SPEECH ENHANCEMENT USING SMALL MICROPHONE ARRAYS WITH OPTIMIZED DIRECTIVITY

Matthias Dörbecker
Institute of Communication Systems and Data Processing,
Aachen University of Technology, Templergraben 55, 52056 Aachen, Germany
Phone: +49 241 806979, Fax: +49 241 8888186
E-mail: matthias@ind.rwth-aachen.de http://www.ind.rwth-aachen.de

ABSTRACT

A common approach to enhance the quality of speech signals disturbed by acoustic background noise is the application of adaptive filtering techniques aiming at the reduction of the noise (e.g. [1]). However, in case of low SNR most of the currently known adaptive techniques result in a poor quality of the processed signal, which is caused by time-variant distortions of the speech signal and by the unnatural character of the remaining noise (e.g. in form of "musical tones"). An alternative approach, which does not affect the speech signal by time-variant distortions, is the application of a microphone array with a fixed directivity pattern aligned to the speaker's position, resulting in a suppression of spatial distributed noise sources (e.g. [4]). Within the scope of this paper it will be shown that – owing to the proposed optimization of the directivity pattern – even an array consisting of only two microphones may offer a performance comparable to state-of-the-art adaptive filtering techniques. Acoustic measurements related to electronic hearing aids confirm that the improvement of the SNR predicted by theory also holds for signals recorded in a real acoustic environment.

1 INTRODUCTION

The structure of an array with two microphones is depicted in Fig. 1. The enhanced signal $\hat{s}(k)$ is obtained by filtering the microphone signals $x_1(k)$ and $x_2(k)$ with time-invariant impulse responses $a_1(k)$ and $a_2(k)$, respectivily, and a subsequent addition of the filtered signals. Conventionally, the filters $a_1(k)$ and $a_2(k)$ are designed to compensate the time delay between the microphone signals, which corresponds to the direction of principal incidence. Owing to the addition with equal phase, the output signal $\hat{s}(k)$ attains a maximum power for sound being incident from a direction within the array's main beam, whereas for every other direction of incidence a reduced output signal emerges because of destructive interference. The resulting structure is often called *delay-and-sum-beamformer*.

However, when only a low number of microphones is used, the delay-and-sum-beamformer results in a poor directivity at low frequencies, as will be outlined in Section 2. Therefore, the delay-and-sum-beamformer is not appropriate to most applications of digital speech communication, where a low number of microphones and small extensions of the array are desirable.

An alternative approach, which provides an improved directivity, is to determine the impulse responses

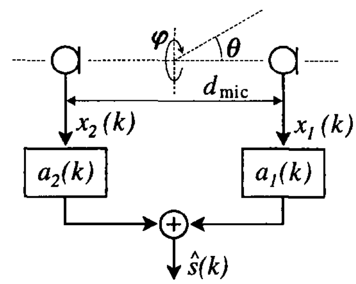

Figure 1: Structure of an array with two microphones

$a_1(k)$ and $a_2(k)$ by a superdirective design known from antenna arrays. While the fundamentals of superdirective arrays have already been described in [2], the superdirective design of microphone arrays has been proposed just in recent years (e.g. [3, 4]). However, a well-known problem arising from the superdirective design is the decreased robustness against random errors of the positions and the transfer functions of the microphones. Since the susceptibility of the array's directivity against random errors increases as the number of microphones is enlarged, the variances of real microphone transfer functions prevent an improvement of the directivity if a large number of microphones is used. On the other hand, for most applications of digital speech communication a low number of microphones is desirable. It has been shown in [5] that in this case the superdirective design results in a susceptibility which is low enough to cope with the variances of real microphone transfer functions. Therefore, in this contribution we focus on arrays consisting of only two microphones.

After a short review of superdirecitive arrays in Section 2, a more flexible design of the array's impulse responses is proposed in Section 3. While the previously mentioned publications [4, 5] refer to idealized acoustic situations with a free space propagation of sound, in Section 4 it is demonstrated that the optimized directivity also holds for real acoustic situations. Relating to the application to electronic hearing aids, it is shown that even the shading of a dummy head in the immediate neighbourhood of the array does not impair the improvement of the directivity. Listening tests described in Section 4 confirm that the two-microphone array provides a significant improvement in terms of speech intelligibility.

2 DESIGN OF SUPERDIRECTIVE ARRAYS

The characteristics of an array can be described by means of the power directivity pattern $\Psi(f, \theta, \varphi)$, which represents the spectral weighting of the acoustic signal as a function of the direction of incidence, that can be expressed by the angles θ and φ as depicted in Fig. 1.

In the following, we focus on equidistantly spaced linear arrays, i.e. the microphones have to be placed on a straight line with a constant distance d_{mic} between two adjacent microphones. This is automatically fulfilled in case of an array consisting of only two microphones. To examine the directivity resulting from the coupling of the microphones, an omnidirectional characteristic is assumed for each individual microphone. This is with no loss in generality, because the results can also be applied to directional microphones.

With these assumptions the array is symmetric relating to a rotation around the array's axis. Therefore, the power directivity pattern is independent of the angle φ according to

$$\Psi(f,\theta) = \Big|\sum_{n=1}^{N} A_n(f)\,\exp(j\beta\, d_{\mathrm{mic}}\cdot(\tfrac{N+1}{2}-n)\cos\theta)\Big|^2, \quad (1)$$

where N is the number of microphones, $A_n(f)$ the transfer function of filter $a_n(k)$, $\beta = 2\pi f/c$ the propagation factor, and c the speed of sound. Alternatively, the directivity can be measured by the gain, which is defined as the ratio of the power directivity pattern for the direction of principal incidence, $\Psi(f,\theta_0)$, relative to the power spectral density of the output signal $\hat{s}(k)$ in case of an omnidirectional incidence of sound. In terms of spherical polar coordinates the gain reads

$$G(f) = \frac{\Psi(f,\theta_0)}{\frac{1}{4\pi}\int\limits_0^{2\pi}\int\limits_0^{\pi}\Psi(f,\theta)\,\sin\theta\; d\theta\, d\varphi}, \quad (2)$$

where θ_0 is the direction of principal incidence. In case of a diffuse noise sound field the gain is equal to the improvement of the SNR. By means of equation (1), an equivalent expression of the gain can be obtained as

$$G(f) = \frac{\Psi(f,\theta_0)}{\sum\limits_{n=1}^{N}\sum\limits_{m=1}^{N} A_n(f)\,A_m^*(f)\,h_{mn}(f)}. \quad (3)$$

The functions $h_{mn}(f)$ are given by

$$\begin{aligned} h_{mn}(f) &= \frac{1}{4\pi}\int\limits_0^{2\pi}\int\limits_0^{\pi}\exp(j\beta\, d_{nm}\cos\theta)\,\sin\theta\; d\theta\, d\varphi \\ &= \begin{cases} \dfrac{\sin(\beta\, d_{nm})}{\beta\, d_{nm}} & \text{for } n \neq m \\ 1 & \text{for } n = m\,, \end{cases} \end{aligned} \quad (4)$$

where d_{mn} denotes the distance between the m-th and the n-th microphone. For the two-microphone array, the distance d_{12} is equal to the spacing d_{mic}.

As described in [2], the design of a superdirective array aims at the maximization of the gain, while the susceptibility against random errors of the microphone transfer functions must not exceed a presupposed upper limit. As a result, the transfer functions

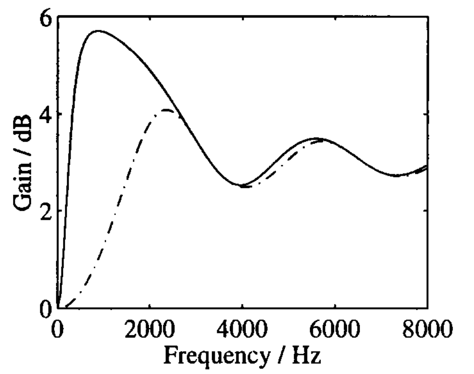

Figure 2: Gain (according to equation (2)) of end-fire arrays with two omnidirectional microphones, placed at a distance of 5 cm
– · – · – delay-and-sum-beamformer
——— superdirective design

$A_n(f)$ of the array filters can be obtained by solving the system of linear equations

$$\begin{aligned} &\sum_{m=1}^{N} h_{nm}(f)\,A_m(f) + \mu\,A_n(f) \\ &\quad = \exp\left(-j\beta\, d_{\mathrm{mic}}\cdot(\tfrac{N+1}{2}-n)\cos\theta_0\right) \end{aligned} \quad (5)$$

for $1 \le n \le N$ [2]. In equation (5), μ denotes an undetermined Lagrangian multiplier, which allows to control the superdirectivity as well as the susceptibility. The choice of a large multiplier $\mu \gg 1$ results in the conventional delay-and-sum-beamformer, which provides a minimum susceptibility at the expense of a poor directivity. On the other hand, the directivity and the susceptibility increase as μ tends towards zero. For the array considered in this contribution, μ has been chosen as 0.02, which provides a superdirectivity close to the optimum value, and a susceptibility which is low enough to cope with the variances of real microphone transfer functions.

The impulse responses $a_n(k)$ are obtained by solving the system of equations (5) for several discrete frequencies $f_\nu = \nu\cdot f_S/M$ with $0 \le \nu \le M/2$, taking the inverse DFT of length M of the $A_n(f_\nu)$, and multiplying the resulting time-domain sequences with a Hamming window. For the simulations described in the following, a sampling frequency of $f_S = 16\,\mathrm{kHz}$ and impulse responses $a_n(k)$ of length 256 have been used.

Fig. 2 depicts the gain of an array with two omnidirectional microphones placed at a distance of $d_{\mathrm{mic}} = 5\,\mathrm{cm}$. The array is designed to have end-fire characteristic, i.e. the direction of principal incidence is in parallel to the array's axis ($\theta_0 = 0$). For this investigation, the influence of the variances of real microphone transfer functions are neglected (i.e. identical microphone transfer functions are assumed).

Obviously, the delay-and-sum-beamformer results in an insufficient gain at low frequencies. It can be stated that for the array considered in this example, the well-known rule of thumb – each doubling of the number of microphones leads to an improvement of 3 dB – is only valid at frequencies above

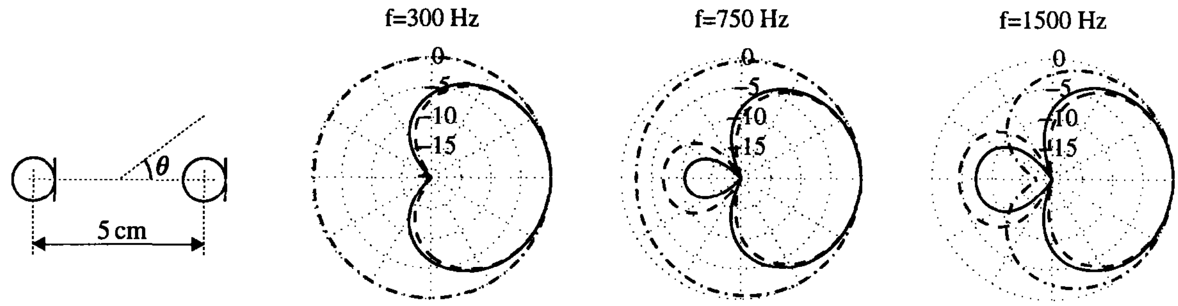

Figure 3: Orientation of the end-fire array and the theoretical directivity pattern $\Psi(f,\theta)$ in dB for different frequencies: $-\cdot-\cdot-$ delay-and-sum-beamformer, $---$ superdirective array, —— optimized array

1500 Hz. Thus if delay-and-sum-beamforming is supposed, quality enhancement at lower frequencies requires that the number of microphones or the dimensions of the array have to be enlarged.

An alternative solution consists in the superdirective design as shown in Fig. 2 by the solid line. The superdirective design yields a significantly improved gain, especially in the frequency range up to 3 kHz, which is of main importance for speech communication. The maximum gain of the superdirective array approaches $20\log_{10} N\,\mathrm{dB} = 6\,\mathrm{dB}$ as μ tends towards zero. For frequencies above 3.3 kHz (i.e. for small wavelengths $\lambda/2 < d_{\mathrm{mic}}$) the superdirective array turns into the delay-and-sum-beamformer.

A similar conclusion can be drawn regarding the directivity pattern $\Psi(f,\theta)$ depicted in Fig. 3. Obviously, the delay-and-sum-beamformer yields a poor directivity at frequencies below 750 Hz. At higher frequencies ($f = 1500\,\mathrm{Hz}$) the delay-and-sum-beamformer yields an excellent attenuation in the rear direction, but the angle of the main beam may be too large for many applications. On the other hand, even with only two microphones the superdirective array provides a significantly improved directivity.

3 OPTIMIZATION OF THE DIRECTIVITY

Although the directivity of the superdirective array is much superior to the delay-and-sum-beamformer, there is still a distinct secondary lobe in the rear direction, especially at higher frequencies ($f \geq 750\,\mathrm{Hz}$ in Fig. 3). The extent of the secondary lobe is caused by the fact that the superdirective design aims at the maximization of the gain. Because of the term $\sin\theta$ in the denominator of equation (2), the rear direction (i.e. $\theta = \pi$) causes a much smaller influence on the gain than any other direction. Consequently, the maximization of the gain does not force a small secondary lobe in the rear direction.

However, in many acoustic situations a small secondary lobe in the rear direction is of greater importance than an optimal SNR in case of an omnidirectional incidence of noise sound. For instance, if the array is applied to an electronic hearing aid, a typical acoustic situation is the so-called *cocktail party situation*. This acoustic situation can be modelled by several noise sources, which are located at the same height where the microphone array is situated. Therefore, the main portion of the noise sound will incidence from a small angle of elevation. For this reason, the maximization of the gain, which refers to an omnidirectional incidence of noise sound, does not yield the optimal directivity pattern for this acoustic situation.

A directivity pattern which is more appropriate to the expected acoustic situation can be obtained by applying a generalized definition of the gain according to

$$G'(f) = \frac{\Psi(f,\theta_0)}{\frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} \Psi(f,\theta)\, w(f,\theta,\varphi)\, \sin\theta \; d\theta\, d\varphi}\;, \qquad (6)$$

where $w(f,\theta,\varphi)$ denotes a weighting function. In case of a uniform weighting $w(f,\theta,\varphi) = 1$, equation (6) represents the conventional definition of the gain. On the other hand, the application of an appropriate non-uniform weighting function enables a definition of the gain, where directions are emphasized, which are expected to contribute more noise sound. For the application to electronic hearing aids it is appropriate to consider only directions within the horizontal plane. This special case suggests a two-dimensional definition of the gain according to

$$G''(f) = \frac{\Psi(f,\theta_0)}{\frac{1}{\pi}\int_0^{\pi} \Psi(f,\theta)\, d\theta}\;. \qquad (7)$$

Therefore, the set of transfer functions $A''_n(f)$, which yields a maximum gain for this acoustic situation, can be determined by solving the system of linear equations (5), where the functions $h_{mn}(f)$ have to be replaced by

$$h''_{mn}(f) = \frac{1}{\pi}\int_0^{\pi} \exp(j\beta d_{nm}\cos\theta)\, d\theta\;. \qquad (8)$$

Since the integral in equation (8) can not be solved in closed form, it has to be approximated numerically. As described in Section 2, the impulse responses of the filters can be obtained by an inverse DFT and appropriate windowing.

The solid lines in Fig. 3 confirm that the optimized design reduces the secondary lobe resulting from the superdirective design by up to 3 dB. Furthermore, it can be observed that the improvement is at the expense of an only marginally enlarged angle of the main beam.

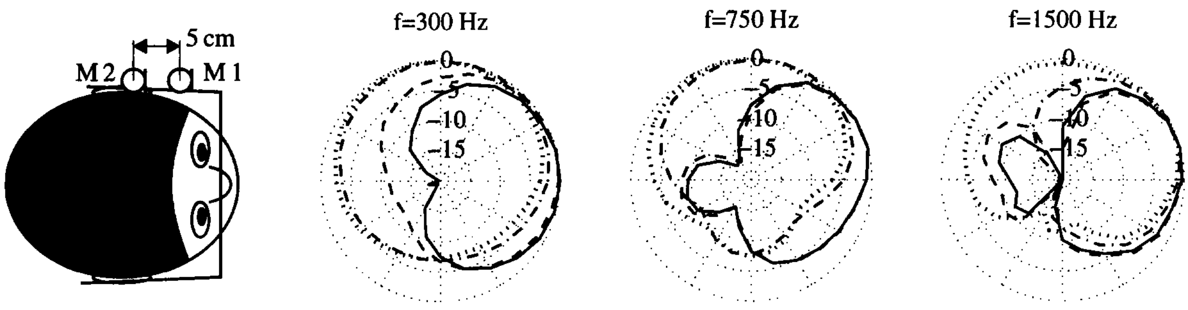

Figure 4: Arrangement of the microphones beneath the dummy head and the measured directivity pattern in dB: ········ single microphone M2 — — — superdirective array — · — · — delay-and-sum-beamformer ——— optimized array

4 PERFORMANCE IN A REAL ACOUSTIC ENVIROMENT

The evaluation described so far refers to an idealized acoustic situation with perfect microphone impulse responses and free space propagation of sound. To examine the influence of a real acoustic environment including not-identical microphones and obstacles in the immediate neighbourhood of the microphones, recordings related to electronic hearing aids have been made. For this purpose, two microphones have been mounted with a spacing of 5 cm at the left ear-piece of a spectacle frame put on a dummy head, as depicted in Fig. 4. The microphone signals have been recorded for different directions of incidence to determine the power directivity pattern $\Psi(f, \theta, \varphi_0)$ for different angles θ, while φ_0 is chosen such that the directivity pattern refers to the horizontal plane. Fig. 4 depicts the directivity pattern of the delay-and-sum-beamformer, the superdirective array, and the optimized design. Additionally, to allow a differentiation between the directivity caused by the shading of the head and the directivity resulting from the coupling of the microphones, the directivity pattern of the single microphone M2 is depicted as a reference.

As predicted by theory, the power directivity pattern of the delay-and-sum-beamformer at low frequencies is almost identical to the directivity which can be obtained by a single microphone. Obviously, the improvement of the directivity resulting from the superdirective design also holds for the signals recorded in the real acoustic situation. Furthermore, it is confirmed that the optimized design provides a higher attenuation of the rear direction at higher frequencies. As a conclusion it can be stated, that the dummy head, being an obstacle in immediate neighbourhood of the array, does not significantly impair the improved directivity.

To evaluate the speech enhancement resulting from the optimized directivity, we refer to an acoustic situation where a single speaker is situated in front of the dummy head. Additionally, several interferent talkers are placed at different, uniformly spaced angles θ around the dummy head in order to achieve a noise sound field without a dominating angle of incidence θ. Informal listening tests confirm a significant reduction of the noise. In contrast to most adaptive noise reduction systems, the array's output signal is of high naturalness because of the absence of time-variant distortions.

5 CONCLUSIONS

In this contribution it has been shown that the superdirective design known from antenna arrays can successfully be applied to microphone arrays. Simulations confirm that in case of a two-microphone array the susceptibility resulting from the superdirective design is low enough to cope with the variances of real microphone transfer functions. Even the shading of a dummy head in the immediate neighbourhood of the array does not significantly impair the improvement of the directivity.

Furthermore, a more flexible new design of the array's impulse responses $a_n(k)$ has been proposed, which allows to consider the expected spatial distribution of the noise sources. As an example, the novel design has been used to reduce the secondary lobe in the rear direction of the two-microphone end-fire array.

Listening tests confirm a significant reduction of spatially distributed noise sources. Owing to the application of time-invariant filtering the output signal is not affected by any time-variant distortions, which results in a very high naturalness. Therefore, the two-microphone array with optimized directivity is a powerful alternative to state-of-the-art noise suppression techniques based on adaptive filtering.

6 REFERENCES

[1] R. Martin and P. Vary, "Combined Acoustic Echo Control and Noise Reduction for Hands-Free Telephony - State of the Art and Perspectives", in *Proc. EUSIPCO*, Sept. 1996, pp. 1107–1110.

[2] E.N. Gilbert and S.P. Morgan, "Optimum Design of Directive Antenna Arrays Subject to Random Variations", *The Bell System Technical Journal*, vol. 34, pp. 637–663, May 1955.

[3] W. Soede, *Improvement of Speech Intelligibility in Noise*, PhD thesis, Technische Universiteit Delft, 1990.

[4] R. Zelinski, "Mikrofon-Arrays mit superdirektiven Eigenschaften zur Sprachsignalübertragung", *Frequenz*, vol. 50, no. 9-10, pp. 198–204, 1996.

[5] M. Dörbecker, "Small Microphone Arrays with Optimized Directivity for Speech Enhancement", submitted to EUROSPEECH-97.

This work has been supported by the German Research Council (DFG).