NUMERICAL NEAR FIELD OPTIMIZATION OF WEIGHTED DELAY-AND-SUM MICROPHONE ARRAYS

Magnus Schäfer, Florian Heese, Jona Wernerus, and Peter Vary

Institute of Communication Systems and Data Processing (ind), RWTH Aachen University {schaefer, heese, wernerus, vary}@ind.rwth-aachen.de

ABSTRACT

A novel concept for the determination of the weighting coefficients for weighted delay-and-sum microphone arrays is introduced in this contribution. The concept is based on a numerical optimization of the reception characteristic of the microphone array. The optimization procedure is shown to improve the reception characteristic in such a way that it closely approximates a target which can be defined according to the application.

Index Terms— Microphone arrays, beamforming, near field, numerical optimization

1. INTRODUCTION

The design of array signal processing systems [1] received continuous interest throughout the last decades with many applications in the radio frequency domain [2] as well as the acoustic domain [3]. A special form of an array signal processing system in the acoustic domain is the linear microphone array which, due to its physical design, can be integrated easily in many communication systems such as video conferencing clients. A well designed microphone array is an efficient way to already achieve a decent signal-to-noise ratio (SNR) directly at the acoustic frontend if the target signal and the acoustic interferers are spatially separated.

Since this spatial separation is usually present in conferencing scenarios, the use of microphone arrays is especially beneficial in such an environment. Furthermore, the reverberation as well as the level of diffuse background noise is usually quite low in normal conference rooms. Hence, a microphone array is an efficient way to simultaneously amplify one target speaker while damping other speakers and background noise.

When designing and parameterizing microphone arrays, the target is usually to generate a certain reception characteristic. In the far field, i.e., at distances from the array that are significantly larger than the physical size of the array setup, there are many known procedures that can be utilized. There are some approaches that are specifically aimed at the near field [4, 5, 6] where the far field designs can only be used to approximately determine the reception characteristic. These approaches however, optimize the reception characteristic on a (semi-)circular arc at one specific distance from the array. A different design was proposed in [7] which allows to define a target region in the near field and modify the constraints for an adaptive beamformer accordingly. No approach is known yet that allows to optimize the reception characteristic for an entire area in the near field of the microphone array simultaneously for different distances and angles.

In Sec. 2, the procedure for the determination of the reception characteristic is introduced which is then utilized in Sec. 3 to optimize the weighting coefficients. An exemplary evaluation of the performance of the procedure follows in Sec. 4 before some concluding remarks are given in Sec. 5.

2. DETERMINATION OF THE RECEPTION CHARACTERISTIC IN THE NEAR FIELD

The proposed optimization procedure for the weighting coefficients relies on the reception characteristic in the near field of the microphone array. The reception characteristic can be determined in a three-step approach by

- simulating or measuring impulse responses between points in the near field and all microphones,
- processing these impulse responses with the microphone array to get an overall filter for every point in the near field, and
- calculating the amplification and damping for every point from these overall filters.

2.1. Impulse Responses in the Near Field

For the determination of the reception characteristic of the microphone array, impulse responses between positions **p** in the near field of the microphone array and all microphones are necessary. These impulse responses can either be simulated (e.g., by the mirror-image method [8]) or measured. When using simulated impulse responses, point sources on an

This work was co-funded by the German federal state North Rhine Westphalia (NRW) and the European Union (European Regional Development Fund: Investing In Your Future).



Fig. 1. Weighted delay-and-sum microphone array

appropriately chosen spatial grid (e.g., in a two-dimensional cartesian coordinate system: $\mathbf{p} = (x \ y)^T$) in the near field can be assumed and impulse responses $h_{\mathbf{p}m}(k)$ (with the discrete time index k and the microphone index m) from every point source to every microphone (located at position \mathbf{p}_m) in the array can be simulated.

With the impulse responses, the microphone signals $x_m(k)$ can be expressed in terms of filtered versions of the assumed source signal s(k).

$$x_{\mathbf{p}m}(k) = h_{\mathbf{p}m}(k) * s(k) \tag{1}$$

The fact that the optimization works in an identical manner with simulated and measured impulse responses makes it very flexible for different practical application scenarios.

2.2. Array Processing

A block diagram of the microphone array can be seen in Fig. 1. It consists of a weighted delay-and-sum setup with different delays τ_m and different weighting coefficients w_m at all M microphones. The output $y_{\mathbf{p}}(k)$ of the microphone array depends on the source location \mathbf{p} and can be calculated according to

$$y_{\mathbf{p}}(k) = \sum_{m=1}^{M} w_m \cdot \delta(k - \tau_m) * x_{\mathbf{p}m}(k).$$
 (2)

For every position, the weighted superposition of the individual signals leads to an effective overall filter g(k) since the output signal can be expressed as a filtered version of the source signal.

$$y_{\mathbf{p}}(k) = \sum_{m=1}^{M} w_m \cdot \delta(k - \tau_m) * \left(h_{\mathbf{p}m}(k) * s(k) \right)$$
(3)

Due to the associative property of the convolution, this can be reformulated to

$$y_{\mathbf{p}}(k) = \sum_{m=1}^{M} w_m \cdot h_{\mathbf{p}m}(k - \tau_m) * s(k) \,. \tag{4}$$

The overall filter $g_{\mathbf{p}}(k)$ for a source at point \mathbf{p} can hence be determined as

$$g_{\mathbf{p}}(k) = \sum_{m=1}^{M} w_m \cdot h_{\mathbf{p}m}(k - \tau_m) \,. \tag{5}$$

2.3. Calculation of the Reception Characteristic

With the frequency transform of the overall filter $g_{\mathbf{p}}(k)$

$$G_{\mathbf{p}}(f) = \mathcal{F}\left\{g_{\mathbf{p}}(k)\right\},\tag{6}$$

the reception characteristic $S_{\mathbf{p}}(f)$ in dB can be calculated at frequency f for every point \mathbf{p} in the vicinity of the microphone array by

$$S_{\mathbf{p}}(f) = 20 \cdot \log_{10} |G_{\mathbf{p}}(f)| .$$
(7)

3. NUMERICAL OPTIMIZATION

The weighted delay-and-sum microphone array offers two degrees of freedom, the factors w_m and the delays τ_m , that have to be set to achieve a certain predefined behaviour of the system. For the delays, this is achieved by adjusting them in such a way that signals from the target direction will be added up coherently in the summation point. For the weighting factors however, a novel numerical optimization scheme is proposed that aims to modify the weighting factors in order to mimic a target for the reception characteristic.

3.1. Definition of the Target

The target $\hat{S}_{\mathbf{p}}(f)$ for the optimization is defined as a spatial distribution of areas of amplification or damping in front of the microphone array. This basically equals the definition of a target SNR for the received signal as the target speaker shall be in the amplified area \mathbb{P}_{high} (target level S_{high}) while all interferers shall be in the damped area \mathbb{P}_{low} (target level S_{low}). The exact choice of both areas and both levels depends on *a priori* knowledge from the application, e.g., in a conferencing scenario, the target speaker shall be amplified while all interfering sources shall be damped. The target can be defined individually for all frequencies but a frequency-independent target is advantageous for many applications.

$$\hat{S}_{\mathbf{p}}(f) = \hat{S}_{\mathbf{p}} = \begin{cases} S_{\text{high}} & \text{for } \mathbf{p} \in \mathbb{P}_{\text{high}} \\ S_{\text{low}} & \text{for } \mathbf{p} \in \mathbb{P}_{\text{low}} \end{cases}$$
(8)

An additional advantage of this concept for the determination of the weighting factors is the fact that the definition of the target areas also allows to include computational complexity considerations into the system design: Larger target areas lead to larger complexity. The majority of the computational complexity within the optimization process lies in the error function which will be introduced in Sec. 3.2. It has to be evaluated only at the points that are in the target area but has to be evaluated frequently within the optimization process.



Fig. 2. Reception characteristic of the microphone array with Chebyshev weighting at 2000 Hz

3.2. Error Function and Optimization

The target of the optimization procedure is to minimize the summed level difference Δ_S between the predefined target \hat{S} and the simulated reception characteristic S. The level difference is summed over all points for which $\hat{S}_{\mathbf{p}}(f)$ is defined according to Eq. 8 and over all frequencies $\lambda \in [\lambda_{\min} \dots \lambda_{\max}]$ for which the reception characteristic shall be optimized.

$$\Delta_{S} = \sum_{\lambda = \lambda_{\min}}^{\lambda_{\max}} \sum_{\mathbf{p} \in (\mathbb{P}_{\operatorname{high}} \cup \mathbb{P}_{\operatorname{low}})} \hat{S}_{\mathbf{p}}(\lambda) - S_{\mathbf{p}}(\lambda) \qquad (9)$$

The optimum weighting factors are determined from this summed level difference in a minimum mean square error (MMSE) sense.

$$\mathbf{w}_{\rm opt} = \arg\min_{\mathbf{w}} \Delta_S^2 \tag{10}$$

The optimization is carried out by an interior-point algorithm [9] under the constraint that the weighting coefficients have to be in the range of $w_{\min} = -1$ and $w_{\max} = 1$. This constraint does not change the relation between the weighting coefficients, it only limits the maximum amplification that is achievable by the array itself. A subsequent scaling of the output y(k) can be applied to map the reception characteristic to a suitable level.

4. PERFORMANCE EXAMPLE

The performance of the novel optimization procedure for the reception characteristic of microphone arrays in the near field can be assessed exemplarily by comparing it to the reception characteristic of unoptimized microphone arrays that utilize, e.g., a Chebyshev weighting $\mathbf{w}_{\mathrm{Cheb}}$. This window is chosen here as a good benchmark since it allows to specify a minimum damping for all sidelobes while at the same time also



Fig. 3. Reception characteristic of the microphone array with optimized weighting at 2000 Hz

minimizing the width of the main lobe. This combination is very advantageous since it maximizes the SNR between a target area and a diffuse noise field.

For both setups, the Chebyshev weighting and the optimized weighting, the same delays τ_m are used to allow for a detailed comparison that only takes the effect of the weighting coefficients into account. Additionally, both weightings are parameterized in such a way that they are supposed to lead to a level difference of 40 dB between the amplified and the damped area. For all presented cases, a microphone array consisting of 8 sensors with a uniform spacing of 3 cm is used which is centered in the origin of the coordinate system.

In a possible application scenario within a video conferencing system, the simulation of the impulse responses can be fairly simple since many conference rooms are not highly reverberant and do not exhibit complicated architecture. Hence, a simple mirror-image approach or even an approximation by a free field model is suitable. The reception characteristic is visualized here (without loss of generality) for a free field setup since this allows for a clearer evaluation of the impact of the weighting coefficients. A comparison of the reception characteristics is given for two different frequencies:

- $\lambda = 2000 \,\text{Hz}$ as a frequency that is right in the center of the operational frequency range of the microphone array
- $\lambda = 500 \,\text{Hz}$ as a representative for the lower frequencies for which the microphone array can be used

The microphone array is designed to amplify sources on the left $(-0.5 \text{ m} \le x < 0 \text{ m} \land 0.2 \text{ m} < y \le 0.8 \text{ m})$ while damping sources on the right $(0 \text{ m} < x \le 0.5 \text{ m} \land 0.2 \text{ m} < y \le 0.8 \text{ m})$. For both dimensions (x and y), the density of the spatial grid is set to 0.01 m leading to 3000 points in \mathbb{P}_{high} and \mathbb{P}_{low} , respectively.

Looking at the performance of the Chebyshev weighting for the 2000 Hz case in Fig. 2, there is already a sig-



Fig. 4. Reception characteristic of the microphone array with Chebyshev weighting at 500 Hz

nificant level difference between the left and the right side showing that the Chebyshev weighting can be used at this frequency with this microphone array. However, the amplified area clearly extends to the area directly to the right of the center (0m < x < 0.25m). In contrast, with the optimized weighting the reception characteristic in Fig. 3 matches the previously defined areas of amplification and damping very well. Especially in the critical transition region around x = 0 m, a more pronounced border between the amplified and the damped area can be observed.

For the 500 Hz case, the reception characteristic of the microphone array with the Chebyshev weighting is depicted in Fig. 4. This reception characteristic strongly resembles the one of a single omnidirectional microphone in the origin of the coordinate system. The reception characteristic for the optimized weighting coefficients can be found in Fig. 5 where, obviously, some level difference between the left and right side can be observed even for this low operational frequency.

5. CONCLUSION

A novel numerical optimization scheme for the weighting factors of a weighted delay-and-sum array was derived in this contribution. The optimization scheme allows to optimize the entire reception characteristic in the vicinity of a microphone array at once. The reception characteristic with optimized weighting factors was shown to match the target characteristic very well.

In a practical application, e.g., within a conferencing system, the optimization scheme is advantageous as it can be used very flexibly due to the fact that it works with simulated as well as measures impulse responses, can be parameterized for lower complexity, and does not rely on any specific microphone array geometry.



Fig. 5. Reception characteristic of the microphone array with optimized weighting at 500 Hz

6. REFERENCES

- S. Haykin and K. J. R. Liu, *Handbook on Array Processing and* Sensor Networks, Adaptive and Learning Systems for Signal Processing, Communications and Control Series. Wiley, 2010.
- [2] S. Haykin, J. Litva, and T. Shepherd, *Radar array processing*, Springer series in information sciences. Springer-Verlag, 1993.
- [3] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*, Digital Signal Processing. Springer, 2001.
- [4] Rodney A. Kennedy, Thushara Abhayapala, Darren B. Ward, and Robert C. Williamson, "Nearfield broadband frequency invariant beamforming," in *IEEE ICASSP*, 1996.
- [5] James G. Ryan and Rafik A. Goubran, "Array optimization applied in the near field of a microphone array," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 2, pp. 173– 176, March 2000.
- [6] Etan Fisher and Boaz Rafaely, "Near-field spherical microphone array processing with radial filtering," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 2, pp. 256–265, February 2011.
- [7] Y.R. Zheng, R.A. Goubran, and M. El-Tanany, "Robust near-field adaptive beamforming with distance discrimination," *Speech and Audio Processing, IEEE Transactions on*, vol. 12, no. 5, pp. 478 – 488, sept. 2004.
- [8] Jont B. Allen and David A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.
- [9] R. H. Byrd, J. C. Gilbert, and J. Nocedal, "A trust region method based on interior point techniques for nonlinear programming," *Mathematical Programming*, vol. 89, no. 1, pp. 149–185, 2000.