

Design of a Ceiling-Microphone Array for Speech Applications with Focus on Transducer Arrangements and Beamforming Techniques

Christian Mortsiefer¹, Jürgen Peissig²

Leibniz Universität Hannover, Institut für Kommunikationstechnik, 30167 Hannover, Deutschland

¹ Email: cmortsiefer@gmail.com ² Email: peissig@ikt.uni-hannover.de

Abstract

A microphone array increases the speech intelligibility in challenging acoustic situations. By creating a highly directive pick-up pattern, influences such as room reflections (i.e. diffuse noise) and other spatial noise sources can be minimized. In the course of this work, various planar microphone arrangements are discussed and subsequently combined with different types of fixed beamforming techniques. A special focus has been taken on spiral and spiral-like topologies that are compared to each other for Delay-and-Sum (DS), Minimum-Variance-Distortionless-Response (MVDR) and a particular case of Filter-and-Sum (FS) beamforming, i.e. the Weighted-Least-Squares (WLS) optimization.

Fundamentals of Microphone Arrays

A general system model according to Fig. 1 can be used to derive the basic equations for a microphone array being composed of an arbitrary microphone topology and a choice of fixed beamforming weights. Considering a

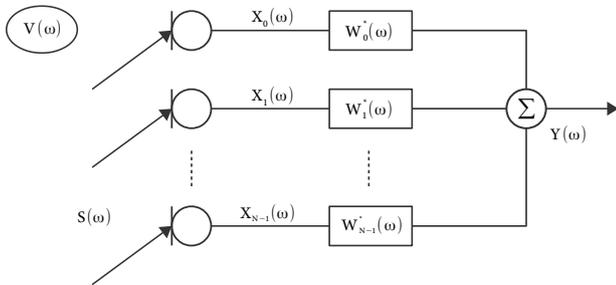


Figure 1: General system model of a microphone array for arbitrary topologies and beamforming weights.

speech source $S(\omega)$ (far-field assumption) and an ideal diffuse noise field $V(\omega)$, the array input spectrum $\mathbf{X}(\omega)$ can be written as $[N \times 1]$ stacked vector

$$\mathbf{X}(\omega) = S(\omega) \mathbf{d}(\omega, \theta, \varphi) + \mathbf{V}(\omega), \quad (1)$$

where N is the number of microphones. The steering vector $\mathbf{d}(\omega, \theta, \varphi)$ represents the acoustic transfer function between the speech source and each microphone for an impinging plane wave from direction (θ, φ) , i.e.

$$\mathbf{d}(\omega, \theta, \varphi) = [a_0 e^{-j\omega\tau_0}, \dots, a_{N-1} e^{-j\omega\tau_{N-1}}]^T, \quad (2)$$

where a_n denotes the n -th microphone sensitivity and τ_n the relative propagation delay to the origin of the coordinate system. This yields a relation to the actual array

topology, since $\tau_n(\theta, \varphi)$ depends on the microphone positions (x_n, y_n) , i.e.

$$\tau_n(\theta, \varphi) = \frac{f_s}{c} (x_n \sin(\theta) \cos(\varphi) + y_n \sin(\theta) \sin(\varphi)), \quad (3)$$

where $c = 343 \frac{m}{s}$ is the speed of sound and f_s is the sampling frequency. For the sake of readability, the dependence on φ is eliminated in the further context. Finally, the array output spectrum $Y(\omega)$ can be expressed as an inner product

$$Y(\omega) = \mathbf{W}^H(\omega) \mathbf{X}(\omega), \quad (4)$$

where $\mathbf{W}_n^*(\omega)$ denotes a set of beamforming weights [1].

Beamforming Techniques

In the course of this work three different beamforming techniques have been considered, the first of which being a DS beamformer [1]. The corresponding weights that steer the directivity pattern in a certain look direction θ_l can be calculated by

$$\mathbf{W}(\omega) = \frac{1}{N} \mathbf{d}(\omega, \theta_l). \quad (5)$$

A beamformer that is known to maximize the directivity index of an arbitrary microphone array can be found by means of the MVDR beamformer. Formally, the weights for MVDR beamforming are derived by minimizing the power spectral density at the array output assuming a certain noise field. The underlying cost function can be formulated as

$$\begin{aligned} \min_{\mathbf{W}(\omega)} \quad & \mathbf{W}^H(\omega) \mathbf{\Gamma}_{VV}(\omega) \mathbf{W}(\omega) \\ \text{subject to} \quad & \mathbf{W}^H(\omega) \mathbf{d}(\omega, \theta_l) = 1, \end{aligned} \quad (6)$$

where $\mathbf{\Gamma}_{VV}(\omega)$ is the coherence matrix for a particular noise field, i.e. an ideal diffuse noise field [1]. The solution to the minimization problem in (6) can be found by applying the method of Lagrange multipliers [2], which eventually yields

$$\mathbf{W}(\omega) = \frac{\mathbf{\Gamma}_{VV}^{-1}(\omega) \mathbf{d}(\omega, \theta_l)}{\mathbf{d}^H(\omega, \theta_l) \mathbf{\Gamma}_{VV}^{-1}(\omega) \mathbf{d}(\omega, \theta_l)}. \quad (7)$$

Finally, a FS beamformer based on the commonly used WLS optimization was considered, where the underlying cost function is given by

$$\min_{\mathbf{W}(\omega)} \sum_i^P F(\omega, \theta_i) \left| \mathbf{W}^H(\omega) \mathbf{d}(\omega, \theta_i) - D(\omega, \theta_i) \right|^2. \quad (8)$$

Based on a number of P discrete directions, this cost function takes the squared absolute difference between the beam pattern $H(\omega, \theta_i) = \mathbf{W}^H(\omega) \mathbf{d}(\omega, \theta_i)$ and a desired directivity pattern $D(\omega, \theta_i)$. The weighting function $F(\omega, \theta_i)$ can be set to assign more or less priority to certain directions denoted by θ_i , where $i = 1, \dots, P$ [3]. The beamforming weights minimizing the WLS cost function can be obtained by setting the gradient in (8) to zero [4], which then yields

$$\begin{aligned} \mathbf{W}(\omega) &= \mathbf{Q}^{-1}(\omega) \mathbf{a}(\omega) \\ \mathbf{Q}(\omega) &= \sum_i^P F(\omega, \theta_i) \mathbf{d}(\omega, \theta_i) \mathbf{d}^H(\omega, \theta_i) \\ \mathbf{a}(\omega) &= \sum_i^P F(\omega, \theta_i) \mathbf{d}(\omega, \theta_i) D^*(\omega, \theta_i). \end{aligned} \quad (9)$$

For further investigations, the desired pattern was set to

$$D(\omega, \theta_i) = \begin{cases} 0 \text{ dB} & i = l \\ -20 \text{ dB} & i \neq l \end{cases}. \quad (10)$$

Transducer Arrangements

A particular form of planar transducer arrangements can be found in spiral shaped designs. The topologies depicted in Fig. 2 have been taken from various popular spiral designs known from literature [5]. Spiral designs offer two advantages. Firstly, a spiral shaped arrangement naturally exhibits unique inter-element spacings, which are helpful to reduce sidelobe levels. Secondly, the influence of similar or equal inter-element spacings pointing in different directions within the plane leads to more or less symmetry with respect to the resulting main lobe. This kind of redundancy is also desirable if the resulting beam patterns are supposed to be of similar shape for different look directions.

Error Assumptions

When a microphone array is realized as an actual device, certain errors occur that should be considered in the design process to assure robustness. In the course of this work, two types of errors have been considered. While a microphone sensitivity error was assumed to be uniformly distributed within the interval of ± 1 dB, microphone displacement of ± 1 mm was considered in terms of a normal distribution ($\sigma^2 = 0.2, \mu = 0$). Principally, these error assumptions can be involved in the underlying system model by the following procedure. Firstly, beamforming weights are calculated based on ideal steering vectors, where the microphone sensitivities a_n share the same value and τ_n is based on the ideal microphone positions. Secondly, a corrupted steering vector is applied, when the actual array output is supposed to be calculated. The corrupted steering vector is obtained by adding a failure to the ideal microphone sensitivities a_n and to the ideal microphone positions (x_n, y_n) .

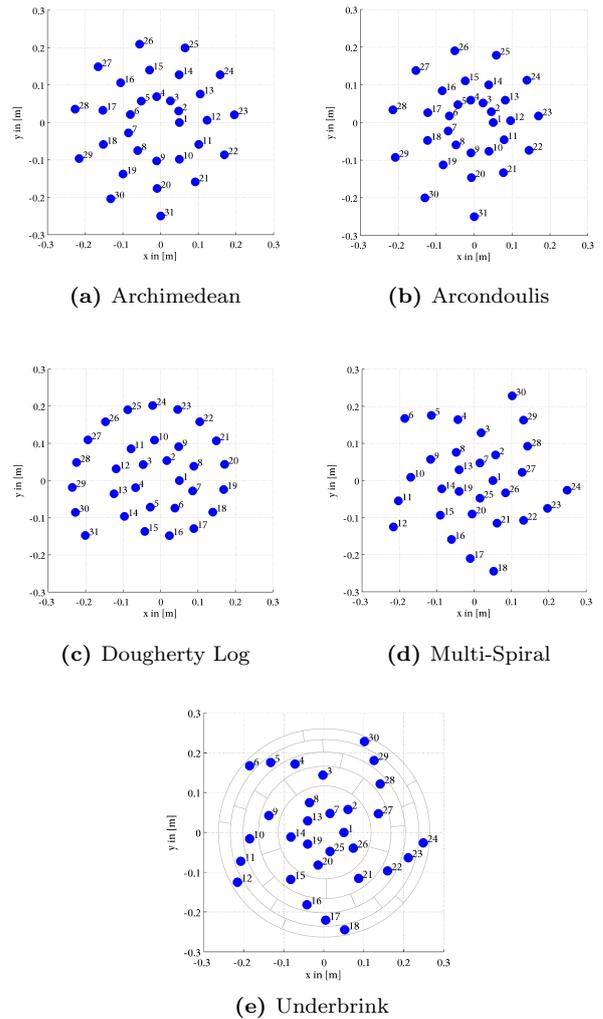


Figure 2: Various spiral shaped topologies.

Simulation Results

An exemplary beam pattern for the Multi-Spiral design in connection with MVDR beamforming is given by Fig. 3. It can be seen that directivity is obtained for a broad frequency range, which is particularly remarkable for frequencies below 500 Hz, as wavelengths are much bigger than the array dimension. For frequencies above 4000 Hz, an increase of sidelobe levels, i.e. grating lobes, can be noticed as a result of spatial aliasing due to the lack of smaller inter-element spacings.

Aiming to contrast the considered topologies and beamforming techniques for multiple steering directions, the analysis of single beam patterns becomes unmanageable. A solution to this problem can be found by applying the following methodology. By computing performance measures, such as Directivity Index (DI) and Maximum Sidelobe Level (MSL), with respect to multiple look directions (i.e. 73 uniformly distributed look directions), a DI/MSL distribution for all considered beamforming techniques and topologies can be obtained. Additionally, the results per look direction are represented by its arithmetic mean over 10 samples of the error distribution and over each octave band.

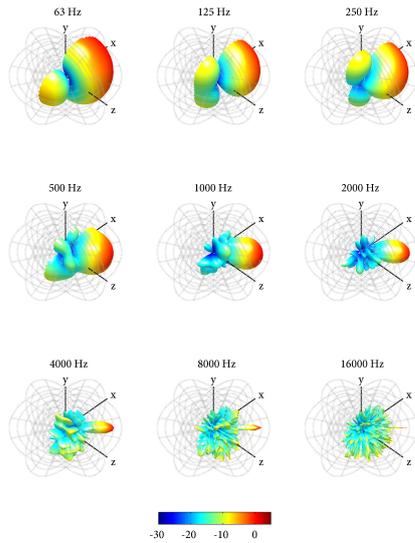


Figure 3: Beam pattern for Multi-Spiral and MVDR beamforming, ($\theta_l = \frac{\pi}{4}$, $\varphi_l = 0$).

Figure 4 illustrates the DI distribution for all considered topologies in connection with DS beamforming. As can be seen, basically no directivity is obtained for frequencies below 500 Hz independent of transducer arrangement, which is a major disadvantage of DS beamforming. The DI distribution for MVDR beamforming is

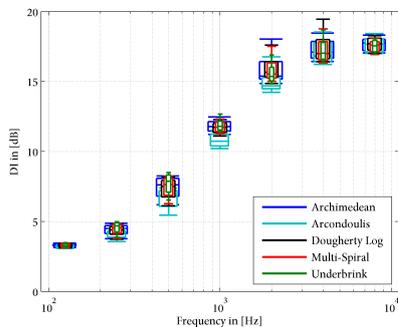
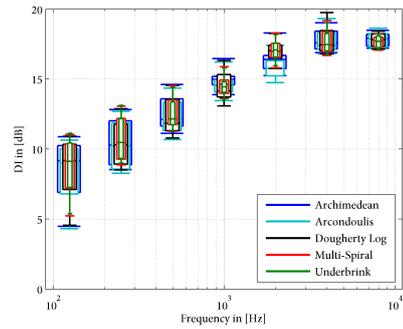
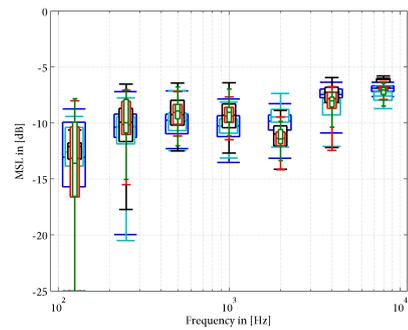


Figure 4: DI distribution for DS beamforming based on a set of 73 uniformly distributed look directions.

shown in Fig. 5a. In comparison to DS beamforming, the obtained DIs (especially at low frequencies) are clearly increased. A topology dependent benefit can be identified around 2000 Hz, where Dougherty Log-Spiral, Multi-Spiral and Underbrink Spiral are superior to the other arrangements. This is true for both, DI and MSL (c.f. Fig. 5b). In the frequency range of 250–1000 Hz, Arcondoulis and Archimedean Spiral achieve lower MSLs than the former. The DI distribution for FS beamforming, i.e. WLS optimization, is depicted in Fig. 6a. Although MVDR and FS beamformer are based on different mathematical approaches, the achieved performance is similar, however FS beamforming provides slightly inferior results. Considering the MSL distribution, FS beamforming provides better sidelobe reduction for frequencies below 500 Hz than MVDR beamforming (c.f. Fig. 6b).



(a) DI distribution for MVDR beamforming



(b) MSL distribution for MVDR beamforming

Figure 5: Performance for MVDR beamforming based on a set of 73 uniformly distributed look directions.

Evaluation Results

For the purpose of objective and subjective validation, a boundary prototype according to the Multi-Spiral arrangement was realized (c.f. Fig. 7). A MATLAB based signal processing was implemented as post-processing architecture by using a time-domain FIR-filter structure ($f_s = 48$ kHz, 512 taps per microphone channel).

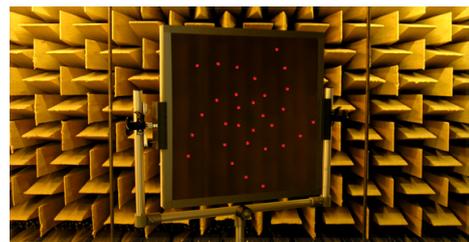
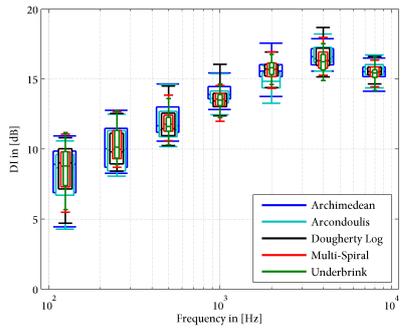


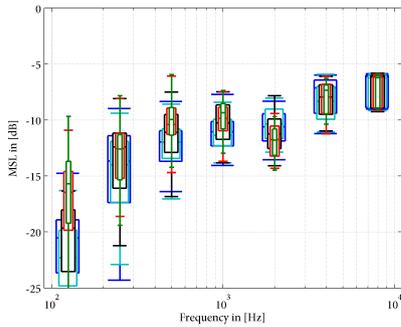
Figure 7: Multi-Spiral prototype for objective and subjective validation.

Polar Diagrams

Acoustic measurements were performed to compare the achieved array performance with its simulated version. Figure 8 shows a comparison of theoretical (green) and measured (red) data as polar diagrams of the xz -plane in the case of MVDR beamforming. In general, the comparison reveals conformity of simulated and measured performance. Discrepancies (c.f. 125–500 Hz) are caused by edge diffraction due to the finite baffle size. The effect of edge diffraction, however, can be disregarded if the device is flush mounted in the ceiling.



(a) DI distribution for FS beamforming



(b) MSL distribution for FS beamforming

Figure 6: Performance for FS beamforming based on a set of 73 uniformly distributed look directions.

Subjective Testing

For subjective validation, an informal AB-Test was conducted, where the array prototype was installed as a ceiling tile, i.e. flush mounted in the ceiling. A B&K head and torso simulator (Type 4128C) served as a potential speaker (speech source). The built-in mouth simulator was used to play back standardized sound material according to EBU-TECH 3253. In addition, a Neumann KH120 loudspeaker facing the corner of the room was used to radiate pink noise (noise source). Based on this setup, recordings were made for three different speaker locations, while the noise source was kept at its position. Eventually, the AB-Test shows that best results in terms of noise suppression and speech intelligibility can be obtained with reference to MVDR beamforming. FS beamforming is slightly inferior. The speech intelligibility is nearly consistent for the applied set of speaker positions. The effect of spatial aliasing, i.e. grating lobes, is audible and varies for different speaker positions.

Discussion and Outlook

In conclusion, MVDR and FS have shown superior results compared to DS beamforming. The question of suitable microphone arrangements is ambiguous, since each considered topology incorporates benefits with respect to certain octave bands. A perceptual evaluation of the different spiral arrangements might be desirable. Further improvements with respect to spatial aliasing could be achieved by increasing the number of smaller inter-element spacings and/or shrinking the array dimension. Low-frequency directivity could be further increased by

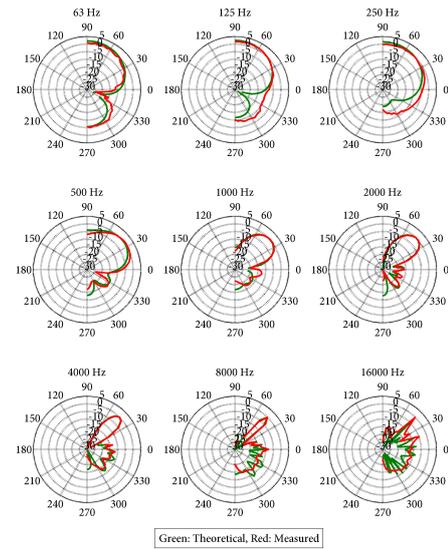


Figure 8: Polar diagrams (zx-plane) for Multi-Spiral and MVDR beamforming, ($\theta_l = \frac{\pi}{4}$, $\varphi_l = 0$).

closer tolerated microphones and by prevention of displacement. Regarding FS beamforming, recent research activities provide many further approaches on alternative optimization techniques (e.g. a Total-Least-Squares cost function) [3]. Finally, beamforming weights could be calculated with respect to measured steering vectors of a physical prototype, aiming to realize a calibration, i.e. equalization procedure [6].

References

- [1] Brandstein, M. and Ward, D.: *Microphone Arrays*. Springer, Berlin, 2001
- [2] Doclo, S.: *Multi-Microphone Noise Reduction and Dereverberation Techniques for Speech Applications*. Ph.D. dissertation, Katholieke Universiteit Leuven, 2003
- [3] Doclo, S. and Moonen, M.: Design of Broadband Beamformers Robust Against Gain and Phase Errors in the Microphone Array Characteristics. *IEEE Transactions on Signal Processing* (2003), 51(10):2511-2526
- [4] Rasumov et al.: Regularization Approaches for Synthesizing HRTF Directivity Patterns. *IEEE Transactions on Acoustics, Speech, and Signal Processing* (2016), 24(2):215-225
- [5] Prime, Z. and Doolan, C.: A Comparison of Popular Beamforming Arrays. *Proceedings of the Annual Conference of the Australian Society*, 2013
- [6] Sydow, C.: Broadband Beamforming for a Microphone Array. *The Journal of the Acoustical Society of America* (1994), 96(2):845-849

The support by Sennheiser electronic GmbH & Co. KG is greatly acknowledged.