Robust Superdirectional Beamforming for Hands-Free Speech Capture in Cars

Edwin Mabande, Adrian Schad, Walter Kellermann
University of Erlangen-Nuremberg, Germany
{emabande,schad,wk}@LNT.de

Abstract

The application of multichannel signal processing algorithms in front-ends of hands-free speech communication systems facilitates a better extraction of a desired source signal and allows for a suppression of unwanted noise and interference in car interiors. In this paper, a novel superdirective beamforming design method called the Robust Least-Squares Frequency-Invariant Beamformer (RLSFIB) is investigated for microphone arrays in cars. The RLSFIB design offers superior spatial selectivity at low frequencies even for small array apertures and some variability in microphone positioning. To cope with the high sensitivity inherent to superdirectional beamformers, this method directly incorporates a white noise gain (WNG) constraint for controlling the beamformer sensitivity. The resulting problem is formulated as a convex optimization problem which can be solved by conventional iterative solvers and the convexity guarantees a globally optimal solution. As the WNG constraint can be chosen freely, the method can match desired robustness levels, and therefore can be adapted to given prior knowledge on microphone mismatch, positioning errors, and microphone self-noise. Design examples for typical car requirements will be presented.

Introduction

Interest in the application of multichannel signal processing algorithms in the front-ends of hands-free systems for in-car use has increased in recent years. This interest has been driven primarily by the enactment of regulations which aim to enhance road safety and the demand for convenience. Multichannel signal processing algorithms facilitate a better extraction of a desired speech signal and suppression of unwanted interference signals such as competing speakers and background noise. Beamforming is a class of such multichannel signal processing algorithms. Applications include in-car communication [1], automatic dialog systems for hands-free telephony as well as for more advanced use cases such as interactive navigation systems.

Data-independent superdirective beamformers can be used to extract the desired speech signal originating from a chosen passenger within a car while suppressing speech from all other passengers and background noise. Although superdirective beamformers are often desirable because they provide high directivity, a well-known problem inherent in these designs is the high sensitivity to spatially white noise, to the mismatch between microphone characteristics, and to microphone positioning errors [2]. This severely limits their application in practice. Since the WNG is a measure of the robustness of a beamformer, a robust superdirectional beamforming design may be obtained by constraining the WNG [2].

In this paper the RLSFIB (see also [3]), which is a data-independent frequency-invariant beamforming design, is investigated for microphone arrays in cars. The RLSFIB design directly constrains the WNG to lie above a given lower limit and thereby ensuring a desired level of robustness. This is accomplished by solving a constrained optimization problem which is convex. For simplification of the following treatment, we use the common assumptions that waves propagate in a free field, that the sources are in the farfield relative to the array, and that all microphones are omnidirectional.

Beamforming

Figure 1 depicts a beamformer with a linear array consisting of N microphones positioned at $\mathbf{p}_n$, $n = 0,\dots,N-1$ in space. The microphone signals are first discretized by the analog-to-digital converter (ADC) before being processed by the beamforming filters.

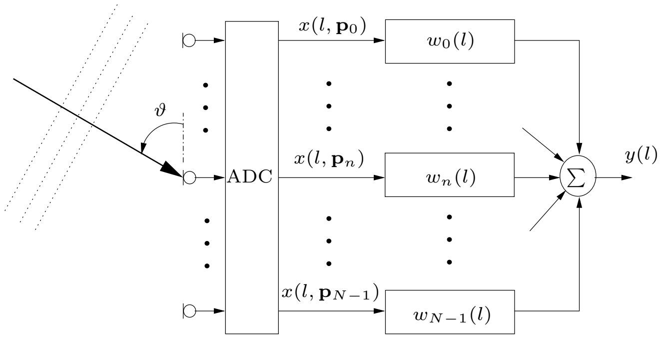

Figure 1: Beamforming with a linear microphone array

A beamformer is characterized here by the beamformer response which describes the wavefield in the farfield produced for a given complex harmonic signal with frequency ω as parameter. The response of the filter-and-sum beamformer depicted in Figure 1 is given by [4]

$$B(\omega,\vartheta) = \mathbf{w}_\mathrm{f}^T(\omega)\mathbf{d}(\omega,\vartheta), \qquad (1)$$

where $\mathbf{d}(\omega,\vartheta) = [\exp(-j\omega\tau_0(\vartheta)),\dots,\exp(-j\omega\tau_{N-1}(\vartheta))]^T$, $\mathbf{w}_\mathrm{f}(\omega) = [W_0(\omega),\dots,W_{N-1}(\omega)]^T$, $W_n(\omega) = \sum_{l=0}^{L-1} w_n(l)\exp(-jl\omega)$ are the frequency responses of the filters $w_n(l)$ and L is the FIR filter length. The time delays due to propagation are given by $\tau_n(\vartheta) = d_n \cos\vartheta/c$, where d_n is the distance of the n-th microphone to the center of the array and ϑ is the angle of arrival relative to the array axis. The magnitude square of the beamformer response is known as the beampattern of the beamformer [5].

The WNG is given by [6]

$$A(\omega) = \frac{|\mathbf{w}_\mathrm{f}^T(\omega)\mathbf{d}(\omega,\vartheta_\mathrm{o})|^2}{\mathbf{w}_\mathrm{f}^H(\omega)\mathbf{w}_\mathrm{f}(\omega)}, \quad (2)$$

where

$$\mathbf{d}(\omega,\vartheta_\mathrm{o}) = [\exp(-j\omega\tau_0(\vartheta_\mathrm{o})),\dots,\exp(-j\omega\tau_{N-1}(\vartheta_\mathrm{o}))]^T$$

denotes the so-called steering vector and the angle ϑ_o denotes the desired look direction. The WNG expresses the gain of the beamformer for the desired signal from the desired look direction relative to the amplification of spatially white noise. Therefore $A(\omega) < 1$ effectively corresponds to an amplification of spatially white noise at frequency ω.

The logarithmic array gain [6], termed the directivity index, of the delay-and-sum beamformer (DSB) reaches its maximum of $10\log N$ if the sensor spacing is $\lambda/2$. Beamforming designs that lead to a directivity index greater than $10\log N$ are called superdirective. The WNG for superdirective beamformers is typically very small, e.g. $A(\omega) < 10^{-3}$, at low frequencies [2]. Consequently these beamformers are highly sensitive to small errors in the array characteristics and they amplify spatially white noise such as microphone self-noise.

Robust Least-Squares Frequency-Invariant Beamformer (RLSFIB)

The idea behind the RLSFIB design is to optimally approximate a desired frequency-independent response, $\hat{B}(\vartheta)$, by $B(\omega,\vartheta)$ in the Least Squares (LS) sense. Typically, a numerical solution is obtained by discretizing the frequency range into P frequencies ω_p, $p = 0,\dots,P-1$ and the angular range into M angles ϑ_m, $m = 0,\dots,M-1$, and solving the resulting set of linear equations numerically. The beamformer design problem then reads:

$$\hat{\mathbf{b}} \stackrel{!}{=} \mathbf{G}(\omega_p)\mathbf{w}_\mathrm{f}(\omega_p),$$

where $\hat{\mathbf{b}} = [\hat{B}(\vartheta_0),\dots,\hat{B}(\vartheta_{M-1})]^T$, $\mathbf{G}(\omega_p) = [\mathbf{d}(\omega_p,\vartheta_0),\dots,\mathbf{d}(\omega_p,\vartheta_{\text{M-1}})]^T$ and $\mathbf{w}_\mathrm{f}(\omega_p) = [W_0(\omega_p),\dots,W_{N-1}(\omega_p)]^T$. Since the number of discretized angles is typically greater than the number of sensors, $M > N$, the problem is therefore overdetermined. The least-squares solution to this problem, which gives the smallest quadratic error by definition, is given by:

$$\min_{\mathbf{w}_\mathrm{f}(\omega_p)} \|\mathbf{G}(\omega_p)\mathbf{w}_\mathrm{f}(\omega_p) - \hat{\mathbf{b}}\|_2^2 \quad (3)$$

In order to ensure that the desired signal from a given angle ϑ_o remains undistorted, the linear constraint

$$\mathbf{w}_\mathrm{f}^T(\omega_p)\mathbf{d}(\omega_p,\vartheta_\mathrm{o}) = 1, \quad (4)$$

must be satisfied. A WNG constraint is incorporated into the design by adding the following quadratic constraint

$$\frac{|\mathbf{w}_\mathrm{f}^T(\omega_p)\mathbf{d}(\omega_p,\vartheta_\mathrm{o})|^2}{\mathbf{w}_\mathrm{f}^H(\omega_p)\mathbf{w}_\mathrm{f}(\omega_p)} \geq \gamma > 0, \quad (5)$$

where γ is the lower bound for the WNG. It is a user-defined parameter which allows direct control of the robustness of the beamforming design. Thus, a robust beamforming design may be ensured by combining (3), (4) and (5) resulting in

$$\min_{\mathbf{w}_\mathrm{f}(\omega_p)} \|\mathbf{G}(\omega_p)\mathbf{w}_\mathrm{f}(\omega_p) - \hat{\mathbf{b}}\|_2^2,$$

subject to

$$\frac{|\mathbf{w}_\mathrm{f}^T(\omega_p)\mathbf{d}(\omega_p,\vartheta_\mathrm{o})|^2}{\mathbf{w}_\mathrm{f}^H(\omega_p)\mathbf{w}_\mathrm{f}(\omega_p)} \geq \gamma, \quad \mathbf{w}_\mathrm{f}^T(\omega_p)\mathbf{d}(\omega_p,\vartheta_\mathrm{o}) = 1, \quad (6)$$

which is a convex problem that is solved for each frequency ω_p. This problem is a convex problem for the following reasons. The unconstrained least-squares problem (3) is a convex function [7]. The sets described by (4) and (5) are convex because their elements lie in a hyperplane and a Euclidean ball, respectively [3]. Since convexity is preserved under intersection, the constrained problem is therefore convex because we minimize a convex function over two convex sets [7].

There is no known analytic solution to (6) and therefore constrained optimization is used. Sequential Quadratic Programming (SQP) methods [8],[9] may be used for this task, e.g. the SQP method implemented in the MATLAB Optimization Toolbox [9] which is used here.

The RLSFIB design may vary from the DSB, $\gamma = N$, to a highly sensitive superdirective beamformer, $\gamma \ll 1$, as desired. This flexibility allows the design to be adapted to any given prior knowledge on microphone mismatch, positioning errors, and microphone self-noise.

Evaluation

Superdirectional beamformers are especially desirable due to their ability to obtain a good spatial selectivity with a small array consisting of few sensors satisfying constraints on space and cost. Geometric constraints often encountered for in-car environments lead to restrictions on array size and varying cost constraints lead to different robustness requirements. The RLSFIB design was therefore evaluated for exemplary in-car array geometries and various WNG values. FIR filters $w_n(l)$

with different lengths were used to approximate the frequency response vectors $[W_n(\omega_0), \ldots, W_n(\omega_P)]$ in the LS sense. The lower and upper cut-off frequencies of 0.2kHz and 5kHz, respectively, were chosen with speech signal capture in mind. The main lobe of the desired frequency response was defined with a 3-dB beamwidth of thirty degrees. Each of the design examples is represented by a figure containing multiple subfigures depicting the beamformer's beampattern and WNG on a logarithmic scale ($A_{\mathrm{dB}} = 10 \log_{10} A$ and $\gamma_{\mathrm{dB}} = 10 \log_{10} \gamma$).

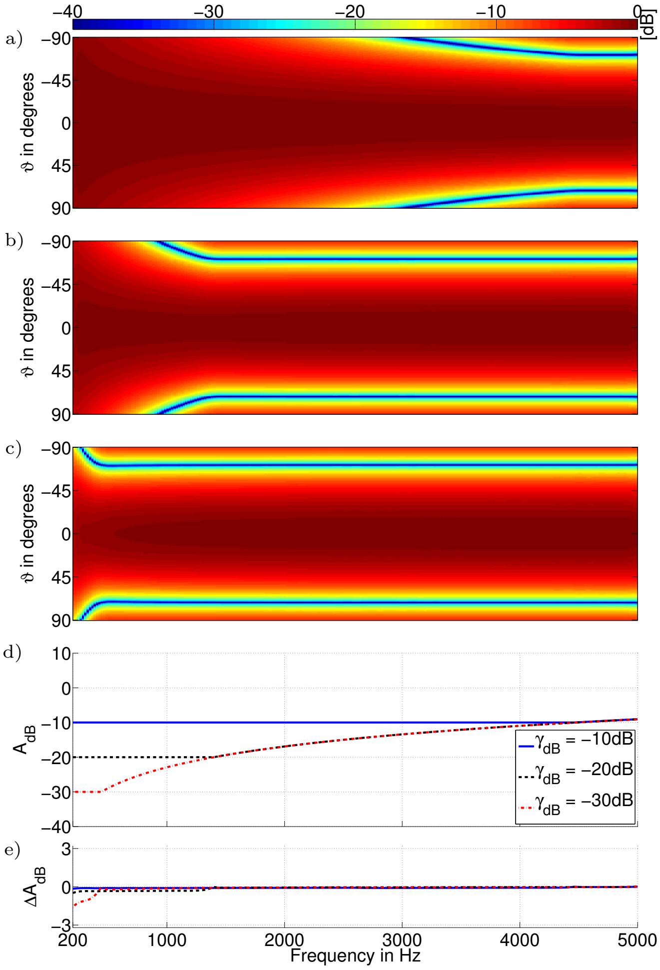

Figure 2: 2-element ULA, $L = 512$; Beampatterns for a) $\gamma_{\mathrm{dB}} = -10$dB, b) $\gamma_{\mathrm{dB}} = -20$dB and c) $\gamma_{\mathrm{dB}} = -30$dB; d) Design-domain-based WNGs; e) WNG deviations due to FIR filter approximations

Figure 2 depicts the results for a two-element uniform linear array (ULA) with $d = 0.008$m and $L = 512$. The main lobe of the beamformer is formed at endfire. The beampatterns and corresponding WNGs for three different values of γ_{dB} are shown. The beampatterns show a markedly improving spatial selectivity across the desired frequency range as the value of γ_{dB} is decreased. Therefore, constraining the WNG leads to the expected trade-off between directivity and sensitivity to errors. The main lobes of all three designs tend to broaden for lower frequencies. The beampattern depicted in Figure 2c shows that although the beamwidth of the main lobe is large due to the small aperture size, good spatial selectivity is still maintained throughout the frequency range. The beampattern is also frequency-invariant above 0.5kHz. However, the constraint of no distortion is violated at the lower frequencies for $\gamma_{\mathrm{dB}} = -20dB$ and $\gamma_{\mathrm{dB}} = -30dB$ as shown in Figures 2b and 2c.

Figure 2d shows that the WNG resulting from the design domain is satisfied perfectly throughout the frequency range for the three different values of γ_{dB}. The FIR filter approximation causes small deviations relative to the ideal behavior in the design domain as depicted in Figure 2e. The WNG constraint is violated slightly for $\gamma_{\mathrm{dB}} = -20dB$ and $\gamma_{\mathrm{dB}} = -30dB$. The FIR filter approximation cannot fully preserve the WNG due to the limited number of degrees of freedom for the FIR filters.

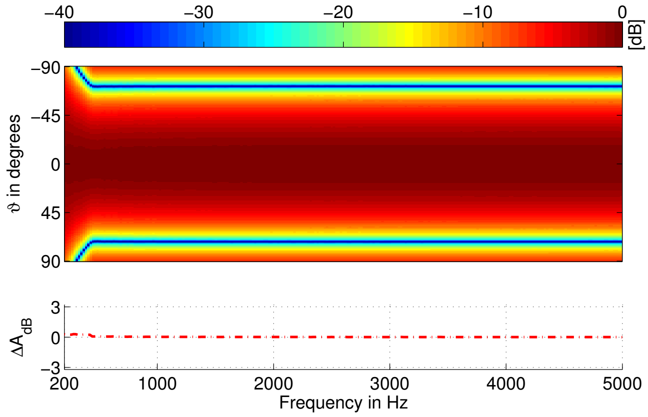

Figure 3: 2-element ULA, $\gamma_{\mathrm{dB}} = -30$dB, $L = 2048$

By selecting filters with larger number of filter coefficients, e.g. $L \geq 2048$, both constraints can be satisfied as shown in Figure 3. This confirms the ability of the RLSFIB design to constrain the WNG efficiently. The phase of the array response is perfectly linear in all the design examples.

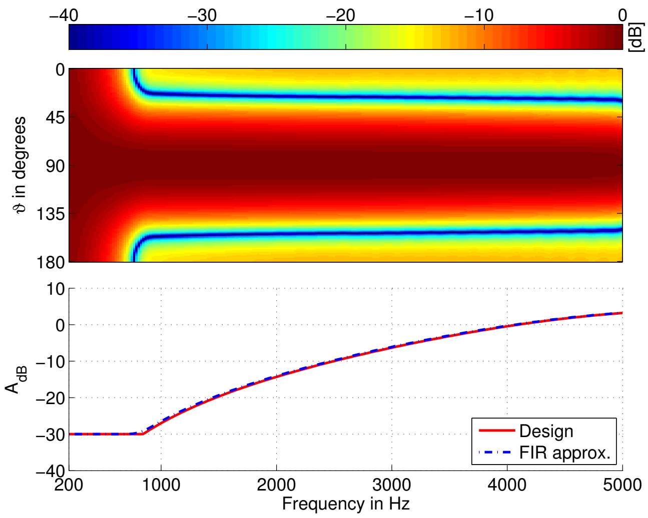

Figure 4: 3-element ULA, $\gamma_{\mathrm{dB}} = -30$dB, $L = 512$

Figure 4 depicts the results for a three-element ULA with $d = 0.02$. The main lobe of the beamformer is formed at broadside with $\gamma = 0.001$ and $L = 512$. The beampattern shows good spatial selectivity and is frequency-invariant above 0.9 kHz. Below this frequency there is a gradual loss in spatial selectivity. The constraint of no magnitude distortion in look direction and the WNG constraint are both fully satisfied.

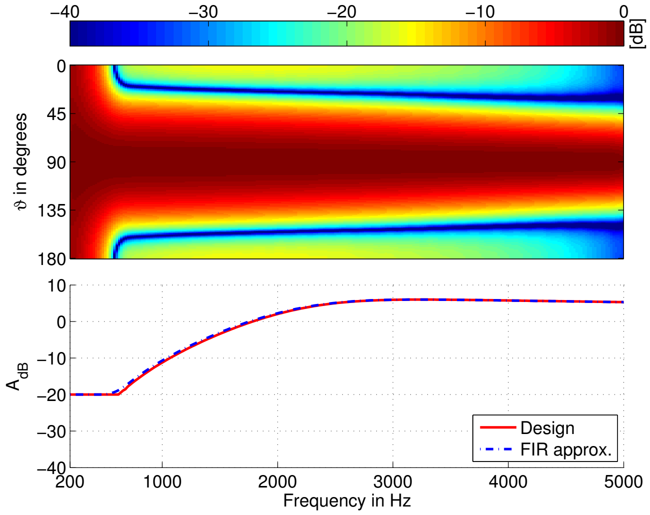

Figure 5: 4-element ULA, $\gamma_{\mathrm{dB}} = -20\mathrm{dB}$, $L = 512$

Another exemplary array geometry used in cars consists of a four-element ULA [1] with microphone spacing $d = 0.03\mathrm{m}$. This spacing was chosen in order to avoid spatial aliasing [5]. This array can be mounted in the lower part of the rear-view mirror in a car. The main lobe of the beamformer is formed at broadside. Figure 5 depicts the results for $\gamma = 0.01$ and $L = 512$. The beampattern shows good spatial selectivity across the desired frequency range although the main beam tends to broaden for frequencies below 0.8kHz. The constraint of no magnitude distortion in look direction and the WNG constraint are both fully satisfied.

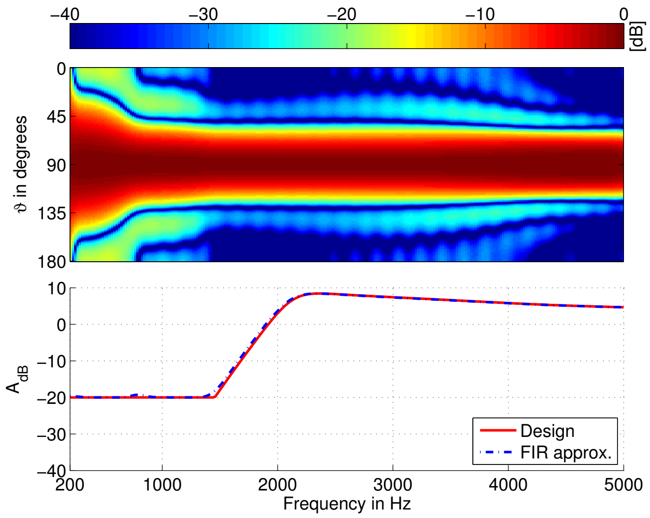

Figure 6: 8-element ULA, $\gamma_{\mathrm{dB}} = -20\mathrm{dB}$, $L = 512$

In the case of larger vehicles such as trucks, buses, vans or public-service vehicles, less restrictive geometrical constraints are often encountered in practice and therefore larger arrays may be used. To this end the RLSFIB design was also investigated for the case of an eight-element ULA with microphone spacing $d = 0.03\mathrm{m}$, $\gamma = 0.01$ and $L = 512$. The results are depicted in Figure 6. The beampattern shows a very good spatial selectivity across the desired frequency range although the main beams tend to broaden below 1kHz. Above this frequency the main lobe is frequency-invariant. The constraint of no magnitude distortion in the look direction and the WNG constraint are both satisfied.

Conclusions

A method which allows full control of the robustness of a least-squares beamformer design is investigated for microphone arrays in cars. The beamformer design has been formulated as a constrained least-squares problem incorporating a linear constraint and a quadratic constraint which in effect constrains the WNG of the resulting design. The constrained least-squares problem is convex and therefore well established methods for convex optimization, such as the SQP methods, may be used to solve the constrained problem. The results shown confirm that the RLSFIB design is capable of providing good spatial selectivity with small arrays consisting of few sensors satisfying constraints on space and cost, while controlling the robustness of the resulting beamformer according to the user's requirements. The flexibility of this design procedure allows the resulting beamformer to be adapted to given prior knowledge on microphone mismatch, positioning errors, and microphone self-noise.

References

[1] E. Hänsler and G. Schmidt, Eds., "Topics in Acoustic Echo and Noise Control," Springer, Berlin, 2006

[2] H. Cox, R.M. Zeskind and T. Kooij, "Practical Supergain," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-34, pp.393-398, June 1986

[3] E. Mabande, A. Schad and W. Kellermann, "Design of Robust Superdirective Beamformers as a Convex Optimization Problem," accepted for IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), April 2009

[4] L.C. Parra, "Steerable Frequency-Invariant Beamforming for Arbitrary Arrays," Journal of the Acoustical Society of America, pp 3839-3847, June 2006

[5] M.S. Brandstein and D.B. Ward, Eds., "Microphone Arrays: Signal Processing Techniques and Applications," Springer, Berlin, 2001

[6] W. Herbordt, "Sound Capture for Human/Machine Interfaces," Springer, Berlin, 2005

[7] S. Boyd and L. Vandenberghe, "Convex Optimization," Cambridge University Press, 2004

[8] R. Fletcher, "Practical Methods for Optimization," John Wiley and Sons Ltd., 1987

[9] "Optimization Toolbox User's Guide", The Mathworks, Inc, 2008