Enhanced spatial analysis of room acoustics using acoustic multiple-input multiple-output (MIMO) systems

Hai Morgenstern* and Boaz Rafaely

*Corresponding author’s address: Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva, 84105, Israel, Israel, hai.morgenstern@gmail.com

Standard acoustic measurements in enclosures typically employ single-input single-output (SISO) acoustic systems. The parameters obtained from these measurements describe features of energy decay and do not characterize spatial attributes of the enclosure. Directional analysis of enclosures became popular with the introduction of microphone and loudspeaker arrays. In particular, spherical arrays have been shown to be highly beneficial for spatial analysis. Spherical microphone arrays facilitate the estimation of the arrival direction of the direct and reflected sound, while the use of both loudspeaker and microphones arrays can support the estimation of both radiation and arrival directions, with the application of conventional beamforming methods. However, when several reflections are attributed to the same time bin in a discrete impulse response, reflection paths may not be uniquely determined by existing beamforming techniques. We present a new method to uniquely determine source and receiver directions for multiple reflections when time separation is unfeasible. The paper presents the formulation of the proposed method, also showing a simulation study to demonstrate the performance of the proposed method.

Published by the Acoustical Society of America through the American Institute of Physics
INTRODUCTION

Standard acoustic parameters, defined in the ISO 3382 standard, are typically measured using an omnidirectional loudspeaker and an omnidirectional microphone [1]. These parameters mainly describe features of energy decay. Spatial parameters, on the other hand, were introduced with the use of arrays, employing the use of loudspeaker, microphone, or both loudspeaker and microphone arrays. Microphone arrays enable the spatial analysis of room impulse responses (RIRs) [2]; for example, various methods for acoustic source localization are employed by the use of microphone arrays in [3–5]. Loudspeaker arrays are able to excite specific parts in a RIR, and have recently been used in room acoustic measurements [6]. Spherical arrays, both microphone and loudspeaker, are suitable for directional room acoustic investigation due to the spherical symmetry which facilitates three-dimensional analysis and convenient beampattern steering [2].

An acoustic multiple-input multiple-output (MIMO) system incorporates the use of both microphone and loudspeaker arrays. Measurements employing such systems may reveal spatial information for both the radiated and received sound fields, and thus improve the spatial analysis of an enclosure. The use of acoustic MIMO systems is relatively new; Farina proposed to use these for room acoustic analysis [7]. In his paper, he focuses mainly on the measurement techniques, and extends Gerzon’s method for recording concert hall acoustics [8]. In a previous work, we formulated the response between a spherical microphone array and a spherical loudspeaker array within an enclosure, enabling directional analysis in both frequency and time domains [9, 10]. More specifically, in [10], two new methods for designing the directivity of both loudspeaker and microphone arrays were proposed. The directivities designed by these methods reveal the directions of radiated and received reflections, at the loudspeaker and microphone arrays respectively.

In the case of a single reflection with a preselected delay, the directions of radiation and arrival are identified as a single pair, corresponding to that single reflection. Multiple reflections with the same delay may occur when the distance travelled from the loudspeaker array to the microphone array is the same for several reflections. In a time-discrete response, either simulated or measured, such multiple reflections are attributed to the same time bin, subject to the imposed temporal resolution. In this case, directions of arrival and radiation can be estimated using existing beamforming techniques, but these directions cannot be distinctly matched to individual reflections.

This contribution proposes a new method for uniquely determining reflection paths in RIRs when time separation is not possible. After estimating all directions of arrival (DOAs) at the microphone array, the array’s beampattern is set to be highly directional with the look direction set to all estimated DOAs; this is done separately for each DOA. Then, the multiple-signal classification (MUSIC) algorithm [11] is employed at the loudspeaker array to identify a single direction of radiation for the corresponding DOA. The paper concludes with a simulation study and discussion.

ACOUSTIC MIMO SYSTEM MODEL

A far-field acoustic MIMO system in an enclosure is considered. As a first step, an ideal, sparse response between source and receiver with infinite spatial resolution is constructed. This response is composed of individual reflection paths, assumed to be frequency-independent. The ideal response is given by:

\[ h(t, \theta_r, \theta_s) = \sum_i a_i \delta(t - \tau_i) \delta(\theta_r - \theta_{r,i}) \delta(\theta_s - \theta_{s,i}), \] (1)

where \( a_i \) is the amplitude of the respective path, the product of the three delta functions selects the respective path from the continuous time, \( t \), and continuous directions \( \theta_s \) and \( \theta_r \), and \( \theta \) is
vector notation for \( \theta = (\theta, \phi) \), the elevation and azimuth angles respectively.

For a spherical loudspeaker array and a spherical microphone array with orders \( N_r \) and \( N_s \), respectively, a discrete-time MIMO RIR matrix is constructed by [10]:

\[
H[j] = \sum_{i=1}^{I} y_r(\theta_{r,i}, \phi_{r,i})a_i y_s(\theta_{s,i}, \phi_{s,i})H = Y_s A Y_r^H, \tag{3}
\]

where \( I \) is the number of reflections attributed to the \( j^{th} \) time bin, \( y_r \) and \( y_s \) are steering vectors for the microphone and loudspeaker arrays, respectively, represented in spherical harmonics, \( \{(\theta_{r,i}, \phi_{r,i})\}_{i=1}^{I} \) and \( \{(\theta_{s,i}, \phi_{s,i})\}_{i=1}^{I} \) are the directions of arrival and radiation for every reflection, respectively, \( \{a_i\}_{i=1}^{I} \) are the amplitudes for each reflection, and \( H \) denotes the conjugate transpose. Eq. (3) is compact writing for eq. (2), in which \( Y_s \) and \( Y_r \) are the steering matrices for microphone and loudspeaker arrays, and \( A = \text{diag}(a_1, a_2, \ldots, a_I) \). The dimension of this MIMO matrix is \( (N_r + 1)^2 \times (N_s + 1)^2 \). For simplicity, the time index, \( j \), is omitted throughout the rest of the paper.

**MUSIC for Spherical Arrays**

MUSIC is a method developed for estimating the DOA of desired signals [11] based on an eigen decomposition of the covariance matrix of an output vector. Due to the inherent nature of the method, it is a suitable candidate for the problem addressed in this paper.

Similar to the system model presented in the previous section, we consider the output vector of a microphone array with order \( N_r \), represented in spherical harmonics. This vector, \( \mathbf{x} \), has dimensions of \((N_r + 1)^2 \times 1\). In free-field conditions and for a pressure field comprised of \( I \) signals, the far-field pressure measured with a spherical microphone can be written as [12]:

\[
\mathbf{x} = \sum_{i=1}^{I} \mathbf{y}(\theta_i, \phi_i)s_i + \mathbf{n}, \tag{4}
\]

where \( \mathbf{y}(\theta_i, \phi_i) = [Y_0^{(s)}(\theta_i, \phi_i), Y_1^{(s)}(\theta_i, \phi_i), \ldots, Y_N^{(s)}(\theta_i, \phi_i)]^T \) is the steering of the \( i^{th} \) signal written in spherical harmonics, \( \mathbf{n} \) is noise, assumed white, and \((\cdot)^T\) denotes the transpose. Eq. (4) can be written in matrix notation with:

\[
\mathbf{x} = \mathbf{Y}(\Phi)\mathbf{s} + \mathbf{n}, \tag{5}
\]

where \( \mathbf{Y}(\Phi) = [\mathbf{y}(\theta_1, \phi_1), \mathbf{y}(\theta_2, \phi_2), \ldots, \mathbf{y}(\theta_I, \phi_I)] \) is the steering matrix, in which every column corresponds to a single signal, and \( \mathbf{s} = [s_1, s_2, \ldots, s_I]^T \) is the concatenated signal vector.

In order to estimate the directions of arrival of the signals, the output covariance matrix is calculated first:

\[
\mathbf{S}_x = E[\mathbf{x}\mathbf{x}^H] = \mathbf{Y}(\Phi)\mathbf{S}_s \mathbf{Y}(\Phi)^H + \mathbf{S}_n, \tag{6}
\]

where \( E[\cdot] \) denotes the expectation, \( \mathbf{S}_s \) and \( \mathbf{S}_n \) are the signal and noise covariance matrices respectively, and \((\cdot)^H\) denotes the conjugate transpose. Now, the output covariance matrix is decomposed using the singular value decomposition (SVD), and is given by [11]:

\[
\mathbf{S}_x = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H, \tag{7}
\]

where \( \mathbf{U} \) is a unitary matrix and \( \boldsymbol{\Lambda} \) is the singular value diagonal matrix. Assuming the signal covariance matrix, \( \mathbf{S}_s \), is of full rank, the signal subspace is assumed to be spanned by \( I \) singular
vectors corresponding to the largest $I$ singular values, and the remaining singular vectors span the noise subspace, denoted $U_n$.

Estimating the directions of the sources is now performed by exploiting the orthogonality of the signals to the noise subspace; the directions are found by finding the peaks of the MUSIC spectrum:

$$P_{mu} (\theta, \phi) = \frac{1}{||U_n^H y(\theta, \phi)||^2}. \tag{8}$$

However, in room acoustics, finding more than one direction using MUSIC is problematic. Usually, a single output vector is given, and thus the covariance matrix is estimated using a single snapshot; i.e., $E\{xx^H\} = xx^H$. In this case, the signal covariance matrix is of unit rank even though there may be more than one source. This shortcoming is usually coped with by smoothing, either in time or in frequency [13].

**MUSIC FOR MIMO WITH SPHERICAL ARRAYS**

In this section, the proposed method is derived in two stages. As mentioned in the last section, estimating the directions of more than one reflection using MUSIC requires some kind of smoothing. Thus, in the first stage, we formulate modal smoothing using the information of the measured MIMO RIR matrix, and find all DOAs at the microphone array. Then, in the second stage, we use this information to relate each arrival direction at the microphone array to a single radiation direction in the loudspeaker array.

A set of acoustic single-input multiple-output (SIMO) systems are defined by taking the different columns of the MIMO RIR matrix, $H$. First, $H$ is represented as:

$$H = [h_{00}, h_{1-1}, ..., h_{N_s N_s}], \tag{9}$$

in which each SIMO system is constructed, following eq. (3), by:

$$h_{\tilde{n} \tilde{m}} = Y_r \mathbf{A} \left[ \begin{array}{c} Y_{\tilde{n} \tilde{m}}^\dagger (\theta_{s,1}, \phi_{s,1}) \\ Y_{\tilde{n} \tilde{m}}^\dagger (\theta_{s,2}, \phi_{s,2}) \\ \vdots \\ Y_{\tilde{n} \tilde{m}}^\dagger (\theta_{s,I}, \phi_{s,I}) \end{array} \right] = Y_r \mathbf{s}(\tilde{n}, \tilde{m}). \tag{10}$$

By denoting $\mathbf{s}(\tilde{n}, \tilde{m})$, as in eq. (10), and by adding white noise, we artificially synthesize $(N_s + 1)^2$ SIMO systems with the same structure as in eq. (5); i.e., $x$ is replaced with $h_{\tilde{n} \tilde{m}}$ and $s$ is replaced with $\mathbf{s}(\tilde{n}, \tilde{m})$. The different vectors synthesized have the same angular information, integrated in $Y_r$. Thus, this smoothing technique is suitable for estimating DOAs.

We now construct the covariance matrix from eq. (6) by taking the average over the synthesized SIMO systems:

$$S_h = \text{Average} \left[ hh^H \right] = \frac{1}{(N_s + 1)^2} \sum_{\tilde{n}=1}^{N_s} \sum_{\tilde{m}=-\tilde{n}}^{\tilde{n}} h_{\tilde{n}\tilde{m}}. \tag{11}$$

Since the covariance matrix is constructed using multiple SIMO systems, and not a single one, the signal and noise subspaces can be separated in the presence of multiple reflections. The noise subspace is constructed using $S_h$, as explained in sec. 3, and the DOAs at the microphone array, $\{(\theta_{r,i}, \phi_{r,i})\}_{i=1}^I$, are found by finding the peaks as in eq. (8). After finding all DOAs, a single radiation direction at the loudspeaker array can now be matched for each estimated DOA.
Initially, the beampattern of the microphone array is set with its look direction pointing to a preselected DOA. This is done by setting the beamforming coefficients of the microphone array as to maximize the array’s directivity index \([14]\); i.e., for a given DOA, \((\theta_r, \phi_r, \theta_i, \phi_i)\), the MIMO RIR from eq. (3) transforms to a multiple-input single-output (MISO) RIR:

\[
\mathbf{h}_i^T = \begin{bmatrix}
Y_0^r(\theta_{r,i}, \phi_{r,i}) & Y_1^r(\theta_{r,i}, \phi_{r,i}) & \cdots & Y_N^r(\theta_{r,i}, \phi_{r,i})
\end{bmatrix} \mathbf{Y}_s^H
= \mathbf{c}_i^T \mathbf{Y}_s^H.
\]  

(12)

Now, MUSIC is applied to the loudspeaker array. This time, a single MISO system, namely \(\mathbf{h}_i\), is sufficient for estimating the direction of radiation; by setting a highly directional beampattern at the microphone array, that coincides with the \(i^{th}\) DOA, a single reflection is excited in the RIR. This can be seen by applying the microphone’s weighting coefficients to eq. (2). Since a single reflection is assumed, a unit rank covariance matrix is sufficient for employing MUSIC. MUSIC is employed at the loudspeaker array, and the \(i^{th}\) direction of radiation is found.

This process is repeated for every estimated DOA. Thus, a single radiation direction is paired to a single DOA, and a unique set of \(I\) reflections is determined.

**Simulation Study of Room Acoustics Analysis**

A room with dimensions of (8m, 9m, 10m) was simulated using McRoomSIM \([15]\). The MIMO RIR matrix, as in eq. (2), was constructed, relating the loudspeaker and microphone arrays, with \(N_s = 4\) and \(N_r = 3\), located at \((x_s, y_s, z_s) = (1m, 1.5m, 2m)\) and at \((x_r, y_r, z_r) = (7m, 6.5m, 6m)\), respectively. For the diagram of the room and the early reflections segment of the omnidirectional response see figures 1a and 1b.

The positions of the loudspeaker and microphone arrays were set as to create two early reflections that coincide at a single discrete time bin. Figures 2a and 2b present the DOAs and directions of radiation, at the microphone and loudspeaker arrays respectively. These directions were estimated by applying MUSIC with modal smoothing, as explained in the last section.

A beampattern that maximizes the DI of the microphone array was realized, with its look direction set to one of the estimated DOAs; i.e. \((\theta_r, \phi_r) = (1.99\text{rad}, 5.66\text{rad})\). A single direction of radiation was calculated using two methods. First, the direction was calculated using the proposed method. This was then compared to a similar method in which instead of using MUSIC to detect a single direction, the maximum directivity index (DI) beamformer (BF) was employed \([16]\). Fig. 3 illustrates the performance of both methods. We can see that the proposed method, MUSIC-MIMO, identifies a single direction of radiation, whereas the maximum DI BF identifies multiple directions. This is due to the inherent nature of the MUSIC algorithm; since MUSIC is an eigen-based method in which the noise subspace is comprised of all singular vectors excluding the vector corresponding to the highest singular value, the directions of other reflections are completely removed.

**Conclusions**

We presented a method to uniquely determine reflections when time separation is unavailable. The method is based on the spatial information acquired by the use of an acoustic MIMO system. MUSIC is employed; using modal smoothing, DOAs are estimated, and then, pairs of directions of radiation and arrival are assembled. A direction of radiation at the loudspeaker array is paired to every estimated DOA by separately setting a directional beampattern at the microphone array, corresponding to each DOA. MUSIC is shown to be a suitable method for the problem addressed due to its ability to estimate a single direction of radiation for each DOA. This
is demonstrated in the previous section. The method improves directional analysis of room acoustics and facilitates room geometry estimation. The method can be also applied in the frequency domain, when multiple reflections are attributed to the same frequency bin.
REFERENCES


