbf{m}} \left\{ (\hat{\mathbf{q}}^{(n)} - e^{\mathbf{Hm}})^T \mathbf{C}_{qq}^{-1}(\hat{\mathbf{q}}^{(n)} - e^{\mathbf{Hm}}) \right\}. \end{aligned} \tag{35}$$

The estimation algorithm terminates when

$$\|\hat{\mathbf{m}}^{(n+1)} - \hat{\mathbf{m}}^{(n)}\| < \varepsilon, \tag{36}$$

where ε is a given threshold.

VI. Experimental Evaluation

In order to evaluate the convergence behavior of the iterative estimation algorithm, we preliminarily conducted a simulation in the room depicted in Fig. 7. The microphone array is denoted by the large circle on the right-bottom corner of the room, the source is in the small circle at the top-right corner, and the numbers that are close to the walls are the indices $k = 1, \ldots, 4$ of the reflective surfaces whose reflection coefficients are to be estimated. The actual values of the reflection coefficients are

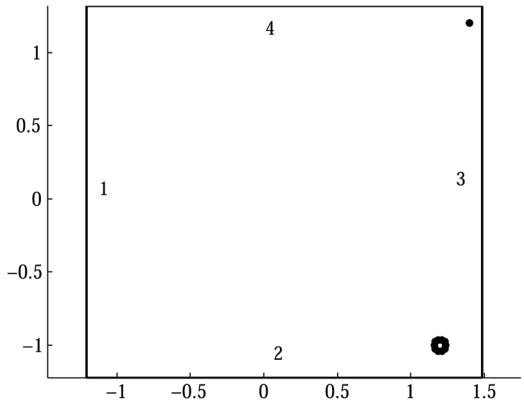

Fig. 7. Room simulation for testing the convergence of the iterative algorithm. The microphone array is depicted with the empty circle in the right-bottom corner and the source is in the top-right corner. Numbers denote the indexes $k = 1, \ldots, 4$ of the walls.

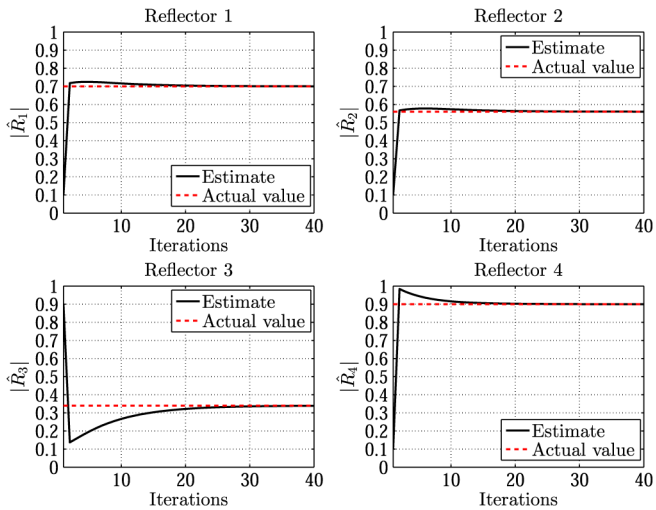

Fig. 8. Simulation results for testing the convergence of the iterative algorithm: the estimated reflection coefficients R_k, $k = 1, \ldots, 4$, at different iterations of the algorithm (black lines) and the actual values of the reflection coefficients (dashed lines).

$R_1 = 0.7$, $R_2 = 0.56$, $R_3 = 0.34$, $R_4 = 0.9$. Note that this is quite a challenging scenario, as the angle of the reflection path is far from the normal-incidence condition, and therefore we cannot write that $g^{(n)}(\theta_k) \approx 1$. This is especially true for the surface $k = 3$, which has the smallest reflection coefficient and the largest incidence angle of the first-order reflection paths. The algorithm has been initialized with values of the reflection coefficients that are purposefully very far from the correct ones. The result of the simulation is shown in Fig. 8, which shows that the estimates converge to the correct values after very few iterations. More specifically, we observe that the algorithm converges after three or four steps for walls 1 and 2. Even if in an unfavorable configuration, the estimates converge to the correct values also for surfaces 3 and 4.

We also conducted experiments with real-world data. Measurements were conducted in a semi-anechoic room at the Chair of Multimedia Communications and Signal Processing at the University of Erlangen-Nuremberg, the setup of which is depicted in Fig. 9. According to the specific experiment, walls

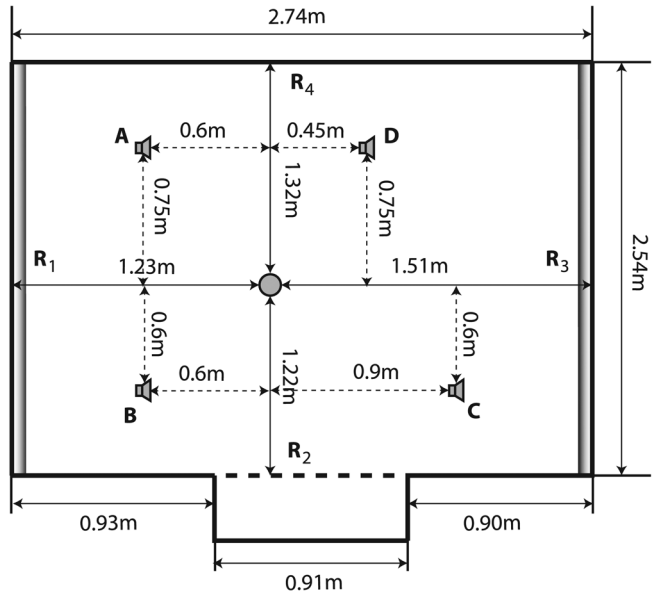

Fig. 9. Experimental measurement setup.

$\mathbf{R}_1$ and $\mathbf{R}_3$ are covered with reflective or semi-reflective materials, whereas walls $\mathbf{R}_2$ and $\mathbf{R}_4$ are absorptive. This is a rather challenging scenario for the reflection coefficient estimation task. In fact, peaks related to the reflection paths bouncing off absorptive walls are strongly attenuated, which renders the estimation of reflection coefficients of absorptive walls quite sensitive to even small measurement and modeling errors. The sound field is captured with a circular microphone array that accommodates ten omnidirectional microphones mounted in a rigid cylindrical baffle with a radius of 0.04 m [40]. The environment is probed by a GENELEC 1029A loudspeaker, the frequency-dependent radiation pattern of which has been provided by the manufacturer. Reflections up to the third order are simulated by the modeling engine, i.e., $L = 3$. The loudspeaker, especially at high frequencies, is strongly directional and is characterized by the maximum directivity in the frontal direction. Reflectors that are located at the rear of the loudspeaker, therefore, would not be adequately probed. In order to overcome this issue, four different orientations (0°, 90°, 180° and 270°) for each position of the loudspeaker are adopted. Measurements are then combined as proposed in [20]. Notice that the use of an omnidirectional loudspeaker would overcome this problem.

In order to test the robustness of the proposed method, the estimation is performed using four different loudspeaker positions, which are marked in Fig. 9 with symbols $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$. For each position, a different and independent estimate is obtained. We expect these estimates to be independent of the source position.

In the first scenario, walls $\mathbf{R}_1$ and $\mathbf{R}_3$ are covered with two distinctly reflective materials. More specifically, $\mathbf{R}_1$ is covered with a 30 mm thick aixFOAM semi-reflective material [41], and $\mathbf{R}_3$ is covered with a 10 mm thick Sonatech reflective material [42], and finally $\mathbf{R}_2$ and $\mathbf{R}_4$ are acoustically similar to thermafleece (see [43, p. 443]). These datasheets describe the reflective properties in subbands in terms of absorption coefficient

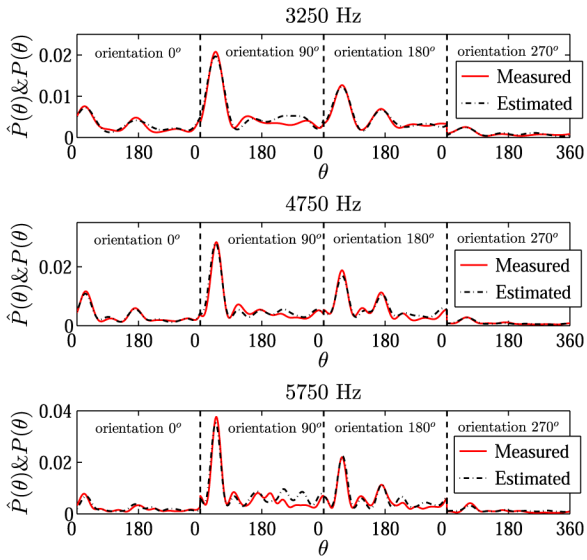

Fig. 10. Examples of matching between modeled and measured pseudospectra at different frequencies for a DAS beamformer; position of the loudspeaker is $\mathbf{A}$ with orientations 0°, 90°, 180° and 270°.

value α according to ISO 354:2003 [9], which is related to the reflection coefficient R by

$$|R| = \sqrt{1-\alpha}.$$

The covariance matrices $\mathbf{C}_{nn}$ and $\mathbf{C}_{qq}$ are assumed to be diagonal and their values are chosen empirically as a trade-off between a good match of $\mathbf{M}\hat{\mathbf{q}}$ to $\hat{\mathbf{P}}$ (smaller variance of $\mathbf{n}$) and a good match of $e^{\mathbf{H}\hat{\mathbf{m}}}$ to $\hat{\mathbf{q}}$ (smaller variance of $\mathbf{q}$).

Estimates were computed using both DAS and MVDR beamformers. The DAS beamformer provided estimates over 10 subbands, whose central frequencies range from 250 to 4750 Hz. The algorithm first estimates the reflection coefficient value at 4750 Hz and then recursively initializes the estimate at the lower subband with the estimate of the upper subband. The choice of starting with the highest subband is dictated by the higher resolution of the pseudospectrum at high frequencies, which enables a more accurate matching between $\hat{\mathbf{P}}$ and $\mathbf{P}$. The pseudospectrum $\hat{\mathbf{P}}$ measured at position $\mathbf{A}$ and the estimated pseudospectrum $\mathbf{M}e^{\mathbf{H}\hat{\mathbf{m}}}$ at three different frequencies are shown in Fig. 10. Note that a very good match can be observed, especially at high frequencies and for relevant peaks.

The estimates of the frequency-dependent reflection coefficients for $\mathbf{R}_1, \ldots, \mathbf{R}_4$ along with the reference values are shown in Fig. 11. We limit the visualization up to $f = 5$ kHz, as reference values for higher frequencies are not available. It can be seen that the estimated values are consistent for all the loudspeaker positions. Furthermore, the estimates for all the walls are close to the values specified in the datasheets, while the values for walls $\mathbf{R}_2$ and $\mathbf{R}_4$ quite closely match the datasheet in [43]. It is worth noting, however, that the reference values for $\mathbf{R}_2$ and $\mathbf{R}_4$ are less accurate than for $\mathbf{R}_1$ and $\mathbf{R}_3$.

In order to further verify the ability of the proposed approach to correctly estimate the reflection coefficients of different reflectors in the environment in the presence of both highly reflective and absorptive surfaces, the experiment has been repeated

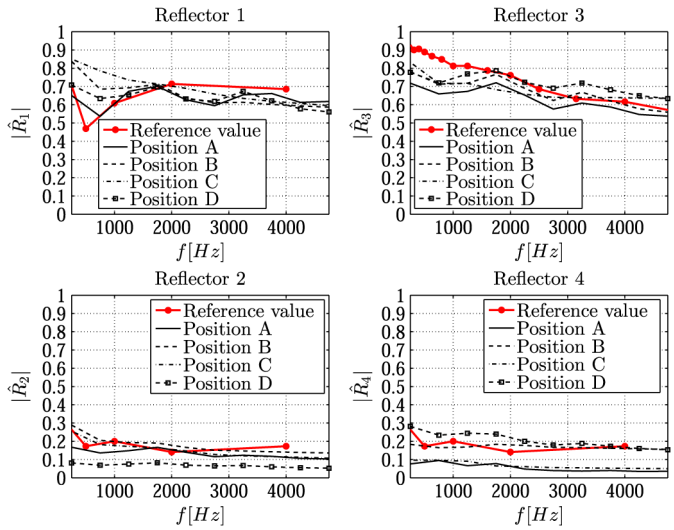

Fig. 11. Estimates using the DAS beamformer in the first scenario. Walls $\mathbf{R}_1$ and $\mathbf{R}_3$ are covered with reflective materials. Wall $\mathbf{R}_2$ and $\mathbf{R}_4$ are absorptive.

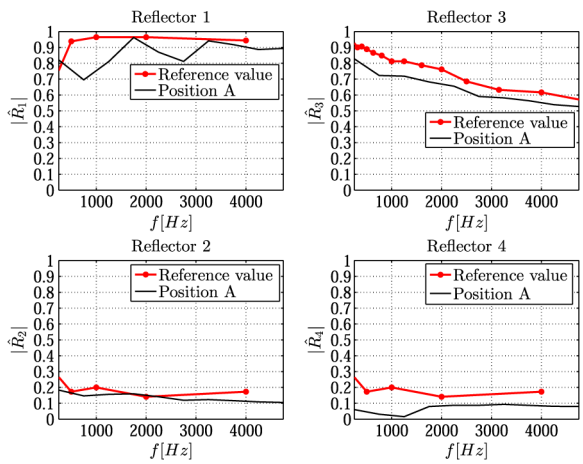

Fig. 12. Estimates using the DAS beamformer for position $\mathbf{A}$ in the second scenario. Wall $\mathbf{R}_1$ is covered with a wooden panel and $\mathbf{R}_3$ is covered with a well-defined, moderately reflective material. Walls $\mathbf{R}_2$ and $\mathbf{R}_4$ are absorptive.

in two other scenarios. In the second scenario, the wall $\mathbf{R}_1$ has been replaced with a wooden panel, while the wall $\mathbf{R}_3$ is covered with Sonatech. The estimates for position $\mathbf{A}$ are shown in Fig. 12. As expected, the estimate of $\mathbf{R}_3$ is reasonably similar to the one obtained in the first scenario, while the estimate of $\mathbf{R}_1$ has values that are close to the typical values for wooden panels found in the literature ([43, p. 442]).

In the third scenario, the wall $\mathbf{R}_1$ is covered with aixFOAM, while all the other walls are absorptive. The estimates at position $\mathbf{A}$ are shown in Fig. 13. The estimate of $\mathbf{R}_1$ is similar to the estimate obtained in the first scenario, whereas the estimate of $\mathbf{R}_3$ has a low value similar to the values estimated for other absorptive walls.

In order to further improve the estimation at lower frequencies and for absorptive walls, the MVDR beamformer is applied. Fig. 14 shows the matching of the modeled and measured MVDR pseudospectra at position A. If we compare the results of the MVDR with those of DAS, we observe that the use of MVDR greatly improves the directivity of the

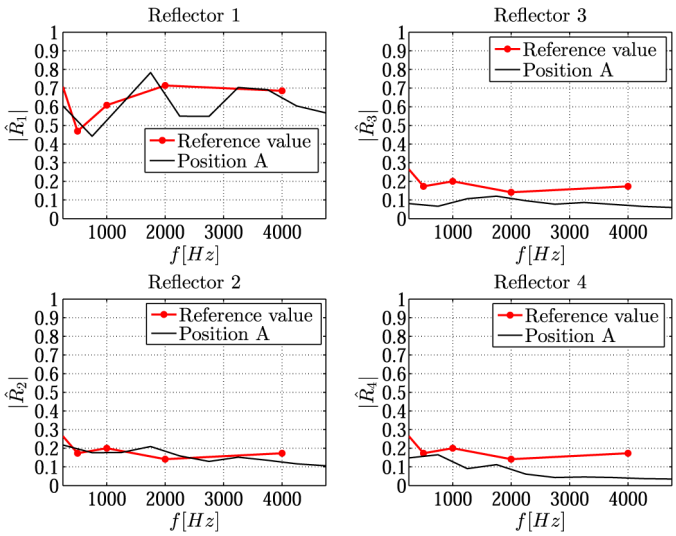

Fig. 13. Estimates using the DAS beamformer for position $\mathbf{A}$ in the third scenario. Wall $\mathbf{R}_1$ is covered with a reflective panel. Walls $\mathbf{R}_2, \ldots, \mathbf{R}_4$ are absorptive.

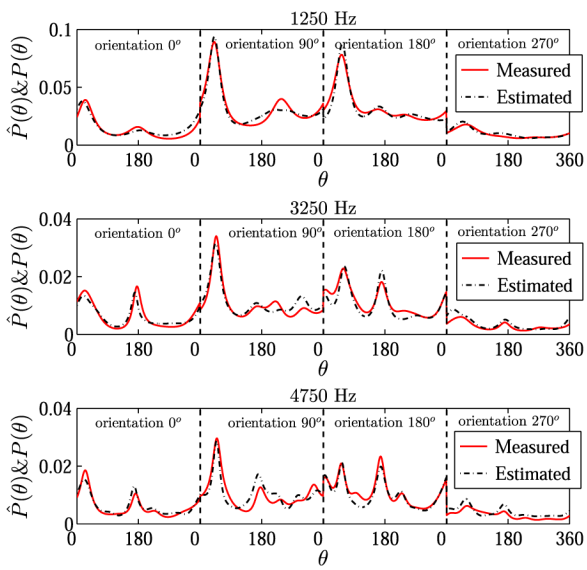

Fig. 14. Examples of matching between modeled and measured pseudospectra at different frequencies for a MVDR beamformer; position of the loudspeaker is $\mathbf{A}$ with orientations 0°,90°, 180° and 270°.

pseudospectra, especially at low frequencies. We also observe a slight improvement in the matching, especially in proximity of secondary peaks and smooth portions. Fig. 15 shows the reflection coefficient estimates in the first scenario obtained using the MVDR beamformer. The results are obtained over the frequency bands ranging from 250 to 5000 Hz, in which the frequency smoothing method is applied. As clearly shown, for all the positions and all the walls, the reflection coefficient values estimated with the MVDR are much closer to the reference values if compared to the estimates obtained using DAS in the same scenario (see Fig. 11 for comparison). This is particularly true in the case of absorptive walls. Note especially that the reflection coefficient of walls $\mathbf{R}_2$ and $\mathbf{R}_4$ in Fig. 15 exhibits a trend that is consistent for all the tested loudspeaker positions, whereas the estimates obtained using DAS are characterized

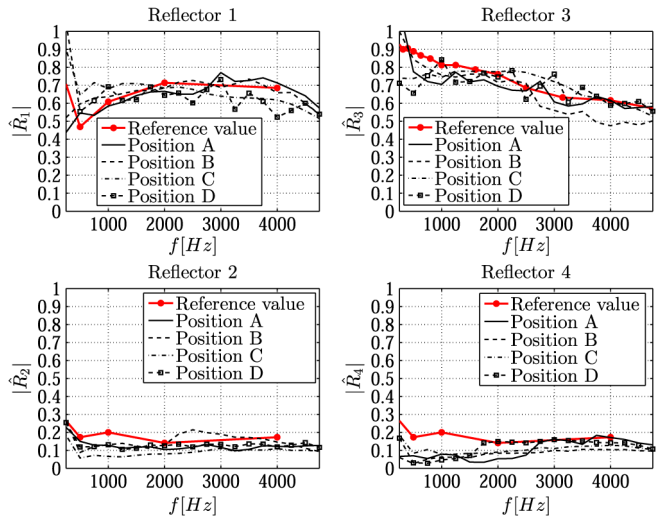

Fig. 15. Estimates using the MVDR beamformer in the first scenario. Walls $\mathbf{R}_1$ and $\mathbf{R}_3$ are covered with reflective materials. Walls $\mathbf{R}_2$ and $\mathbf{R}_4$ are absorptive.

TABLE I
ESTIMATION VARIANCE ACROSS SOURCE POSITIONS IN DIFFERENT FREQUENCY BANDS.

	750 Hz	1750 Hz	2750 Hz	3750 Hz	4750 Hz
DAS					
$var\|\hat{R}_1\|$	0.0108	0.0001	0.0007	0.0006	0.0006
$var\|\hat{R}_2\|$	0.0036	0.0022	0.0013	0.0012	0.0012
$var\|\hat{R}_3\|$	0.0008	0.0022	0.0022	0.0015	0.0025
$var\|\hat{R}_4\|$	0.0044	0.0071	0.0053	0.0049	0.0042
MVDR					
$var\|\hat{R}_1\|$	0.0056	0.0003	0.0006	0.0039	0.0006
$var\|\hat{R}_2\|$	0.0007	0.0005	0.0017	0.0007	0.0003
$var\|\hat{R}_3\|$	0.0009	0.0004	0.0052	0.0050	0.0012
$var\|\hat{R}_4\|$	0.0010	0.0019	0.0008	0.0011	0.0002

by a larger variance among different positions. To support this claim, the variance of estimated reflection coefficients across source positions for different walls and beamforming methods is provided in Table I as function of frequency. Notice that MVDR appears to be more stable with respect to DAS especially for low frequencies and absorbing walls.

VII. CONCLUSIONS

In this manuscript, we have presented a methodology for the in-situ estimation of the reflection coefficients of planar walls. The proposed technique is based on a two-step procedure that first compares measured and modeled pseudospectra, then performs the estimation of the reflection coefficients in frequency bands. Based on the EM algorithm, we devised and developed an iterative estimation scheme that enables an accurate estimation of reflection coefficients even at low frequencies and for absorptive materials, as proven by the experimental results using materials with known acoustical properties. In particular, the use of MVDR beamformers pseudospectrum has been proven beneficial for absorptive materials, as it led to more consistent estimates than for the DAS when varying the source position.

In the manuscript, we considered environments characterized by large planar surfaces with homogeneous material properties for each surface. In case of an inhomogeneous surface, we could split the surface into separate regions characterized by distinct

reflection coefficients. Thus, however, the number of unknown reflection coefficients would increase and, as a consequence, the amount of available data (acquired reflection paths) per unknown would decrease. Moreover, discontinuities of the material create diffractive waves on the edges of each portion of material. Diffractive waves may have opposite polarity with respect to the reflected wave, thus decreasing its apparent amplitude. Such diffractive phenomena arise also for curved and/or small surfaces. We should keep in mind, therefore, that the accuracy of the estimation algorithm for small, curved or inhomogeneous surfaces could suffer from inaccuracies. This is especially true at low frequencies, where the diffractive waves become more relevant. Ensuring the robustness of the estimation procedure for such cases is an interesting topic of future research.

REFERENCES

[1] T. Betlehem and T. Abhayapala, "Theory and design of sound field reproduction in reverberant rooms," *J. Acoust. Soc. Amer.*, vol. 117, pp. 2100–2111, 2005.

[2] A. Canclini, D. Marković, F. Antonacci, A. Sarti, and S. Tubaro, "A room-compensated virtual surround system exploiting early reflections in a reverberant room," in *Proc. EURASIP Conf. Signal Process. (EUSIPCO)*, 2012.

[3] H. Chung, H. Shim, N. Hahn, S. Chon, and K. Sung, "Sound reproduction method by front loudspeaker array for home theater applications," *IEEE Trans. Consumer Electron.*, vol. 58, no. 2, pp. 528–534, May 2012.

[4] F. Ribeiro, C. Zhang, D. Florencio, and D. Ba, "Using reverberation to improve range and elevation discrimination for small array sound source localization," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 7, pp. 1781–1792, Sep. 2010.

[5] M. Delany and E. Bazley, "Acoustical properties of fibrous absorbent materials," *Appl. Acoust.*, vol. 3, no. 2, pp. 105–116, Apr. 1970.

[6] Y. Liu and F. Jacobsen, "Measurement of absorption with a pu sound intensity probe in an impedance tube," *J. Acoust. Soc. Amer.*, vol. 118, pp. 2117–2120, 2005.

[7] *Determination of sound absorption coefficient and impedance in impedance tubes - Part 1: Method using standing wave ratio*, ISO Std. 10 534–1, 1996.

[8] *Determination of sound absorption coefficient and impedance in impedance tubes - Part 2: Transfer-function method*, ISO Std. 10 534–2, 1998.

[9] *Measurement of sound absorption in a reverberation room*, ISO Std. 354, 2003.

[10] E. Mommertz, "Angle-dependent in-situ measurements of reflection coefficients using a subtraction technique," *Appl. Acoust.*, vol. 46, pp. 251–265, 1995.

[11] C. Nocke, "In-situ acoustic impedance measurement using a free-field transfer function method," *Appl. Acoust.*, vol. 59, no. 3, pp. 253–264, 2000.

[12] J. Ducourneau, V. Planeau, and A. Nejade, "Design of a multipolar weighting for acoustic antennae," *Appl. Acoust.*, vol. 70, pp. 484–492, 2009.

[13] J. Rathsam and B. Rafaely, "Analysis of in-situ acoustic absorption using a spherical microphone array," *Proc. Meetings Acoust.*, vol. 6, 2009.

[14] R. Lanoye, G. Vermeir, W. Lauriks, R. Kruse, and V. Mellert, "Measuring the free field acoustic impedance and absorption coefficient of sound absorbing materials with a combined particle velocity-pressure sensor," *J. Acoust. Soc. Amer.*, vol. 119, no. 5, pp. 2826–2831, 2006.

[15] G. Dutilleux, F. C. Sgard, and U. R. Kristiansen, "Low-frequency assessment of the in situ acoustic absorption of materials in rooms: An inverse problem approach using evolutionary optimization," *Int. J. Numer. Meth. Eng.*, vol. 53, no. 9, pp. 2143–2161, 2002.

[16] M. Müller-Trapet and M. Vorländer, "Signal processing for hemispherical measurement data," in *Proc. Meetings Acoust.*, 2013, vol. 19, ASA.

[17] *Measurement of sound absorption properties of road surfaces in situ - Part 1: Extended surface method*, ISO Std. 13, 2002.

[18] P. Stoica and R. Moses, "Spatial methods," in *Introduction to Spectral Analysis*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1997, ch. 6, pp. 232–238.

[19] D. Johnson and D. Dudgeon, *Array Signal Processing: Concepts and Techniques*, ser. Prentice-Hall signal processing series. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[20] D. Markovic, C. Hofmann, F. Antonacci, K. Kowalczyk, A. Sarti, and W. Kellermann, "Reflection coefficient estimation by pesudospectrum matching," in *Proc. Int. Workshop Acoust. Echo Noise Cancell. (IWAENC)*, 2012.

[21] H. Kuttruff, *Room acoustics*, 4th Ed. ed. London, U.K.: Spon, 2000.

[22] A. Brutti, M. Omologo, and P. Svaizer, "Inference of acoustic source directivity using environment awareness," in *Proc. EURASIP Signal Process. Conf. (EUSIPCO)*, 2012.

[23] P. C. Meuse and H. F. Silverman, "Characterization of talker radiation pattern using a microphone array," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1994, pp. 257–264.

[24] F. Antonacci, M. Foco, A. Sarti, and S. Tubaro, "Fast tracing of acoustic beams and paths through visibility lookup," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 4, pp. 812–824, May 2008.

[25] H. van Trees, *Optimum array processing: Detection, estimation, and modulation theory*. New York, NY, USA: Wiley, 2002.

[26] B. V. Veen and K. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, pp. 4–24, Apr. 1988.

[27] O. Frost, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.

[28] J. Bitzer and K. Simmer, "Superdirective microphone arrays," in *Microphone Arrays: Signal Processing Techniques and Applications*. Berlin, Germany: Springer, 2001, ch. 2, pp. 19–37.

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

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

[31] T. D. Abhayapala, "Broadband source localization by modal space processing," in *Advances in Direction-of-Arrival Estimation*. Norwood, MA, USA: Artech House, 2006.

[32] H. Sun, E. Mabande, K. Kowalczyk, and W. Kellermann, "Localization of distinct reflections in rooms using spherical microphone array eigenbeam processing," *J. Acoust. Soc. Amer.*, vol. 131, no. 2, pp. 2828–2840, 2012.

[33] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *J. R. Statist. Soc.. Ser. B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[34] M. Feder, "Parameter estimation of superimposed signals using the em algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 4, pp. 477–489, Apr. 1988.

[35] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, ser. ser. Prentice Hall Signal Processing Series. Englewood Cliffs, NJ, USA: Prentice-Hall PTR, 1993, vol. 1.

[36] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate analysis*. New York, NY, USA: Academic, 1979.

[37] F. Antonacci, J. Filos, M. Thomas, E. Habets, A. Sarti, P. Naylor, and S. Tubaro, "Inference of room geometry from acoustic impulse responses," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 10, pp. 2683–2695, Dec. 2012.

[38] J. Filos, M. Thomas, F. Antonacci, A. Sarti, and P. Naylor, "Robust inference of room geometry from acoustic measurements using the hough transform," in *Proc. 19th Eur. Signal Process. Conf. (EUSIPCO'11)*, 2011, pp. 161–165.

[39] J. D. Gibbons and S. Chakraborti, *Nonparametric statistical inference*. Boca Raton, FL, USA: CRC, 2003, vol. 168.

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

[41] AixFoam, [Online]. Available: http://www.aixfoam.com/absorption-foam-sh0061-felt-lamination

[42] SonaTech, [Online]. Available: http://www.sonatech.de/konfiguration/dateienpdf/SONATECH_PUR_SKIN_ProspektSeite6.pdf

[43] T. Cox and P. D'Antonio, *Acoustic Absorbers and Diffusers: Theory, Design and Application*. London, U.K.: Taylor & Francis, 2009.