NUMERICAL NEAR FIELD OPTIMIZATION OF A NON-UNIFORM SUB-BAND FILTER-AND-SUM BEAMFORMER

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

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

ABSTRACT

A novel near field filter-and-sum beamformer using non uniform frequency sub-bands is presented. The concept is based on numerical optimization of the reception characteristic of the microphone array. In order to improve the reception characteristic over frequency and space, a non uniform filterbank is utilized to subdivide the frequency range. Individual optimization processes for each sub-band result in a clearly improved reception characteristic. The new system is able to closely approximate a target (independently of the frequency) which can be defined according to the application.

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

1. INTRODUCTION

In audio-visual communication systems such as video conferencing the desired speech signal may be degraded by non-stationary (e.g., competing speakers) and stationary interfering signals. In order to improve the listening comfort and to ensure high audio quality, speech enhancement techniques are required. Single microphone algorithms might show satisfactory results for stationary noise reduction, e.g., [1, 2], but have limited performance in complex scenarios (e.g., when suppressing competing speakers). Since the desired and interfering signals are usually spatially separated, speech enhancement techniques utilizing multichannel microphone arrays, such as beamformers, are appropriate. They can amplify a target speaker efficiently while simultaneously damping other speakers and background noise. When designing microphone arrays, a specific spatial reception characteristic is usually the desired target.

Beamformer realizations can be classified into fixed and adaptive [3, 4]. Fixed beamforming algorithms are independent of the input data. They can realize robust directional gains with moderate numerical complexity. Typical representatives are the (weighted) sum-and-delay as well as the filter-and-sum beamformer.

Adaptive beamforming algorithms are well suited for cancelling moving interferers. Among various adaptive beamforming groups, the minimum variance distortionless response (MVDR), the multichannel Wiener filter (MWF) and the linearly constrained minimum variance (LCMV) are the most common [5, 6].

Furthermore beamformers can be realized operating in the time domain or on several sub-bands, e.g., [7, 8, 9, 10]. There are several advantages using a sub-band beamformer compared to a conventional full-band beamformer such as an overall shorter filter degree or an improved reception characteristic with respect to the operational frequency.

1.1. Relation to prior work

For the far field case, i.e., if the distance to the array is significantly larger than its geometric setup, many beamformer design methods are known, e.g., [11, 12]. There are also procedures known, aiming specifically at the near field, where the far field assumption provides only an approximation in the best case, e.g., [12, 13, 14, 15]. However, these approaches optimize the reception characteristic limited by several design constraints, e.g., only on a (semi-)circular arc at one specific distance from the array. To circumvent this limitations, we extend our recently proposed method [16] in this contribution which allows to optimize the reception characteristic for an entire area in the near field of the microphone array simultaneously for different distances and angles.

The work in [16] considered a weighted delay-and-sum array with full-band processing as basis for the optimization. In [17] we compared two recent different approaches for designing microphone array beamformers, a numerical near field optimized beamformer and a linearly constrained minimum variance beamformer, using real audio recordings.

In this contribution we extend the concept for the numerical near field optimization of a reception characteristic to the class of sub-band filter-and-sum beamformers.

1.2. Organization of the paper

The remainder of this paper is organized as follows. In Sec. 2, a brief overview of the proposed system is given. Sec. 3 describes the numerical optimization procedure based on the target reception characteristic. Experimental results are shown in Sec. 4 and conclusions are drawn in Sec. 5.

2. SYSTEM OVERVIEW

Since the target reception characteristic of a microphone array is frequency dependent our delay-and-sum beamformer approach [16] is extended by a non uniform filterbank [18] which subdivides each microphone signal into N frequency sub-bands. The optimization of the filter-and-sum units is carried out in these frequency bands to gain more degrees of freedom for the optimization procedure.

A simplified block diagram of the proposed microphone array system is depicted in Fig. 1. It consists of a filterbank with N sub-bands followed by different filter-and-sum units represented by the impulse responses $\mathbf{h}_n^m$ $m \in \{1, \dots, M\}$, $n \in \{1, \dots, N\}$ with n denoting the sub-band index and m the microphone index at all

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).

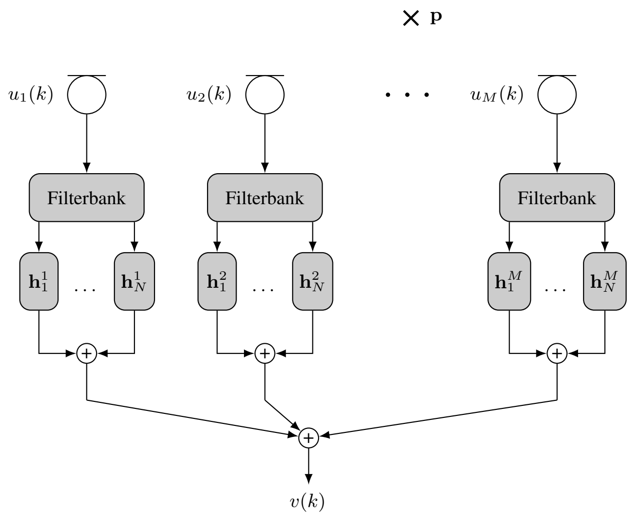

Fig. 1. Modified filter-and-sum beamformer with M Microphones and N non-uniform sub-bands

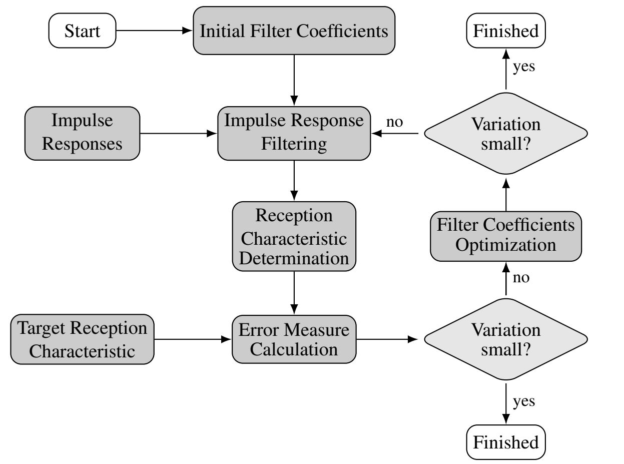

Fig. 2. Block diagram of the optimization process for each frequency sub-band

M microphones. The samples $u_m(k)$ are obtained by analog-digital conversion with sampling frequency of $f_s = 48$ kHz, where k is the discrete time index.

A point source $s(k)$ is assumed to be at position $\mathbf{p}$ on an appropriately chosen spatial grid (e.g., in a two-dimensional Cartesian coordinate system: $\mathbf{p} = (x\ y)^T$). With the impulse responses $h_{\mathbf{p}m}(k)$ from the point source to each microphone the microphone signal $u_m(k)$ can be expressed as:

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

The output $v(k)$ depends on the source location $\mathbf{p}$ and can be calculated according to:

$$v(k) = \sum_{m=1}^{M} \sum_{n=1}^{N} h_n^m(k) * \left(h_n^{\mathrm{FB}}(k) * u_m(k) \right), \tag{2}$$

where $h_n^{\mathrm{FB}}(k)$ represents the filterbank and $h_n^m(k)$ the FIR sub-band filters of length L.

3. NUMERICAL OPTIMIZATION

The basic concept of the numerical optimization process for each sub-band is depicted in Fig. 2. The procedure consists of an iterative minimization of an error measure, which is the summed level difference between the predefined target reception characteristic and the simulated one, based on the current state of the filter coefficients.

Since the optimization works in an identical manner for simulated and measured impulse responses it is very flexible with respect to different practical application scenarios.

3.1. Determination of the Reception Characteristic in the Near Field

The reception characteristic can be calculated by the following three steps:

- simulating or measuring impulse responses between points on an appropriately chosen spatial grid in the near field and all microphones,
- processing these impulse responses with the modified filter-and-sum beamformer (see Fig. 1) 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.

Since the output signal $v(k)$ for each source location $\mathbf{p}$ can be expressed as a filtered version of the source signal:

$$v(k) = \sum_{m=1}^{M} \sum_{n=1}^{N} h_n^m(k) * h_n^{\mathrm{FB}}(k) * h_{\mathbf{p}m}(k) * s(k), \tag{3}$$

the overall filter $g_{\mathbf{p}}(k)$ can be calculated as:

$$g_{\mathbf{p}}(k) = \sum_{m=1}^{M} \sum_{n=1}^{N} h_n^m(k) * h_n^{\mathrm{FB}}(k) * h_{\mathbf{p}m}(k). \tag{4}$$

The frequency transform of the overall filter $g_{\mathbf{p}}(k)$ results in:

$$G_{\mathbf{p}}(f) = \mathcal{F}\left\{ g_{\mathbf{p}}(k) \right\} . \tag{5}$$

Therefore 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:

$$S_{\mathbf{p}}(f) = 20 \cdot \log_{10} \left| G_{\mathbf{p}}(f) \right| . \tag{6}$$

3.2. Definition of the Target

The target reception characteristic $\hat{S}_{\mathbf{p}}(f)$ is defined as a spatial distribution of areas with defined amplification or damping in front of the microphone array. It can be defined individually for all frequencies but a frequency-independent target is suitable for many applications. This corresponds to a given SNR for the received signal as the target speaker should be in the amplified region $\mathbb{P}_{\mathrm{high}}$ (target level S_{high}) while all interfering signals are assumed to be in the damped area $\mathbb{P}_{\mathrm{low}}$ (target level S_{low}).

$$\hat{S}_{\mathbf{p}}(f) = \hat{S}_{\mathbf{p}} = \begin{cases} S_{\mathrm{high}} & \text{for } \mathbf{p} \in \mathbb{P}_{\mathrm{high}} \\ S_{\mathrm{low}} & \text{for } \mathbf{p} \in \mathbb{P}_{\mathrm{low}} \end{cases} \tag{7}$$

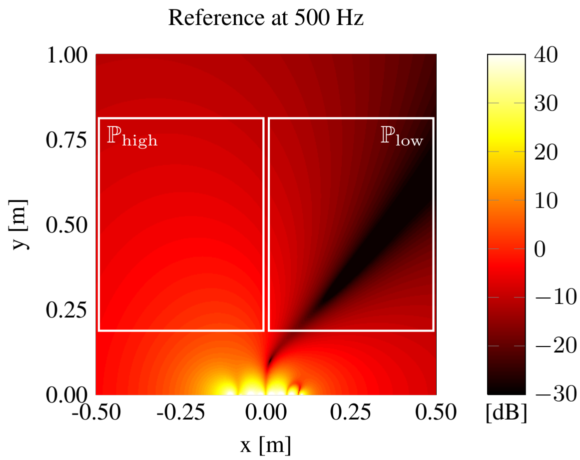

Fig. 3. Reception characteristic of the delay-and-sum microphone array with optimized weighting at 500 Hz [16]

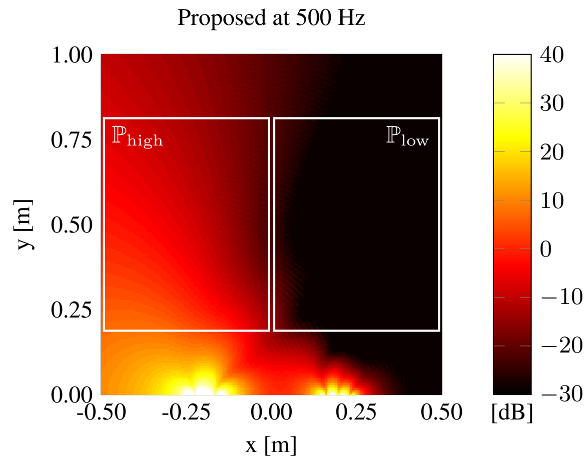

Fig. 4. Reception characteristic of the proposed microphone array with optimized filters at 500 Hz

The precise choice of the areas and levels depends on *a priori* knowledge from the application, e.g., a conferencing scenario, where the target speaker should be amplified while all interfering sources shall be attenuated.

3.3. Error Function

The target of the optimization procedure is to minimize an error measure, which is the summed level difference Δ_S between the predefined target $\hat{S}$ and the simulated reception characteristic S. For all points where $\hat{S}_{\mathbf{p}}(f)$ is defined according to Eq. 7 and over all frequencies f_i $(i \in \{i_{\min}, \ldots, i_{\max}\})$ for which the reception characteristic shall be optimized the level difference is summed:

$$\Delta_S(n) = \sum_{i=i_{\min}}^{i_{\max}} \sum_{\mathbf{p}\in\left(\mathbb{P}_{\text{high}}\cup\mathbb{P}_{\text{low}}\right)} \hat{S}_{\mathbf{p}}(f_i) - S_{\mathbf{p}}(f_i), \tag{8}$$

where $f_{i_{min}}$ and $f_{i_{max}}$ denote the lower and upper edge frequencies of sub-band n.

3.4. Optimization Procedure

The optimum filter coefficients for each sub-band n are determined based on this summed level difference in a minimum mean square error (MMSE) sense by:

$$\left[\mathbf{h}_n^1, \ldots, \mathbf{h}_n^M\right]_{\text{opt}} = \arg\min_{\mathbf{h}} \Delta_S(n)^2 \,. \tag{9}$$

The optimization is carried out by an iterative interior-point algorithm [19] with the constraint that the filter coefficients have to be in the range of -1 and 1. This constraint only limits the maximum amplification that is achievable by the array itself and does not change the relation between the filter coefficients. A subsequent scaling of the output $v(k)$ can be applied to map the reception characteristic to a desired level.

4. RESULTS

The proposed microphone array system is compared with the optimized near field weighted delay-and-sum beamformer [16]. The optimization scheme also allows to optimize the entire reception characteristic in the vicinity of a microphone array at once. The reception characteristics shown here are calculated for a free field setup to allow a better comparison of both algorithms.

For both setups, a level difference of 40 dB was chosen between the amplified and the damped area. The microphone arrays are designed to amplify sources on the left ($-0.5\,\text{m} \le x < 0\,\text{m} \wedge 0.2\,\text{m} < y \le 0.8\,\text{m}$) while damping sources on the right ($0\,\text{m} < x \le 0.5\,\text{m} \wedge 0.2\,\text{m} < y \le 0.8\,\text{m}$). For both dimensions ($x$ and y), the density of the spatial grid is set to $0.01\,\text{m}$ leading to 3000 points in $\mathbb{P}_{\text{high}}$ and $\mathbb{P}_{\text{low}}$, respectively.

The microphone array consists of $M = 8$ sensors with a spacing of 3 cm between the sensors and is centered at the origin of the coordinate system. Spatial aliasing can be expected for frequencies greater than approx. 5600 Hz. For the proposed system the number of sub-bands was set to $N = 6$ using a non uniform filter bank [18]. The frequency range of each sub-band can be seen in Table 1. For simplicity all sub-band filters have been realized as FIR filter. The degree of the filter-and-sum units $\mathbf{h}_n^m$ was set to $L = 8$. A comparison of the reception characteristics is given for two different operating frequencies:

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

Band	Frequency range [Hz]		Band	Frequency range [Hz]	
1	1	268	4	1549	2614
2	268	839	5	2614	4731
3	839	1549	6	4731	12049

Table 1. Filterbank sub-bands

In Fig. 3 the performance based on optimized weighting according to [16] for the 500 Hz case is shown. A noticeable level difference between the right and left side can be seen. However, the target for the damped area is only partly achieved. The reception characteristic for the proposed system (see Fig. 4) fits the previously

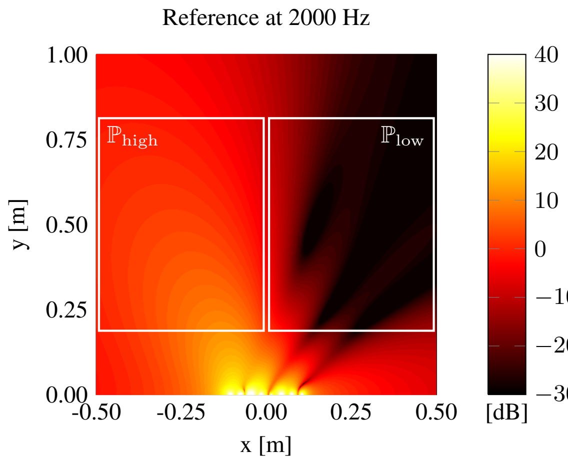

Fig. 6. Reception characteristic of the delay-and-sum microphone array with optimized weighting at 2000 Hz [16]

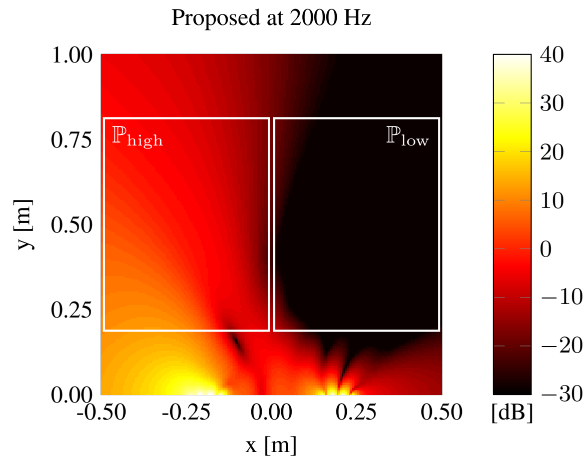

Fig. 7. Reception characteristic of the proposed microphone array with optimized filters at 2000 Hz

defined areas of damping and amplification very well even at this low frequency.

Comparing the performance for the 2000 Hz case (Fig. 6 and 7) the difference of the reception characteristics is smaller. Both algorithms provide a significant level difference between $\mathbb{P}_{\text{high}}$ and $\mathbb{P}_{\text{low}}$. But especially in the critical border region at $x = 0$ m the new beamformer matches the predefined target reception characteristic even better.

As an additional performance example the proposed algorithm was also evaluated using measured impulse responses for the optimization procedure. Therefore the Speech & Acoustic Lab of the Faculty of Engineering at Bar-Ilan University, with controllable reverberation time, was utilized to create an audio-database using an 8-channel microphone array.[1]

According to a typical scenario, e.g., a video conference, the reverberation time was set to 160 ms. The density of the spatial grid (in polar coordinates) for this setup is given by angles from $0°$ to $180°$ in $15°$ steps for 1 m and 2 m radii. $\mathbb{P}_{\text{low}}$ in polar coordinates maps to all radii for angles from $0°$ to $90°$, $\mathbb{P}_{\text{high}}$ from $91°$ to $180°$, respectively. The remaining setup was not changed. Fig. 5 depicts the transfer function for the center of regions $\mathbb{P}_{\text{high}}$ and $\mathbb{P}_{\text{low}}$. A significant level difference (in average approx. 14.5 dB) can be seen over the complete frequency range up to the spatial alias frequency of approx. 5600 Hz. This also confirms the performance of the proposed system independently of the operational frequency.

[1] The audio-database includes impulse responses as well as audio recordings (speech and noise). The measurements were taken for different microphone setups and different reverberation times. The database will be made available to the research community in the near future.

5. CONCLUSION

A novel filter-and-sum beamformer based on numerical near field optimization was presented. The beamformer consists of a non uniform filterbank with FIR filters in the sub-bands which are optimized to achieve improved beamforming. The proposed approach combines the advantages of decoupled sub-band filter optimization with a frequency resolution according to the auditory system. The optimization scheme allows to realize a predefined reception characteristic which can be freely chosen according to the application. The proposed system provides a distinct spatial separation independently of the frequency. Switching between different reception characteristics, e.g., for speaker selection in a conference scenario, can be easily achieved by several pre-computed filter sets.

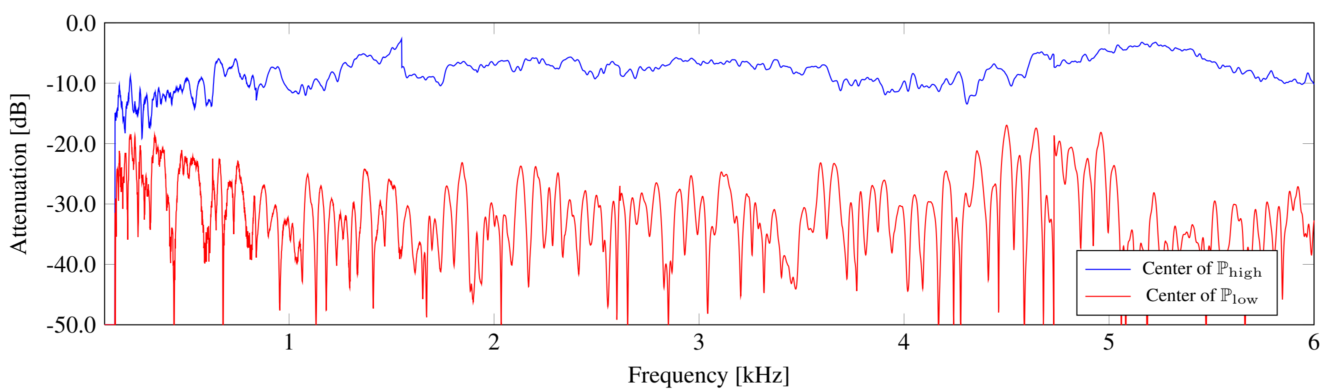

Fig. 5. Transfer function of the microphone array with optimized filters for the center of regions $\mathbb{P}_{\text{high}}$ and $\mathbb{P}_{\text{low}}$

6. REFERENCES

[1] P. Vary and R. Martin, *Digital Speech Transmission - Enhancement, Coding & Error Concealment*. John Wiley & Sons, Ltd., Jan. 2006.

[2] Y. Ephraim and D. Malah, “Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator,” vol. 32, no. 6, pp. 1109–1121, 1984.

[3] S. Haykin and K. Liu, *Handbook on Array Processing and Sensor Networks*. Wiley-IEEE Press, 2010, vol. 60.

[4] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*. Springer Verlag, 2001.

[5] B. D. Van Veen and K. M. Buckley, “Beamforming: A Versatile Approach to Spatial Filtering,” vol. 5, no. 2, pp. 4–24, Apr. 1988.

[6] S. Markovich, S. Gannot, and I. Cohen, “Multichannel Eigenspace Beamforming in a Reverberant Environment with Multiple Interfering Speech Signals,” vol. 17, no. 6, pp. 1071–1086, Aug. 2009.

[7] Y. Zhao, W. Liu, and R. Langley, “Subband design of fixed wideband beamformers based on the least squares approach,” *Signal Processing*, vol. 91, no. 4, pp. 1060–1065, 2011.

[8] F. Lorenzelli, A. Wang, and K. Yao, “Broadband array processing using subband techniques,” in *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*, vol. 5. IEEE, 1996, pp. 2876–2879.

[9] J. De Haan, N. Grbic, I. Claesson, and S. Nordholm, “Design of oversampled uniform dft filter banks with delay specification using quadratic optimization,” in *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP’01). 2001 IEEE International Conference on*, vol. 6. IEEE, 2001, pp. 3633–3636.

[10] S. Nordholm, S. Low, I. Claesson, and K. Yiu, “Non-uniform optimal subband beamforming: An evaluation on real acoustic measurements,” in *Image and Signal Processing, 2008. CISP’08. Congress on*, vol. 5. IEEE, 2008, pp. 747–751.

[11] D. Ward, R. Kennedy, and R. Williamson, “Theory and Design of Broadband Sensor Arrays with Frequency Invariant Far-field Beam Patterns,” *The Journal of the Acoustical Society of America*, vol. 97, p. 1023, 1995.

[12] S. Doclo and M. Moonen, “Design of Far-field and Near-field Broadband Beamformers using Eigenfilters,” *Signal Processing*, vol. 83, no. 12, pp. 2641–2673, 2003.

[13] R. A. Kennedy, T. Abhayapala, D. B. Ward, and R. C. Williamson, “Nearfield Broadband Frequency Invariant Beamforming,” in *IEEE ICASSP*, 1996.

[14] J. G. Ryan and R. 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.

[15] E. Fisher and B. 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.

[16] M. Schäfer, F. Heese, J. Wernerus, and P. Vary, “Numerical Near Field Optimization of Weighted Delay-and-Sum Microphone Arrays.” IWAENC, Sep. 2012.

[17] F. Heese, M. Schäfer, P. Vary, E. Hadad, S. M. Golan, and S. Gannot, “Comparison of Supervised and Semi-supervised Beamformers Using Real Audio Recordings,” in *Proceedings of IEEE 27-th Convention of Electrical and Electronics Engineers in Israel (IEEEI)*. IEEE, Nov. 2012.

[18] H. W. Löllmann, “Allpass-Based Analysis-Synthesis Filter-Banks: Design and Application,” Dissertation, IND, RWTH Aachen, Nov. 2011. [Online]. Available: http://darwin.bth.rwth-aachen.de/opus3/volltexte/2011/3843/pdf/3843.pdf

[19] 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.