ACOUSTICS 2012
HONG KONG

Joint spherical beam forming for directional analysis of reflections in rooms

Hai Morgenstern[1], *Franz Zotter*[2], *Boaz Rafaely*[3]

[1,3] *Ben-Gurion University of the Negev, Beer-Sheva*

[2] *University of Music and Performing Arts, Graz*

Abstract: This contribution presents a new approach for analysing spatial directions in room impulse responses captured with source and receiver of adjustable directivity. A distinct peak in a room impulse response is usually associated with an acoustic path length of direct or reflected sound. Given the ability to modify the directivity of source and receiver by spherical beamforming, beam coefficients can be adjusted as to emphasize the peak at a preselected time instant. We present a new approach to jointly optimize the coefficients for both source and receiver under the constraint of a unit peak amplitude while minimizing the energy of the entire response. The beam pattern described by these coefficients highlights the dominant acoustic path directions of the corresponding path length at the source and the receiver.
Keywords: Room acoustic analysis, room impulse responses, MIMO systems, spherical arrays

1. Introduction

Standard acoustic measurements typically employ omni-directional loudspeaker and microphone [1]. Recently, spherical microphone arrays have been used in room acoustic studies [2], showing improved spatial analysis. Directional loudspeakers have been used as a new tool in room acoustic measurements [3], showing advantages over an omni-directional loudspeaker. The special benefit of directional loudspeakers is their ability to excite specific parts of acoustic room response compared to omni-directional loudspeakers. Spherical loudspeaker arrays are suitable candidates for directional sources in room acoustic investigation, due to their spherical symmetry in radiating sound in three-dimensions, and flexibility in design [4].

Systems that incorporate both spherical microphone arrays and spherical loudspeaker arrays have been studied only very recently. Farina propsed the use of such a system for room acoustic analysis [5], detailing a method for recording concert hall data, such as multiple room impulse responses. Zotter et al. showed how to model spherical loudspeaker arrays as sources with adjustable directivity [6]. Moreover, a compact spherical array of microphones have been applied to analyse the acoustic path directions observed from receiver side [2,7]. Morgenstern and Rafaely have presented a more theoretical analysis of a multiple-input multiple-output (MIMO) system [8] analysing system properties and behavior. Although such systems have been presented and investigated in general, no specific method has been proposed to arrive at an improved spatial analysis of concert hall or room acoustics.

This work potentially improves analysis of reflection paths in enclosed sound fields by a new signal processing method, requiring the use of acoustic MIMO systems based on spherical arrays. Acoustic reflections are emphasized by directional loudspeakers as soon as the directivity is aligned with the reflection path. Obviously, alignment with a reflection path is a fairly discrete geometrical condition. Mignot et al. describe room responses as *sparse* in direction and time and discuss suitable computational means to analyse acoustic reflections [9]. Particularly the early part of the room response is assumed to consist of such directionally sparse reflections at the bounding surfaces of the room. Each reflection path has a length that corresponds to a time-delay in the room response.

Previously Morgenstern and Rafaely formulated the room response observed between spherical loudspeaker and

[1]haimorg@post.bgu.ac.il
[2]zotter@iem.at
[3]br@ee.bgu.ac.il

spherical microphone array as a frequency-domain MIMO matrix, of which they suggested to consider the singular values to observe new aspects [8]. By contrast, the present paper expands on a time-domain formulation and a simplified concept of a far-field approximation. This emphasizes the natural sparseness of the room responses in time and angle and yields two new algorithms to form optimal beams that carry valuable information about acoustic reflection paths.

2. Sparse model of a room response

An idealized, sparse response from one source to one receiver in a room can be regarded to be composed of individual reflection paths. Each path is characterized by its amplitude, path length, and source and receiver directions, yielding the characterization of the i^{th} path $\{\mathrm{a}_i,\ \tau_i,\ \theta_{\mathrm{R},i},\ \theta_{\mathrm{S},i}\}$. The directions θ_R, θ_S are Cartesian unit vectors and the propagation delay τ expresses the path length. The spatial impulse response from source to receiver is a sum over all propagation paths. Each path is composed of a product of three Dirac delta functions selecting the respective path from the continuous time t and continuous directions $\boldsymbol{\theta}_\mathrm{R}$, $\boldsymbol{\theta}_\mathrm{S}$. The impulse response can therefore be written as:

$$h(t,\boldsymbol{\theta}_\mathrm{R},\boldsymbol{\theta}_\mathrm{S}) = \sum_i \mathrm{a}_i\ \delta(t-\tau_i)\ \delta(\boldsymbol{\theta}_\mathrm{R}-\theta_{\mathrm{R},i})\ \delta(\boldsymbol{\theta}_\mathrm{S}-\theta_{\mathrm{S},i}). \tag{1}$$

Given source and receiver by their directivities $g_\mathrm{S}(\boldsymbol{\theta}_\mathrm{S})$ and $g_\mathrm{R}(\boldsymbol{\theta}_\mathrm{R})$, the impulse response is obtained by integration over both, $h(t) = \int_{\mathbb{S}^2}\int_{\mathbb{S}^2} g_\mathrm{R}(\boldsymbol{\theta}_\mathrm{R})\ h(\cancel{\tau}\,t, \boldsymbol{\theta}_\mathrm{R},\boldsymbol{\theta}_\mathrm{S})\ g_\mathrm{S}(\boldsymbol{\theta}_\mathrm{S})\ \mathrm{d}\boldsymbol{\theta}_\mathrm{R}\,\mathrm{d}\boldsymbol{\theta}_\mathrm{S}$. There is no attempt to represent wall reflections with impulse responses other than delta functions and acoustic near-fields at low frequencies, to maintain simplicity.

Source and receiver directivities specified in spherical harmonics are computed by $g_\mathrm{X}(\boldsymbol{\theta}_\mathrm{X}) = \sum \gamma_{nm}^{(\mathrm{X})}\, Y_n^m(\boldsymbol{\theta}_\mathrm{X})$, and the transform identity $\int_{\mathbb{S}^2} Y_n^m(\boldsymbol{\theta})\,\delta(\boldsymbol{\theta}-\theta_i)\,\mathrm{d}\boldsymbol{\theta} = Y_n^m(\theta_i)$ converts the integral in Eq. (1) into a summation

$$h(t) = \sum_{n,m,n',m'} \gamma_{nm}^{(\mathrm{R})}\ h_{nm}^{n'm'}(t)\ \gamma_{n'm'}^{(\mathrm{S})}\ , \qquad \text{with} \qquad h_{nm}^{n'm'}(t) = \sum_i \mathrm{a}_i\ \delta(t-\tau_i)\ Y_n^m(\theta_{\mathrm{R},i})\ Y_{n'}^{m'}(\theta_{\mathrm{S},i}). \tag{2}$$

A Finite resolution MIMO-matrix is now formulated. The modeled impulse responses $h_{nm}^{n'm'}$ can be individually measured using spherical arrays of loudspeakers and microphones. In practice, this is only possible within limited orders, $0 \leq n' \leq \mathrm{N_S}$, $0 \leq n \leq \mathrm{N_R}$, according to directional resolution limits of source and receiver; the other index is mathematically limited as $-n' \leq m' \leq n'$, $-n \leq m \leq n$. Eq. (2) can be first written in matrix form as sums over both pairs of indices nm and $n'm'$ and then vectorized:

$$h(t) = \left[\gamma_{nm}^{(\mathrm{R})}\right]_{nm}^{\mathrm{T}}\ \left[h_{nm}^{n'm'}(t)\right]_{nm}^{n'm'}\ \left[\gamma_{nm}^{(\mathrm{S})}\right]_{nm} =: \boldsymbol{\gamma}_\mathrm{R}^\mathrm{T}\,\boldsymbol{H}(t)\,\boldsymbol{\gamma}_\mathrm{S}, \tag{3}$$

where $\boldsymbol{\gamma}_\mathrm{X}$ contains the $(\mathrm{N_X}+1)^2$ directivity or beamforming coefficients of source (X=S) and receiver (X=R). It is also wise to point out an alternative vectorization of the problem that is used for joint beamforming later on

$$h(t) = \left[h_{nm}^{n'm'}(t)\right]_{nmn'm'}^{\mathrm{T}}\ \left[\gamma_{nm}^{(\mathrm{R})}\,\gamma_{n'm'}^{(\mathrm{S})}\right]_{nmn'm'} =: \boldsymbol{h}(t)^\mathrm{T}\,\boldsymbol{c}. \tag{4}$$

Notice the relations $\boldsymbol{h}(t) = \mathrm{vec}\{\boldsymbol{H}(t)\}$ and $\boldsymbol{c} = \mathrm{vec}\{\boldsymbol{\gamma}_\mathrm{R}\,\boldsymbol{\gamma}_\mathrm{S}^\mathrm{T}\}$.

Moreover, the time resolution is limited in practice. For simplicity we assume time segments $T_j = j\,\varepsilon + \left[-\frac{\varepsilon}{2};\ \frac{\varepsilon}{2}\right)$ along which information is gathered by integration $h(j) = \int_{T_j} h(t)\,\mathrm{d}t$ for the discrete-time index $j \in \mathbb{Z}$; this also defines a discrete-time version of $h_{nm}^{n'm'}(t)$. For long segment lengths ε several reflections might superimpose, cf. Eqs. (2) (3),

$$\boldsymbol{H}(j) = \sum_{\tau_i \in T_j} \mathrm{a}_i \left[Y_n^m(\theta_{\mathrm{R},i})\right]_{nm} \left[Y_n^m(\theta_{\mathrm{S},i})\right]_{nm}^\mathrm{T} =: \sum_{\tau_i \in T_j} \mathrm{a}_i\,\boldsymbol{y}(\theta_{\mathrm{R},i})\,\boldsymbol{y}(\theta_{\mathrm{S},i})^\mathrm{T}. \tag{5}$$

Note the similarity of this sum to a singular value decomposition. We conclude that the number of superimposed reflections determines the rank of the matrix, $\mathrm{rank}\,\{\boldsymbol{H}(j)\} \leq |\{\tau_i \in T_j\}|$, apart from the common size limits of the rank $(\min\{\mathrm{N_S}, \mathrm{N_R}\}+1)^2$. Nevertheless, for different i the vectors $\boldsymbol{y}(\theta_{\mathrm{R},i})$ and $\boldsymbol{y}(\theta_{\mathrm{S},i})$ need not be orthonormal.

3. Joint beamforming by singular value decomposition at the instant j_0

In this section we focus on one time instant j_0 which contains at least one reflection path. The goal here is to find beamforming weights that enforce a unity amplitude at this time instant $h(j_0) = 1$. In order to find suitable weights, we use the representation given in Eq. (4) to formulate the least-squares problem

$$\min\ \boldsymbol{c}^{\mathrm{T}}\boldsymbol{c} \tag{6}$$

$$\text{s. t. } \boldsymbol{h}(j_0)^{\mathrm{T}}\,\boldsymbol{c} = 1. \tag{7}$$

It yields the solution $\boldsymbol{c} = \boldsymbol{h}(j_0)/\|\boldsymbol{h}(j_0)\|^2$ and minimizes the directivities in directions that do not support the transmission from source to receiver. However, the solution does not directly reveal the joint beamforming coefficients: it is related to their products, and we may define it in matrix form

$$\boldsymbol{c} = \mathrm{vec}\{\boldsymbol{\gamma}_{\mathrm{R}}\,\boldsymbol{\gamma}_{\mathrm{S}}^{\mathrm{T}}\} =: \mathrm{vec}\{\boldsymbol{C}\}. \tag{8}$$

Accordingly, the matrix form of the solution is $\boldsymbol{C} = \boldsymbol{H}(j_0)/\|\boldsymbol{H}(j_0)\|_{\mathrm{F}}^2$. More than one solution might exist, depending on the entries of $\boldsymbol{C}$; in this case specified by the rank of $\boldsymbol{H}(j_0)$, cf. Eq. (5). In general, suitable pairs of solutions for $\boldsymbol{\gamma}_{\mathrm{R}}$ and $\boldsymbol{\gamma}_{\mathrm{S}}$ can be defined by left and right singular vectors of

$$\boldsymbol{C} = \boldsymbol{U}\,\boldsymbol{S}\,\boldsymbol{V}^{\mathrm{T}} = \sum_i s_i\,\boldsymbol{u}_i\,\boldsymbol{v}_i^{\mathrm{T}} \tag{9}$$

to be $\hat{\boldsymbol{\gamma}}_{\mathrm{R},i} = \boldsymbol{u}_i$, and $\hat{\boldsymbol{\gamma}}_{\mathrm{S},i} = \boldsymbol{v}_i$, and their significance is ranked by the singular value s_i. To meet the criterion the solution needs to be re-normalized $h(j_0) = 1$

$$\begin{aligned} \boldsymbol{\gamma}_{\mathrm{R}} &= \boldsymbol{u}_i/\sqrt{a} \\ \boldsymbol{\gamma}_{\mathrm{S}} &= \boldsymbol{v}_i/\sqrt{a}, \qquad\qquad \text{with} \qquad a := \boldsymbol{u}_i^{\mathrm{T}}\,\boldsymbol{H}(j_0)\,\boldsymbol{v}_i. \end{aligned} \tag{10}$$

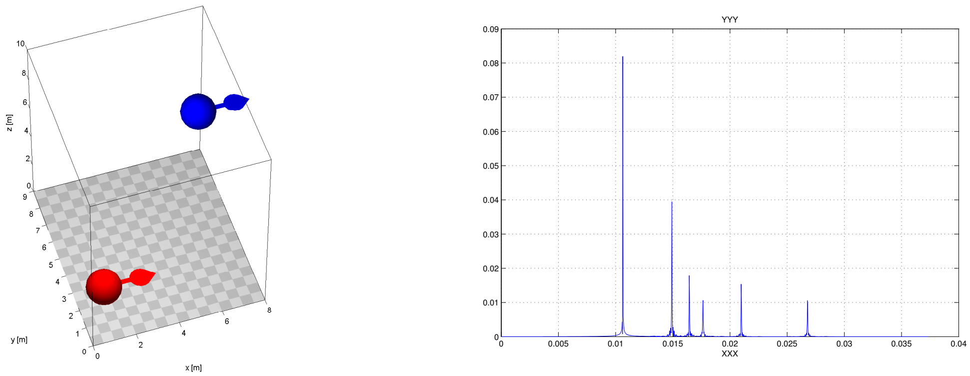
Case study using simulated room response. A room with dimensions of of (8m, 9m, 10m) was simulated using McRoomSIM[10]. The directivity matrix, as in eq. (5), relating order-limited source and receiver, with $N_s = 4$ and $N_r = 3$, located at $(x_s, y_s, z_s) =$ (1m, 1.5m, 2m) and at $(x_r, y_r, z_r) =$ (7m, 6.5m, 6m), respectively, was constructed.

For the diagram of the room and the early reflections segement of the omni-directional response see figs. 1(a) and 1(b). Fig. 2(a) demonstrates the directivity by plotting $\boldsymbol{u}_1$, $\boldsymbol{v}_1$ for the direct path, singular vectors of $\boldsymbol{H}(j_0)$. Figs. 2(b) and 2(c) demonstrate two directivity patterns corresponding to $\boldsymbol{u}_1$, $\boldsymbol{v}_1$ and to $\boldsymbol{u}_2$, $\boldsymbol{v}_2$: two sets of singular vectors of $\boldsymbol{H}(j_1)$, where j_1 corresponds to a time instant in which two reflections are superimposed.

4. Joint beamforming by constrained minimization for all instants j

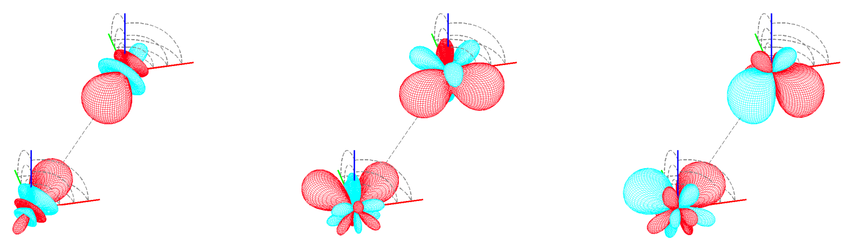
It is possible to achieve $h(j_0) = 1$ for one time instant of an impulse response using optimal beamforming coefficients for source and receiver. Moreover, it is possible to minimize the energy of the entire impulse response $\min\ \sum_j h^2(j)$, for all time instances, subject to this constraint by requiring

$$\begin{gathered} \min\ \boldsymbol{c}^{\mathrm{T}}\,\boldsymbol{G}\,\boldsymbol{c}, \\ \text{subject to } \boldsymbol{h}(j_0)^{\mathrm{T}}\,\boldsymbol{c} = 1 \end{gathered} \qquad\qquad \text{with} \quad \boldsymbol{G} := \sum_j \boldsymbol{h}(j)\ \boldsymbol{h}(j)^{\mathrm{T}}. \tag{11}$$



(a) (b)

Figure 1: a) Diagram of the simulated room; red and blue sphere represent source and receiver, respectively, b) Impulse response between omni-directional directivity patterns at source and receiver.



(a) (b) (c)

Figure 2: Resulting beam patterns of joint beamforming for one time instant: (a) for the instant of the direct sound, (b) and (c) and two coinciding early reflections, i.e. having the same path length.

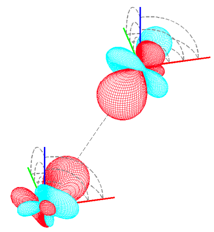
The solution of such a problem is known from minimum variance distortionless resposnes (MVDR) beamforming[11] and yields

$$\boldsymbol{c} = \frac{\boldsymbol{G}^{-1}\,\boldsymbol{h}(j_0)}{\boldsymbol{h}(j_0)^{\mathrm{T}}\,\boldsymbol{G}^{-1}\,\boldsymbol{h}(j_0)}. \tag{12}$$

In our case the matrix to be inverted is frequently numerically singular and therefore unstable, but it can always be regularized. This is done using the eigendecomposition $\boldsymbol{G} = \boldsymbol{Q}\,\mathrm{diag}\{[\lambda_i]_i\}\,\boldsymbol{Q}^{\mathrm{H}}$ to define the regularized inverse as $\boldsymbol{G}^{-1} \approx \boldsymbol{Q}^{\mathrm{H}}\,\mathrm{diag}\{[\frac{1}{\lambda_i+\alpha}]_i\}\,\boldsymbol{Q}$ with the regularization parameter α. Once more, the above solution does only deliver products of the joint beamforming weights. Therefore the same procedure has to be applied as specified in Eqs. (9) and (10) of the previous section to find the weights.

Case study using simulated room response The same room as in sec. 3 was simulated. Fig. 3(a) demonstrates the optimized beamformer derived as in sec. 4, constraining the amplitude of the direct sound while minimizing the response. The figure shows the graphical interpretation of the optimization process; i.e., the main lobes of both source and receiver point in the directions of the constrained time instance while nulls are placed in the directions interfering reflections.

Influence of the regularization parameter α and the relation between the solutions In the limit of a large parameter α, the above matrix inverse is approximated as scaled identity matrix $\lim_{\alpha\to\infty}\boldsymbol{Q}^{\mathrm{H}}\,\mathrm{diag}\{[\frac{1}{\lambda_i+\alpha}]_i\}\,\boldsymbol{Q} = \frac{1}{\alpha}\,\boldsymbol{I}$. In this case, the above solution yields the minimization of the entire beam patterns instead by the optimal solution $\boldsymbol{c} = \boldsymbol{h}(j_0)/\|\boldsymbol{h}(j_0)\|^2$, which is the same as described in the previous section.



(a)

Figure 3: Result of joint beamforming minimizing the squared impulse response: (a) unity constraint for the time instance of the direct sound.

5. Conclusions and further work

We presented two methods to obtain jointly optimal directivities for a source and a receiver array in a room, regarding the room impulse response. This facilitates room geometry estimation and spatial analysis of room reflections. The first joint beamformer yields a pair of minimal directivities under the unit response constraint at one time instant. Its beam patterns support directions of reflections with suitable delay. The second proposed beamformer minimizes the room response instead and hereby suppresses reflection paths with other delays. This method requires regularization. We showed that the amount regularization links the two beamformers: in the limit of a large regularization parameter the two beamformers (sec. 3 and 4) become equal.

Further work is expected to include methods to reduce the effect of matrix regularization, and for developing improved formulations for computing the coefficients that facilitate best spatial separation of room reflections.

References and links

1. I. Standard, "3382. acoustics–measurement of the reverberation time of rooms with reference to other acoustical parameters", International Standards Organization (1997).
2. M. Park and B. Rafaely, "Sound-field analysis by plane-wave decomposition using spherical microphone array", J. Acoust. Soc. Am. **118**, 3094–3103 (2005).
3. S. Tervo, J. Patynen, and T. Lokki, "Acoustic reflection path tracing using a highly directional loudspeaker", in *Applications of Signal Processing to Audio and Acoustics, 2009. WASPAA'09. IEEE Workshop on*, 245–248 (IEEE) (2009).
4. F. Zotter, "Analysis and synthesis of sound-radiation with spherical arrays", Ph.D. thesis, University of Music and Performing Arts, Graz, chapter VI, 111–152 (2009).
5. A. Farina, "Room impulse responses as temporal and spatial filters", in *The 9th Western Pacific Acoustics Conference, Seoul, Korea* (2006).
6. F. Zotter, A. Sontacchi, and R. Holdrich, "Modeling a spherical loudspeaker system as multipole source", Fortschritte der Akustik **33**, 221 (2007).
7. A. O'Donovan, R. Duraiswami, and D. Zotkin, "Imaging concert hall acoustics using visual and audio cameras", in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, 5284–5287 (IEEE) (2008).
8. H. Morgenstern and B. Rafaely, "Analysis of acoustic mimo systems in enclosed sound fields", in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on* (IEEE) (2012).
9. R. Mignot, L. Daudet, and F. Ollivier, "Compressed sensing for acoustic response reconstruction: Interpolation of the early part", in *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop on*, 225–228 (IEEE) (2011).
10. A. Wabnitz, N. Epain, C. Jin, and A. van Schaik, "Room acoustics simulation for multichannel microphone arrays", in *Proceedings of the International Symposium on Room Acoustics* (2010).
11. H. Van Trees, *Optimum array processing* (Wiley Online Library) (2002).