Practical Aspects of the Calibration of Spherical Microphone Arrays

Till Rettberg*, Karim Helwani*, Sascha Spors*, and Herbert Buchner◇,

*Quality and Usability Lab, Telekom Innovation Laboratories, ◇Machine Learning Group

Technische Universität Berlin, 10587 Berlin, Germany

till.rettberg@campus.tu-berlin.de, karim.helwani@telekom.de

Introduction

Spherical microphone arrays have been widely studied for acoustic scene analysis and beamforming. Spatial sound field analysis typically requires two distinct steps: A calibration stage equalizes spatio-temporal deviations of the microphones. Then a decoupling stage extracts orthogonal spatial components of the wave field. For spherical microphone arrays, the decomposition of the measured sound pressure into spherical harmonics has been shown to be a powerful framework, which allows plane wave decomposition analytically in the spherical wave spectrum [1]. The term *modal* (or sometimes *phase mode*) *beamforming* has been coined in literature. However, the approach usually requires an ideal sensor geometry, exact microphone placement and pre-equalization.

In practice, these preconditions may not always be met. For such cases, a different approach has been proposed in [2]: A microphone array in a sound field of finite countable plane waves from fixed incidence angles is modeled as a multiple input/multiple output (MIMO) system. Capturing the array responses to synthesized plane waves from the desired directions under free field conditions, directional calibration filters can be obtained by MIMO system inversion. Not relying on specific array geometry or pre-equalization, this method presents an integrated calibration that performs compensation and decoupling of the spatial components in one step.

This paper evaluates the practical utilization of the latter approach for a 64-channel microphone array. A frequency-selective inversion is proposed and the resulting calibration filters are compared with simulated modal beamfilters with respect to achievable directivity and robustness.

Integrated Calibration via MIMO System Inversion

The system theory-based approach presented in [2] considers a linear $Q \times P$ MIMO system of P microphones and Q synthesized plane waves from fixed incidence angles. It is described by a block-wise Toeplitz matrix $\mathbf{H}$. Then, an ideal calibration filter $\mathbf{B}$ is desired, that retrieves the system input from the convolutive mixture captured by the microphones, apart from modeling delays, denoted by the block-diagonal matrix $\mathbf{C}$ and $\mathbf{D}$ denoting the synthesis operator for the virtual loudspeaker array [2] :

$$\mathbf{D} \cdot \mathbf{H} \cdot \mathbf{B} = \underbrace{\text{Bdiag}\left\{[0,\ldots,0,1,0,\ldots,0]^T,\ldots,[0,\ldots,0,1,0,\ldots,0]^T\right\}}_{\mathbf{C}} \quad (1)$$

$$\Rightarrow \mathbf{B} = \mathbf{H}^{-1} \cdot \mathbf{D}^{-1} \cdot \mathbf{C}. \quad (2)$$

According to the MIMO inverse theorem [5], this system can be exactly inverted, if the transfer functions of the individual signal paths do not share common zeros in the z-domain. Unfortunately, this is not the case for general microphone/speaker placement in free field. It can be assumed however, that a sufficiently good approximation can be found for a restricted frequency range. Therefore, a frequency-selective narrowband inversion is proposed: With a MIMO-DFT matrix $\mathbf{F}$ of sufficiently large DFT size $L \to \infty$, $\mathbf{H}$ is block-wise diagonalized. A suitable permutation matrix $\mathbf{P}$ resorts the transformed system matrix into blockdiagonal form, of $P \times Q$ submatrices $\underline{\tilde{\mathbf{H}}}_l$ with $l = 1, \ldots, L$, i.e.,

$$\begin{aligned}\underline{\tilde{\mathbf{H}}} &:= \mathbf{P}^T \cdot \mathbf{F}^H \cdot \mathbf{H} \cdot \mathbf{F} \cdot \mathbf{P} \quad (3)\\ &= \text{Bdiag}\left[\underline{\tilde{\mathbf{H}}}_1,\ldots,\underline{\tilde{\mathbf{H}}}_L\right].\\ \Rightarrow \underline{\tilde{\mathbf{H}}}^{-1} &= \text{Bdiag}\left[\underline{\tilde{\mathbf{H}}}_1^{-1},\ldots,\underline{\tilde{\mathbf{H}}}_L^{-1}\right] \quad (4)\end{aligned}$$

Exact bin-wise inverses $\underline{\tilde{\mathbf{H}}}_l^{-1}$ however do not exist, since $P > Q$ must hold [5], so we aim for the least-squares solution

$$\underline{\tilde{\mathbf{H}}}_l^{\dagger} := \left[\underline{\tilde{\mathbf{H}}}_l^{\mathrm{H}}\underline{\tilde{\mathbf{H}}}_l\right]^{-1}\underline{\tilde{\mathbf{H}}}_l^{\mathrm{H}}. \quad (5)$$

Array and Calibration Measurement

The array uses an equal area sampling scheme [4], which has been shown to offer good sampling efficiency and being well-conditioned for low orders $N \leq 8$ [6]. The layout consists of 65 nodes, the bottom position accommodates the cable outlet. A rigid steel sphere of 7.5cm radius houses 64 omnidirectional back-electret condenser capsules (*AKG CE20/18PS*). The sphere body slightly deviates from an ideal sphere geometry, details can be found in [7].

A calibration measurement was conducted in the large anechoic chamber of the Institute of Fluid Mechanics and Engineering Acoustics at TU Berlin. The results presented here are derived from a measurement series of 180 virtual loudspeaker positions in the array's equatorial plane with a radius of 6m. This distance is arguably sufficient to approximate plane wave with a single loudspeaker even for a reasonable frequency range regarding the array's small dimensions. Hence, the synthesis operator $\mathbf{D}$ in eq. 1 becomes the unit matrix.

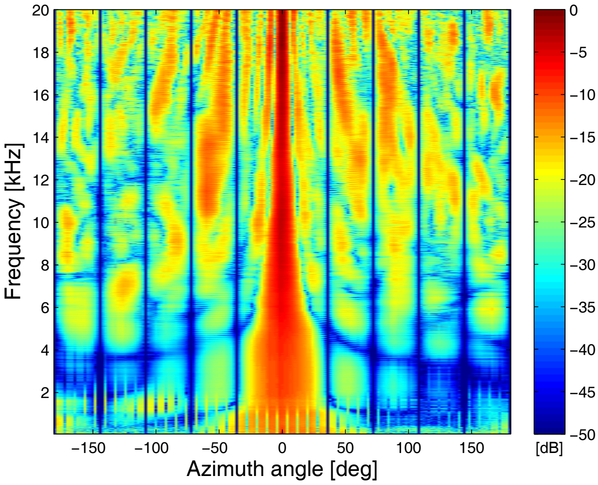


Figure 1: Directivity Pattern obtained by MIMO inversion, using 10 speaker positions in azimuth direction.

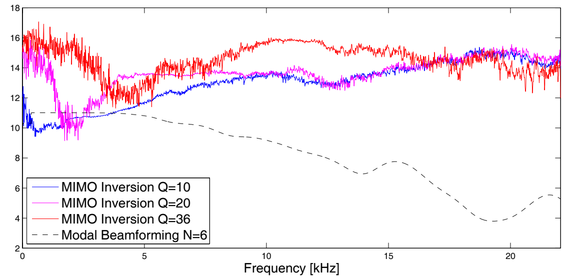
From the measured loudspeaker positions, different sets of equi-spaced incidence angles were selected. This paper presents the resulting beamfilters for $Q = 10, 20$ and 36 plane wave sources, a more extensive review including a discussion of numeric stability is given in [7]. The beampattern for $Q = 10$ sources is depicted in figure 1: The decoupled incidence directions are visible as notches of maximal suppression across the whole frequency range, spatial aliasing deteriorates the pattern from 5 kHz upwards. Directivity is maintained for high frequencies, it can be argued that the inversion exploits the individual microphone characteristics, which are no longer onmidirectional for higher frequencies. For reference, modal beamfilters of maximum order $N = 6$ have been simulated for an ideal spherical array of the same dimensions and sampling layout. A quadrature scheme proposed in [3] was employed, minimizing the discrete spherical harmonics' orthonormality error. Common measures of beamformer directivity and robustness in a diffuse noise field are the *directivity index (DI)* and the *white noise gain (WNG)*:

$$WNG(\theta_0, kr) = 10 \log_{10} \left(\frac{|\mathbf{dW}|^2}{\mathbf{W}^H \mathbf{W}} \right), \tag{6}$$

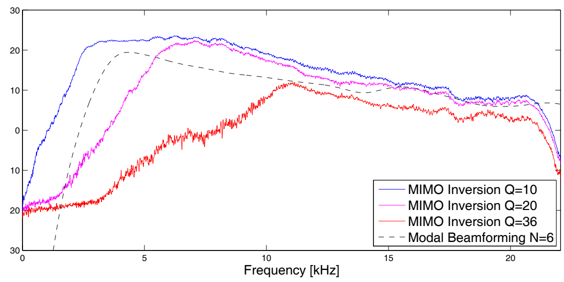
$$DI(\theta_0, kr) := 10 \log_{10} \left(\frac{K |D(\theta_0)|^2}{\sum\limits_{k=1}^{K} |D(\theta_k)|^2} \right) \tag{7}$$

with beampattern D at K discrete evaluation positions, column vector $\mathbf{W}$ of complex weights for each microphone and row vector $\mathbf{d}$ denoting the complex pressure caused by a plane wave from steering direction θ_0 [3].

Figure 2 illustrates the general trade-off between achievable directivity and robustness. The negative WNG at



(a) Directivity Index



(b) White Noise Gain

Figure 2: Directivity Index and White Noise Gain of three beamfilter sets with different numbers of decoupled sources and an order-6 modal beamformer.

frequencies for higher Q corresponds to ill-conditioned submatrices $\tilde{\mathbf{H}}_1$ and to an increased noise floor in the beampattern [7].

Conclusion

The filters obtained from inversion show overall satisfying performance. Robustness is comparable with the modal beamformer, pronounced directivity is achieved for even higher frequencies. Systematically balancing this trade-off by WNG-constrained optimization of the source number Q is suggested for possible future improvements.

References

[1] B.Rafaely: "Plane wave decomposition of the sound field on a sphere by spherical convolution", Journal of the Acoustical Society of America, Vol.116, July 2004

[2] K.Helwani, H.Buchner and S.Spors: "Calibration of Microphone Arrays with Arbitrary Geometries", German Annual Conference on Acoustics (DAGA), March 2011

[3] Z.Li: "The Capture and Recreation of 3D Auditory Scenes", Dissertation, University of Maryland, College Park, 2005

[4] P.Leopardi: "A Partition of the Unit Sphere into Regions of Equal Area and Small Diameter", El. Transactions on Numerical Analysis, Vol.25, 2006

[5] M.Miyoshi and Y.Kaneda: "Inverse Filtering of Room Acoustics", IEEE Transactions on Audio, Speech, and Language Processing, Vol.36, Feb.1988

[6] F.Zotter: "Analysis and Synthesis of Sound-Radiation with Spherical Arrays", Dissertation, University of Music and Performing Arts, Graz, Sept.2009

[7] T.Rettberg: "Acoustical Scene Analysis using a non-ideal spherical microphone array", Diploma Thesis, Technical University Berlin, 2012