A STEREO INPUT-OUTPUT SUPERDIRECTIVE BEAMFORMER FOR DUAL CHANNEL NOISE REDUCTION

Thomas Lotter, Bastian Sauert, and Peter Vary

Institute of Communication Systems and Data Processing (ind)
RWTH Aachen University, Templergraben 55, D-52056 Aachen, Germany
E-mail: {lotter | sauert | vary}@ind.rwth-aachen.de

ABSTRACT

This contribution presents a stereo input-output beamformer for dual channel noise reduction. The computationally very efficient beamformer adapts superdirective filter design techniques to binaural input signals to optimally enhance signals from a given spatial direction. The beamformer outputs a stereo enhanced signal and thus preserves the spatial impression. Experiments in a real environment using a dummy head and various speech sources indicate that the proposed algorithm is capable of improving the speech intelligibility significantly.

1. INTRODUCTION

Hearing aids suffer from reduced intelligibility and quality of speech in the presence of noise or multiple speech sources. Single microphone speech enhancement algorithms, e.g. the Minimum Mean Square Error (MMSE) estimator of the speech Discrete Fourier Transform (DFT) amplitudes [1] can significantly improve the speech quality, but show very little improvements in terms of speech intelligibility or reduction of instationary speech interferers. With multiple microphones spatial information can be exploited by fixed (e.g. [2]) or adaptive (e.g. [3], [4]) beamforming to reduce interferers and reverberation causing only little speech distortion. Beamforming for binaural input signals, i.e. signals recorded by single microphones at the left and right ear, has found significantly less attention than beamformers for microphone arrays. In [5] a system is proposed that suppresses lateral noise sources by comparing the interaural level and phase differences to reference values for the frontal direction. However, the system suffers from its susceptibility to reverberation. In [6], the Griffiths-Jim adaptive beamformer [3] has been applied to binaural noise reduction in subbands. However, the algorithm requires a voice activity detection (VAD) which will frequently fail especially at low signal to noise ratios.

In this contribution we present a robust fixed beamformer for reduction of acoustical background noise or interfering speech with binaural microphones. The well known superdirective design criterion (e.g. [2]) is combined with an appropriate binaural model and a postprocessing scheme is introduced in order to obtain an enhanced stereo signal instead of a mono output.

The remainder of the paper is organized as follows: Section 2 summarizes the bases for the proposed algorithm, i.e. the superdirective filter design and the applied binaural model. Section 3 introduces the proposed stereo input-output superdirective beamformer and finally in Section 4 the performance of the proposed beamformer is examined for different multi-taker scenarios in a real environment.

2. BASES FOR BINAURAL BEAMFORMER

The left part of Figure 1 shows a time signal s arriving at the left and right ear microphones from the angle θ_s in the horizontal plane. The time signals at the left and right microphone are denoted by y_l, y_r. The microphone signal spectra can be expressed by the head related transfer functions (HRTFs) towards left and right ear $D_\mathrm{l}(\omega)$, $D_\mathrm{r}(\omega)$. 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

$$Y_\mathrm{l}(\omega_k) = D_\mathrm{l}(\omega_k)S(\omega_k) \quad \text{and} \quad Y_\mathrm{r}(\omega_k) = D_\mathrm{r}(\omega_k)S(\omega_k). \quad (1)$$

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

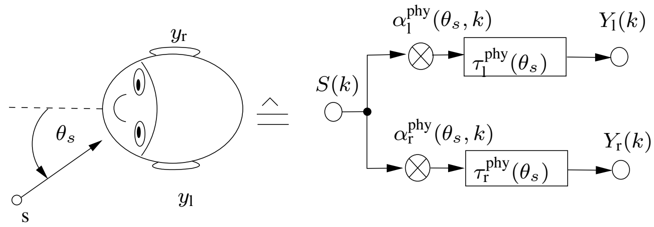

Figure 1: Acoustic transfer of source from θ_s towards the left and right ear.

The acoustic transfer vector $\boldsymbol{D} = [D_\mathrm{l}, D_\mathrm{r}]^T$ consists of delays τ depending on the angle of arrival and of frequency and angle dependent amplifications α due to the shadowing effect of the head.

$$\boldsymbol{D}(\theta_s, k) = [\alpha_\mathrm{l}(\theta_s, k)e^{-j\omega_k\tau_\mathrm{l}(\theta_s)}, \alpha_\mathrm{r}(\theta_s, k)e^{-j\omega_k\tau_\mathrm{r}(\theta_s)}]^T \quad (2)$$

2.1. Binaural Model

To obtain $\boldsymbol{D}$ for given values of θ, the parametric binaural model from [7], which consists of cascaded blocks for interaural time differences (ITDs) and interaural intensity differences (IIDs), is applied. For both ITDs and IIDs a spherical head of radian a is assumed resulting in the transfer function

$$D_\mathrm{mod}(\theta, \omega) = \frac{1 + j\frac{\gamma_\mathrm{mod}(\theta)\omega}{2\omega_0}}{1 + j\frac{\omega}{2\omega_0}} \cdot e^{-j\omega\tau_\mathrm{mod}(\theta)} \quad (3)$$

with $\omega_0 = \frac{c}{a}$, where c is the speed of sound. The model is determined by the angle dependent parameters γ_mod and τ_mod with

$$\gamma_\mathrm{mod}(\theta) = (1 + \beta_\mathrm{min}/2) + (1 - \beta_\mathrm{min}/2)\cos\left(\frac{\theta - \pi/2}{\theta_\mathrm{min}} 180^\circ\right) \quad (4)$$

$$\tau_\mathrm{mod}(\theta) = \begin{cases} -\frac{a}{c}\cos(\theta - \pi/2) & ; \quad -\frac{\pi}{2} \leq \theta < 0 \\ \frac{a}{c}|\theta| & ; \quad 0 \leq \theta < \frac{\pi}{2} \end{cases} . \quad (5)$$

Empirically the parameters are set to $\beta_{\text{min}} = 0.1$, $\theta_{\text{min}} = 150^\circ$ and $a = 0.0875\text{m}$.

2.2. DFT domain Superdirective Beamforming

A superdirective beamformer can efficiently be realized in the DFT domain. Let $y_m(i)$ denote the m-th microphone signal sampled at regular time instances i. After segmentation into frames of length L and windowing with a function $h(i)$, e.g. Hann window, the DFT coefficient of microphone m, frame λ and frequency bin k is calculated with:

$$Y_m(\lambda, k) = \sum_{i=0}^{L-1} y_m(\lambda R + i) h(i) e^{-j2\pi k i/L}, \quad m \in \{\text{l}, \text{r}\}. \qquad (6)$$

For the computation of the next DFT, the window is shifted by R samples. For the sake of brevity the index λ is omitted in the following. A fixed beamformer achieves a spatial directivity by summing up the spectral coefficients after complex multiplication with coefficients W_m^*. An enhanced spectral coefficient is thus obtained by

$$Z(k) = \boldsymbol{W}^H(k)\boldsymbol{Y}(k) = W_\text{l}^*(k)Y_\text{l}(k) + W_\text{r}^*(k)Y_\text{r}(k). \qquad (7)$$

The objective of the standard superdirective design of the weight vector $\boldsymbol{W}$ is to maximize the output SNR. This is achieved by minimizing the output energy with the constraint of an unfiltered signal from a desired direction. The Minimum Variance Distortionless Response (MVDR) approach can be written as (e.g. [2])

$$\min_{\boldsymbol{W}} \boldsymbol{W}^H \boldsymbol{\Phi} \boldsymbol{W} \ \text{ w.r.t } \boldsymbol{W}^H \boldsymbol{D} = 1, \qquad (8)$$

$$\rightarrow \quad \mathbf{W} = \frac{\boldsymbol{\Phi}^{-1}\boldsymbol{D}}{\boldsymbol{D}^H \boldsymbol{\Phi}^{-1} \boldsymbol{D}}. \qquad (9)$$

$\boldsymbol{\Phi}$ is a matrix of cross power spectral densities. Assuming an isotropic noise field the elements of $\boldsymbol{\Phi}$ only depend on the distance r_{mn} between microphone m and n.

$$\Phi_{mn}(k) = \text{si}(\omega_k r_{mn}/c). \qquad (10)$$

3. BINAURAL MODEL BASED SUPERDIRECTIVE DESIGN

Figure 2 shows a block diagram of the proposed superdirective stereo input-output beamformer in the frequency domain. First the input DFT coefficients are according to (7) summed after complex multiplication with coefficients $\boldsymbol{W}$, which are designed using the binaural model. The enhanced Fourier coefficients Z then serve as reference for the calculation of weight factors G, which generate binaural enhanced output spectra $\hat{S}_\text{l}$, $\hat{S}_\text{r}$ via multiplication with the input spectra Y_l, Y_r. The enhanced binaural time signal is finally synthesized via IDFT and overlap add.

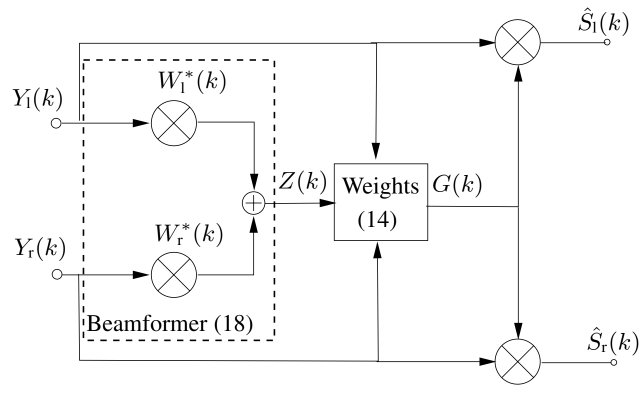

Figure 2: Superdirective binaural input-output beamformer

3.1. Superdirective Coefficients

The superdirective design rule requires the transfer vector for the desired direction $\boldsymbol{D}(\theta_s, k) = [D_\text{l}(\theta_s, k),\ D_\text{r}(\theta_s, k)]^T$ and the matrix of cross power spectral densities $\boldsymbol{\Phi}(k)$ as inputs. The transfer vector can be extracted from (3) with $D_\text{l}(\theta_s, k) = D_{\text{mod}}(\theta_s, \omega_k)$ and $D_\text{r}(\theta_s, k) = D_{\text{mod}}(\pi - \theta_s, \omega_k)$. On the other hand, the 2×2 cross power spectral density matrix $\boldsymbol{\Phi}(k)$ 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

$$\boldsymbol{\Phi}(k) = \begin{pmatrix} 1 & \Gamma_{\text{lr}}(k) \\ \Gamma_{\text{lr}}(k) & 1 \end{pmatrix} \qquad (11)$$

with the coherence function

$$\Gamma_{\text{lr}}(k) = \frac{\Phi_{\text{lr}}(k)}{\sqrt{\Phi_{\text{ll}}(k)\Phi_{\text{rr}}(k)}}. \qquad (12)$$

The head related coherence function is much lower than the value, that could be expected from (10) when only taking the microphone distance between left and right ear into account [8]. It can be calculated by averaging a number J of equidistant HRTFs across the horizontal plane, $\theta_j = 2\pi \cdot j/J$.

$$\Gamma_{\text{lr}}(k) = \frac{\sum_{j=1}^{J} D_\text{l}(\theta_j, k) D_\text{r}^*(\theta_j, k)}{\sqrt{\left(\sum_{j=1}^{J} |D_\text{l}(\theta_j, k)|^2\right)\left(\sum_{j=1}^{J} |D_\text{r}(\theta_j, k)|^2\right)}}. \qquad (13)$$

3.2. Binaural Output Generation

A binaural beamformer, that outputs a mono signal would be unacceptable, because the noise reduction benefit is consumed by the loss of spatial hearing. To output a binaural signal the enhanced spectral coefficients $Z(k)$ are used as a reference to calculate spectral weights for both sides. Real weight factors $G(k)$ are desirable in order to minimize distortions from the frequency domain filter. In addition, a distortionless response for a signal from the desired direction should be guaranteed. To fulfill the first demand, the weights can be calculated by comparing the spectral amplitude of the beamformer output to the sum of both input spectral amplitudes:

$$G(k) = \frac{|Z(k)|}{|Y_\text{l}(k)| + |Y_\text{r}(k)|} \qquad (14)$$

To fulfill the distortionless response with (14) the MVDR design rule according to (8) has to be modified with a correction factor:

$$\min_{\boldsymbol{W}} \boldsymbol{W}^H \boldsymbol{\Phi} \boldsymbol{W} \ \text{ w.r.t. } \boldsymbol{W}^H \boldsymbol{D} = \text{corr}(\theta_s, k). \qquad (15)$$

The correction factor $\text{corr}(\theta, k)$ is to be determined in the following. Assuming, that a single desired signal arrives from θ_s, i.e. , $\mathbf{Y}(k) = \mathbf{D}(\theta_s, k)S(k)$ and consequently $|Y_\text{l}(k)| = \alpha_\text{l}^{\text{phy}}(k)|S(k)|$, $|Y_\text{r}(k)| = \alpha_\text{r}^{\text{phy}}(k)|S(k)|$, and the coefficient vector $\boldsymbol{W}$ to be designed for this angle θ_s, then by applying (7) and (15), (14) can be written as

$$G(\theta_s, k) = \frac{|\text{corr}(\theta_s, k)S(k)|}{\alpha_\text{l}^{\text{phy}}(\theta_s, k)|S(k)| + \alpha_\text{r}^{\text{phy}}(\theta_s, k)|S(k)|}. \qquad (16)$$

The demand $G \overset{!}{=} 1$ for a signal from θ_s yields

$$\text{corr}(\theta_s, k) = \alpha_\text{l}^{\text{phy}}(\theta_s, k) + \alpha_\text{r}^{\text{phy}}(\theta_s, k). \qquad (17)$$

The design of the superdirective coefficient vector $\boldsymbol{W}(\theta_s, k)$ for frequency bin k and desired angle θ_s is therefore:

$$\mathbf{W}(\theta_s, k) = \left(\alpha_\mathrm{l}^\mathrm{phy}(\theta_s, k) + \alpha_\mathrm{r}^\mathrm{phy}(\theta_s, k)\right) \cdot \frac{(\boldsymbol{\Phi}^{-1}(k))\boldsymbol{D}(\theta_s, k)}{\boldsymbol{D}^H(\theta_s, k)(\boldsymbol{\Phi}^{-1}(k))\boldsymbol{D}(\theta_s, k)}. \quad (18)$$

3.3. Directivity Evaluation

The performance of a beamformer in terms of spatial directivity can be described by the directivity pattern $\Psi(\theta_s, \theta, k)$. It is defined as the squared transfer function for a signal, that arrives from a certain spatial direction θ when the beamformer is designed for angle θ_s. In the case of the stereo input-output binaural beamformer the squared transfer function equals the squared weight factor G^2 according to (14) that is applied to the spectral coefficients.

$$\Psi(\theta_s, \theta, k)/\mathrm{dB} = 20 \log_{10}(G(k)) \quad (19)$$

which can be written as

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

Figure 3 shows the beam pattern for three different frequencies when the beamformer is steered towards $\theta_s = 0°$.

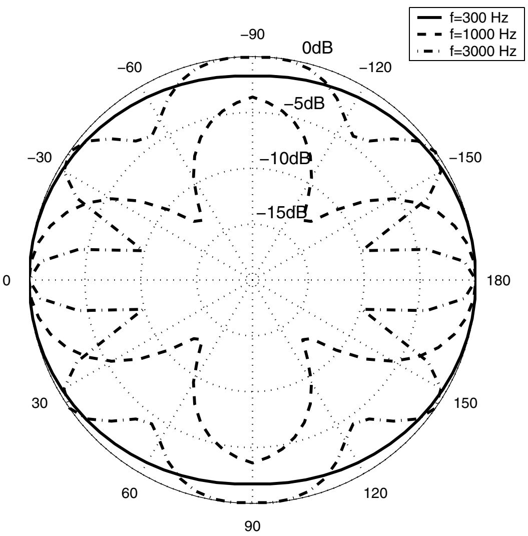

Figure 3: Beam pattern $\Psi(\theta_s = 0°, \theta, f)$ of superdirective stereo input-output beamformer for DFT bins corresponding to 300Hz, 1000Hz and 3000Hz.

In this case, the superdirective design leads to a simple delay-and-sum beamformer. 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 4 shows the beam pattern for $\theta_s = -60°$. Here, the directivity is much higher compared to that of the frontal desired direction, especially signals from the opposite side will highly be attenuated. The main lobe is comparably large at all plotted frequencies.

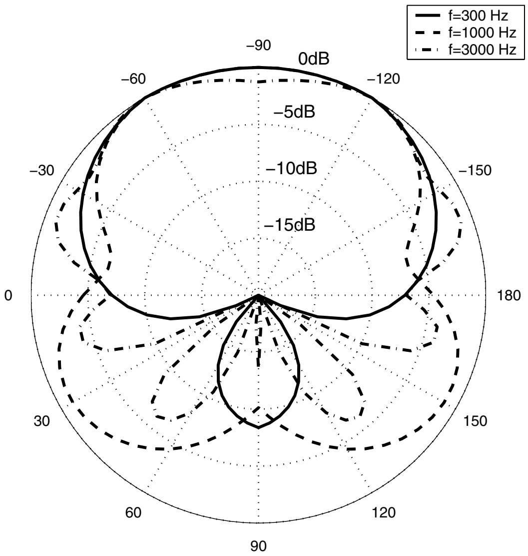

Figure 4: Beam pattern $\Psi(\theta_s = -60°, \theta, f)$ of superdirective stereo input-output beamformer for DFT bins corresponding to 300Hz, 1000Hz and 3000Hz.

4. PERFORMANCE IN REAL ENVIRONMENT

In order to evaluate the performance of the beamformer in a realistic environment, experiments with speech sources and a dummy head were executed in a conference room (reverberation time $T_0 \approx$ 800ms) with two source-target distances as depicted in Figure 5. In the first scenario, the speech sources were located within a short

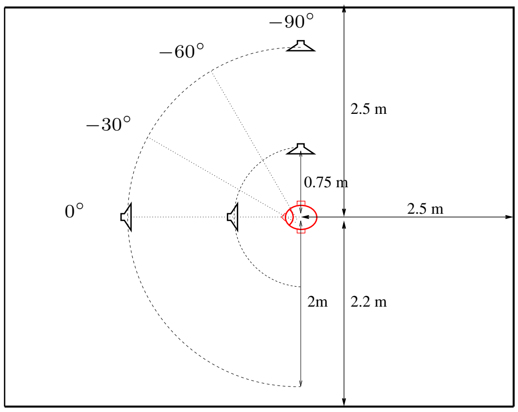

Figure 5: Experimental setup inside conference room (reverberation time $T_0 \approx$ 800ms).

distance of 0.75m to the head. Also, the head was positioned at least 2.2m away from the nearest wall. In the second scenario, the loudspeakers were moved 2m away from the dummy head. Thus, the two scenarios differ significantly in the direct-to-reverberation ratio. For a realistic scenario, the hearing aids were attached above the ears of the dummy head without taking special precautions to match exact positions. In the experiments, a desired speech source towards which the beamformer is steered is disturbed by an interfering speech source.

Since the goal of the binaural noise reduction filter is to improve the speech intelligibility, the performance of the algorithm is judged using an intelligibility weighted gain. First, a spectral noise reduction gain is calculated as the difference between the power spectral density attenuation of the undesired source subtracted by the attenuation of the desired source. To judge the intelligibility improvement the frequency dependent noise reduction gain is grouped into critical bands, which are weighted according to the ANSI's speech intelligibility index standard [9].

Figure 6 plots the performance of the superdirective stereo input-output beamformer in terms of speech intelligibility weighted gain for a desired speech source from $\theta_s = 0°$ and speech interferers from variable directions. The two plots in Figure 6 show the gain when all sources were located 0.75m and 2m away from the dummy head.

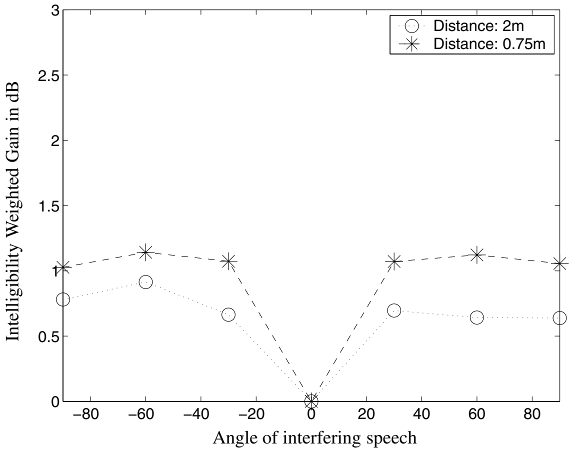

Figure 6: Intelligibility weighted gain for speech from $\theta_s = 0°$ and interfering speech from other directions in conference room (distance to dummy head 0.75m and 2m respectively).

The binaural input-output superdirective beamformer only delivers about 0.6-1.2dB intelligibility weighted improvement because of its comparably low directivity towards the frontal direction as depicted in Figure 3. Due to the decreased direct-to-reverberation ratio the overall performance is lower for the 2m distance from the dummy head.

Figure 7 plots the intelligibility weighted noise reduction gain of the superdirective stereo input-output beamformer when the beamformer is steered towards $\theta_s = -60°$. Results for desired sources and interfering sources from a distance of 0.75m and 2m from the dummy head are shown. For low angular differences to the desired direction, i.e. , $\theta_s = -90°$ and $\theta_s = -30°$, the gain mostly stays below 1dB. When the interfering speech source is located at the other side, the superdirective beamformer achieves the highest intelligibility weighted gain, with values of nearly 3dB. Due to the decreased direct to reverberation ratio at the distance of 2m the gain remains below 2dB.

5. CONCLUSION

We have presented an efficient superdirective beamformer for dual-channel noise reduction. The beamformer applies a binaural model to design superdirective filter coefficients. By modification of the design rule and the introduction of a simple post processing scheme a binaural distortionless output is obtained. Experimental results in a real environment show, that the beamformer can achieve a large improvement in terms of intelligibility weighted gain especially for lateral desired directions.

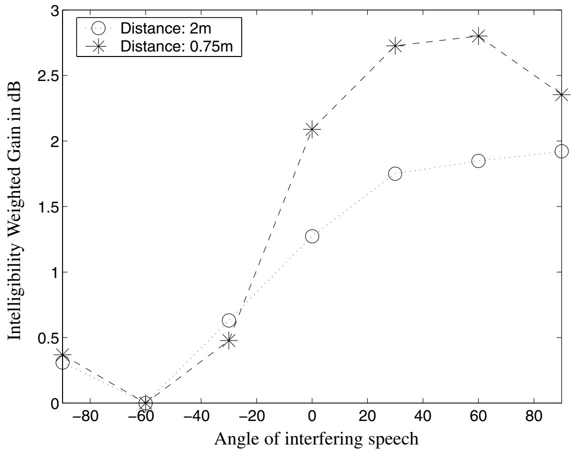

Figure 7: Intelligibility weighted gain for speech from $\theta_s = -60°$ and speech interferer from other directions in conference room (distance to dummy head 0.75m and 2m respectively).

REFERENCES

[1] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech and Signal Processing*, vol. 32, pp. 1109–1121, December 1984.

[2] J. Bitzer and K. Simmer, "Superdirective microphone arrays," in *Microphone Arrays* (M. Brandstein and D. Ward, eds.), pp. 19–38, Springer Verlag, 2001.

[3] L. Griffiths and C. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propagation*, vol. AP-30, pp. 27–34, January 1982.

[4] 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 Verlag, 2003.

[5] V. Hohmann, J. Nix, G. Grimm, and T. Wittkop, "Binaural Noise Reduction for Hearing Aids," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, (Orlando, USA), May 2002.

[6] D. Campbell and P. Shields, "Speech Enhancement using Subband Adaptive Griffiths-Jim signal processing," *Speech Communication, Special Issue: Speech Processing for Hearing Aids*, vol. 39, pp. 97–110, 2003.

[7] C. Brown and R. Duda, "A structural Model for Binaural Sound Synthesis," *IEEE Trans. Speech and Audio Processing*, vol. 6, pp. 476–488, September 1998.

[8] M. Doerbecker, *Mehrkanalige Signalverarbeitung zur Verbesserung akustisch gestoerter Sprachsignale am Beispiel elektronischer Hoerhilfen*. PhD thesis, ISBN 3-86073-439-3, RWTH Aachen, 1998.

[9] A. S.-. American National Standard, "Methods for Calculation of the Speech Intelligibility Index," *ANSI S3.5-1997*, 1997.