Improving Beamforming for Distant Speech Recognition in Reverberant Environments Using a Genetic Algorithm for Planar Array Synthesis

Hannes Pessentheiner[†], *Gernot Kubin*[†], *and Harald Romsdorfer*[*]

[†]Signal Processing and Speech Communication Laboratory,
Graz University of Technology, Graz, Austria, www.spsc.tugraz.at
[*]Synvo GmbH, Leoben, Austria, www.synvo.com
{hannes.pessentheiner, gernot.kubin}@tugraz.at, romsdorfer@synvo.com

Abstract

Beamforming is crucial for hands-free mobile terminals and voice-enabled automated home environments based on distant-speech interaction to mitigate causes of system degradation, e.g., interfering noise sources or competing speakers. Among the abilities and complexity of a beamformer its performance also depends on the microphone array geometry. This paper highlights a major disadvantage in beamforming—a high sensitivity to reflections from the ceiling and the floor at frequencies above 2000 Hz—by using a horizontal uniform circular array with a center microphone in reverberant environments, e.g., a conference room, and a steering direction close to the array plane. Furthermore, it presents an improved beamformer and a planar array synthesized by a modified genetic algorithm, which exhibits a reduced sensitivity to floor and ceiling reflections and an improved directivity at lower frequencies in combination with a beamformer.

Index Terms: array signal processing, beamforming, planar array, genetic algorithm, non-dominated sorting, multiple objective optimization, crowded tournament selection, convex-optimization based beamformer

1 Introduction

In acoustic array signal processing, beamforming takes advantage of interference to change the directionality of a microphone array (MA). It provides the ability to enhance acoustic signals from a certain direction, the steering direction, and to reduce causes of system degradation, e.g., interfering noise sources, room reverberation, or competing speakers. Beamforming is fundamental for robust voice-enabled automated home environments based on distant speech interaction. The right beamformer (BF) is as important as the right choice of the MA; both choices depend on the operating environment, e.g., a reverberant conference room or lecture hall, rooms of a house, etc. Due to its symmetry, a uniform circular array (UCA) yields a directivity pattern stability for the entire array plane [1]—the array plane is the x-y-plane with $z = 0$. It is a prerequisite for acoustic localization of emitting sources in all azimuthal directions of this plane without suffering from significantly varying deviations in localization accuracy. Both, the uniform linear array (ULA) and the UCA exhibit an up-down–ambiguity, but in comparison to a ULA, a UCA does not exhibit a front-back–ambiguity, which is one of its key benefits. However, the use of a UCA with or without a center microphone introduces a high sensitivity to reflections from the ceiling and the floor above 2000 Hz, and a performance loss in distant speech recognition systems (DSR) (see Fig. 1) in case of reverberant environments. Fig. 2 illustrates this sensitivity with clearly recognizable lobes outside the array plane, where almost all lobes feature a low attenuation. A way to address this problem is the use of a genetic-algorithm–based synthesis [2][3] of planar arrays which employs the three-dimensional main-to-side-lobe ratio (MSR) and 3dB-beamwidth (3dB-BW), and the directivity index (DI) of the three-dimensional directivity pattern (see Fig. 2) as quality measures for the fitness evaluation of the offspring—the evolved MA—generated by the genetic algorithm (GA). The resulting MAs can be applied to, e.g., DSRs illustrated in Fig. 1.

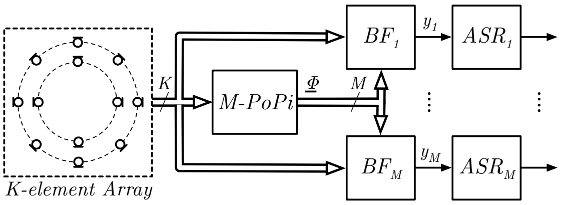

Figure 1: *Our distant speech recognition system (DSR) consists of a K-element planar array (PA) with omni-directional microphones, a Multiband-Position-Pitch algorithm (M-PoPi) for speaker localization, a set of beamformers (BF) and automatic speech recognition systems (ASR) for M detected acoustic sources (see [4] for more details).*

2 The Genetic Algorithm

The GA performs global search in complex optimization problems [2] and is able to solve deadlocked optimization issues where algorithms such as the gradient descent method or Newton's method suffer from missing abilities, e.g., to escape from a local optimum. Evolution, the key part of the GA, yields—depending on its number of generations and its initial population (a set of individuals)—the global optimum or its approximation. In comparison to other optimization algorithms, the GA does not require

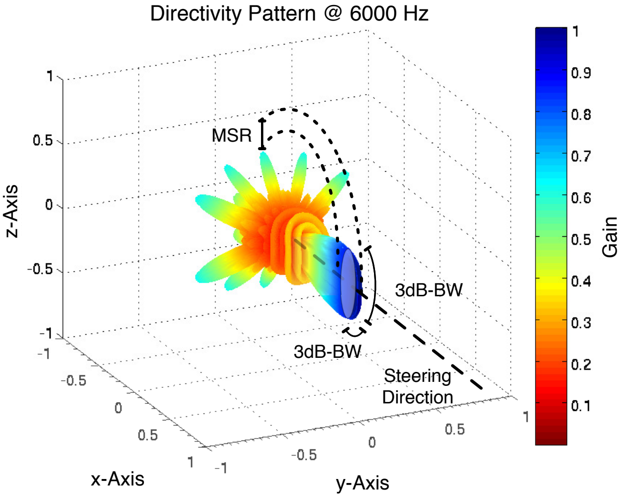

Figure 2: *Three-dimensional directivity pattern of a UCA-delay-and-sum-beamformer–combination with its lobes outside the array plane for $f = 6000$ Hz based on an array diameter of* 0.30 *m and 24+1 omni-directional microphones around the x-y-plane. The additional microphone (+1) corresponds to a center microphone.*

This work was partially funded by the DIRHA European project (FP7-ICT-2011-7-288121) and supported by the K-Project AAP in the context of COMET - Competence Centers for Excellent Technologies by BMVIT, BMWFJ, Styrian Business Promotion Agency (SFG), Province of Styria - Government of Styria ("Abteilung 3: Wissenschaft und Forschung" as well as "Abteilung 14: Wirtschaft und Innovation"). The programme COMET is conducted by the Austrian Research Promotion Agency (FFG).

complex mathematical formulations. The basic GA is based on three simple techniques: selection, crossover, and mutation [5].

Selection depends on an individual's fitness: the better the fitness the higher the probability of being selected for reproduction. A simple implementation of a GA requires a single fitness criterion, i.e., the evaluation of one feature that represents the individual's performance. But selection by means of a single feature may lead to individuals which are optimally adapted to a single task only; they are not flexible enough to cope with other tasks. However, real engineering problems require flexible individuals used for different tasks, e.g., on the one hand the use of an MA-BF–combination—let's assume this as an individual—shall reduce the acoustic presence of competing speakers, but on the other hand additional tasks are to reduce reverb as well as interfering and disturbing reflections from the ceiling and the floor [4]. Thus, it's necessary to evaluate different features of an individual to determine its overall fitness related to more than a single task accomplishment. Its overall fitness has to be improved by, e.g., increasing (maximizing) its ability to reduce noise and/or decreasing (minimizing) the 3dB-BW.

Crossover is the main distinguishing feature of a GA and a prerequisite for generating offspring sharing features from two parents. Let's assume an individual represented by a vector $\mathbf{p}$ consisting of N microphone radii r_i and N azimuthal angles φ_i alternatingly arranged,

$$\mathbf{p} = (r_1, \varphi_1, r_2, \varphi_2, ..., r_N, \varphi_N)^T , \tag{1}$$

where each data pair (r_i, φ_i) describes the position of a microphone in the array plane. The basic GA requires the binary coding of this vector, which yields a bit string representing an individual used to optimize the cost. One or more randomly chosen crossover points divide the original bit string into two or more substrings, which are replaced alternately by substrings from another individual, i.e., an individual swaps its even or odd substrings with another one.

Mutation maintains diversity by inverting randomly selected bits of a bit string and improves the ability to escape from local optima. It ensures the population against permanent fixation at any particular bit [5]. A mutated individual may differ from all other individuals entirely. A high mutation probability yields a primitive random search which reduces the benefits of a GA, whereas a lower one increases them.

Beyond the techniques mentioned above there are several other strategies to improve the performance of a GA, especially in case of MA optimization, e.g., by considering non-dominated sorting, crowding distances, and special elitist replacement strategies [6].

2.1 Multi-Objective Optimization

As mentioned before, many engineering problems are subject to multiple objectives, e.g., minimizing the 3dB-BW and maximizing the MSR and the DI in all three spatial dimensions. There are several approaches to incorporate these measures into the fitness evaluation of an individual. One multi-objective optimization approach is to combine these measures to a composite measure by weighting and summing up their values. But this approach raises the question of whether proper weighting coefficients can be determined in advance. In practice, the selection can be difficult due to missing information of, e.g., limits and non-linear correlation between the measures [7]. Another approach for the fitness evaluation is to alter the weights assigned to each measure [8] or the objectives randomly after each generation [9]. An alternative approach is a rank-based fitness assignment method to find solutions in a Pareto optimal set, i.e., each element of a Pareto optimal set is optimal in the sense that no objective can be improved without making at least one objective worse off. Each solution of a Pareto optimal set is non-dominated with respect to each other [8].

2.2 Non-Dominated Sorting

The non-dominated–sorting algorithm compares the individuals' evaluated objective measures of a population to divide them into several layers within the performance space. The mass of each individual—it represents the overall performance in a certain manner—depends on the values of all considered objective measures and the rank of the layer it belongs to [10]; the measures' values define the shape of all layers in the performance space. The algorithm determines whether an individual is dominated by another one. It finds all members of each non-dominated layer for a rank-assignment, i.e., individuals of the first non-dominated layer are elements of rank one and dominate all other individuals, whereas individuals of the second layer dominate higher ranked individuals, but they are dominated by individuals of the first layer, and so on [6].

2.3 Crowded Tournament Selection

An individual's rank defines the affiliation to a group of individuals, but it does not reveal any details about their differences or similarities among themselves. This missing knowledge may affect the diversity of a population negatively. To maintain diversity, i.e., to prevent getting caught in a local optimum, the selection of individuals in an area of a high population density has to be avoided. Selecting solutions in an area with a high density may lead to a population where an individual resembles all other population members. The crowding distance maintains diversity of non-dominated solutions and tries to select (widely) differing individuals which increases diversity and provides information about the density of solutions surrounding one particular solution. The crowding distance of a special solution is the average distance of its two neighboring solutions. It plays a significant role in the selection and replacement strategy (crowded tournament selection) of a GA; the most crowded areas are most likely to be replaced to maintain diversity [11].

2.4 Symmetry Constraint

It is necessary to achieve a certain directivity pattern stability all over the array plane in order to localize emitting sources in all azimuthal directions of that plane without falling below a certain localization accuracy. A symmetry constraint reduces the variations in directivity for different angles. For instance, optimizing the microphone positions in a restricted domain only, i.e., a sector of a disc where the disc represents the initial domain, and adding rotated copies to the remaining initial domain leads to a reduction of these variations. Without a symmetry constraint, the MA would be designed in a way that the MA-BF–combination only achieves a high performance in a single direction—the steering direction—due to the use of a BF. A change of the steering direction after optimizing the MA will lead to a performance and accuracy loss in beamforming and source localization.

3 Objective Measures

In this paper the 3dB-BW, the MSR, and the DI of the three-dimensional directivity pattern are quality measures for the fitness evaluation of the offspring—the evolved MA—generated by the GA. The 3dB-BW is the angular range of a three-dimensional directivity-pattern's main lobe in which the gain is larger than -3dB. It is expressed in steradians and exhibits a range from 0 to 4π. The MSR is the gain-ratio between the peak of the main lobe and the peak of the side lobe with the lowest attenuation. It is expressed in dB. The DI for a given steering direction is the ratio of the received power from this direction to the received average power from all directions. It is a measure in dB and represents the directivity of a MA. See [12] for an efficient implementation of this measure.

4 Convex-Optimization–based BF

In our experiments, we employed a delay-and-sum (DS) and a convex-optimization–based beamformer (CVX). Due to existing but varying implementations of the CVX, we are going to discuss it in detail. The BF design is a modification of the design mentioned in [4]. It constrains the white noise gain (WNG) to be larger than a lower limit γ. It considers the three-dimensional undistorted signal response from the steering direction (φ_s, θ_s) and null-placements in different directions as constraints. The BF design is based on least squares computations that approximate a desired three-dimensional directivity pattern

$$\hat{b}(f,\varphi,\theta) = \sum_{n=1}^{N} w_n(f) e^{i\frac{2\pi f}{c} r_n \cdot \eta(\varphi,\theta,\varphi_n,\theta_n)} \qquad (2)$$

with

$$\eta(\varphi,\theta,\varphi_n,\theta_n) = \sin(\theta)\sin(\theta_n)\cos(\varphi-\varphi_n)+\cos(\theta)\cos(\theta_n). \qquad (3)$$

Alternatively, this can be expressed in vector notation according to $\hat{\mathbf{B}}(f) = \mathbf{G}(f)\cdot[\mathbf{w}(f)\otimes\mathbf{I}]$, where f represents the frequency, φ and θ are steering direction dependent angles, φ_n and θ_n are the angles of a microphone with index n, N is the number of microphones, c is the sound velocity, r_n is the distance between a microphone and the center of the coordinate system, and $\mathbf{w}(f) = (w_1(f), w_2(f), ..., w_N(f))^T$ is the BF coefficient vector. Moreover, $\mathbf{I}$ is the identity matrix, $\otimes$ denotes the Kronecker product, and $\mathbf{G}(f)$ is an $(N_\theta \times [N \cdot N_\varphi])$ capturing response matrix according to $G_{l,m,n}(f) = e^{i\frac{2\pi f}{c} r_n \cdot \eta(\varphi_m,\theta_l,\varphi_n,\theta_n)}$, where N_φ is the number of discretized azimuthal angles φ_m, and N_θ is the number of discretized elevation angles θ_l. The BF assumes the same desired response for all frequencies, i.e. $\hat{\mathbf{B}}(f) = \hat{\mathbf{B}}$, and

$$\arg\min_{\mathbf{w}(f)} \|\mathbf{G}(f)\cdot[\mathbf{w}(f)\otimes\mathbf{I}]-\hat{\mathbf{B}}\|_F \qquad (4)$$

subjected to the WNG, the undistorted desired signal, and the optional null-placement constraints

$$\frac{|\mathbf{w}^T(f)\mathbf{d}(f)|^2}{\mathbf{w}^H(f)\mathbf{w}(f)} \geq \gamma,\ \ \mathbf{w}^H(f)\mathbf{d}(f) = 1,\ \ \mathbf{w}^H(f)\mathbf{V}(f) = \mathbf{0}, \qquad (5)$$

where $\mathbf{d}(f) = (d_1(f), d_2(f), ..., d_N(f))^T$ represents the capturing model of the steering direction, and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_S]$ is a matrix which consists of vectors $\mathbf{v}(f) = (v_1(f), v_2(f), ..., v_{M-1}(f))^T$ that describe the sound capture model of the competing speakers, S is the number of nulls, $(\cdot)^T$ is the transpose, $(\cdot)^H$ is the Hermitian-transpose, and $\|\cdot\|_F$ is the Frobenius norm.

5 Experimental Framework

Thus far, the basics of the GA, its extensions, and the objective measures considered in our experiments have been addressed. Now, let's proceed with details about the initial MA and our established GA-implementation, which is similar to the fast elitist and binary-coded NSGA-II [6]. The initial MA consists of 24 omni-directional microphones, whereas three microphones are randomly positioned on a sector of a disc with a diameter of 0.30 m on the x-y-plane. The disc consists of 8 sectors, where evolution takes place in only one sector due to the symmetry constraint. All other sectors are rotated copies of the evolution sector. We implemented the DS and CVX without null-steering to determine whether the MA-BF–combination leads to the same results in both cases, i.e., to find out if the microphone positions converge to the same global optimum. For performance evaluation of the MA-BF–combination, we considered the whole disc with its rotated copies of sectors and a center microphone (CM). Our GA-implementation consists of the control parameters shown in Tab. 1 and the following strategies:

(1) generate the initial population
(2) evaluate the objective measures of each individual
(3) execute non-dominated sorting w.r.t. the objective measures
(4) execute tournament selection, crossover, mutation

Parameters	Values
Number of Generations	100
Population Size	300
Tournament Size	2
Bits per coordinate	32
Number of Crossover points	20
Crossover Probability	0.95
Mutation Probability	0.02

Table 1: Details about the settings of our modified GA.

(5) evaluate the objective measures of the new population
(6) execute non-dominated sorting
(7) execute crowded tournament selection and elitist-replacement
(8) repeat steps (4) to (7) until termination criterion is met

6 Results & Discussion

The optimized MA, which consists of 25 microphones, is evaluated and compared with the performance of a UCA, which consists of as many microphones as the optimized MA. Fig. 3a illustrates the optimized array—a first-ranked individual of the last generation. Experiments with both BFs yielded the same geometries with slightly different spacings. For the evaluation, we chose a first-ranked MA which was almost identical in both cases. As appears from the figure, the resulting MA is a concentric circular array (CCA) which consists of a UCA with $r_1 = 0.0717m$ and a circular array with $r_2 = 0.1492m$ and pairwise equidistant microphone spacings. For more information about the microphone spacings of the optimized array, please refer to Fig. 3a. As one can see in the figure, the absolute angular distances $\Delta\phi_{21} = |\phi_2 - \phi_1|$ and $\Delta\phi_{32} = |\phi_3 - \phi_2|$ are different, which results in an asymmetric array. We expect identical distances by increasing the number of generations. Experiments with a higher number of generations are in progress. The evaluation of the 3dB-BW does not exhibit any noteworthy improvements or degradations w.r.t. the increase of the main lobe width. However, there are significant improvements of the MSR at lower frequencies (up to 4700 Hz in case of the CVX and 3150 Hz in case of the DS) and higher frequencies (starting around 4550 Hz), but also small degradations at frequencies between 3150-4550 Hz (see Fig. 3b). Despite these degradations, the overall performance in terms of the MSR increased in case of the DS and CVX. Graphical evaluations of the three-dimensional directivity patterns revealed a decrease in size of lobes outside the array plane and an increase in attenuation of at least 3.194 dB (compare Fig. 2 with Fig. 4). Thus, we accomplished the goal of our experiment: an increased attenuation of reflections from the ceiling and the floor. And besides that, we increased the MSR at lower and higher frequencies, whereas the increase at higher frequencies results from the increase in attenuation outside the array plane. As shown in Fig. 3c, the use of a DS in combination with the optimized MA results in a higher directivity factor due to an overall increase in directivity between 1000-3350 Hz, whereas the CVX yields a smaller directivity factor due to an overall decrease in directivity between 1600-4800 Hz.

7 Conclusion

Our experiments show that beamforming can be improved by using a genetic algorithm based on multi-objective optimization, non-dominated sorting, crowding distances, and elitist replacement. We were able to improve the directional behaviour outside the array plane without suffering from significant degradations. Graphical evaluations of the three-dimensional directivity patterns revealed a decrease in the size of lobes outside the array plane and, thus, an increase in attenuating reflections from the ceiling and the floor by at least 3.194 dB. Evolution with both beamformers yielded similar array geometries. To compare the BFs' real-world behaviour with different planar array geome-

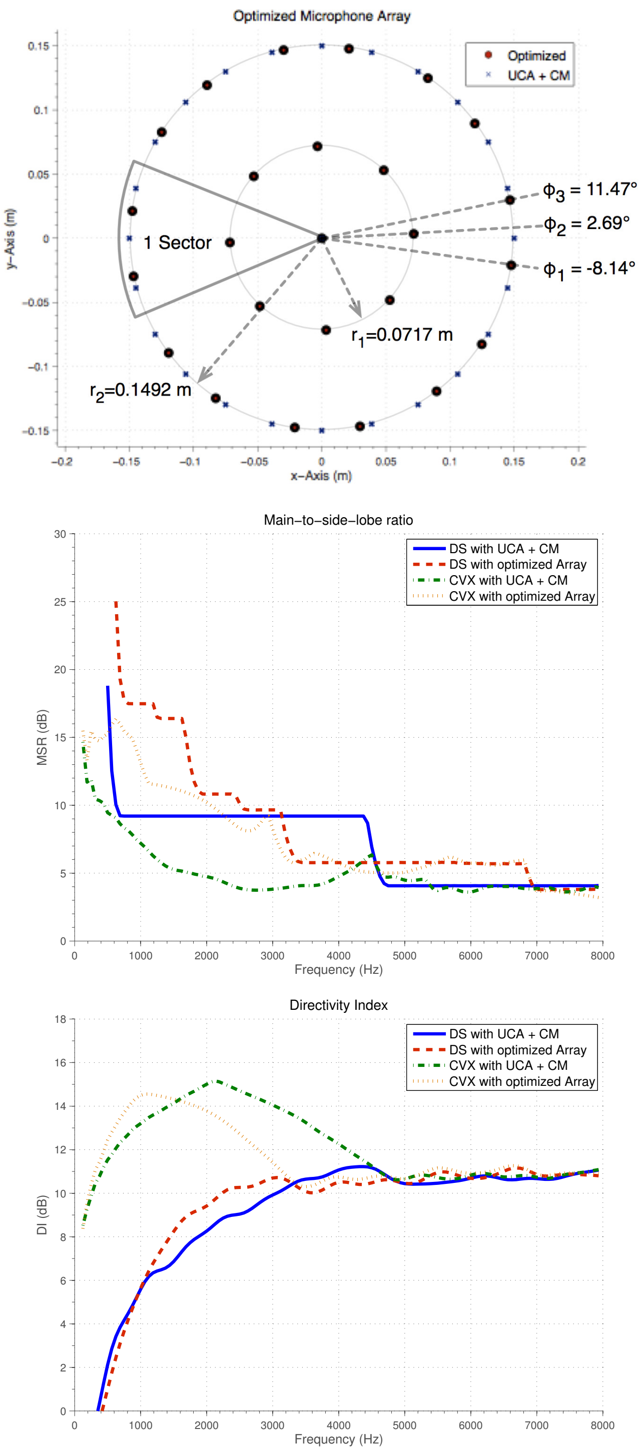

Figure 3: *This figure illustrates (a) the evolved microphone array and its corresponding (b) main-to-side-lobe ratios and (c) directivity indices.*

tries, we are going to perform experiments using our DSR [4] (see Fig. 1) in real reverberant environments, e.g., an office or cocktail party room. Further experiments with our DSR, new evolved array structures and beamformers, e.g., the proposed CVX, and experiments with a word recognizer in reverberant environments are in progress and will be reported.

References

[1] Wax, M. and Sheinvald, J., “Direction finding of coherent signals via spatial smoothing for uniform circular arrays,” IEEE Trans. on Antennas and Propagation, 42(5):613–620, 1994.

[2] Man, K. F., Tang, K. S., and Kwong, S., “Genetic Algorithms: Concepts and Applications,” IEEE Trans. on Industrial Electronics, 43(5):519-534, October 1996.

[3] Yu, J. and Yu, F., “Evolutionary Algorithm for Microphone Array Optimization,” Applied Mechanics and Materials (Volumes 143-144), Electrical Information and Mechatronics and Applications, pp. 287-292. December 2011.

[4] Pessentheiner, H., Petrik, S., and Romsdorfer, H., “Beamforming Using Uniform Circular Arrays for Distant Speech Recognition in Reverberant Environments and Double-Talk Scenarios,” Interspeech 2012, Portland, Oregon, September 2012.

[5] Mitchell, M., “An Introduction to Genetic Algorithms,” 5th Edition (1999), The MIT Press, February 1998.

[6] Deb, K. et al., “A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II,” IEEE Trans. on Evolutionary Computation, 6(2):182-197, April 2002.

[7] Konak, A., Coit, D. W., and Smith, A. E., “Multi-objective optimization using genetic algorithms: A tutorial,” Reliability Engineering and System Safety, 91(9):002-1007, January 2006.

[8] Fonseca, C. M. and Fleming, P. J., “Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion, and Generalization,” Genetic Algorithms: Proc. of the 5th International Conference, pp. 416-423, San Mateo, CA: Morgan Kaufmann, July 1993.

[9] Ishibuchi, H. and Murata, T., “A multi-objective genetic local search algorithm and its application to flowshop scheduling,” IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 28(3):392-403, August 1998.

[10] Nobahari, H., Nikusokhan, M., and Siarry, P., “Non-dominated Sorting Gravitational Search Algorithm,” Proc. of the 2011 International Conference on Swarm Intelligence, ICSI 2011, pp. 1-10, Chongqing, China, June 2011.

[11] Raquel, C. R., and Naval, C. P. Jr., “An Effective Use of Crowding Distance in Multiobjective Particle Swarm Optimization,” Proceedings of the 2005 Conference on Genetic and Evolutionary Computation, pp. 257-264, Washington DC, USA, June 2005.

[12] Nuttall, A. H. and Cray, B. C., “Efficient Calculation of Directivity Indices for Certain Three-Dimensional Arrays,” NUWC-NPT Technical Report 11, Naval Undersea Warfare Center Division, Newport, Rhode Island, USA, July 1996.

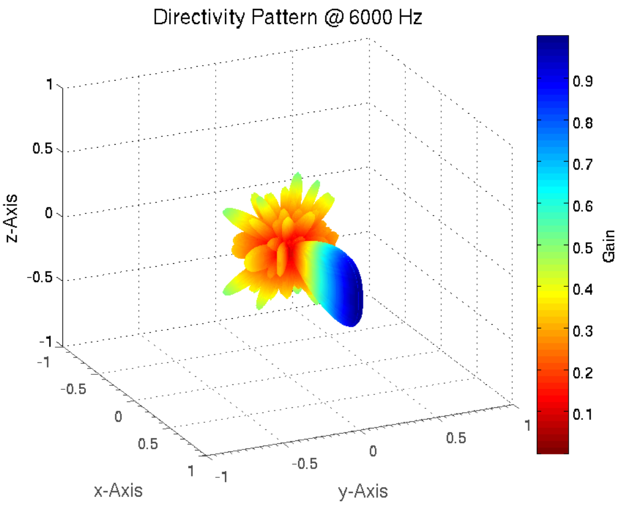

Figure 4: *Three-dimensional directivity pattern of an optimized UCA-DS–combination with its lobes outside the array plane for* $f = 6000$ *Hz based on an max. array diameter of* 0.30 *m and 25 omni-directional microphones.*