On the Estimation of Acoustic Reflection Coefficients from In-Situ Measurements using a Spherical Microphone Array

Sascha Spors and Till Rettberg

Institute of Communications Engineering, University of Rostock, Germany, Email: Sascha.Spors@uni-rostock.de

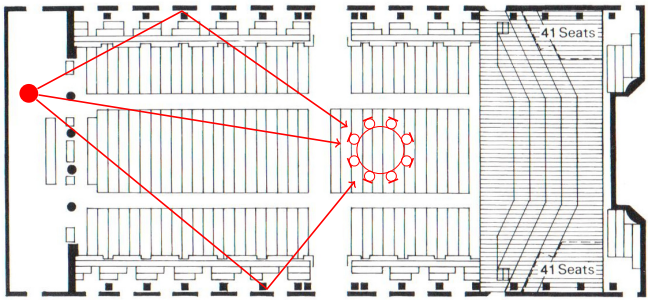


Figure 1: Estimation of reflection coefficients by analysis of the acoustic properties of the early reflections (floor plan [4]).

Introduction

The acoustic reflection coefficients of surfaces are important parameters for geometric modeling of acoustic enclosures. They determine the frequency and incidence-angle dependent reflection characteristics of the walls and finally the overall perceptual impression of a room. The reflection coefficient of a particular material is typically determined by measuring a probe of the material in a controlled laboratory setup. This is often not possible for existing rooms, since either material samples may not be available or their exact specification. A method which allows for the in-situ estimation of acoustic reflection coefficients is therefore of practical interest. Various methods have been proposed for this purpose, e.g. [1, 2, 3]. Typically one or more sources and microphones distributed at distinct locations in the room are used. We focus in this contribution on methods that utilize (rigid) spherical microphone arrays as these are commonly used for the capture of spatial sound in virtual acoustics.

Figure 1 depicts the underlying scenario. The spatio-temporal room impulse responses (SRIRs) from a source to a spherical microphone array are captured in the environment of interest. Under the assumption of a specular reflection model, the early reflections that have only been reflected once contain the desired information on the acoustic characteristics of a particular wall. The basic concept of the proposed approach is to apply beamforming techniques in order to extract the spectral properties of the direct sound and the mirror image sources. As prerequisite, this requires the localization of direct sound and early reflections. Reflection coefficients are in general frequency and incidence-angle dependent. A consequence of the static setup is that the reflection coefficients can only be estimated for the given incidence angles. Aspects which only have been considered so far in the state-of-the art are the (i) automated localization of image sources, (ii) estimation of frequency dependent reflection coefficients and (iii) the impact of technical limitations (e.g. number of microphones, sensor noise, ...).

For ease of illustration, the given examples are limited to (image) sources located in the horizontal plane were the original sound source and the microphone array are contained in. The derived approach however can be applied straightforwardly to the general case.

Approach

A wideband approach using the pre-captured SRIRs from a source to a spherical microphone array is developed. The underlying signal model is given by the mirror image source model [5]

$$P(\mathbf{x},\omega) = \sum_{q=0}^{Q-1} A_q(\alpha_q,\beta_q,\omega)\,\frac{1}{r_q} e^{-j\frac{\omega}{c}r_q}, \qquad (1)$$

where $r_q = ||\mathbf{x}-\mathbf{x}_q||$ denotes the distance between the q-th source at $\mathbf{x}_q$ and the position $\mathbf{x}$ of the microphone, $A_q(\alpha,\beta,\omega)$ the reflection coefficient, and α_q, β_q the elevation and azimuth angle of the incidence sound at the reflective surface. The index zero denotes the direct sound.

The estimation of acoustic reflection coefficients is performed in two stages:

1. localization of direct sound and early reflections,
2. estimation of source levels and reflection coefficients.

The two stages and the subsequent signal processing operations are depicted in Figure 2. The captured SRIRs are first pre-processed. This may involve operations like high-pass filtering to cope for the low-frequency limits of array processing and temporal truncation removing the higher-order mirror image sources. The pre-processed SRIRs are then decomposed into their incidence directions using a plane wave decomposition (PWD) [6] in the first stage. Inspection of the PWD in the time/incidence-angle domain reveals that sources generate maxima at a particular time-instant and incidence-angle (cf. to Figure 4a). The time-instants of the maxima relate to the distances of the sources by the speed of sound. These properties of the PWD are exploited to localize the direct sound and early reflections. The acoustic properties of the direct sound and early reflections are then extracted by beamforming into the estimated incidence-angles of the localized sources and temporal windowing.

This research was supported by grant SP 1295/6-1 of the Deutsche Forschungsgemeinschaft (DFG).

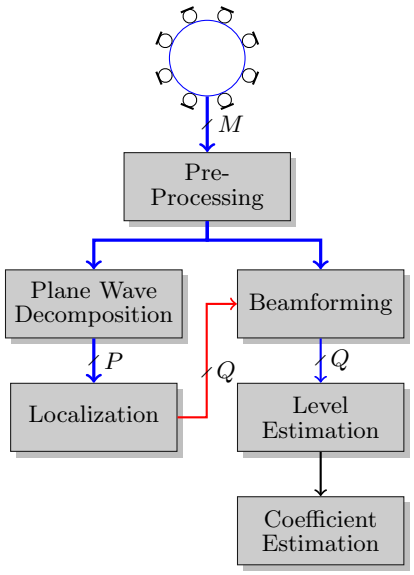


Figure 2: Signal processing blocks for the estimation of reflection coefficients (M: # of microphones, P: # of plane waves, Q: # of sources).

Localization of Sources

First a PWD of the SRIRs is computed as basis for the localization of direct sound and early reflections. The PWD is realized by beamforming into all directions. However, different beamforming methods can be used for this purpose. A matched-filter beamformer [6, 7] is optimized with respect to its signal-to-noise (SNR) ratio at the output. Since noise is a limiting factor for the localization of sources, this technique is chosen.

Matched-Filter Beamformer

A matched-filter is a linear filter which is used to detect signal templates of finite duration in a noisy observed signal. It is assumed that the time-alignment of the templates is not known and that the templates are superimposed by additive white Gaussian noise. The matched-filter results from optimizing for a maximum SNR at its output. It is given as the time-reversed conjugate of the signal template. This principle is applied to beamforming for spherical microphone arrays by transforming the incident sound field on the sphere into its spherical harmonics expansion coefficients (spherical Fourier transform) [6]

$$\mathring{P}_n^m(\omega) = \int_0^{2\pi} \int_0^{\pi} P(\theta,\phi,\omega)\,[Y_n^m(\theta,\phi)]^* \sin(\theta) d\theta d\phi \quad (2)$$

where θ, ϕ denote the elevation and azimuth angle on the sphere of radius R and $Y_n^m(\theta,\phi)$ the spherical harmonics defined as

$$Y_n^m(\theta,\phi) = \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}} P_n^m(\cos\theta) e^{im\phi}, \quad (3)$$

where $P_n^m(\cos\theta)$ denote the associated Legendre functions or degree $m \in -n,\ldots,n$ and order $n \in \mathbb{N}$. As we are aiming at a far-field beamformer, the signal template equals the sound field of a unit-amplitude plane wave impinging onto the north-pole ($\theta = 0, \phi = 0$) of the sphere.

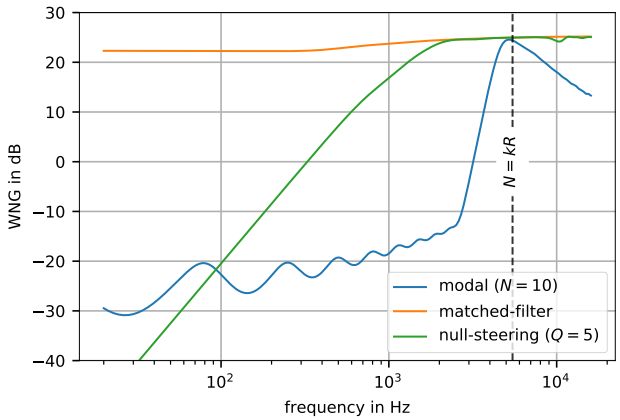


Figure 3: White-noise gain (WNG) of a modal, matched filter and null-steering beamformer (rigid array, $R = 0.1$ m, $N = 10$, Lebedev grid, modal order of simulated soundfield $N_{\text{sf}} = 32$, $\theta_{\text{pw}} = 90^{\circ}$, $\phi_{\text{pw}} = 0^{\circ}$, Tikhonov regularization β=0.01).

Its zonal spherical harmonics expansion coefficients are given as

$$\mathring{b}_n(r,\omega) = 4\pi i^n \left[j_n(\frac{\omega}{c}r) - \frac{j_n'(\frac{\omega}{c}R)}{h_n^{(2)\prime}(\frac{\omega}{c}R)} h_n^{(2)}(\frac{\omega}{c}r) \right], \quad (4)$$

where $j_n(\cdot)$ and $h_n^{(2)}(\cdot)$ denote the n-th order spherical Bessel and Hankel function of second kind, respectively. Primed quantities denote derivatives with respect to the argument. A PWD $\bar{P}(\theta,\phi,\omega)$ is then computed by performing an inverse spherical Fourier transform of $\mathring{P}_n^m(\omega) \cdot \mathring{b}_n(R,\omega)$

$$\bar{P}(\theta,\phi,\omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \mathring{P}_n^m(\omega) \cdot \mathring{b}_n(R,\omega) Y_n^m(\theta,\phi). \quad (5)$$

Inverse temporal Fourier transform yields the time-domain PWD $\bar{p}(\theta,\phi,t)$ which is used in the remainder for localization of sources. As example, Figure 4a shows the PWD of multiple point sources. The white-noise-gain (WNG) quantifies the attenuation of spatio-temporal uncorrelated Gaussian noise by a beamformer. Figure 3 shows the WNG of a matched-filter beamformer in comparison to a modal beamformer optimized for maximum directivity. It is evident that the matched-filter beamformer attenuates the sensor noise over the entire frequency range.

Localization

Feature detection techniques from computer vision are applied for the identification and localization of sources in the PWD $\bar{p}(\theta,\phi,t)$. The Laplacian of Gaussian (LoG) [8] approach is chosen among the commonly applied blob detection techniques. The magnitude $|\bar{p}(\theta,\phi,t)|$ of the PWD serves as 'input image'. In the LoG approach the input is first filtered by a Gaussian kernel of different scales, then the scale-normalized Lapacian operator is applied to the filtered input and the simultaneous maxima/minima of the operator output at the different scales are detected.

The approach has been modified to match the specific properties of the PWD. Point sources exhibit an amplitude decay over distance. In order to obtain comparable levels for the image sources, the amplitude decay is compensated. The periodicity of the PWD with respect to the elevation and azimuth angle is considered by using a periodic convolution for the filtering with the Gaussian kernel. The kernel itself is modified such to account for the asymmetric spatio-temporal structure of the PWD of a point source. The direct sound is identified as the source with the closest distance. The example depicted in Figure 4 shows the localization of five point sources by the outlined method. White Gaussian noise with a peak-signal-to-noise ratio (PSNR) of 60 dB is added to the microphone signals in order to simulate equipment noise. The PWD, as well as the localized source positions are shown in Figure 4a, the pre-processed PWD filtered by the Gaussian kernel at different scales in Figure 4b. Here the source locations are clearly visible as dark spots in all scales.

Estimation of Reflection Coefficients

From the mirror image source model (1) two conclusions can be drawn for the estimation of the reflection coefficients from the PWD: (i) the level of the direct sound has to be considered and (ii) the amplitude decay of the sources has to be compensated.

Null-Steering Beamformer

The far-field null-steering beamformer is a data-dependent beamformer which cancels out the mutual interference between plane waves incident from different directions. Once the source locations have been estimated, a linear system of equations is constructed

$$\mathbf{P}(\omega) = \mathbf{A}(\omega) \cdot \bar{\mathbf{P}}(\omega), \tag{6}$$

where the $M \times 1$ and $Q \times 1$ vectors $\mathbf{P}(\omega)$ and $\bar{\mathbf{P}}(\omega)$ are composed from the microphone signals $P(\theta, \phi, \omega)$ and the plane wave decomposition $\bar{P}(\hat{\theta}_q, \hat{\phi}_q, \omega)$ for the estimated source locations $\hat{\theta}_q, \hat{\phi}_q, \hat{r}_q$. The $M \times Q$ matrix $\mathbf{A}(\omega)$ contains the response of a unit-amplitude plane wave on the microphone array for the estimated source locations. Since typically $Q < M$, (6) is overdeterminded and solved for the least-squares solution. This approach is also known as linearly constrained minimum-variance beamformer [6]. The resulting WNG depends on the source arrangement and suffers if the desired source is close to a null. Towards low frequencies it generally deteriorates and regularization may be necessary. The WNG of a null-steering beamformer (without regularization) is shown in Figure 3.

Level Estimation

In order to extract the acoustic properties of one particular (image) source, the time-domain output $\bar{p}(\hat{\theta}_q, \hat{\phi}_q, t)$ of the null-steering beamformer is windowed symmetrically around the time-instance which coincides with the estimated distance $\hat{r}_q$ of that source. Figure 5 shows the output signal of the beamformer and the corresponding window. The level of the source is estimated by computing the square root of the power of the windowed signal.

Coefficient Estimation

In order to estimate the reflection coefficients of the surfaces creating the mirror image sources, the windowed time-domain output of the null-steering beamformer is fed into a fractional octave-band filter bank. In building acoustics a bandwith of one octave is common, starting from 125 Hz [9]. The level is computed for each frequency band. These levels are then normalized with respect to the estimated level of the direct sound and the level decay.

Simulation Results

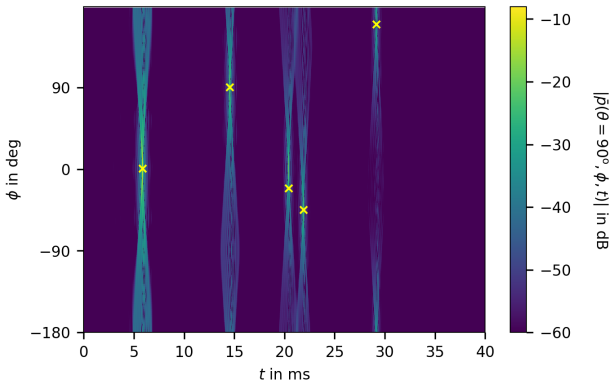
The presented approach is evaluated by numerical simulation of the sound field in a medium sized room. The mirror image source model [5] with frequency-dependent reflection coefficients is used for this purpose. The reflection coefficients of floor and ceiling were set to fully absorbing, the reflection coefficients of the side walls according to the reflection properties of a 1/4" wood panel [9]. The size of the room is $10 \times 11 \times 3$ m, the source is located at $(7, 8, 1.8)$ m and the microphone array at $(1.1, 4.5, 1.8)$ m. The pressure on a rigid sphere with radius $R = 0.1$ m is simulated for $M = 170$ microphones distributed on its surface according to the Lebedev grid. Spatio-temporal white Gaussian noise with a PSNR of 60 dB is added to the microphone signals. A rectangular window of 6.3 ms length is used to window the output of the null-steering beamformer. The room simulation, as well as the beamforming and coefficient estimation algorithms were implemented on basis of the Sound Field Analysis toolbox for `numpy`[1]. The localization using the modified LoG approach is implemented on basis of the `scikit-image`[2] toolbox.

Figure 6 shows the matched-filter PWD computed from the simulated microphone signals and the localized sources. The direct sound and the mirror image sources are localized accurately, despite the additive noise. Some near-field effects can be observed at low frequencies for the direct sound. Figure 7 shows the true and estimated levels for the direct sound, as well as the true and estimated reflection coefficients of the side walls. Deviations are present for the direct sound at low frequencies which can be accounted to near-field effects. For the mirror image sources located at higher distances these deviations are significantly lower.

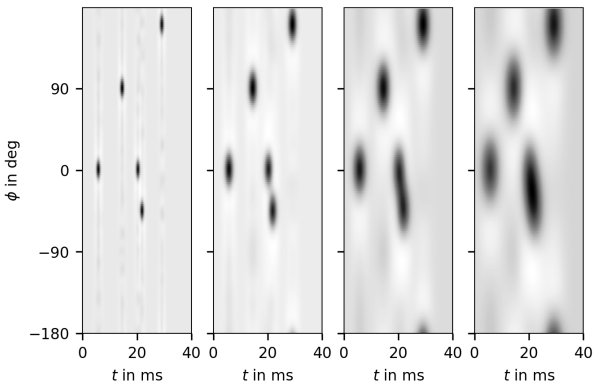
Conclusions

The paper presents an approach to the localization of mirror image sources and reflection coefficient estimation on the basis of impulse responses captured in-situ by a spherical microphone array. The aim of the approach is to gain additional insights on the acoustic environment in the scope of virtual acoustics were microphone arrays of

[1] https://github.com/spatialaudio/sfa-numpy
[2] http://scikit-image.org/



(a) Plane wave decomposition and localized sources (yellow crosses).



(b) Pre-processed PWD filtered by multiscale Gaussian kernel.

Figure 4: Localization of multiple point sources (rigid array, $R = 0.1$ m, $N = 10$, Lebedev grid, PSNR = 60 dB, $N_{\mathrm{sf}} = 32$, $f_{\mathrm{s}} = 32$ kHz, $Q = 5$).

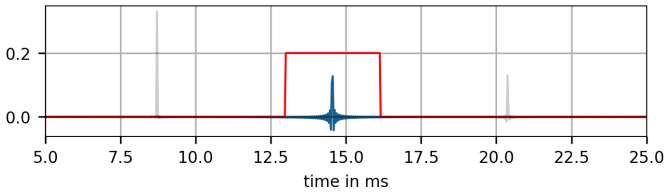


Figure 5: Windowed output of null-steering beamformer (rigid array, $R = 0.1$ m, $N = 10$, Lebedev grid, $N_{\mathrm{sf}} = 32$, $Q = 5$, $f_{\mathrm{s}} = 32$ kHz, 1/4" wood panel).

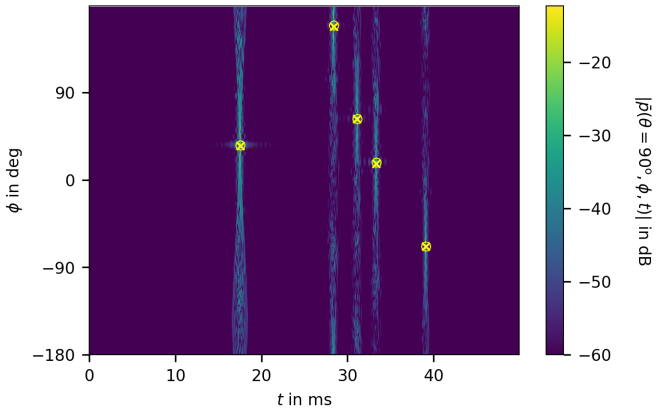


Figure 6: Plane wave decomposition, localized direct sound and mirror image sources (yellow circles: true positions, yellow crosses: estimated positions).

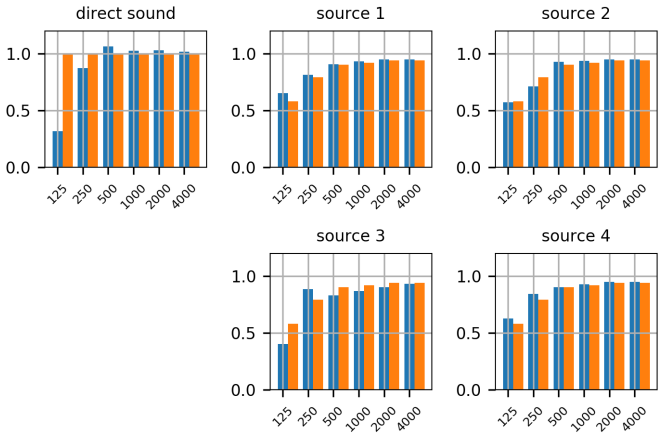


Figure 7: Estimated reflection coefficients (true coefficients: orange bars, estimated coefficients: blue bars).

moderate size are frequently applied. In additional simulations the approach turned out to be quite robust. However some parameters of the localization algorithm and the regularization of the null-steering beamformer may have to be optimized in a particular application. Preliminary results with measured datasets where promising. A detailed evaluation was not possible with the available datasets due to a lack of ground truth data.

Future work may include the application of a near-field null-steering beamformer and the combination of the (image) source localization with room geometry estimation, for instance using [10], in order to link the estimated reflection coefficients to the room geometry.

References

[1] J. Rathsam and B. Rafaely, "Analysis of absorption in situ with a spherical microphone array," *Applied Acoustics*, vol. 89, pp. 273–280, 2015.

[2] M. Tamura, "Spatial fourier transform method of measuring reflection coefficients at oblique incidence. I: Theory and numerical examples," *Journal of the Acoustical Society of America*, vol. 88, no. 5, pp. 2259–2264, November 1990.

[3] D. Markovic, K. Kowalczyk, F. Antonacci, C. Hofmann, A. Sarti, and W. Kellermann, "Estimation of acoustic reflection coefficients through pseudospektrum matching," *Transactions on Audio, Speech and Language Processing*, vol. 22, no. 1, pp. 125–137, January 2014.

[4] L. Beranek, *Concert Halls and Opera Houses: Music, Acoustics, and Architecture*. Springer, 2003.

[5] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[6] B. Rafaely, *Fundamentals of Spherical Array Processing*. Springer, 2015.

[7] S. Spors, T. Rettberg, and F. Winter, "A comparison of modal and spatially matched-filter beamforming for rigid spherical microphone arrays in the context of data-based binaural synthesis," *Journal of the Acoustical Society of America*, vol. 141, no. 5, p. 3855, June 2017.

[8] T. Lindeberg, *Scale-Space Theory in Computer Vision*. Springer, 1994.

[9] M. Long, *Architectural Acoustics*. Elsevier, 2014.

[10] P. Annibale, J. Filos, P. A. Naylor, and R. Rabenstein, "Geometric inference of the room geometry under temperature variations," in *International Symposium on Communications, Control and Signal Processing*, Rome, Italy, May 2012.