HIGH RESOLUTION IMAGING OF ACOUSTIC REFLECTIONS WITH SPHERICAL MICROPHONE ARRAY

Lucio Bianchi, Marco Verdi, Fabio Antonacci, Augusto Sarti, Stefano Tubaro

Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano
Piazza Leonardo da Vinci 32 – 20133 Milano, Italy

ABSTRACT

This paper proposes a methodology for the accurate visualization of acoustic reflections in a room from acoustic measurements by a spherical microphone array. The goal is to provide insight on the relationship between architectural and acoustic features. This task requires high-resolution acoustic images. In this contribution, we achieve this goal by introducing two main modifications of existing approaches. At first, we adopt an explicit model that takes into account the scattering due to the rigid spherical surface where the microphone capsules are hosted. Then, we obtain an estimate of the acoustic power coming from a grid of directions using a spectral analysis approach based on the matching of array data covariance matrix. We conclude this manuscript by showing applications of the devised methodology in real world cases.

Index Terms— Acoustic reflections, spherical microphone array, space-time audio processing.

1. INTRODUCTION

Early reflections have an important influence on the perception of sound in enclosed spaces [1, 2]. They can be regarded as a set of spatio-temporal events occurring at distinct time instants and originating from distinct spatial locations. Undesired reflections coming from walls and obstacles could represent a problem. For this reason, researchers in acoustics have worked in the past on an accurate determination of the direction of arrival of the early reflections in an impulse response (see, e.g., [3, 4, 5, 6] and references therein). Furthermore, the determination of the direction of the early reflections is important also in applications of room equalization and correction [7, 8, 9], in the identification and control of industrial and aerospace noise sources [10] and for speech enhancement [11].

This paper describes a methodology for estimating the temporal and spatial distribution of reflections in an enclosed space. Particular attention is devoted to the data visualization paradigm, chosen carefully to provide an accessible and intuitive source of information. We base our analysis on the impulse responses acquired by exciting the environment with an impulsive sound source [12].

All the state-of-the-art imaging techniques use a microphone array to acquire several impulse responses simultaneously, by steering the response of the array towards different directions. In this context, spherical arrays represent the most widely used geometry. In [3] an open spherical microphone array is used to compute the impulse responses over a grid of directions and results are shown on a cartographic-like map. In [13, 14] a rigid spherical microphone array is described and the steering is achieved through the decomposition of the incident sound field in terms of spherical harmonics. High-resolution images can be obtained within this framework by performing a sub-space pre-processing [15].

In [16] a panoramic projection of the visual image of the enclosure is layered with the magnitude of directional impulse response, obtained through a beamforming based on spherical harmonic decomposition. In [17, 18] the previous approach is refined by using measured directional array transfer functions to deconvolve the acquired signals. The main issue of the latter two approaches lies in the fact that a preliminary measurement session in an anechoic room is required, in order to determine the response of each microphone in the array to sound waves impinging from different directions. In this paper we adopt a rigid spherical microphone array configuration, and we formulate the directional imaging problem as a spectral analysis problem, representing the sound field as a spectrum of plane waves propagating in the environment. A state-of-the-art spectral analysis tool [19, 20], based on the covariance matrix of array data, is used to obtain an high resolution estimate of the plane waves spectrum. Since the magnitude plane wave spectrum is related to the sound intensity coming from a specific direction [21, Sec. 4.4], we obtain an estimate of the intensity of the reflections coming from a specific direction. Finally, we map the reflection intensity diagram on a panoramic visual image of the environment.

The rest of the paper is structured as follows. In Section 2 we formulate the imaging problem as a spectral analysis problem and we introduce the scattering model for the rigid sphere. In Section 3 we review the spectral estimation approach. In Section 4 we show some experimental results. Finally, Section 5 draws some conclusions.

2. DATA MODEL AND PROBLEM FORMULATION

In this section we introduce the data model adopted throughout the rest of the manuscript. We consider the sound field generated by a single sound source in a reflective environment and acquired by a spherical microphone array. Under the assumption of far-field propagation (i.e., the excitation source and the reflectors are considered to be in the far field with respect to the microphone array [16]), we can parametrize the reflections according to their angle of incidence on the microphone array.

Let \( y(t) \in \mathbb{R}^{M \times 3} \) denote the impulse response vector acquired by a spherical microphone array composed by \( M \) microphones. In Sec. 4 we discuss the impulse response measurement technique adopted in this work; however, we remark that the proposed approach is independent on the particular technique adopted. Our analysis is based on short-time segments of the acquired impulse responses, each weighed by a suitable window function. Let also \( y(\omega_l, t_m) \in \mathbb{C}^{M \times 1} \) be the \( m \)-th bin of its Short-Time Discrete Fourier Transform [22] for the segment centered in \( t_m \), i.e. \( \omega_l = 2\pi f_l/F_s \), being \( f_l \) and \( F_s \) the temporal frequency and the temporal sampling frequency, respectively. We denote by \( c \) the speed of sound. With
the above notation at hand, the signal acquired by the microphone array can be written as
\[ y(\omega_l, t_n) = \sum_{q=1}^{Q} a(\gamma_q, \omega_l) s_q(\omega_l, t_n) + v(\omega_l, t_n), \] (1)
where \( v(\omega_l) \in \mathbb{C}^{M \times 1} \) is an additive noise term; \( \gamma_q = (\theta, \phi)_q \in \Gamma \) and \( s_q \in \mathbb{C} \) are the unknown parameters of the \( q \)th reflection, i.e. the direction of arrival and the associated signal, respectively; and \( a(\cdot) : \Gamma \rightarrow \mathbb{C}^{M \times 1} \) is the propagation function. We denote with \( \theta \) the polar angle and with \( \phi \) the azimuth angle. Figure 1a shows the reference frame.

The function \( a(\gamma_q, \omega_l) \) encodes the modifications undertaken by a planar wave field component impinging on the microphone array from direction \( \gamma_q \). More formally, the \( m \)th component of \( a(\gamma_q, \omega_l) \) represents the pressure generated at the \( m \)th microphone due to an impulsive excitation coming from direction \( \gamma_q \). Denoting by \( x_m \in \mathbb{R}^2 \) the position vector relative to the \( m \)th microphone located on the surface of a rigid sphere of radius \( x \), and by \( d_q \in \mathbb{R}^3 \) the unit vector in the direction \( \gamma_q \), we can write [23, Eq. (10)]
\[ a(\gamma_q, \omega_l)_m = -j \frac{i}{(\omega_l/c)^2} \sum_{\mu=0}^{\infty} \frac{\omega_l^2 + 1}{\omega_l^2} P_\mu(\chi_{x_m, d_q}), \] (2)
where \( i \) is the imaginary unit; \( P_\mu(\cdot) \) is the Legendre polynomial of degree \( \mu \) [24, Chap. 18]; \( \chi_{x_m, d_q} \) is the standard inner product in \( \mathbb{R}^3 \); and \( h'_\mu(\omega_l/x/c) \) is the first derivative of the spherical Hankel function of first kind and order \( \mu \) [24, Sec. 10.47].

In this work we aim at estimating the parameters \( \gamma_q \) and \( |s_q|^2 \) associated with each reflection observed in the \( m \)th frame of the acquired signals. However, the total number of reflections \( Q \) and their associated directions of arrival are unknown a-priori. Hence we adopt a non-parametric estimation method [25, Sec. 6.3] on a predefined angular grid. We denote by \{\( \gamma_q \}_q \}, \{u, v\} \in \mathbb{Z}^2 the element of a grid that covers \( \Gamma \) and we assume that each \( \gamma_q \) is close to a grid point, as depicted in Fig. 1b; in other words, there exist angles \( (\theta_1, \phi_1), \ldots, (\theta_T, \phi_T) \) such that \( \gamma_q \approx \gamma_\hat{u} \). We define \( U = T \Psi \) as the total number of grid points. We let \( a_w(\omega_l) = a(\gamma_q, \omega_l) \) and \( s_\hat{u}(\omega_l, t_n) = \begin{cases} s_q(\omega_l, t_n), & \text{if } \hat{u} = q \\ 0, & \text{otherwise} \end{cases} \) (3)

Using this notation, the model in (1) can be rewritten as
\[ y(\omega_l, t_n) = A(\omega_l)S(\omega_l, t_n) + e(\omega_l, t_n), \] (4)
where we have defined \( A(\omega_l) = [a_{w(1)}(\omega_l), \ldots, a_{w(T)}(\omega_l)] \) and \( \{S(\omega_l, t_n)\}_u = s_\hat{u}(\omega_l, t_n) \) so that \( A(\omega_l) \in C^{M \times U} \) and \( S(\omega_l, t_n) \in C^{U \times 1} \).

In the following sections we show how to compute an accurate estimation of \( |s_\hat{u}(\omega_l, t_n)|^2 \) given the array data, and how we associate them to a panoramic visual image of the environment under test in order to provide an high-resolution reflection intensity map. We remark that the estimation problem based on the data model in Eq. (4) is of widespread use in many application fields; in particular, this estimation problem is widely studied in the context of spectral analysis. This fact allows us to resort to widely studied spectral analysis approaches to solve the estimation problem.

3. PARAMETER ESTIMATION

In this section we briefly review the estimation method introduced in [19, 20] and we show how it can be applied to the problem at hand. In order to simplify the notation, in this section we omit the dependency of the data on \( \omega_l \) and \( t_n \).

Under the assumption of uncorrelated noise and sources, the covariance matrix of the array data can be written as [25, Eq. (6.4.3)]
\[ R = E[\tilde{y}y^H] = A^H T A + V, \] (5)
where \( T \) is the propagation function, the problem in (8) can be equivalently minimized problem in (9) is convex and has a global minimum, i.e. by finding the set \( \{\hat{\sigma}_w\}_w \) that minimizes [19, Eq. (19)]
\[ \min_{\{\hat{\sigma}_w\}} \| \hat{R}^{-1/2}(y y^H - R) \|_F^2, \] (8)
where \( \hat{R}^{-1/2} \) denotes the Frobenius norm for matrices. By expanding the cost function, the problem in (8) can be equivalently formulated as [19, Eq. (22)]
\[ \min_{\{\hat{\alpha}_w\}} \{\hat{R}^{-1}y - 1\} + \sum_{w=1}^{U+M} h_w^2 \hat{\sigma}_w, \] (9)
where \( h_w = \| \hat{a}_w \|/\| y \| \) and \( \hat{\sigma}_w \) denotes the matrix trace. The minimization problem in (9) is convex and has a global minimum [19, Sec. III-A].

In the following we review the iterative algorithm derived in [19, Sec. III-B] to solve (9). Consider the modified problem
\[ \min_{\{\hat{\alpha}_w\}, B} \{\hat{R}B^H \hat{\Sigma} B\} + \sum_{w=1}^{U+M} h_w^2 \hat{\sigma}_w \quad \text{s.t.} \quad \hat{A} \hat{B} = y y^H. \] (10)
The minimization over $\mathbf{B}$ for a fixed set $\{\tilde{\sigma}_w\}$ is given by $\dot{\mathbf{B}} = \Sigma A^H R^{-1} y y^H$ [19, Appendix A]. Substituting $\mathbf{B}$ into (10) yields the original problem in (9). This considerations allows to conclude that the sets $\{\tilde{\sigma}_w\}$ obtained from (9) and (10) must be identical.

Upon defining $\mathbf{B} = [\beta_1, \ldots, \beta_W]^T$, the problem (10) for a fixed set $\{\beta_w\}$ can be rewritten as

$$\min_{\{\beta_w\}} \sum_{w=1}^W \|\beta_w\|^2 + h_w^2 \tilde{\sigma}_w. \quad (11)$$

It is shown in [19] that the minimizer of (11) is $\tilde{\sigma}_w = \|\beta_w\|/h_w$. Since the cost function in (10) is convex in both $\tilde{\sigma}_w$ and $\beta_w$, a cyclic minimization over $\mathbf{B}$ and $\{\tilde{\sigma}_w\}$ leads to a global minimum. The $i$th iteration involves the following operations:

$$\mathbf{B}^{(i)} = \Sigma^{(i-1)} A R^{-1} (i-1) y y^H$$

$$\tilde{\sigma}_w^{(i)} = \|\beta_w^{(i)}\|/h_w, \quad w = 1, \ldots, W$$

$$R^{(i)} = A \Sigma^{(i)} A^H.$$ 

Following [19], for the initialization of (12) we use the beamforming estimator $\tilde{\sigma}_w^{(0)} = \|\tilde{a}_w y^H\|^2/\|\tilde{a}_w\|^4$. The algorithm is stopped when the condition $\|\Sigma^{(0)} - \Sigma^{(i-1)}\|/\|\Sigma^{(i-1)}\| < \tau$ is satisfied for a given value of the threshold $\tau$.

4. EXPERIMENTAL VALIDATION

In this section we show some experimental results of the proposed high-resolution imaging approach. All the experiments are performed with mh acoustics' Eigenmike<sup>®</sup> spherical microphone array, composed of $M = 32$ capsules mounted on a rigid sphere, whose exact locations can be found in [26]. The technique can be straightforwardly applied to any rigid spherical microphone array.

The angular grid $\{u\}_{u=1}^U$ is designed to uniformly cover the spherical angular region of interest $\Gamma$. For this purpose, the angular axes $\theta$ and $\phi$ are uniformly sampled over $\Gamma$ and $\Psi$ points, respectively. In this setting, the double index $u$ can be conveniently sorted as $u = (v - 1)\Psi + \psi$, with $v = 1, \ldots, V$, $\psi = 1, \ldots, \Psi$ and $u = 1, \ldots, U$.

In all the experiments, we consider a sampling frequency $f_s = 44.1$ kHz and we consider frames of length $1.5$ ms. Impulse responses are measured using exponential sine sweeps [27]. The acquired impulse responses are processed with an octave pass-band filter centered at $f_c = 4$ kHz, as recommended in [17], for the purpose of a better identification of reflections. All the frequency bins in the passband of the filter concur to the generation of a single acoustic image, with the product of their geometric and harmonic means as suggested in [28]. The threshold $\tau$ is set to $10^{-12}$.

Finally, the power estimates in $\Sigma$ are mapped onto a 2-D image using the equiangular projection [29] (i.e. the azimuth angle is uniformly mapped to the horizontal axis while the polar angle is mapped to the vertical axis of a 2-D plot).

4.1. Validation in a controlled environment

The first set of experiments is conducted in an acoustically controlled environment ($T_{90} = 50$ ms), according to the geometry depicted in Fig. 2a. The acoustic scene consists of a sound source placed at an height of $1.3$ m, the spherical microphone array placed at the same height and distant $1.6$ m from the sound source and a reflective panel placed $1$ m behind the array. Reflections from the ceiling and from one wall (the one behind the sound source) are highly damped through the use of absorbing panels. Reflections from the floor are damped by a thick carpet. Figure 2b shows the impulse response recorded by the first capsule of the microphone array. The peaks are labeled according to the acoustic paths shown in Fig. 2a. We observe that the location of the peaks in Fig. 2b is consistent with the length of the acoustic paths in Fig. 2a; in particular $l_{13} = 1.6$ m ($\sim 4.7$ ms), $l_{14} = 3.05$ m ($\sim 8.8$ ms), $l_{24} = 3.6$ m ($\sim 10.5$ ms), $l_{34} = 4.4$ m ($\sim 12.8$ ms). Acoustic paths reflected by the ceiling and the rear wall are not shown because they are so damped that their associated peaks are not visible in the impulse response.

Figure 2: Geometry and four relevant acoustic paths (Fig. 2a). Impulse response recorded by the first capsule (Fig. 2b); peaks are labeled according to the acoustic paths shown in Fig. 2a.

4.2. Experiments in a real-world acoustic environment

The last set of experiments are conducted in the Auditorium “Giovanni Arvedi”, located in the Museo del Violino (Violin Museum), Cremona, Italy.<sup>1</sup> Dimensions of the hall are $14$ m $\times$ $35$ m for a maximum height of $14$ m, the volume of the hall being $5300$ m$^3$. The sound source is placed in the center of the stage area (in the position marked by the cross) on a support of height $1.2$ m, while the microphone array is placed in correspondence of the black dot in Fig. 4.

<sup>1</sup>We thank Museo del Violino Fondazione Stradivari and its staff for their availability during the measurement session held in the Auditorium G. Arvedi.
Figure 3: Panoramic views of acoustic reflections in the acoustically controlled environment. The values of the power estimates expressed in dB are mapped to color scale.

The distance between the source and the center of the microphone array is 10.4 m.

The results of the application of the proposed analysis methodology to impulse responses acquired with the setup described above are shown in Fig. 5. The two acoustic images are computed from two different time segments; the first one, Fig.5a, centered in $t = 30.32$ ms, corresponds to the direct acoustic path from the sound source to the center of the microphone array; the other, Fig. 5b, centered in $t = 33.59$ ms, corresponds to the propagation time along the acoustic path generated from the sound source and diffracted by the railing. The acquired impulse responses do not exhibit other relevant events from the frontal direction, except the direct and diffusive paths. This is due to the design of the environment, which does not contain planar surfaces between the source and the seats, thus enabling only diffusive paths. We observe that the proposed methodology is able to discriminate acoustic events occurring within $\sim 3$ ms, and coming from close directions ($\sim \text{deg } 15$).

These results demonstrate the direct applicability of the proposed approach to a challenging real-world scenario, in which the acoustic properties of the environment are not controlled.

5. CONCLUSIONS

In this paper we have proposed a methodology for the accurate visualization of acoustic reflections in a room. We have introduced two modifications of state-of-the-art schemes, consisting in an explicit model for the scattering due to the rigid sphere that hosts the microphone array; and the adoption of a spectral analysis approach to detect the directional energy distribution impinging on the array. We have validated the proposed approach in an acoustically controlled environment and in an auditorium.
6. REFERENCES


