Hindawi Publishing Corporation
EURASIP Journal on Applied Signal Processing
Volume 2006, Article ID 63297, Pages 1–14
DOI 10.1155/ASP/2006/63297

Dual-Channel Speech Enhancement by Superdirective Beamforming

Thomas Lotter and Peter Vary

Institute of Communication Systems and Data Processing, RWTH Aachen University, 52056 Aachen, Germany

Received 31 January 2005; Revised 8 August 2005; Accepted 22 August 2005

In this contribution, a dual-channel input-output speech enhancement system is introduced. The proposed algorithm is an adaptation of the well-known superdirective beamformer including postfiltering to the binaural application. In contrast to conventional beamformer processing, the proposed system outputs enhanced stereo signals while preserving the important interaural amplitude and phase differences of the original signal. Instrumental performance evaluations in a real environment with multiple speech sources indicate that the proposed computational efficient spectral weighting system can achieve significant attenuation of speech interferers while maintaining a high speech quality of the target signal.

1. INTRODUCTION

Speech enhancement by beamforming exploits spatial diversity of desired speech and interfering speech or noise sources by combining multiple noisy input signals. Typical beamformer applications are hands-free telephony, speech recognition, teleconferencing, and hearing aids. Beamformer realizations can be classified into fixed and adaptive.

A fixed beamformer combines the noisy signals of multiple microphones by a time-invariant filter-and-sum operation. The combining filters can be designed to achieve constructive superposition towards a desired direction (delay-and-sum beamformer) or in order to maximize the SNR improvement (superdirective beamformer), for example, [1]. As practical problems such as self-noise and amplitude or phase errors of the microphones limit the use of optimal beamformers, constrained solutions have been introduced that limit the directivity to the benefit of reduced susceptibility [2–4]. Most fixed beamformer design algorithms assume the desired source to be positioned in the far field, that is, the distance between the microphone array and the source is much greater than the dimension of the array. Near-field superdirectivity [5] additionally exploits amplitude differences between the microphone signals. Adaptive beamformers commonly consist of a fixed beamformer steered towards a desired direction and a time-varying branch, which adaptively steers beamformer spatial nulls towards interfering sources. Among various adaptive beamformers, the Griffiths-Jim beamformer [6], or extensions, for example, in [7, 8], is most widely known. Adaptive beamformers can be considered less robust against distortions of the desired signal than fixed beamformers.

Beamforming for binaural input signals, that is, signals recorded by single microphones at the left and right ear, has found significantly less attention than beamformers for (linear) microphone arrays. An important application is the enhancement of speech in a difficult multitalker situation using binaural hearing aids.

Current hearing aids achieve a speech intelligibility improvement in difficult acoustic condition by the use of independent small endfire arrays, often integrated into behind-the-ear devices with low microphone distances around 1-2 cm. When hearing aids are used in combination with eyeglasses, larger arrays are feasible, which can also form a binaural enhanced signal [9].

Binaural noise reduction techniques get into attention, when space limitation forbids the use of multiple microphones in one device or when the enhancement benefits of two independent endfire arrays are to be combined with binaural processing benefit. In contrast to an endfire array, a binaural speech enhancement system must work with a dual-channel input-output signal, at best without modification of the interaural amplitude and phase differences in order not to disturb the original spatial impression.

Enhancement by exploiting coherence properties [10] of the desired source and the noise [3, 11] has the ability to

reduce diffuse noise to a high degree, however fails in suppressing sound from directional interferers, especially unwanted speech. Also, due to the adaptive estimation of the instantaneous coherence in frequency bands, musical tones can occur. In [12, 13], a noise reduction system has been proposed, that applies a binaural processing model of the human ear. To suppress lateral noise sources, the interaural level and phase differences are compared to reference values for the frontal direction. Frequency components are attenuated by evaluation of the deviation from reference patterns. However, the system suffers severely from susceptibility to reverberation. In [14], the Griffiths-Jim adaptive beamformer [6] has been applied to binaural noise reduction in subbands, and listening tests have shown a performance gain in terms of speech intelligibility. However, the subband Griffiths-Jim approach requires a voice activity detection (VAD) for the filter adaptation which can cause cancellation of the desired speech when the VAD frequently fails especially at low signal-to-noise ratios.

In [15], a two-microphone adaptive system is presented with the core of a modified Griffiths-Jim beamformer. By lowband-highband separation, a tradeoff is provided between array-processing benefit and binaural benefit by the choice of the cutoff frequency. In the lower band, the binaural signal is passed to the respective ear. The directional filter is only applied to the high-frequency regions, whose influence to sound localization and lateralization is considered less significant. Both adaptive algorithms from [14, 15] have the ability to adaptively cancel out an interfering source. However, the beamformer adaptation procedure needs to be coupled to a voice activity detection (VAD) or correlation-based measure to counteract against possible target cancellation.

In this contribution, a full-band binaural input-output array that applies a binaural signal model and the well-known superdirective beamformer as core is presented [16]. The dual-channel system thus comprises the advantages of a fixed beamformer, that is, low risk of target cancellation and computational simplicity.

To deliver an enhanced stereo signal instead of a mono output, an efficient adaptive spectral weight calculation is introduced, in which the desired signal is passed unfiltered and which does not modify the perceptually important interaural time and phase differences of the target and residual noise signal. To further increase the performance, a well-known Wiener postfilter is also adapted for the binaural application under consideration of the same requirements.

The rest of the paper is organized as follow. In Section 2, the binaural signal model is introduced as a basis for the beamformer algorithm. Section 3 includes the proposed superdirective beamformer with dual-channel input and output as well as the adaptive postfilter. Finally, in Section 4 performance results are given in a real environment.

2. BINAURAL SIGNAL MODEL

For the derivation of binaural beamformers, an appropriate signal model is required. The microphone signals at the left and right ears do not only differ in the time difference depending on the position of the source relative to the head. Furthermore, the shadowing effect of the head causes significant intensity differences between the left- and right-ear microphone signals. Both effects are described by the *head-related transfer functions* (HRTFs) [17].

Figure 1(a) shows a time signal s arriving at the microphones from the angle θ_S in the horizontal plane. The time signals at the left and right microphones are denoted by y_l, y_r. The microphone signal spectra can be expressed by the HRTFs towards left and right ears $D_\mathrm{l}(\omega)$, $D_\mathrm{r}(\omega)$. As the beamformer will be realized in the DFT domain, a DFT representation of the spectra is chosen. At discrete DFT frequencies ω_k with frequency index k, the left- and right-ear signal spectra are given by

$$Y_\mathrm{l}(\omega_k) = D_\mathrm{l}(\omega_k)S(\omega_k), \qquad Y_\mathrm{r}(\omega_k) = D_\mathrm{r}(\omega_k)S(\omega_k). \tag{1}$$

Here, $S(\omega_k)$ denotes the spectrum of the original signal s. For brevity, the frequency index k is used instead of ω_k.

The acoustic transfer functions are illustrated in Figure 1. The shadowing effect of the head is described by multiplication of each spectral coefficient of the input spectrum $S(k)$ with an angle and frequency-dependent physical amplitude factors $\alpha_\mathrm{l}^\mathrm{phy}, \alpha_\mathrm{r}^\mathrm{phy}$ for the left- and right-ear side. The physical time delays $\tau_\mathrm{l}^\mathrm{phy}, \tau_\mathrm{r}^\mathrm{phy}$, that characterize the propagation time from the origin to the left and right ears, are approximately considered to be frequency-independent. The HRTF vector $\mathbf{D}$ can thus be written by

$$\mathbf{D}(\theta_s,k) = \left[\alpha_\mathrm{l}^\mathrm{phy}(\theta_s,k)e^{-j\omega_k\tau_\mathrm{l}^\mathrm{phy}(\theta_s)}, \alpha_\mathrm{r}^\mathrm{phy}(\theta_s,k)e^{-j\omega_k\tau_\mathrm{r}^\mathrm{phy}(\theta_s)}\right]^T. \tag{2}$$

For convenience, the physical transfer function can be normalized to that of zero degree. With $\alpha^\mathrm{phy}(0^\circ,k) := \alpha_\mathrm{l}^\mathrm{phy}(0^\circ, k) = \alpha_\mathrm{r}^\mathrm{phy}(0^\circ,k)$ and $\tau^\mathrm{phy}(0^\circ) := \tau_\mathrm{l}^\mathrm{phy}(0^\circ) = \tau_\mathrm{r}^\mathrm{phy}(0^\circ)$, the normalized amplitude factors $\alpha_\mathrm{l}^\mathrm{norm}$, $\alpha_\mathrm{r}^\mathrm{norm}$ and time delays $\tau_\mathrm{l}^\mathrm{norm}$, $\tau_\mathrm{r}^\mathrm{norm}$, respectively, can be written as

$$\begin{aligned}
\alpha_\mathrm{l}^\mathrm{norm}(\theta_S,k) &= \frac{\alpha_\mathrm{l}^\mathrm{phy}(\theta_S,k)}{\alpha_\mathrm{l}^\mathrm{phy}(0^\circ,k)},\\
\tau_\mathrm{l}^\mathrm{norm}(\theta_S) &= \tau_\mathrm{l}^\mathrm{phy}(\theta_S) - \tau_\mathrm{l}^\mathrm{phy}(0^\circ),\\
\alpha_\mathrm{r}^\mathrm{norm}(\theta_S,k) &= \frac{\alpha_\mathrm{r}^\mathrm{phy}(\theta_S,k)}{\alpha_\mathrm{r}^\mathrm{phy}(0^\circ,k)},\\
\tau_\mathrm{r}^\mathrm{norm}(\theta_S) &= \tau_\mathrm{r}^\mathrm{phy}(\theta_S) - \tau_\mathrm{r}^\mathrm{phy}(0^\circ).
\end{aligned} \tag{3}$$

The transfer vector $\mathbf{D}$ or the amplitudes $\alpha_\mathrm{l}^\mathrm{phy}$, $\alpha_\mathrm{r}^\mathrm{phy}$ and time delays $\tau_\mathrm{l}^\mathrm{phy}$, $\tau_\mathrm{r}^\mathrm{phy}$ as well as their normalized versions are in the following obtained by two different approaches. Firstly, a database of measured head-related impulse responses is used

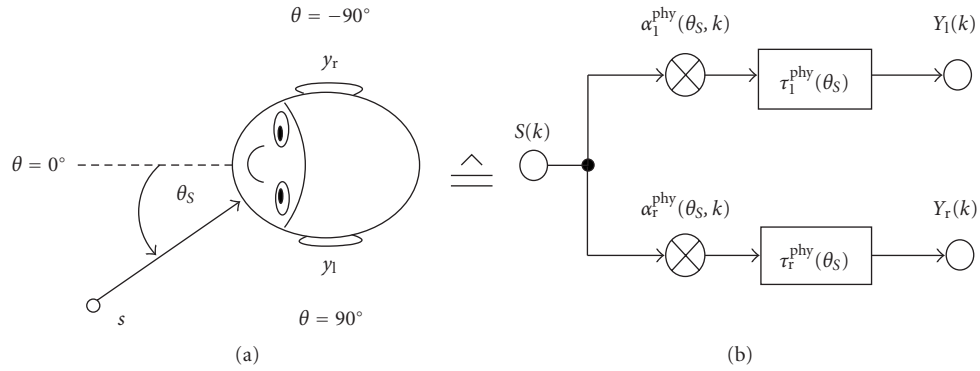

Figure 1: Acoustic transfer of a source from θ_S towards the left and right ears.

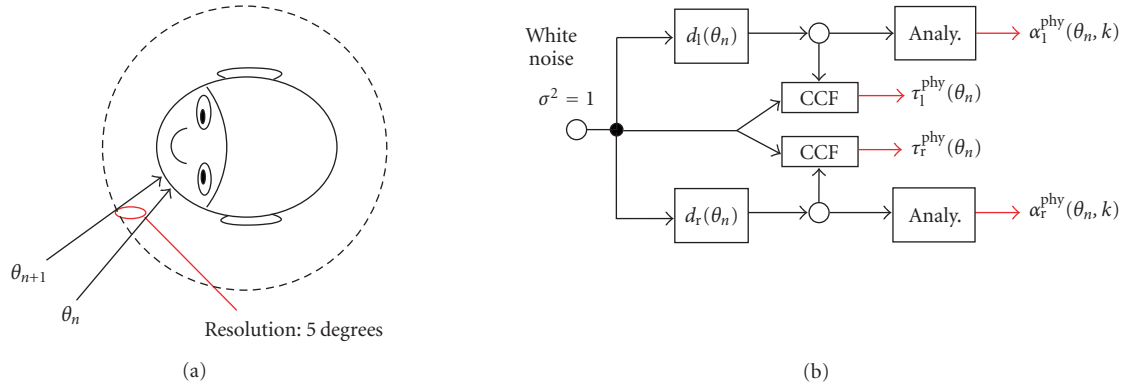

Figure 2: Generation of physical binaural transfer cues α_l^{phy}, α_r^{phy}, τ_l^{phy}, τ_r^{phy} using a database of head-related impulse responses.

to extract the transfer vectors for a number of relevant spatial directions. Secondly, a binaural model is applied to approximate transfer vectors.

2.1. HRTF database

The first approach to extract interaural time differences and amplitude differences is to use a database of head-related impulse responses, for example, [18]. This database comprises recordings of head-related impulse responses $d_l(\theta_n, i)$, $d_r(\theta_n, i)$ with time index i for several spatial directions with in-the-ear microphones using a Knowles Electronics Manikin for Auditory Research (KEMAR) head. For a given resolution of the azimuths, for example, 5 degrees, the values of $\alpha_l^{phy}, \alpha_r^{phy}, \tau_l^{phy}, \tau_r^{phy}$ are determined according to Figure 2. White noise is filtered with the impulse responses $d_l(\theta_n), d_r(\theta_n)$ for the left and right ears. A maximum search of the cross-correlation function of the output signals delivers the relative time differences τ_l^{phy}, τ_r^{phy}. The left- and right-ear delays can then be calculated using (3). For the extraction of the amplitude factors α_l^{phy}, α_r^{phy}, a frequency analysis is performed. Here, the same analysis should be applied as that of the frequency-domain realization of the beamformer.

2.2. HRTF model

Using binaural cues extracted from a database delivers fixed HRTFs. The real HRTFs will however vary greatly between the persons and also on a daily basis depending on the position of the hearing aids. An adjustment of the beamformer to the user without the demand to measure the customers HRTFs is desirable. This can be achieved by using a parametric binaural model.

In [19], binaural sound synthesis is performed using a two filter blocks that approximate the interaural time differences (ITDs) and the interaural intensity differences (IIDs), respectively, of a spherical head. Useful results have been obtained by cascading a delay element with a single-pole and single-zero head-shadow filter according to

$$D_{\text{mod}}(\theta, \omega) = \frac{1 + j(\gamma_{\text{mod}}(\theta)\omega/2\omega_0)}{1 + j(\omega/2\omega_0)} \cdot e^{-j\omega\tau_{\text{mod}}(\theta)}, \quad (4)$$

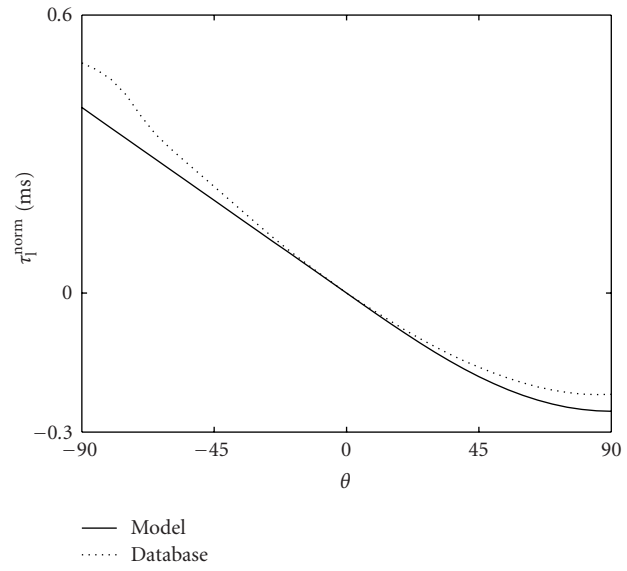

FIGURE 3: Normalized time differences of left ear $\tau_l^{norm}(\theta)$ using the HRTF database and the binaural model, respectively.

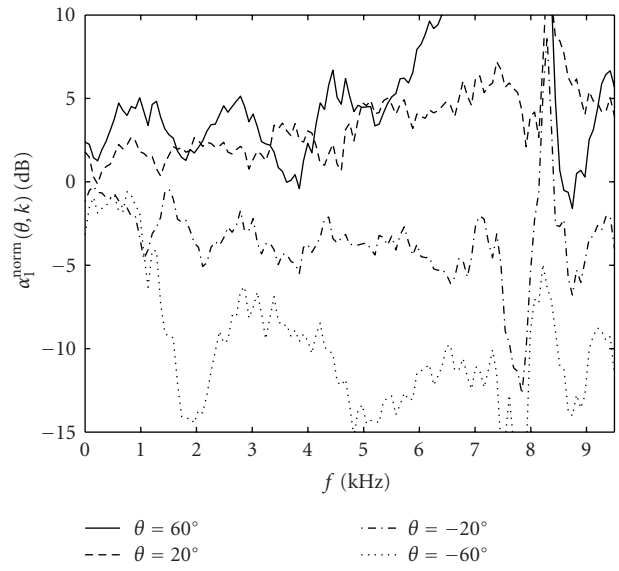

FIGURE 4: Normalized amplitude factors $\alpha_l^{norm}(\theta, k)$ for different azimuth angles extracted from database.

with $\omega_0 = c/a$, where c is the speed of sound and a is the radius of the head. The model is determined by the angle-dependent parameters γ_{mod} and τ_{mod} with

$$\gamma_{mod}(\theta) = \left(1 + \frac{\beta_{min}}{2}\right) + \left(1 - \frac{\beta_{min}}{2}\right)\cos\left(\frac{\theta - \pi/2}{\theta_{min}} 180°\right),$$

$$\tau_{mod}(\theta) = \begin{cases} -\dfrac{a}{c}\cos(\theta - \pi/2), & -\dfrac{\pi}{2} \le \theta < 0, \\ \dfrac{a}{c}|\theta|, & 0 \le \theta < \dfrac{\pi}{2}. \end{cases} \tag{5}$$

The parameters of the model are set to $\beta_{min} = 0.1$, $\theta_{min} = 150°$, which produces a fairly good approximation to the ideal frequency response of a rigid sphere (see [19]). The transfer vector $\mathbf{D} = [D_l, D_r]^T$ can be extracted from (4) with

$$D_l(\theta_s, k) = D_{mod}(\theta_s, \omega_k), \qquad D_r(\theta_s, k) = D_{mod}(\pi - \theta_s, \omega_k). \tag{6}$$

The model provides the radius of the spherical head a as parameter. It is set to 0.0875 m, which is commonly considered as the average radius for an adult human head.

2.3. Comparison of HRTF extraction methods

Figure 3 shows the normalized time differences τ_l^{norm} in dependence of the azimuth angle extracted from the HRTF database and by applying the binaural model. While the model-based approach delivers smaller absolute values, the time differences are very similar.

Figure 4 plots the normalized amplitude factors α_l^{norm} over the frequency for different azimuths using the HRTF database, while Figure 5 shows the normalized amplitude

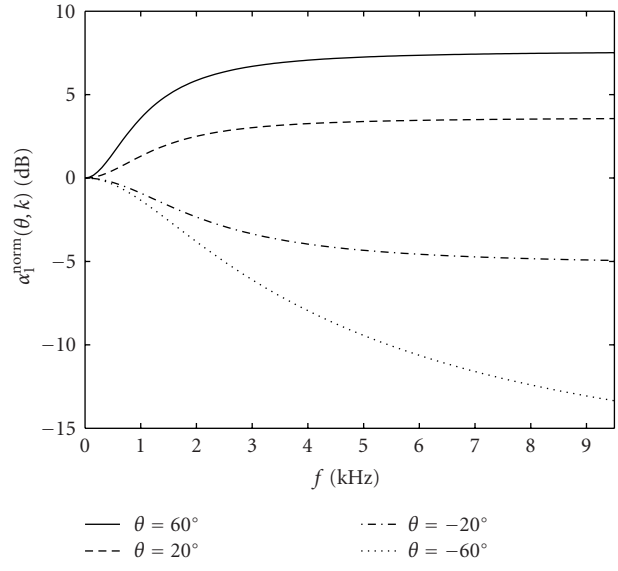

FIGURE 5: Normalized amplitude factors $\alpha_l^{norm}(\theta, k)$ for different azimuth angles extracted from the binaural model.

factors obtained by the HRTF model. The model-based approach delivers amplitude values that interpolate the angle- and frequency-dependent amplitude factors of the KEMAR head, or in other words the fine structure of the HRTF is not considered by the simple model.

Due to the high variance between persons, measurements of the targets person's HRTFs should at best be provided to a binaural speech enhancement algorithm. However, we think that a strenuous and time-consuming measurement for several angles is not feasible for many application scenarios, for example, not during the hearing aid fitting process. In case

of the target person's HRTFs being unknown to the binaural algorithm, the fine structure of a specific HRTF cannot be exploited. Therefore, we prefer the model-based approach, which can be customized to some extent with little effort by choosing a different head radius, for example, during the hearing aid fitting process. In the following, the dual-channel input-output beamformer design will be illustrated only with underlying the model-based HRTF.

3. SUPERDIRECTIVE BINAURAL BEAMFORMER

In this section, the superdirective beamformer with Wiener postfilter is adapted for the binaural application. The proposed fixed beamformer uses superdirective filter design techniques in combination with the signal model to optimally enhance signals from a given desired spatial direction compared to all other directions. The enhancement of the beamformer and postfilter is then exploited to calculate spectral weights for left- and right-ear spectral coefficients under the constraint of the preservation of the interaural amplitude and phase differences.

3.1. Superdirective beamformer design in the DFT domain

Consider a microphone array with M elements. The noisy observations for each microphone m are denoted as $y_m(i)$ with time index i. Since the superdirective beamformer can efficiently be implemented in the DFT domain, noisy DFT coefficients $Y_m(k)$ are calculated by segmenting the noisy time signals into frames of length L and windowing with a function $h(i)$, for example, Hann window including zero-padding. The DFT coefficient of microphone m, frame λ, and frequency bin k can then be calculated with

$$Y_m(k,\lambda) = \sum_{i=0}^{L-1} y_m(\lambda R + i)h(i)e^{-j2\pi ki/L}, \quad m \in \{1,\ldots,M\}. \tag{7}$$

For the computation of the next DFT, the window is shifted by R samples. These parameters are chosen to $N = 256$ and $R = 112$ at a sampling frequency of $f_s = 20\,\text{kHz}$. For the sake of brevity, the index λ is omitted in the following.

In the DFT domain, the beamformer is realized as multiplication of the input noisy DFT coefficients Y_m, $m \in \{1,\ldots,M\}$, with complex factors W_m. The output spectral coefficient is given as

$$Z(k) = \sum_m W_m^*(k)Y_m(k) = \mathbf{W}^H\mathbf{Y}. \tag{8}$$

The objective of the superdirective design of the weight vector $\mathbf{W}$ is to maximize the output SNR. This can be achieved by minimizing the output energy with the constraint of an unfiltered signal from the desired direction. The minimum variance distortionless response (MVDR) approach can be written as (see [1–3])

$$\min_{\mathbf{W}} \mathbf{W}^H(\theta_S,k)\mathbf{\Phi}_{MM}(k)\mathbf{W}(\theta_S,k) \quad \text{w.r.t,} \quad \mathbf{W}^H(\theta_S,k)\mathbf{D}(\theta_S,k) = 1. \tag{9}$$

Here $\mathbf{\Phi}_{MM}$ denotes the cross-spectral-density matrix,

$$\mathbf{\Phi}_{MM}(k) = \begin{pmatrix} \Phi_{11}(k) & \Phi_{12}(k) & \cdots & \Phi_{1M}(k) \\ \Phi_{21}(k) & \Phi_{22}(k) & \cdots & \Phi_{2M}(k) \\ \vdots & \vdots & \vdots & \vdots \\ \Phi_{M1}(k) & \Phi_{M2}(k) & \cdots & \Phi_{MM}(k) \end{pmatrix}. \tag{10}$$

If a homogenous isotropic noise field is assumed, then the elements of $\mathbf{\Phi}_{MM}$ are determined only by the distance d_{mn} between microphones m and n [10]:

$$\Phi_{mn}(k) = \text{si}\left(\frac{\omega_k d_{mn}}{c}\right). \tag{11}$$

The vector of coefficients can then be determined by gradient calculation or using Lagrangian multipliers to

$$\mathbf{W}(\theta_S,k) = \frac{\mathbf{\Phi}_{MM}^{-1}(k)\mathbf{D}(\theta_S,k)}{\mathbf{D}^H(\theta_S,k)\mathbf{\Phi}_{MM}^{-1}(k)\mathbf{D}(\theta_S,k)}. \tag{12}$$

If a design should be performed with limited superdirectivity to avoid the loss of directivity by microphone mismatch, the design rule can be modified by inserting a tradeoff factor μ_s [3],

$$\mathbf{W}(\theta_S,k) = \frac{(\mathbf{\Phi}_{MM}^{-1}(k) + \mu_s\mathbf{I})\mathbf{D}(\theta_S,k)}{\mathbf{D}^H(\theta_S,k)(\mathbf{\Phi}^{-1}(k) + \mu_s\mathbf{I})\mathbf{D}(\theta_S,k)}. \tag{13}$$

If $\mu_s \to \infty$, then $\mathbf{W} \to 1/\mathbf{D}^H$, that is, a delay-and-sum beamformer results from the design rule. A more general approach to control the tradeoff between directivity and robustness is presented in [4].

The directivity of the superdirective beamformer strongly depends on the position of the microphone array towards the desired direction. If the axis of the microphone array is the same as the direction of arrival, an *endfire* array with higher directivity than for a *broadside* array, where the axis is orthogonal to the direction of arrival, is obtained.

3.1.1. Binaural superdirective coefficients

In the binaural application $M = 2$ microphones are used, the spectral coefficients are indexed by l and r to express left and right sides of the head. The superdirective design rule according to (13) requires the transfer vector for the desired direction $\mathbf{D}(\theta_s,k) = [D_\text{l}(\theta_s,k), D_\text{r}(\theta_s,k)]^T$ and the matrix of cross-power-spectral densities $\mathbf{\Phi}_{22}$ as inputs for each frequency bin k. The transfer vector can be extracted from (4) according to (6). On the other hand, the 2×2 cross-power-spectral density matrix $\mathbf{\Phi}_{22}(k)$ can be calculated using the head related coherence function. After normalization by $\sqrt{\Phi_\text{ll}(k)\Phi_\text{rr}(k)}$, where $\Phi_\text{ll}(k) = \Phi_\text{rr}(k)$, the matrix is

$$\mathbf{\Phi}_{22}(k) = \begin{pmatrix} 1 & \Gamma_\text{lr}(k) \\ \Gamma_\text{lr}(k) & 1 \end{pmatrix}, \tag{14}$$

with the coherence function

$$\Gamma_\text{lr}(k) = \frac{\Phi_\text{lr}(k)}{\sqrt{\Phi_\text{ll}(k)\Phi_\text{rr}(k)}}. \tag{15}$$

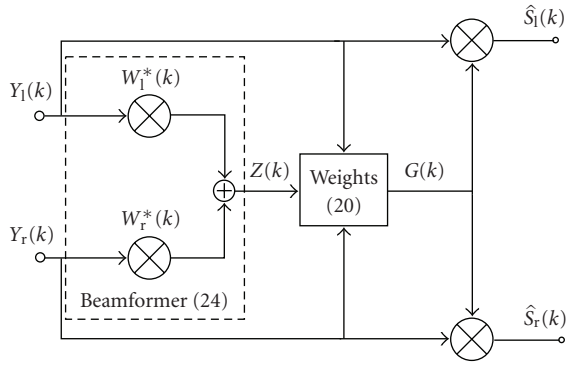

Figure 6: Superdirective binaural input-output beamformer.

The head-related coherence function is much lower than the value that could be expected from (11) when only taking the microphone distance between left and right ears into account [3]. It can be calculated by averaging a number N of equidistant HRTFs across the horizontal plane, $0 \le \theta < 2\pi$,

$$\Gamma(k) = \frac{\sum_{n=1}^{N} D_l(\theta_n,k) D_r^*(\theta_n,k)}{\sqrt{\left(\sum_{n=1}^{N} |D_l(\theta_n,k)|^2\right)\left(\sum_{n=1}^{N} |D_r(\theta_n,k)|^2\right)}}. \tag{16}$$

In this work, an angular resolution of 5 degrees in the horizontal plane is used, that is, $N = 72$.

3.1.2. Dual-channel input-output beamformer

A beamformer that outputs a monaural signal would be unacceptable, because the benefit in terms of noise reduction is consumed by the loss of spatial hearing. We therefore propose to utilize the beamformer output for the calculation of spectral weights. Figure 6 shows a block diagram of the proposed superdirective stereo input-output beamformer in the frequency domain.

In analogy to (8), the input DFT coefficients are summed after complex multiplication by superdirective coefficients,

$$Z(k) = \mathbf{W}^H(k)\mathbf{Y}(k) = W_l^*(k)Y_l(k) + W_r^*(k)Y_r(k). \tag{17}$$

The enhanced Fourier coefficients Z can then serve as reference for the calculation of weight factors G (as defined in the following), which output binaural enhanced spectra $\hat{S}_l$, $\hat{S}_r$ via multiplication with the input spectra Y_l, Y_r. Afterwards, the enhanced dual-channel time signal is synthesized via IDFT and overlap add.

Regarding the weight calculation method, it is advantageous to determine a single real-valued gain for both left- and right-ear spectral coefficients. By doing so, the interaural time and amplitude differences will be preserved in the enhanced signal. Consequently, distortions of the spatial impression will be minimized in the output signal. Real-valued weight factors $G_{\text{super}}(k)$ are desirable in order to minimize distortions from the frequency-domain filter. In addition, a distortionless response for the desired direction should be guaranteed, that is, $G_{\text{super}}(\theta_s,k) \stackrel{!}{=} 1$.

To fulfil the demand of just one weight for both left- and right-ear sides, the weights are calculated by comparing the spectral amplitudes of the beamformer output to the sum of both input spectral amplitudes,

$$G_{\text{super}}(k) = \frac{|Z(k)|}{|Y_l(k)| + |Y_r(k)|}. \tag{18}$$

To avoid amplification, the weight factor is upper-limited to one afterwards. To fulfil the distortionless response of the desired signal with (18), the MVDR design rule according to (13) has to be modified with a correction factor $\text{corr}_{\text{super}}$:

$$\begin{aligned} &\min_{\mathbf{W}} \mathbf{W}^H(\theta_S,k)\mathbf{\Phi}_{MM}(k)\mathbf{W}(\theta_S,k) \quad \text{w.r.t.,} \\ &\mathbf{W}^H(\theta_S,k)\mathbf{D}(\theta_S,k) = \text{corr}_{\text{super}}(\theta_S,k). \end{aligned} \tag{19}$$

$\text{corr}_{\text{super}}(\theta,k)$ is to be determined in the following. Assuming that a desired signal s arrives from θ_s, that is, $\mathbf{Y}(k) = \mathbf{D}(\theta_s,k)S(k)$ and consequently $|Y_l(k)| = \alpha_l^{\text{phy}}(\theta_S,k)|S(k)|$, $|Y_r(k)| = \alpha_r^{\text{phy}}(\theta_S,k)|S(k)|$. Also assume that the coefficient vector $\mathbf{W}$ has been designed for this angle θ_s. Then, after insertion of (17) into (18), we obtained

$$G_{\text{super}}(k) = \frac{|\text{corr}_{\text{super}}(\theta_s,k)S(k)|}{\alpha_l^{\text{phy}}(\theta_s,k)|S(k)| + \alpha_r^{\text{phy}}(\theta_s,k)|S(k)|}. \tag{20}$$

The demand $G_{\text{super}} \stackrel{!}{=} 1$ for a signal from θ_S yields

$$\text{corr}_{\text{super}}(\theta_s,k) = \alpha_l^{\text{phy}}(\theta_s,k) + \alpha_r^{\text{phy}}(\theta_s,k). \tag{21}$$

The design of the superdirective coefficient vector $\mathbf{W}(\theta_s,k)$ for frequency bin k and desired angle θ_s with tradeoff factor μ_s is therefore

$$\begin{aligned} \mathbf{W}(\theta_s,k) = &\left(\alpha_l^{\text{phy}}(\theta_s,k) + \alpha_r^{\text{phy}}(\theta_s,k)\right) \\ &\cdot \frac{(\mathbf{\Phi}_{MM}^{-1}(k) + \mu_s\mathbf{I})\mathbf{D}(\theta_s,k)}{\mathbf{D}^H(\theta_s,k)(\mathbf{\Phi}_{MM}^{-1}(k) + \mu_s\mathbf{I})\mathbf{D}(\theta_s,k)}. \end{aligned} \tag{22}$$

3.1.3. Directivity evaluation

Now, the performance of the beamformer is evaluated in terms of spatial directivity and directivity gain plots. The directivity pattern $\Psi(\theta_s,\theta,k)$ is defined as the squared transfer function for a signal that arrives from a certain spatial direction θ if the beamformer is designed for angle θ_s.

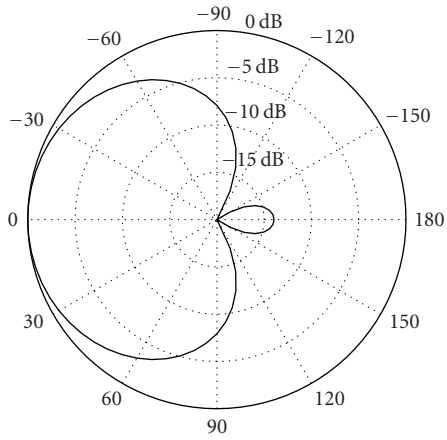

Figure 7: Beam pattern (frequency-independent) of typical delay-and-subtract beamformer applied in a single behind-the-ear device. Parameters are microphone distance: $d_{\text{mic}} = 0.01$ m and internal delay of beamformer for rear microphone signal: $\tau = (2/3) \cdot (d_{\text{mic}}/c)$.

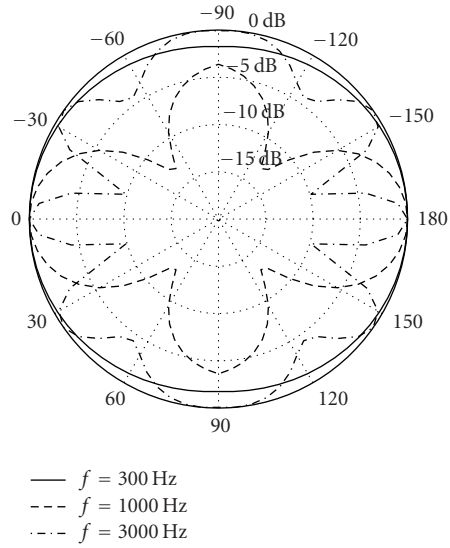

Figure 8: Beam pattern $\Psi(\theta_s = 0^\circ, \theta, f)$ of superdirective binaural input-output beamformer for DFT bins corresponding to 300 Hz, 1000 Hz, and 3000 Hz (special case of broadside delay-and-sum beamformer).

As a reference, Figure 7 plots the directivity pattern of a typical hearing aid first-order delay-and-subtract beamformer integrated, for example, in a single behind-the-ear device. In the example, the rear microphone signal is delayed 2/3 of the time, which a source from $\theta_S = 0^\circ$ needs to travel from the front to the rear microphone, and is subtracted from the front microphone signal. The approach is limited to low microphone distances, typically lower than 2 cm, to avoid spectral notches caused by spatial aliasing. Also, the lower-frequency region needs to be excluded, because of its low signal-to-microphone-noise ratio caused by the subtract operation.

The behind-the-ear endfire beamformer can greatly attenuate signals from behind the hearing-impaired subjects but cannot differentiate between left- and right-ear sides. The dual-channel input-output beamformer behaves the opposite. Due to the binaural microphone position, the directivity shows a front-rear ambiguity.

In the case of the stereo input-output binaural beamformer, the directivity pattern is determined by the squared weight factors G^2_{super}, according to (18), that are applied to the spectral coefficients

$$\Psi(\theta_s, \theta, k)/\text{dB} = 20 \log_{10} \left(G_{\text{super}}(\theta_s, \theta, k) \right), \tag{23}$$

which can be written as

$$\Psi(\theta_s, \theta, k)/\text{dB} = 20 \log_{10} \left(\frac{\left| \mathbf{W}^H(\theta_s, k) \mathbf{D}(\theta, k) \right|}{\alpha_{\text{l}}^{\text{phy}}(\theta, k) + \alpha_{\text{r}}^{\text{phy}}(\theta, k)} \right). \tag{24}$$

Figure 8 shows the beam pattern for the desired direction $\theta_s = 0^\circ$. In this case, the superdirective design leads to the special case of a simple delay-and-sum beamformer, that is, a broadside array with two elements. Thus, the achieved directivity is low at low frequencies. At higher frequencies, the phase difference generated by a lateral source becomes significant and causes a narrow main lobe along with sidelobes due to spatial aliasing. However, the side lobes are of lower magnitude due to the different amplitude transfer functions.

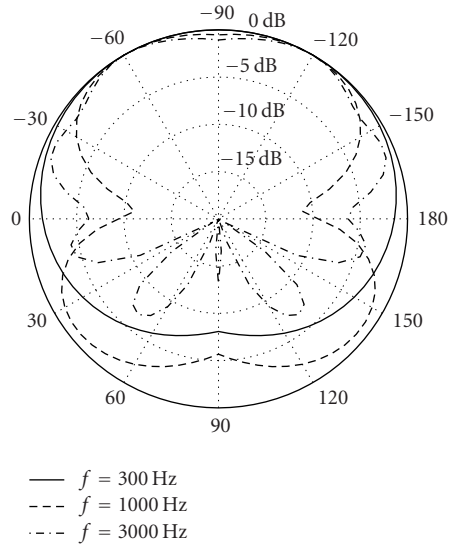

Figure 9: Beam pattern $\Psi(\theta_s = -60^\circ, \theta, f)$ of superdirective binaural input-output beamformer for DFT bins corresponding to 300 Hz, 1000 Hz, and 3000 Hz (design parameter $\mu_s = 10$, which corresponds to a low degree of superdirectivity).

Figure 9 shows the directivity pattern for the desired angle $\theta_s = -60°$. The design parameter was set to $\mu_s = 10$, that is, low degree of superdirectivity. Hence, approximately a delay-and-sum beamformer with amplitude modification is obtained. Because of significant interaural differences, the directivity is much higher compared to that of the frontal desired direction, especially signals from the opposite side will be highly attenuated. The main lobe is comparably large at all plotted frequencies.

Figure 10 shows that the directivity if the design parameter is adjusted for a maximum degree of superdirectivity, that is, $\mu_s = 0$. As expected, the directivity further increases especially for low frequencies and the main lobe becomes more narrow.

To measure the directivity of the dual-channel input-output system in a more compact way, the overall gain can be considered. It is defined as the ratio of the directivity towards the desired direction θ_s and the average directivity. As only the horizontal plane is considered, the average directivity can be obtained by averaging over $0 \le \theta < 2\pi$ with equidistant angles at a resolution of 5 degrees, that is, $N = 72$. The directivity gain DG is given as

$$\mathrm{DG}(\theta_s, k) = \frac{\Psi(\theta_s, \theta_s, k)}{(1/N)\sum_{n=1}^{N} \Psi(\theta_s, \theta_n, k)}. \tag{25}$$

Figure 11 depicts the directivity gain as a function of the frequency for different desired directions with low degree of superdirectivity. The gain increases from 0 dB to up to 4–5.5 dB below 1 kHz depending on the desired direction. Since the microphone distance between the ears is comparably high with 17.5 cm, phase ambiguity causes oscillations in the frequency plot.

Towards higher frequencies, the interaural amplitude differences gain more influence on the directivity gain. For $\theta_S = 0°$, unbalanced amplitudes of the spectral coefficients of left- and right-ear sides decrease the gain in (18) towards high frequencies due to the simple addition of the coefficients in the numerator, while the denominator is dominated by one input spectral amplitude for a lateral signal. For lateral desired directions however, the interaural amplitude differences are exploited in the numerator with (18) resulting in directivity gain values up to 5 dB.

Figure 12 shows the directivity for the case that the coefficients are designed with respect to high degree of superdirectivity. Now, even at low frequencies, a gain of up to nearly 6 dB can be accomplished.

3.2. Multichannel postfilter

The superdirective beamformer produces the best possible signal-to-noise ratio for a narrowband input by minimizing the noise power subject to the constraint of a distortionless response for a desired direction [20]. It can be shown [21] that the best possible estimate in the MMSE sense is the multichannel Wiener filter, which can be factorized into the superdirective beamformer followed by a single-channel Wiener postfilter. The optimum weight vector $\mathbf{W}_{\mathrm{opt}}(k)$ that

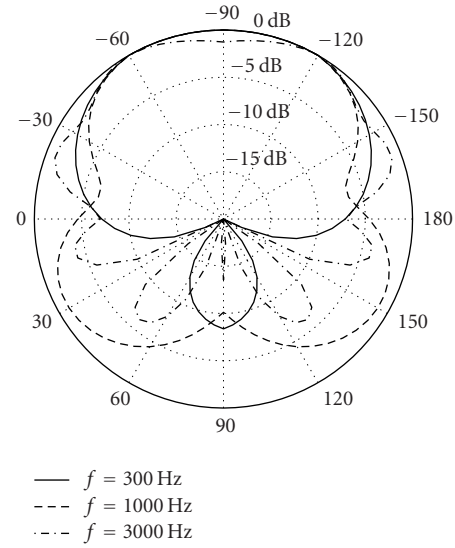

Figure 10: Beam pattern $\Psi(\theta_s = -60°, \theta, f)$ of superdirective binaural input-output beamformer for DFT bins corresponding to 300 Hz, 1000 Hz, and 3000 Hz (design parameter $\mu_s = 0$, i.e., maximum degree of superdirectivity).

transforms the noisy input vector $\mathbf{Y}(k) = \mathbf{S}(k) + \mathbf{N}(k)$ into the best scalar estimate $S(k)$ is given by

$$\mathbf{W}_{\mathrm{opt}}(k) = \underbrace{\frac{\Phi_{\mathrm{ss}}(k)}{\Phi_{\mathrm{ss}}(k) + \Phi_{\mathrm{nn}}(k)}}_{\text{Wiener filter}} \cdot \underbrace{\frac{\Phi_{MM}^{-1}(k)\mathbf{D}(\theta_S, k)}{\mathbf{D}^H(\theta_S, k)\Phi_{MM}^{-1}(k)\mathbf{D}(\theta_S, k)}}_{\text{MVDR beamformer}}. \tag{26}$$

Possible realizations of the Wiener postfilter are based on the observation that the noise correlation between the microphone signals is low [22, 23]. An improved performing algorithm is presented in [21], where the transfer function H_{post} of the postfilter is estimated by the ratio of the output power spectral density Φ_{zz} and the average input power spectral density of the beamformer Φ_{yy} with

$$H_{\mathrm{post}}(k) = \frac{\Phi_{zz}(k)}{\Phi_{yy}(k)} = \frac{\Phi_{zz}(k)}{(1/M)\sum_{i=1}^{M} \Phi_{ii}(k)}. \tag{27}$$

3.2.1. Adaptation to dual-channel input-output beamformer

In the following, the dual-channel input-output beamformer is extended by also adapting the formulation of the postfilter according to (27) into the spectral weighting framework.

The goal is to find spectral weights with similar requirements as for the beamformer gains. Again, only one postfilter weight is to be determined for both left- and right-ear spectral coefficients in order not to disturb the original spatial impression, that is, the interaural amplitude and phase differences. Secondly, a source from a desired direction θ_S should pass unfiltered, that is, the spectral postfilter weight for a signal from that direction should be one.

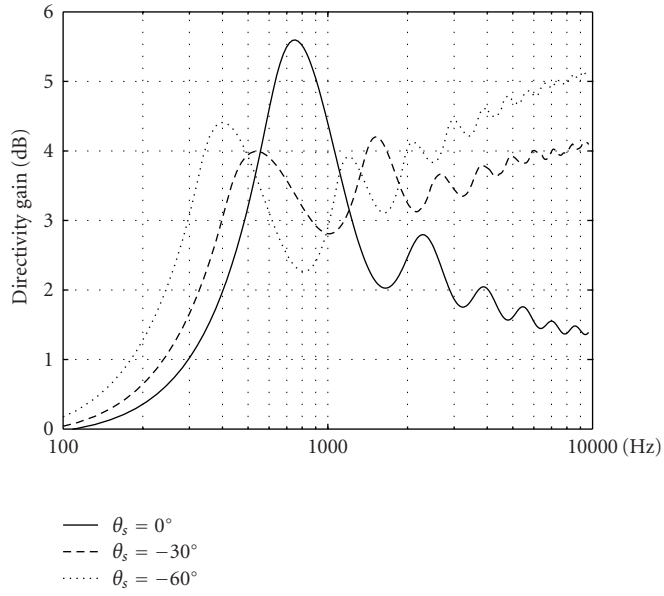

FIGURE 11: Directivity gain according to (25) of superdirective stereo input-output beamformer for desired direction $\theta_s = 0°$ (solid), $\theta_s = 30°$ (dashed), and $\theta_s = -60°$ (dotted) for low degree of superdirectivity ($\mu_s = 10$).

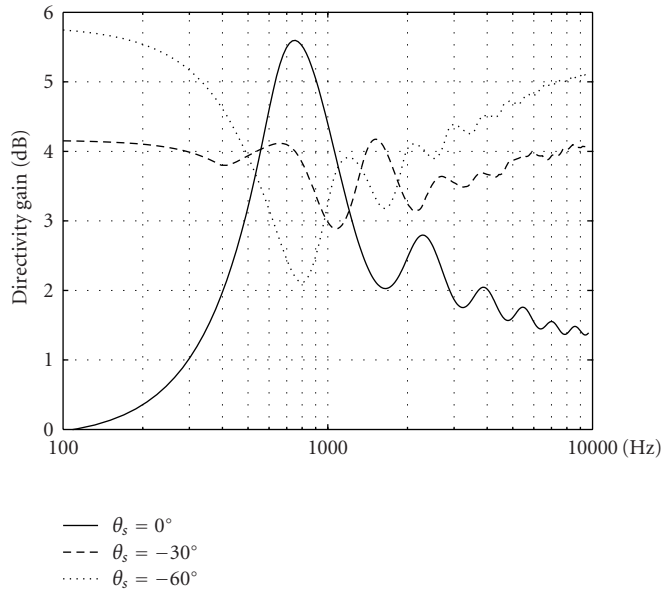

FIGURE 12: Directivity gain according to (25) of superdirective stereo input-output beamformer for desired direction $\theta_s = 0°$ (solid), $\theta_s = -30°$ (dashed), and $\theta_s = -60°$ (dotted) for high degree of superdirectivity ($\mu_s = 0$).

In analogy to the optimal MMSE estimate according to (26) weights, G_{post} postfilter weights are multiplicatively combined with the beamformer weights G_{super} according to (18) to the resulting weights $G(k)$:

$$G(k) = G_{\text{super}}(k) \cdot G_{\text{post}}(k). \tag{28}$$

To realize the postfilter according to (27) in the spectral weighting framework, weights are calculated with

$$G_{\text{post}}(k) = \frac{2\left|Z(k)\right|^2}{\left|Y_{\text{l}}(k)\right|^2 + \left|Y_{\text{r}}(k)\right|^2} \cdot \text{corr}_{\text{post}}\left(\theta_S, k\right). \tag{29}$$

The desired angle- and frequency-dependent correction factor $\text{corr}_{\text{post}}$ will guarantee a distortionless response towards a signal from the desired direction θ_S. For a signal from θ_S, (29) can be rewritten as

$$G_{\text{post}}(k) = \frac{2\left|\mathbf{W}^H(\theta_S,k)\mathbf{D}(\theta_S,k)S(k)\right|^2}{\left|Y_{\text{l}}(k)\right|^2 + \left|Y_{\text{r}}(k)\right|^2} \cdot \text{corr}_{\text{post}}(\theta_s,k). \tag{30}$$

Since the beamformer coefficients have been designed with respect to $\mathbf{W}(\theta_S,k)^H\mathbf{D}(\theta_S,k) = \alpha_{\text{l}}^{\text{phy}}(\theta_S,k) + \alpha_{\text{r}}^{\text{phy}}(\theta_S,k)$, the spectral weights can be reformulated as

$$\begin{aligned} G_{\text{post}}(k) &= \frac{2\left|S(k)\right|^2\left(\alpha_{\text{l}}^{\text{phy}}(\theta_s,k) + \alpha_{\text{r}}^{\text{phy}}(\theta_s,k)\right)^2}{\left(\alpha_{\text{l}}^{\text{phy}}(\theta_s,k)\right)^2\left|S(k)\right|^2 + \left(\alpha_{\text{r}}^{\text{phy}}(\theta_s,k)\right)^2\left|S(k)\right|^2} \\ &\quad \cdot \text{corr}_{\text{post}}(\theta_s,k) \\ &= \frac{2\left(\alpha_{\text{l}}^{\text{phy}}(\theta_s,k) + \alpha_{\text{r}}^{\text{phy}}(\theta_s,k)\right)^2}{\left(\alpha_{\text{l}}^{\text{phy}}(\theta_s,k)\right)^2 + \left(\alpha_{\text{r}}^{\text{phy}}(\theta_s,k)\right)^2} \cdot \text{corr}_{\text{post}}(\theta_s,k). \end{aligned} \tag{31}$$

Demanding $G_{\text{post}}(k) = 1$ gives

$$\text{corr}_{\text{post}}(\theta_S,k) = \frac{\left(\alpha_{\text{l}}^{\text{phy}}(\theta_s,k)\right)^2 + \left(\alpha_{\text{r}}^{\text{phy}}(\theta_s,k)\right)^2}{2\left(\alpha_{\text{l}}^{\text{phy}}(\theta_s,k) + \alpha_{\text{r}}^{\text{phy}}(\theta_s,k)\right)^2}. \tag{32}$$

Consequently, after insertion of (32) into (29), the resulting postfilter weight calculation for combination with the dual-channel input-output beamformer according to (18), (22) can finally be written as

$$\begin{aligned} G_{\text{post}}(k) &= \frac{\left|Z(k)\right|^2}{\left|Y_{\text{l}}(k)\right|^2 + \left|Y_{\text{r}}(k)\right|^2} \\ &\quad \cdot \frac{\left(\alpha_{\text{l}}^{\text{phy}}(\theta_s,k)\right)^2 + \left(\alpha_{\text{r}}^{\text{phy}}(\theta_s,k)\right)^2}{\left(\alpha_{\text{l}}^{\text{phy}}(\theta_s,k) + \alpha_{\text{r}}^{\text{phy}}(\theta_s,k)\right)^2}. \end{aligned} \tag{33}$$

Again, to avoid amplification, the postfilter weight should be upper-limited to one. Figure 13 shows a block diagram of the resulting system with stereo input-output beamformer plus Wiener postfilter in the DFT domain. After the dual-channel beamformer processing, the postfilter weights are calculated according to (33) and are multiplicatively combined with the beamformer gains according to (28). The dual-channel output spectral coefficients $\hat{S}_{\text{l}}(k)$, $\hat{S}_{\text{r}}(k)$ are generated by multiplication of left- and right-side input coefficients $Y_{\text{l}}(k)$, $Y_{\text{r}}(k)$ with the respective weight $G(k)$. Finally, the binaural enhanced time signals are resynthesized using IDFT and overlap add.

4. PERFORMANCE EVALUATION

In this section, the performance of the dual-channel input-output beamformer with postfilter is evaluated by a multitalker situation in a real environment. The performance of

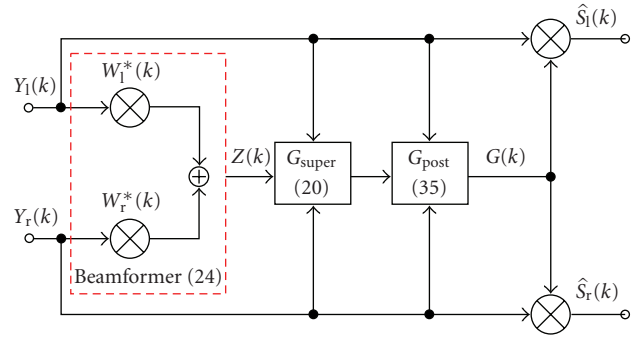

Figure 13: Superdirective input-output beamformer with postfiltering.

the system depends on various parameters of the real environment in which it is applied in. First of all, the unknown HRTFs of the target person, for example, a hearing-impaired person will deviate from the binaural model or from a preevaluated HRTF database. The noise reduction performance of the system, that relies on the erroneous database, will thus decrease. Secondly, reverberation will degrade the performance.

In order to evaluate the performance of the beamformer in a realistic environment, recordings of speech sources were made in a conference room (reverberation time $T_0 \approx 800$ ms) with two source-target distances as depicted in Figure 14. All recordings were performed using a head measurement system (HMS) II dummy head with binaural hearing aids attached above the ears without taking special precautions to match exact positions. In the first scenario, the speech sources were located within a short distance of 0.75 m to the head. Also, the head was located at least 2.2 m away from the nearest wall. In the second scenario, the loudspeakers were moved 2 m away from the dummy head. Thus, the recordings from the two scenarios differ significantly in the direct-to-reverberation ratio. In the experiments, a desired speech source s_1 arrives from angle θ_{S_1} towards which the beamformer is steered and an interfering speech signal s_2 arrives from angle θ_{S_2}. The superdirectivity tradeoff factor was set to $\mu_s = 0.5$.

Firstly, the spectral attenuation of the desired and unwanted speech for one source-interferer configuration, $\theta_{S_1} = -60°$, $\theta_{S_2} = 30°$, at a distance of 0.75 m from the head is illustrated. The theoretical behavior of the beamformer without postfilter for that specific scenario is indicated by Figure 12. The desired source should pass unfiltered, while the interferer from $\theta_{S_2} = 30°$ should be frequency-dependently attenuated. A lower degree of attenuation is expected at $f = 1000$ Hz due to spatial aliasing.

Figure 15 plots the measured results in the real environment. The attenuation of the interfering speech source varies mainly between 2–7 dB, while the desired source is also attenuated by 1–2 dB, more or less constant over the frequency. At frequencies below 700 Hz, the superdirectivity already allows a significant attenuation of the interferer. Due to spatial aliasing, the attenuation difference is very low around 1200 Hz. At

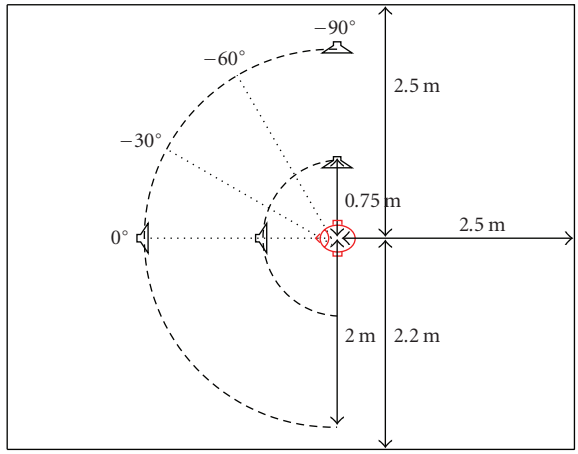

FIGURE 14: Recording setup inside conference room (reverberation time $T_0 \approx 800$ ms).

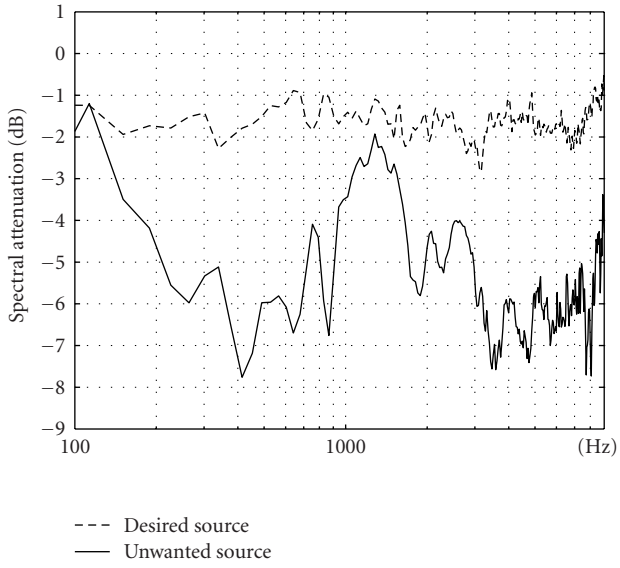

FIGURE 15: Spectral attenuation of desired source from $\theta_{S_1} = -60°$ and unwanted source from $\theta_{S_2} = 30°$.

high frequencies, the attenuation difference rises again because the beamformer can exploit the significance of the interaural amplitude differences here.

4.1. Intelligibility-weighted improvement

To judge the benefit of the frequency-dependent noise reduction gain like exemplarily shown in Figure 15, a speech intelligibility-weighted noise reduction gain is applied in the following.

For its calculation, a spectral noise reduction gain is determined as the difference between the power spectral density attenuation of the undesired source subtracted from the

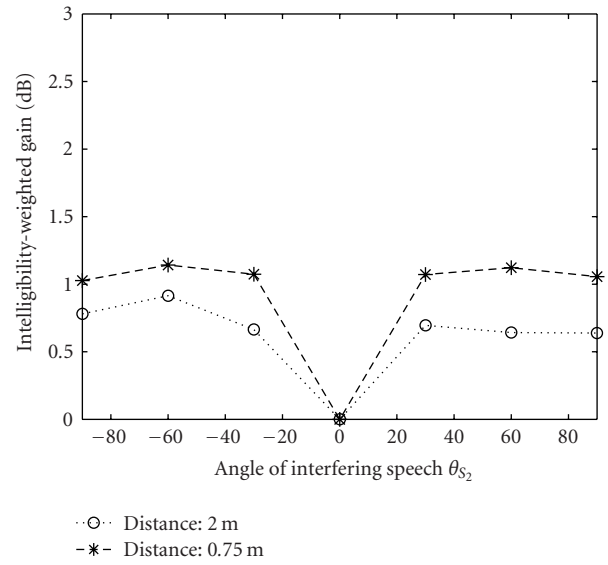

FIGURE 16: Intelligibility-weighted gain according to (35) of superdirective stereo input-output beamformer for speech from $\theta_{S_1} = 0°$ and interfering speech from other directions (distance to dummy head 0.75 m and 2 m, resp.).

attenuation of the desired source, that is,

$$\mathrm{AG}(\omega) = 10\log_{10}\frac{\Phi_{s_2s_2}(\omega)}{\Phi_{\hat{s}_2\hat{s}_2}(\omega)} - 10\log_{10}\frac{\Phi_{s_1s_1}(\omega)}{\Phi_{\hat{s}_1\hat{s}_1}(\omega)}. \quad (34)$$

Here, $\Phi_{s_1s_1}(\omega)$ is the power spectral density of the desired speech and $\Phi_{s_2s_2}(\omega)$ that of the unwanted speech. $\Phi_{\hat{s}_1\hat{s}_1}$ is the power spectral density of the remaining components of s_1 after beamformer processing according to Figure 6 or Figure 13 and $\Phi_{\hat{s}_2\hat{s}_2}$ is that of the remaining components of s_2. To judge the intelligibility improvement that is achieved by the frequency-dependent reduction gain, $\mathrm{AG}(\omega)$ is grouped into critical bands $\overline{\mathrm{AG}}(\omega_b)$ and is weighted according to the ANSI's speech intelligibility index standard [24].

The intelligibility-weighted gain can then be calculated by

$$\mathrm{IWG} = \sum_{b=1}^{K} a_b \overline{\mathrm{AG}\,(w_b)}, \quad (35)$$

where a_b are the weights according to [24] in the respective critical frequency band b. After evaluation of (34) and (35) for the left- and right-microphone signals, the intelligibility-weighted gains of left and right ears are averaged.

Figure 16 plots the performance of the superdirective binaural input-output beamformer in terms of speech intelligibility-weighted gain for a desired speech source from 0° and speech interferers from variable directions. The two plots in Figure 16 show the gain when all sources were located 0.75 m and 2 m away from the dummy head.

The binaural input-output superdirective beamformer only delivers about 0.6–1.2 dB intelligibility-weighted improvement because of its comparably low directivity towards the frontal direction as depicted in Figure 8. Due to the

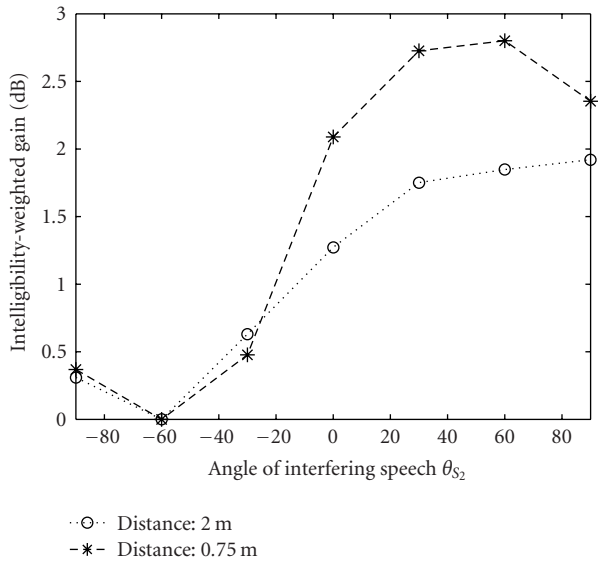

Figure 17: Intelligibility-weighted gain according to (35) of superdirective stereo input-output beamformer ($\mu_s = 0$) for speech from $\theta_{S_1} = -60°$ and speech interferer from other directions (distance to dummy head 0.75 m or 2 m, resp.).

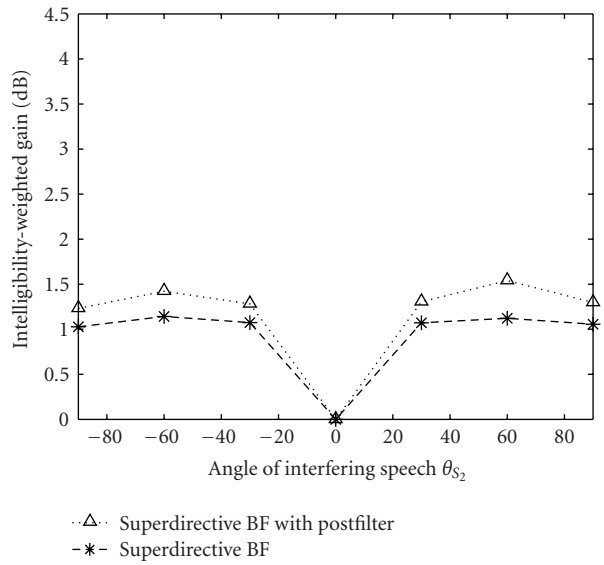

Figure 18: Intelligibility-weighted gain according to (35) of superdirective stereo input-output beamformer with and without postfilter for speech from $\theta_{S_1} = 0°$ and speech interferer from other directions (distance to dummy head 0.75 m).

decreased direct-to-reverberation ratio, the overall performance is all the time lower for the 2 m distance from the dummy head.

Figure 17 plots the performance of the superdirective binaural input-output beamformer when the desired signal arrives from $\theta = -60°$. Results for desired sources and interfering sources from a distance of 0.75 m and 2 m from the dummy head are shown.

For low angular differences to the desired direction, that is, $\theta = -90°$ and $\theta = -30°$, the gain mostly stays below 1 dB. When the interfering speech source is located at the other side, the superdirective beamformer achieves the highest intelligibility-weighted gain, whose value is nearly 3 dB.

Due to the decreased direct-to-reverberation ratio at the distance of 2 m, the gain remains below 2 dB.

Now, the influence of the additional binaural postfilter for the superdirective input-output beamformer is examined. Figures 18 and 19 show the intelligibility-weighted noise reduction gain of the superdirective stereo input-output beamformer with and without postfilter for desired directions $\theta_S = 0°$ and $\theta_S = -60°$.

The postfilter provides an additional intelligibility-weighted gain which is proportional to that obtained by the superdirective binaural beamformer. For a desired signal from a lateral direction, the absolute improvement is much larger than for the frontal direction.

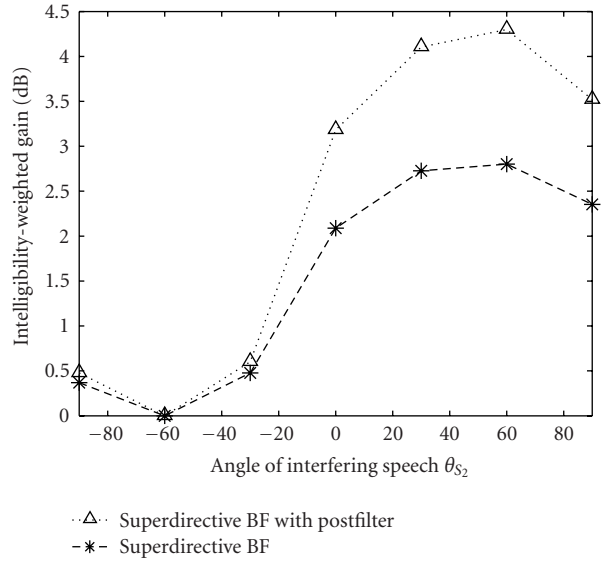

Figure 19: Intelligibility-weighted gain according to (35) of superdirective stereo input-output beamformer (μ_s) with and without postfilter for speech from $\theta_{S_1} = -60°$ and speech interferer from other directions (distance to dummy head 0.75 m).

4.2. Improvements for both ear sides

Figure 20 plots the intelligibility-weighted gains independently for both ear sides when applying the beamformer with postfiltering. The left part of Figure 20 shows the performance for a lateral desired direction, $\theta_{S_1} = -60°$, while the right part plots that for the frontal direction $\theta_{S_1} = 0°$.

Due to the formulation of the beamformer with postfilter as a single spectral weight for both left- and right-ear spectral coefficients, the proposed algorithm almost delivers the same improvement for the good- and bad-ear sides.

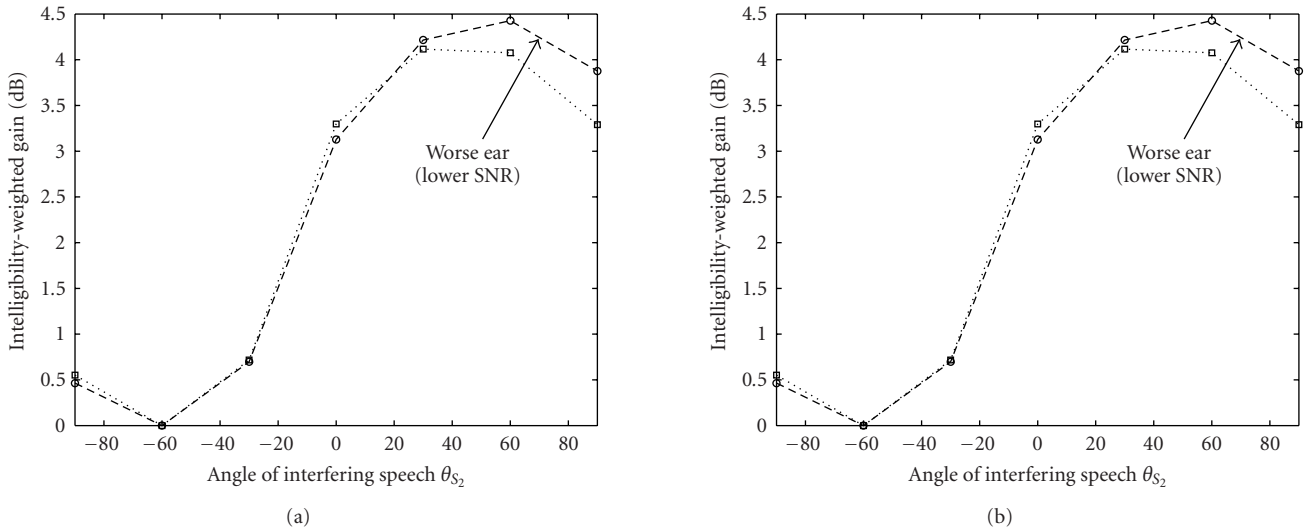

FIGURE 20: Intelligibility-weighted gain for left ear (dashed line) and right ear (dotted line): (a) $\theta_{S_1} = -60°$; (b) $\theta_{S_1} = 0°$.

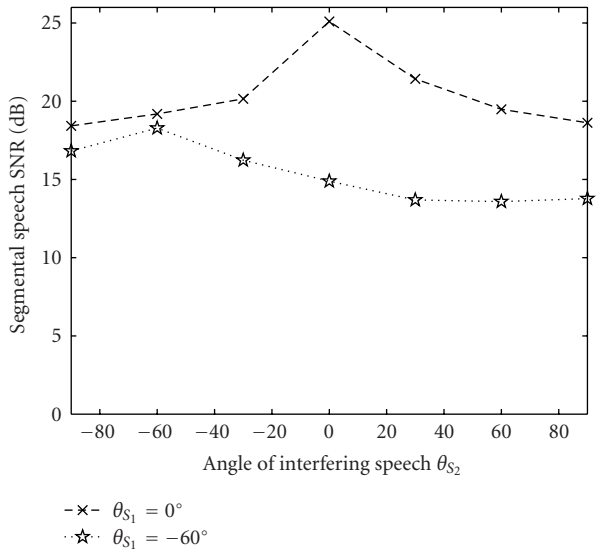

FIGURE 21: Segmental speech SNR of target signal for two different desired directions.

The improvement at the ear side with the lower SNR is only slightly higher.

4.3. Speech quality of target source

To measure the speech quality of the target signal after processing, the segmental SNR is measured. Again, the target speech was mixed with interferers from other directions. The speech quality was then determined by applying the resulting filter on the target signal alone and calculating the segmental speech SNR between input and filtered output. Figure 21 plots the segmental speech SNR for the two considered desired angles, $\theta_{S_1} = -60°$ and $\theta_{S_1} = 0°$. The speech quality of the target source is somewhat degraded due its attenuation caused by imperfect knowledge of the HRTFs as also depicted in Figure 15, however the speech SNR is always high at 15–25 dB. For the lateral desired direction, the target attenuation is always higher than for the frontal direction.

5. CONCLUSION

We have presented a dual-channel input-output algorithm for binaural speech enhancement, which consists of a superdirective beamformer and a postfilter with an underlying binaural signal model, and consists of a simple spectral weighting scheme. The system perfectly preserves the interaural amplitude and phase differences and passes a source from a desired direction unfiltered in principle.

Instrumental measurements and informal listening tests indicate a significant attenuation of speech interferers in a real environment, while the target source is only slightly degraded. The amount of interference rejection depends on the spatial position of the desired and unwanted sources, a high rejection of interfering noise or speech can particularly be expected for lateral desired directions. The proposed algorithms can efficiently be realized in real time.

REFERENCES

[1] J. Bitzer and K. U. Simmer, "Superdirective microphone arrays," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. S. Brandstein and D. B. Ward, Eds., chapter 2, pp. 19–38, Springer, Berlin, Germany, 2001.

[2] E. N. Gilbert and S. P. Morgan, "Optimum design of directive antenna arrays subject to random variations," *Bell System Technical Journal*, vol. 34, pp. 637–663, 1955.

[3] M. Dörbecker, *Mehrkanalige Signalverarbeitung zur Verbesserung akustisch gestörter Sprachsignale am Beispiel elektronischer Hörhilfen*, Ph.D. thesis, Rheinisch-Westfälische Technische Hochschule Aachen, Aachen, Germany, 1998.

[4] S. Doclo and M. Moonen, "Design of broadband beamformers robust against gain and phase errors in the microphone array characteristics," *IEEE Transactions on Signal Processing*, vol. 51, no. 10, pp. 2511–2526, 2003.

[5] W. Tager, "Near field superdirectivity (NFSD)," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '98)*, vol. 4, pp. 2045–2048, Seattle, Wash, USA, May 1998.

[6] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Transactions on Antennas and Propagation*, vol. 30, no. 1, pp. 27–34, 1982.

[7] O. Hoshuyama and A. Sugiyama, "Robust adaptive beamforming," in *Microphone Arrays*, M. S. Brandstein and D. B. Ward, Eds., pp. 87–110, Springer, Berlin, Germany, 2001.

[8] W. Herbordt and W. Kellermann, "Adaptive beamforming for audio signal acquisition," in *Adaptive Signal Processing—Applications to Real-World Problems*, J. Benesty and Y. Huang, Eds., pp. 155–194, Springer, Berlin, Germany, 2003.

[9] J. G. Desloge, W. M. Rabinowitz, and P. M. Zurek, "Microphone-array hearing aids with binaural output. I. Fixed-processing systems," *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 6, pp. 529–542, 1997.

[10] G. W. Elko, "Spatial coherence functions for differential microphones in isotropic noise fields," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. S. Brandstein and D. B. Ward, Eds., chapter 4, pp. 61–86, Springer, Berlin, Germany, 2001.

[11] V. Hamacher, "Comparison of advanced monaural and binaural noise reduction algorithms for hearing aids," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '02)*, vol. 4, pp. 4008–4011, Orlando, Fla, USA, May 2002.

[12] V. Hohmann, J. Nix, G. Grimm, and T. Wittkop, "Binaural noise reduction for hearing aids," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '02)*, vol. 4, pp. 4000–4003, Orlando, Fla, USA, May 2002.

[13] T. Wittkop, S. Albani, V. Hohmann, J. Peissig, W. S. Woods, and B. Kollmeier, "Speech processing for hearing aids: noise reduction motivated by models of binaural interaction," *Acustica United with Acta Acustica*, vol. 83, no. 4, pp. 684–699, 1997.

[14] D. R. Campbell and P. W. Shields, "Speech enhancement using sub-band adaptive Griffiths–Jim signal processing," *Speech Communication*, vol. 39, no. 1-2, pp. 97–110, 2003, Special issue on speech processing for hearing aids.

[15] D. P. Welker, J. E. Greenberg, J. G. Desloge, and P. M. Zurek, "Microphone-array hearing aids with binaural output. II. A two-microphone adaptive system," *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 6, pp. 543–551, 1997.

[16] T. Lotter, *Single and multimicrophone speech enhancement for hearing aids*, Ph.D. thesis, Rheinisch-Westfälische Technische Hochschule Aachen, Aachen, Germany, 2004.

[17] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, Cambridge, Mass, USA, 1983.

[18] B. Gardner and K. Martin, "HRTF measurement of a KEMAR dummy-head microphone," Tech. Rep. 280, MIT Media Laboratory Perceptual Computing, Davos, Switzerland, May 1994, http://sound.media.mit.edu/KEMAR.html.

[19] C. P. Brown and R. O. Duda, "A structural model for binaural sound synthesis," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 5, pp. 476–488, 1998.

[20] R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays*, John Wiley & Sons, New York, NY, USA, 1980.

[21] K. U. Simmer, J. Bitzer, and C. Marro, "Post-filtering techniques," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. S. Brandstein and D. B. Ward, Eds., chapter 3, pp. 39–60, Springer, Berlin, Germany, 2001.

[22] R. Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '88)*, vol. 5, pp. 2578–2581, New York, NY, USA, April 1988.

[23] C. Marro, Y. Mahieux, and K. U. Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 3, pp. 240–259, 1998.

[24] ANSI S3.5-1997 American National Standards Institute, "Methods for Calculation of the Speech Intelligibility Index," ANSI S3.5-1997, 1997.

Thomas Lotter received the Dipl. Ing. degree in electrical engineering in 2000 from the Aachen University of Technology, RWTH Aachen. He received the Ph.D. degree from the RWTH Aachen University in 2004 after working at the Institute of Communication Systems and Data Processing in the area of single and multimicrophone speech enhancement. In 2004, he joined Siemens Audiological Engineering Group in Erlangen, Germany, with focus on wireless hearing aid applications. His main interests include speech enhancement, signal processing for wireless systems, wireless standards, and audio coding.

Peter Vary received the Dipl. Ing. degree in electrical engineering in 1972 from the University of Darmstadt, Darmstadt, Germany. In 1978, he received the Ph.D. degree from the University of Erlangen-Nuremberg, Germany. In 1980, he joined Philips Communication Industries (PKI), Nuremberg, where he became the Head of the Digital Signal Processing Group. Since 1988, he has been a Professor at Aachen University of Technology, Aachen, Germany, and the Head of the Institute of Communication Systems and Data Processing. His main research interests are speech coding, channel coding, error concealment, adaptive filtering for acoustic echo cancellation and noise reduction, and concepts of mobile radio transmission.