

MICROPHONE POSITION OPTIMIZATION FOR PLANAR SUPERDIRECTIONAL BEAMFORMING

Ina Kodrasi^{1,2}, Thomas Rohdenburg², Simon Doclo¹

ina.kodrasi@uni-oldenburg.de

¹University of Oldenburg, Institute of Physics, Signal Processing Group, Oldenburg, Germany

²Fraunhofer IDMT, Project Group Hearing, Speech and Audio Technology, Oldenburg, Germany

ABSTRACT

The performance of a fixed beamformer highly depends on the position of the microphones in the array. In this paper, different heuristic optimisation approaches for arbitrary planar arrays and an exhaustive search approach for structured array geometries are presented to optimise the microphone positions for a superdirective beamformer, aiming at maximizing the mean directivity index for several steering angles of interest. Through the derivation of an upper bound on the achievable performance, it is shown that the proposed approaches generate configurations with a near-optimal performance. In addition, the theoretical results are validated using real measurements, demonstrating the practical usability of the proposed methods.

Index Terms— superdirective beamforming, microphone position, directivity index

1. INTRODUCTION

In many speech communication applications, the microphone signals are corrupted by background noise and reverberation. The objective of a fixed (data-independent) beamformer is to obtain spatial focusing on the speech source, thereby reducing noise and reverberation not coming from the same direction as the speech source. Of the different types of fixed beamformers that are available, the superdirective beamformer [1, 2] which optimizes the array gain for a diffuse noise field, remains a popular choice.

Although (superdirective) beamformers are conventionally designed for a fixed array geometry, the position of the microphones in the array has an important influence on the achievable array gain. It can be shown that the array gain can be expressed as a non-linear function of the microphone positions, which exhibits many local maxima, such that we have to resort to either exhaustive search or heuristic optimisation approaches to find the (globally) optimal microphone array configuration. Optimization methods for obtaining optimal microphone positions for a linear microphone array steering towards one specific angle have been proposed in [3, 4]. However, many (multimedia) applications nowadays rely on the ability to steer towards multiple angles, for which planar arrays [5] are more suitable. Applying exhaustive search to optimize the microphone positions for arbitrary planar arrays is typically infeasible in practise, due to the huge number of possible array configurations.

This paper discusses several approaches for obtaining optimal microphone positions for a planar superdirective beamformer that should be able to steer towards multiple angles. For maximizing the proposed non-linear cost function, defined as the mean (frequency-weighted) directivity index for the considered steering angles, we propose three heuristic algorithms. In addition, exhaustive search is

applied to structured array geometries in order to evaluate the limitation imposed by considering only such pre-defined array geometries. Theoretical knowledge about the maximum directivity index that can be achieved using a uniform linear array (ULA) is used to show that both proposed approaches yield array configurations with near-optimal performance. When computational complexity is an issue, the preferred approach to be used is the genetic algorithm.

2. CONFIGURATION AND DEFINITIONS

Consider the linear microphone array configuration depicted in Fig. 1, consisting of N omnidirectional microphones, a desired speech source $S(\omega)$ located in the far-field at an angle θ_s , and a noise field $V(\omega)$. Let l_n denote the distance between the n -th microphone and the reference point, arbitrarily chosen here as the first microphone. The signal received at the n -th microphone is equal to

$$Y_n(\omega) = d_n(\omega, \theta_s) S_r(\omega) + V_n(\omega),$$

where $S_r(\omega)$ represents the speech component of the signal received at the reference point, $V_n(\omega)$ the noise component of the n -th microphone signal, and

$$d_n(\omega, \theta) = e^{-j\omega\tau_n(\theta)}. \quad (1)$$

The delay $\tau_n(\theta)$ in number of samples is equal to $l_n \sin \theta \frac{f_s}{c}$, with c the speed of sound propagation and f_s the sampling frequency. The stacked vector of the microphone signals $\mathbf{Y}(\omega) = [Y_0(\omega) Y_1(\omega) \dots Y_{N-1}(\omega)]^T$ can hence be written as

$$\mathbf{Y}(\omega) = \mathbf{d}_s(\omega) S_r(\omega) + \mathbf{V}(\omega),$$

where $\mathbf{d}_s(\omega)$ is the steering vector for the speech source defined as

$$\mathbf{d}_s(\omega) = \mathbf{d}(\omega, \theta_s) = [d_0(\omega, \theta_s) d_1(\omega, \theta_s) \dots d_{N-1}(\omega, \theta_s)]^T,$$

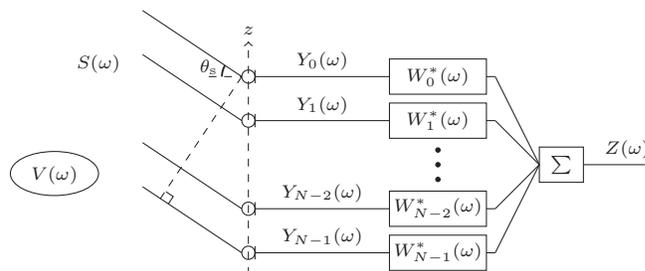


Fig. 1. Linear microphone array configuration

and $\mathbf{V}(\omega)$ is the stacked vector of noise signals defined similarly as $\mathbf{Y}(\omega)$. The output signal $Z(\omega)$ is obtained by filtering and summing the microphone signals, i.e.

$$Z(\omega) = \mathbf{W}^H(\omega)\mathbf{Y}(\omega) = \mathbf{W}^H(\omega)\mathbf{d}_s(\omega)S_r(\omega) + \mathbf{W}^H(\omega)\mathbf{V}(\omega),$$

with $\mathbf{W}(\omega)$ the stacked filter coefficient vector of the beamformer.

The performance measures commonly used to evaluate beamformers are the directivity index (DI) and the white noise gain (WNG). The directivity index is defined as the ability to suppress a (spherically or cylindrically) isotropic noise field, i.e.

$$DI_{\theta_s}(\omega) = 10 \log_{10} \frac{|\mathbf{W}^H(\omega)\mathbf{d}_s(\omega)|^2}{\mathbf{W}^H(\omega)\mathbf{\Gamma}_{VV}^{\text{diff}}(\omega)\mathbf{W}(\omega)} \quad (2)$$

where $\mathbf{\Gamma}_{VV}^{\text{diff}}(\omega)$ represents the diffuse noise coherence matrix. For a spherically isotropic noise field, the coherence between two microphones i and j at a distance l_{ij} is equal to

$$\Gamma_{V_i V_j}(\omega) = \frac{\sin(\omega f_s l_{ij}/c)}{\omega f_s l_{ij}/c}, \quad (3)$$

whereas for a cylindrically isotropic noise field, the coherence is equal to

$$\Gamma_{V_i V_j}(\omega) = J_0\left(\frac{\omega l_{ij}}{c}\right), \quad (4)$$

with J_0 the zero-th-order Bessel function of the first kind.

The WNG is defined as the ability of the beamformer to suppress spatially uncorrelated noise (e.g. self-noise of the microphones) for which the noise coherence matrix $\mathbf{\Gamma}_{VV}^{\text{unc}}(\omega) = \mathbf{I}_N$, with \mathbf{I}_N the $N \times N$ -dimensional identity matrix, i.e.

$$WNG(\omega) = 10 \log_{10} \frac{|\mathbf{W}^H(\omega)\mathbf{d}_s(\omega)|^2}{\mathbf{W}^H(\omega)\mathbf{W}(\omega)}.$$

For the sake of clarity, the frequency-domain parameter ω will be omitted where possible in the remainder of the paper.

3. SUPERDIRECTIVE BEAMFORMING

A superdirective beamformer maximizes the directivity index in (2) and imposes a unity gain in the direction of the speech source, i.e. $\mathbf{W}^H \mathbf{d}_s = 1$ [1]. The filter coefficients then can be computed as

$$\mathbf{W} = \frac{(\mathbf{\Gamma}_{VV}^{\text{diff}})^{-1} \mathbf{d}_s}{\mathbf{d}_s^H (\mathbf{\Gamma}_{VV}^{\text{diff}})^{-1} \mathbf{d}_s}.$$

It is well known that superdirective beamformers are sensitive to uncorrelated noise, especially at low frequencies [1, 6]. A commonly used technique to limit the amplification of uncorrelated noise is to impose a WNG constraint, i.e. $\mathbf{W}^H \mathbf{W} \leq \beta$. When such a constraint is imposed, the filter coefficients can be computed as

$$\mathbf{W} = \frac{(\mathbf{\Gamma}_{VV}^{\text{diff}} + \mu \mathbf{I}_N)^{-1} \mathbf{d}_s}{\mathbf{d}_s^H (\mathbf{\Gamma}_{VV}^{\text{diff}} + \mu \mathbf{I}_N)^{-1} \mathbf{d}_s} \quad (5)$$

where the parameter μ is (iteratively) computed such that the inequality constraint $\mathbf{W}^H \mathbf{W} \leq \beta$ is satisfied. Using the filter coefficients in (5), the directivity index of a superdirective beamformer steering towards θ_s takes the form

$$DI_{\theta_s}(\omega) = 10 \log_{10} \frac{[\mathbf{d}_s^H (\mathbf{\Gamma}_{VV}^{\text{diff}} + \mu \mathbf{I}_N)^{-1} \mathbf{d}_s]^2}{\mathbf{d}_s^H (\mathbf{\Gamma}_{VV}^{\text{diff}} + \mu \mathbf{I}_N)^{-1} \mathbf{\Gamma}_{VV}^{\text{diff}} (\mathbf{\Gamma}_{VV}^{\text{diff}} + \mu \mathbf{I}_N)^{-1} \mathbf{d}_s} \quad (6)$$

which obviously depends on θ_s and the microphone positions, cf. (1), (3), and (4). Since in many applications it is important to steer towards different directions, we will focus here on maximizing the mean directivity index in (6) for multiple steering angles θ_s^i , $i = 0, \dots, n-1$. Therefore, if the vector \mathbf{p} represents the microphone positions in the array, we define the cost function to be maximized as the mean (frequency-weighted) directivity index over all considered steering angles θ_s^i , i.e.,

$$DI_m(\mathbf{p}) = \sum_{i=0}^{n-1} \int_0^\pi DI_{\theta_s^i}(\omega) F(\omega) d\omega \quad (7)$$

where $F(\omega)$ denotes a frequency weighting applied to the directivity index. The non-linear dependence of the steering vector and the noise coherence matrix on the microphone positions, implies that $DI_m(\mathbf{p})$ is a non-linear function of \mathbf{p} . Further, the dimension of \mathbf{p} reflecting the number of microphones might be large. Maximizing this multi-dimensional non-linear function is a tedious problem for which we propose several approaches in the next section.

4. OPTIMAL BEAMFORMER DESIGN

A straightforward way to optimize the microphone positions \mathbf{p} such that the mean directivity index DI_m in (7) is maximized, is through exhaustive search. However, for a planar array configuration, supposing there are n_x , n_y possible x - and y - positions for each of the N microphones, the total number of combinations is $\binom{m}{N}$, where $m = n_x n_y$ and $\binom{m}{N}$ denotes the binomial coefficient. Even for a small number of possible positions and a small number of microphones, the computational complexity for an exhaustive search is extremely high. For example, optimizing an array configuration for $N = 8$ microphones, where each microphone can occupy 90 different positions, would result in approximately 80 billion combinations. One method for reducing this computational load is through applying exhaustive search only to structured array geometries, as discussed in Section 4.1. To overcome the limitation of considering only structured array geometries, we also propose heuristic approaches for arbitrary array configurations in Section 4.2.

4.1. Exhaustive search on structured array geometries

The advantage of defining structured array geometries consists in significantly reducing the dimension of the parameter search space. Consider for example the unequally spaced rectangular or symmetric configuration shown in Fig. 2(a) and 2(b). In these array configurations, the distances d_1 , d_2 , d_3 , and d_4 determine the exact position of all sensors. As a result, using an exhaustive search approach to optimize these distances for such structured array geometries becomes feasible. Alternatively, one could obviously also consider several other structured configurations in an attempt to cover a wide range of the search space, such as a circular or an equally spaced cross-like array geometry, where the optimization problem reduces to finding the optimal radius or inter-sensor distance.

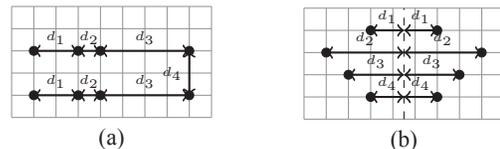


Fig. 2. Structured array configurations: (a) unequally spaced rectangular and (b) symmetric

4.2. Heuristic approaches

To optimize multi-dimensional non-linear functions, heuristic algorithms have been proposed that obtain a near-optimal solution (i.e. local extremum) at a significantly reduced computational cost. The heuristic approaches we present are the sifting, dynamic programming, and genetic algorithm approach, which for the previously mentioned example, reduce the computational load to approximately 500, 15000, and 3000 combinations respectively.

Sifting approach. This approach optimizes the position of one sensor at a time. Starting from an initial configuration \mathbf{p}_0 , the position of the first sensor is varied throughout the grid until a configuration \mathbf{p}_1 is found such that $DI_m(\mathbf{p}_1) > DI_m(\mathbf{p}_0)$. In the second step, the position of the second sensor is varied until a configuration \mathbf{p}_2 is found such that $DI_m(\mathbf{p}_2) > DI_m(\mathbf{p}_1)$. This process is repeated until all sensors have been accounted for and typically, the final configuration \mathbf{p}_{N-1} will be highly dependent on the initial configuration \mathbf{p}_0 .

Dynamic programming approach. In this approach which has been proposed in [7] for linear arrays but can also be applied to planar arrays, it is assumed that the optimal position of a sensor only depends on the position of the next sensor. The procedure works as follows. The first element of the array can be placed in any of the possible locations. Likewise, the second element can also be placed anywhere. For each location of the second element, there is a position of the first element that yields the highest value of the cost function while all other combinations are disregarded. The same iterative procedure is followed until every microphone has been accounted for.

Genetic algorithm. The input to this algorithm is a set of potential solutions, which the algorithm attempts to improve through quantitatively evaluating each candidate. A typical genetic algorithm requires encoding of the search space as well as a so-called fitness function which evaluates each solution. To apply it to microphone position optimization, we have encoded the microphone configurations \mathbf{p} into an array of bits and defined the fitness function as the cost function $DI_m(\mathbf{p})$. The search for an optimal solution in a genetic algorithm typically relies on the basis of crossover and mutation [8]. Crossover refers to randomly selecting one point in the array bits of two configurations and the corresponding parts are exchanged to create two new configurations. On the other hand, mutation is the occasional random inversion of a bit done with the purpose of reintroducing useful genetic material that might be lost. Once we define the space encoding, the fitness function and the mutation probability, an initial population sample is provided and the genetic algorithm attempts to improve it until a certain number of iterations that can be user-defined.

5. SIMULATIONS

In this section, the proposed optimization approaches are used to design linear and planar microphone arrays using $N = 8$ microphones, sampling frequency $f_s = 16000$ Hz, filter length $L = 64$, and WNG constraint -15 dB. Furthermore, the weights $F(\omega)$ applied to the directivity index in (7) are defined according to the speech intelligibility weighting curve in [9]. Whereas in Section 5.1 a linear array for a single steering angle is designed, in Section 5.2 a planar array is designed for multiple steering angles. The beamformer designs are based on the assumption of a cylindrically isotropic noise field in order to be able to evaluate the proposed methods with real measurements, cf. Section 5.3. However, the methods presented can be generalized to beamformers designed under any noise field assumption.

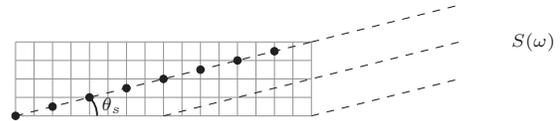


Fig. 3. Optimal linear array placed in the direction of the steering angle

5.1. Theoretical linear arrays

It has been theoretically proved in [2] that the maximum directivity index for a cylindrical noise field and a single steering angle θ_s , is equal to $10 \log_{10} [2(N-1)]$ when no WNG constraint is imposed, i.e. 11.8 dB for $N = 8$. This maximum directivity index value is achieved by a ULA placed in the direction of the steering angle as illustrated in Fig. 3. When imposing a WNG constraint of -15 dB, an exhaustive search approach can be used to determine the optimal inter-sensor distance for the ULA \mathbf{p}_1 , that attains the maximum directivity index for $\theta_s = 0^\circ$. Calculating the performance of this configuration yields $DI_m(\mathbf{p}_1) = 9.3$ dB.

Now suppose we would like to design a microphone array to steer towards θ_s^1 and θ_s^2 . The optimal configuration to steer towards θ_s^1 would be the ULA placed in the direction of θ_s^1 , while the optimal configuration to steer towards θ_s^2 would be the ULA placed in the direction of θ_s^2 . This can obviously not be achieved in practice since the same configuration needs to be used to steer at both angles of interest. As a result, the performance value $DI_m = 9.3$ dB can be considered as a loose upper bound on the achievable performance when designing linear as well as planar microphone arrays to steer towards single or multiple angles¹.

5.2. Theoretical planar arrays

In this section, the proposed optimization approaches are used to design planar arrays to steer towards 3 steering angles, namely $\theta_s = -45^\circ$, $\theta_s = 0^\circ$, and $\theta_s = 45^\circ$. The x - and y - position of the sensors are chosen to be full integer values in the intervals $x \in [1\text{cm}, 18\text{cm}]$, $y \in [1\text{cm}, 5\text{cm}]$, therefore a total of 90 possible sensor positions are defined for each of the $N = 8$ sensors.

The exhaustive search algorithm is applied to the configurations shown in Fig. 2 to optimize the distances d_1 , d_2 , d_3 , and d_4 . For the genetic algorithm, the encoding of the search space is done through arrays of 80 bits, where the first 5 bits denote the x - position of the first sensor, the next 5 bits denote the y -position of the first sensor and so on.

Table 1 illustrates the optimal performance values obtained through the exhaustive search approach, whereas Table 2 presents these values when the heuristic approaches are applied. One notices that all considered optimization approaches lead to very similar performance values, while the configurations \mathbf{p}_{opt} that they yield are different. The latter confirms that the cost function in (7) has many local maxima, several of which are found by the considered optimization methods. Among all considered approaches, the sifting approach yields the configuration \mathbf{p}_{opt} shown in Fig. 4 with the highest $DI_m(\mathbf{p}_{\text{opt}}) = 9.0$ dB. This value is only 0.3 dB lower than the loose theoretical upper bound derived in Section 5.1, which confirms that even though there is no guarantee that the proposed approaches determine the global optimal configuration, they at least determine configurations whose performance is very close to the

¹Note that the optimal performance value for planar arrays steering towards multiple angles could only be determined through exhaustive search, which is not feasible in practise, cf. Section 4.

Table 1. Performance of optimal microphone configurations using the exhaustive search algorithm

Method	Unequal rectangular	Symmetric
$DI_m(\mathbf{p}_{\text{opt}})$	8.9 dB	8.9 dB

Table 2. Performance of optimal microphone configurations using the heuristic approaches

Method	Dynamic	Genetic	Sifting
$DI_m(\mathbf{p}_{\text{opt}})$	8.9 dB	8.9 dB	9.0 dB

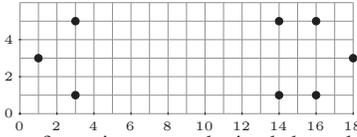


Fig. 4. Array configuration \mathbf{p}_{opt} obtained through the sifting approach

global optimum. When computational complexity is an issue, the preferred approach to be used is the genetic algorithm.

5.3. Practical beamformer designs

The optimal array configurations in Section 5.2 were determined using free-field conditions and ideal microphones. To validate their performance in practise, we have mounted these configurations on a hard surface using uncalibrated microphones. This setup has been placed in an anechoic room and measurements of the (real) steering vectors \mathbf{d}_s have been made for every 5° , using a loudspeaker placed 2.5 m away from the setup in order to mimic far-field conditions. The (real) noise coherence matrix and the beamformer filter coefficients have been recalculated based on the measured steering vectors.

As an example, for the array configuration found through the sifting approach, Fig. 5(a) depicts the theoretical and measured DI for all 3 steering angles. The theoretical and measured DI exhibit a very similar pattern. Furthermore, the measured beam patterns illustrated in Fig. 5(b) for all 3 steering angles, show that such a configuration offers a very good spatial selectivity for multiple angles. Therefore, the proposed theoretical optimization approaches result in array configurations with a high performance in practice as well.

6. CONCLUSION

In this paper we have described several computationally efficient approaches for optimizing the microphone positions for planar superdirective beamformers, that are required to steer towards several angles. Through the derivation of an upper bound on the achievable performance, it is shown that the proposed methods find near-optimal configurations. Furthermore, we have demonstrated their usability for designing a superdirective beamformer using 8 microphones to steer towards 3 steering angles.

7. REFERENCES

- [1] J. Bitzer and K. U. Simmer, "Superdirective microphone arrays," in *Microphone Arrays: Signal Processing Techniques and Applications*, chapter 2, pp. 19–37. Springer Verlag, 2001.
- [2] G. Elko, "Superdirectional microphone arrays," in *Acoustic Signal Processing for Telecommunication*, chapter 10, pp. 181–237. Kluwer Academic Publishers, Boston, 2000.

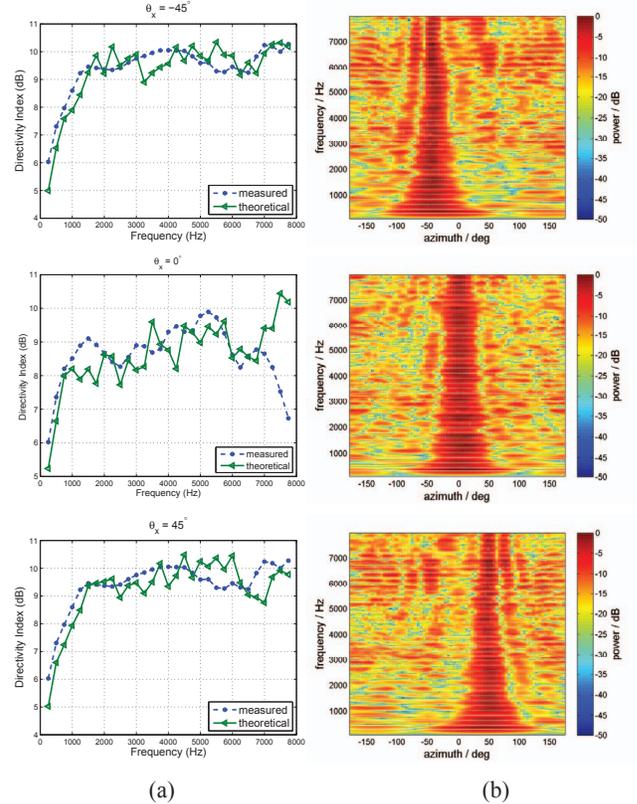


Fig. 5. (a) Theoretical and measured directivity index and (b) measured beam patterns of the optimal array configuration found through the sifting approach

- [3] M. Kajala and M. Hämäläinen, "Broadband beamforming optimization for speech enhancement in noisy environments," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, Oct. 1999, pp. 19–22.
- [4] S. Gergen, N. Madhu, and R. Martin, "A new performance criterion for microphone array geometries and filter-and-sum beamforming," in *Proc. ITG Conference on Speech Communication*, Bochum, Germany, Oct. 2010.
- [5] R. Martin, A. Petrovsky, and T. Lotter, "Planar superdirective microphone arrays for speech acquisition in the car," in *Proc. European Conference on Speech Communication and Technology*, Aalborg, Denmark, Sep. 2001, pp. 2623–2626.
- [6] S. Doclo and M. Moonen, "Superdirective beamforming robust against microphone mismatch," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 2, pp. 617–631, Feb. 2007.
- [7] M. I. Skolnik, G. Nemhauser, and J. W. Sherman, "Dynamic programming applied to unequally spaced arrays," *IEEE Transactions on Antennas and Propagation*, vol. 12, no. 1, pp. 35–43, Jan. 1964.
- [8] C. R. Reeves and J. E. Rowe, *Genetic algorithms - principles and perspectives*, chapter 2, pp. 19–60, Kluwer Academic Publishers, Massachusetts, 2003.
- [9] American National Standards Institute, *American National Standard Specification for Octave-Band and Fractional-Octave-Band Analog and Digital Filters*, 1986.