DIRECTION OF ARRIVAL ESTIMATION BASED ON THE DUAL DELAY LINE APPROACH FOR BINAURAL HEARING AID MICROPHONE ARRAYS

Stefan Goetze[1], *Thomas Rohdenburg*[2], *Volker Hohmann*[2], *Birger Kollmeier*[2], *and Karl-Dirk Kammeyer*[1]

[1]University of Bremen
Dept. of Communications Engineering
28334 Bremen, Germany
goetze@ant.uni-bremen.de

[2]University of Oldenburg
Medical Physics Group
26111 Oldenburg, Germany
thomas.rohdenburg@uni-oldenburg.de

ABSTRACT

Multi-channel beamformer algorithms are promising solutions for noise reduction in hearing aids as they exploit the spatial distribution of the interfering signals and therefore in general lead to less signal distortion than single channel algorithms. Beamformers need *a priori* information about the microphone array and the direction of arrival of the target speech source. For head-worn arrays it is usually assumed that the user physically steers the arrays' look direction toward the desired speech source. This may become unsatisfying for the hearing aid user for high directivity beamformers with a small main lobe and when the target signal source is moving. In this contribution an automatic steering (electronic control of the look direction) is applied based on the dual delay line approach after Liu et al. [1]. This approach is modified to be applicable for head-mounted hearing-aid arrays. We show that the original free-field approach does not work on a head-mounted array because of the inappropriate propagation model. If we apply the true HRTF or a spherical head propagation model, the estimate is reliable within $\pm 8^\circ$ degree mean estimation error for an input SNR of 10dB or higher. However, for lower SNR the method seems to be not robust enough.

Index Terms— Direction of Arrival (DOA), Head Related Transfer Function (HRTF), Noise Reduction, Beamforming

1. INTRODUCTION

In modern hearing aids multiple microphones are applied to reduce ambient noise by exploiting spatial information. Many contributions in the literature either assume a fixed look direction to zero degree or the Direction of Arrival (DOA) to be perfectly known. In the first case steering is accomplished by head movements to the desired source. However it has been shown by several authors that a steering mismatch due to a wrong estimation of the DOA severely degrades the beamformer performance [2, 3]. In this contribution the dual delay line approach after Liu et al. [1] is extended by the consideration of head shadowing effect to work with binaural beamforming algorithms for digital hearing aids. The performance of the system is analyzed in interaction with a binaural noise reduction scheme consisting of a fixed Minimum Variance Distortionless Response (MVDR) beamformer and a binaural post-filter.

The remainder of this paper is organized as follows: In Section 2 the proposed DOA estimation technique is reviewed for free-field assumptions of [1] and extended to work with Head Related Transfer Functions (HRTFs). In Section 3 the binaural noise reduction scheme is described. Simulation results for both, DOA estimation and noise reduction performance are presented in Section 4 and Section 5 gives some final conclusions.

Notation: Vectors and matrices are printed in boldface while scalars are printed in italic. k is the discrete time index, m the discrete frequency index and ℓ the discrete block index, respectively. The superscripts ^{T}, *, and ^{H} denote the transposition, the complex conjugation and the Hermitian transposition respectively.

2. ESTIMATION OF DIRECTION OF ARRIVAL

For noise reduction by microphone arrays a reliable estimate of the DOA of the desired sound source is a crucial point. The performance of beamforming noise reduction techniques is often heavily degraded if DOA estimation errors occur, especially if adaptive algorithms are applied [3].

2.1. Free-field assumptions

For the free-field assumption the dual delay line approach after Liu et al. [1] is promising because the spatial resolution can be directly influenced by choosing an appropriate number of sectors I. It will be briefly reviewed in the following with a somewhat modified notation and the specific problems caused by the shadow effects of the human head will be pointed out.

As depicted in Fig. 1 two microphones capture the sound signals $x_0[k]$ and $x_1[k]$ at two spatial positions $\mathbf{p}_0$ and $\mathbf{p}_1$. The time signals are multiplied by a Hann window $w[k]$ and transformed into the frequency domain

$$x^{(\ell)}[m] = \sum_{k=0}^{L_{\mathrm{DFT}}-1} x[\ell L_{\mathrm{Bl}} + k] w[k] e^{-j2\pi km/L_{\mathrm{DFT}}}. \quad (1)$$

Here L_{DFT} and L_{Bl} are the DFT-length and the block length, respectively. An appropriate zero-padding can be applied to reduce cyclic convolutions effects. For the reason of better readability the block index ℓ is omitted in the remainder if it is not necessary. Following [1] we divide the azimuth range of interest $\Phi = -90^\circ..90^\circ$ into I sectors as depicted in Figure 1.

For each sector i which corresponds to an angle Φ_i a propagation vector $\mathbf{d}[m, \Phi]$ for the left and the right channel can defined as

$$\mathbf{d}[m, \Phi] = \begin{bmatrix} |d_0[m, \Phi]| e^{-j2\pi m \frac{f_s}{M} \tau(\Phi_{i,0})} \\ |d_1[m, \Phi]| e^{-j2\pi m \frac{f_s}{M} \tau(\Phi_{i,1})} \end{bmatrix}. \quad (2)$$

For free field assumptions the absolute values of (2) equal one for all discrete frequencies ($|d_i[m, \Phi]| = 1, \forall m, \Phi$) and the differ-

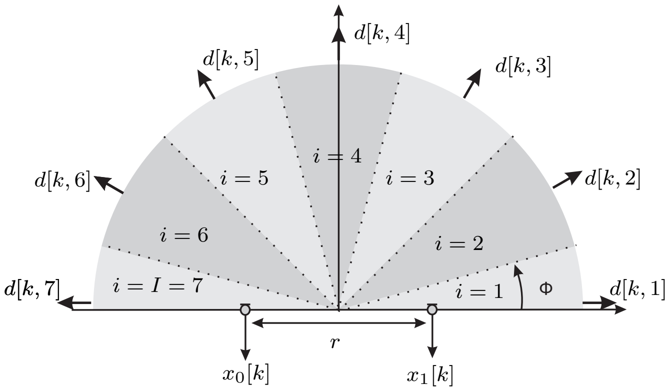

Fig. 1. Dual-microphone setup with $I = 7$ possible DOA sectors.

ence between the signal of the left channel $x_0[k]$ and the right channel $x_1[k]$ is just a time delay $\Delta\tau = \tau(\Phi_{i,0}) - \tau(\Phi_{i,1}) = \frac{r \cos \Phi_i}{c}$. Here r and $c = 344$ m/s are the inter-microphone distance and the speed of sound, respectively.

The microphone signals can be defined as a superposition of the desired signal $s[m]$ multiplied by the corresponding propagation vector $\mathbf{d}[m, \Phi]$ and some ambient noise $n[m]$:

$$x_0[m, \Phi] = s[m] \cdot d_0[m, \Phi] + n_0[m] \quad (3)$$

$$x_1[m, \Phi] = s[m] \cdot d_1[m, \Phi] + n_1[m] \quad (4)$$

Thus the desired direction of arrival can be obtained by

$$\Phi_{\mathrm{opt}}[m] = \arg \min_{\Phi}[m] \left\{ \Delta x[m, \Phi] \right\} \quad (5)$$

with

$$\Delta x[m, \Phi] = \left| x_0[m, \Phi]/d_0[m, \Phi] - x_1[m, \Phi]/d_1[m, \Phi] \right| . \quad (6)$$

Replacing $x_0[m, \Phi]$ (3) and $x_1[m, \Phi]$ (4) in (5) the minimization leads to a minimum of

$$v[m, \Phi] = \left| n_0[m]/d_0[m, \Phi] - n_1[m]/d_1[m, \Phi] \right| \quad (7)$$

at the angle $\Phi[m] = \Phi_{\mathrm{opt}}[m]$. For free field assumptions the minimum of (7) gives a good estimate of the desired direction for a moderate noise level. Hence if head shadow effects have to be taken into account which results in a non-flat absolute value of the propagation factor ($|d_i[m, \Phi]| \neq 1$) the estimate fails completely.

2.2. Robustness improvements

For improving the robustness of the DOA estimation an averaging in time direction

$$\Delta\tilde{x}^{(\ell)}[m, \Phi] = \alpha \cdot \Delta x^{(\ell-1)}[m, \Phi] + (1 - \alpha) \cdot \Delta x^{(\ell)}[m, \Phi] \quad (8)$$

and in frequency direction

$$\hat{\Phi}_{\mathrm{opt}} = \frac{1}{L_{\mathrm{DFT}}} \sum_{m=0}^{L_{\mathrm{DFT}}-1} \Phi_{\mathrm{opt}}[m] \quad (9)$$

can be applied. Furthermore the maximum tracking speed of the DOA estimator should be limited to a certain threshold by

$$|\hat{\Phi}_{\mathrm{opt}}^{(\ell-1)} - \hat{\Phi}_{\mathrm{opt}}^{(\ell)}| < \xi \quad (10)$$

to avoid short but high estimation errors. This would lead to annoying artifacts if the beamformer steers to a completely wrong direction for a short period.

2.3. Head Shadowing Effects

If microphones are used which are mounted near the human head, e.g., on the frame of eyeglasses or in behind-the-ear (BTE) hearing-aids the free field assumption becomes invalid and the true Head Related Transfer Functions (HRTFs) have to be taken into account. For simulations 6-channel HRTFs were measured in an anechoic room using two three-channel BTE hearing aid shells mounted on a Brüel & Kjær (B&K) dummy head. Since in general HRTFs are unique for every human person they are not available for real-world DOA estimation. Thus head models have to be applied to estimate the HRTFs. In this contribution a head model by Duda [4, 5] is used which is a simple but effective parametric model that estimates the characteristics of a sphere. The interaural time difference (ITD) cues are modeled by Woodworth and Schlosberg's frequency independent (ray-tracing) formula. The gross magnitude characteristics of the HRTF spectrum, namely the interaural level difference (ILD) cues, are covered by a first order IIR head shadow filter which also accounts for an additional frequency dependent delay for low frequencies [5]. Near-field effects and interference effects that introduce ripples in the frequency response which are quite prominent on the shadowed side are incorporated and described in [4].

If a DOA estimator has to work near the human head shadowing effects have to be taken into account. As it is shown in Fig. 2 the HRTFs have strong level differences for different angles and thus the free-field assumption, where only the phase of the propagation factor is considered leads to wrong DOA estimates.

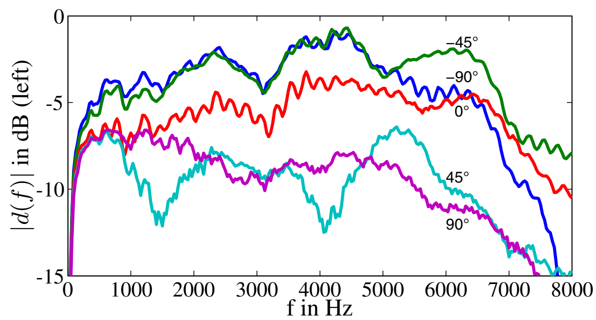

Fig. 2. Absolute values of Head Related Transfer Functions (HRTFs) of left channel.

3. MULTI-CHANNEL NOISE REDUCTION

Fig. 3 shows the system model of the multi-channel noise reduction scheme used in this paper. The discrete microphone signals $x_i[k], i = 1..6$ are transformed into the frequency domain by the Short Time Fourier Transform (STFT) (1). The DOA estimator feeds the MVDR beamformer with the propagation vector $\mathbf{d}[m, \hat{\Phi}_{\mathrm{opt}}]$ corresponding to the estimated angle $\hat{\Phi}_{\mathrm{opt}}$. The monaural beamformer output is further processed by the binaural post-filter $\mathbf{H}_{\mathrm{Bin}}[m]$ to generate binaural output [3, 6] which is transformed back into time domain by the Inverse Short Time Fourier Transform (STFT^{-1}). The multi-channel algorithms used here are designed using the well-known constraint Minimum Variance Distortionless Response (MVDR) solution [7]:

$$\mathbf{W}[m] = \frac{\mathbf{\Gamma}_{NN}^{-1}[m]\mathbf{d}[m]}{\mathbf{d}^H[m]\mathbf{\Gamma}_{NN}^{-1}[m]\mathbf{d}[m]} \quad (11)$$

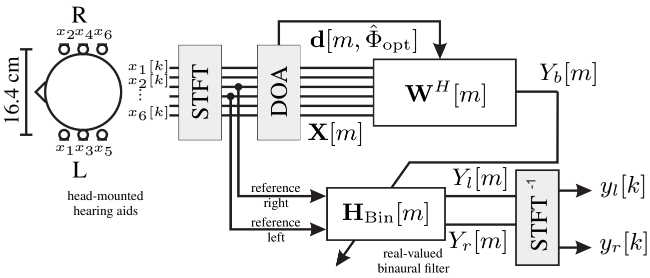

Fig. 3. Signal model and beamformer setup.

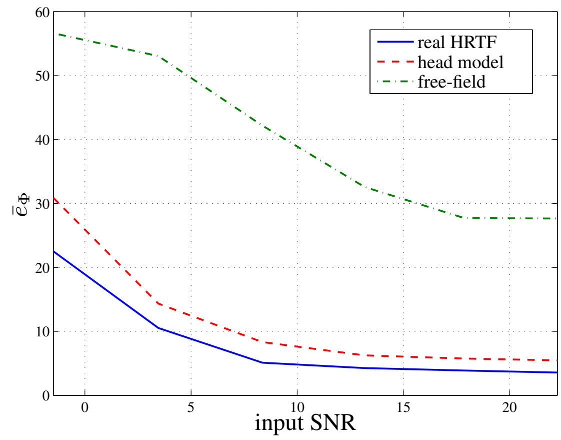

Fig. 4. Estimation error for a DOA estimator for different assumptions for the propagation vector over the input SNR.

This solution allows to include different assumptions about the wave propagation of the target signal (included in the propagation vector $\mathbf{d}$), and the characteristics of the noise field as described by its cross power spectral density matrix $\mathbf{\Gamma_{NN}}[m]$. Although the beamformer is steered adaptively by the DOA estimator to variable directions, it is referred to as a *fixed* beamformer, as it is fixed in terms of the expected noise field. If the beamformer should optimally reduce noise from an arbitrary direction the beamformer coefficients can be designed with an isotropic noise field characteristic. For a diffuse noise field the cross power spectral density matrix $\mathbf{\Gamma_{NN}}[m]$ depends on the underlying propagation model and can be estimated by integrating the propagation vectors over all directions. For the free-field assumption the isotropic noise field $\mathbf{\Gamma_{NN}}[m]$ can be solved analytically: in 3-D the correlation can be described by a sinc-function [7], in cylindrical coordinates by a bessel-function. Due to the spatial filtering effect of the head the correlation between bilateral microphone signals is much lower than in free-field. Since the output of the beamformer is monaural we define a binaural post-filter according to [6]. The binaural post-filter $\mathbf{H}_{\text{Bin}}[m]$ controlled by the beamformer output is real-valued and therefore it preserves the interaural phase-difference between the two reference inputs from the left and right hearing-aid [3, 6].

4. SIMULATION RESULTS

The performance of the proposed algorithms for DOA estimation and for binaural noise reduction based on the imperfect real-world DOA estimates will be evaluated in the following. For simulations diffuse noise signals were generated by summing up speech-colored random noise filtered with measured HRTFs from all directions to simulate a 2D-isotropic noise field. A moving speaker was added for different input SNRs. The block length for all simulations was chosen to $L_{\text{Bl}} = 256$ with an overlap of 128 samples at a sampling frequency of $f_s = 16\text{kHz}$. The FFT-length was 512 samples, which means a zero padding factor of two. The number of possible angles was chosen to $I = 37$ which leads to a resolution of 5° for a range of $\Phi = -90^\circ..90^\circ$. The threshold for the maximum tracking speed of the algorithm was fixed to $\xi = 5^\circ$.

Fig. 4 shows the mean estimation error of the DOA estimator

$$\bar{e}_\Phi = \frac{1}{|\mathcal{A}|} \sum_{\mathcal{A}} \Phi - \hat{\Phi} \tag{12}$$

for different input SNRs. Here Φ and $\hat{\Phi}$ are the true and the estimated direction of arrival, respectively. $\mathcal{A}$ is the set of frames where speech is present and $|\mathcal{A}|$ its cardinality.

It can be seen from Fig. 4 that an estimation of the direction of arrival drastically fails if free-field assumptions are made (dash-dotted line). The use of the (in practice unknown) true HRTFs (solid line) lead to the best DOA estimates. The estimation using the head model according to eq. (5) only leads to a slight degradation and thus is a feasible approximation for the unknown true HRTF.

For low input SNR (< 8dB) the estimation errors increase thus DOA estimation based on the dual delay line approach becomes unreliable. This is a general problem since the approach is based on looking for and comparing signal powers from different directions. For low SNR the signal power difference between clean speech + noisy speech from the desired direction and noisy speech from other directions is not sufficient for a reliable estimate. This result was also reported by other authors, e.g. [8]. Thus for low input SNR other DOA estimation methods should be applied, see e.g. [9] for an overview.

In Fig. 5 and 6 the performance of the binaural noise reduction scheme relying on real DOA estimates is evaluated by means of the Signal to Noise Ratio Enhancement (SNRE) and the Perceptual Similarity Measure (PSM) [10]. PSM is a speech quality measure from PEMO-Q [10] which estimates the perceptual similarity between the processed signal and a clean speech reference. This measure has shown a high correlation with subjective overall quality ratings [11]. Here the PSM is measured between the clean speech component at the left (right) reference microphone and the left (right) output of the binaural post-filter.

Fig. 5 shows the segmental SNRE between the left (right) output of the binaural post-filter and the left (right) reference channel. The SNRE is the difference of the Signal to Noise Ratio (SNR) at the output of the noise reduction scheme and a reference input SNR. It can be seen from Fig. 5 that if the binaural noise reduction scheme relies on DOA estimates based on free-field assumptions hardly any SNR enhancement is achieved (dash-dotted line). Although the use of true HRTFs leads to the best results (solid line), relying on the head model (dashed line) is capable of improving noisy speech when a head-mounted noise reduction device is applied. Fig. 5 gives the impression that the sound quality improvement increases for lower input SNRs. From Fig. 4 it is clear that this impression is misleading because mean DOA estimation errors at input SNRs lower than 5 dB are not satisfactory.

In Fig. 6 the PSM is shown which better reflects the perceived audio quality. Here it can be seen that the overall sound quality decreases drastically for lower input SNR. Again the results for the

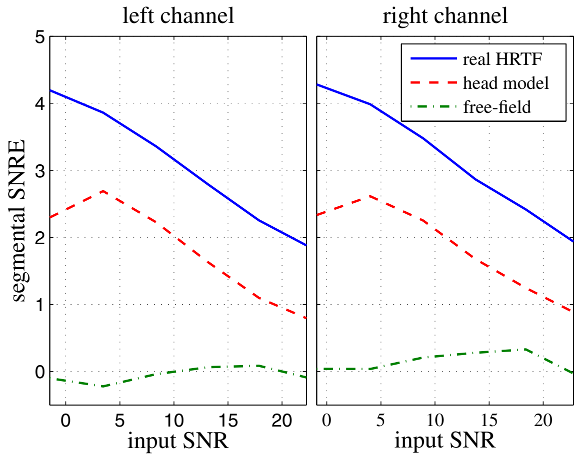

Fig. 5. SNRE of the beamformer steered by the DOA estimate for different input SNR.

head model give a good approximation for the real HRTFs, while free-field assumptions lead to a much lower sound quality.

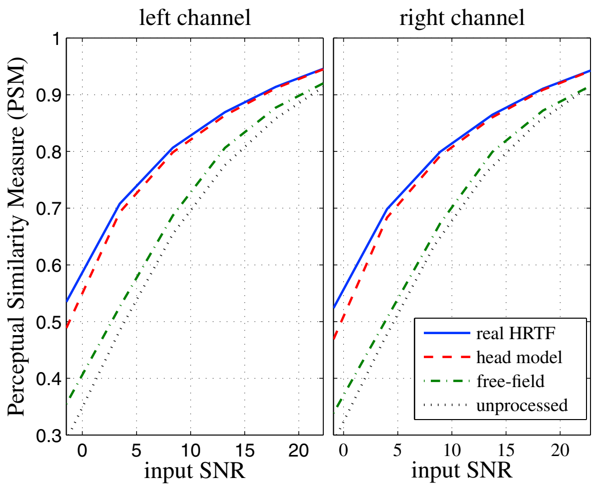

Fig. 6. PSM of the beamformer steered by the DOA estimate for different input SNR.

The so-called ΔPSM [11], which is the difference between the dotted line (unprocessed) and the particular PSM curve shows the quality improvement achieved by the processing. We see that for low input SNR the ΔPSM values are higher, which means that the improvement is better, but that the overall quality of the output signal is very poor. The ΔPSM values match with the SNRE curves from Fig. 5 but Fig. 6 additionally shows the overall quality and thus is more appropriate to compare the different methods.

5. CONCLUSIONS

In this work we analyzed the direction of arrival estimation method after Liu which is based on the delay line approach for the purpose of DOA estimation for hearing aid applications. It could be shown that the underlying free-field assumptions do not lead to satisfactory results and head related transfer functions have to be considered. Since in general it is impossible to estimate the true HRTFs, simulations based on a head model were performed, which showed good results for moderate input SNR. However, for low SNR environments the delay line approach is not capable to deliver reliable results and thus further methods need to be investigated for comparison.

6. REFERENCES

[1] C. Liu, B.C. Wheeler, D. O'Brian, R.C. Bilger, C.R.Lansing, and A.S. Feng, "Localization of Multiple Sound Sources with Two Microphones," *Journal of the Acoustical Society of America (JASA)*, vol. 108, no. 4, pp. 1888–1905, Oct. 2000.

[2] J. E. Greenberg and P. M. Zurek, "Microphone Array Hearing Aids," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. S. Brandstein and D. Ward, Eds., chapter 11, pp. 229–253. Springer, 2001.

[3] T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Robustness Analysis of Binaural Hearing Aid Beamformer Algorithms by Means of Objective Perceptual Quality Measures," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, USA*, Oct. 2007.

[4] R.O. Duda and W.L. Martens, "Range Dependence of the Response of a Spherical Head Model," *Journal of the Acoustical Society of America (JASA)*, vol. 104, no. 5, pp. 3048–3058, Nov. 1998.

[5] P.C. Brown and R.O. Duda, "A Structural Model for Binaural Sound Synthesis," *IEEE Trans. on Speech and Audio Processing*, vol. 6, no. 5, pp. 476–488, Sept. 1998.

[6] T. Lotter and P. Vary, "Dual Channel Speech Enhancement by Superdirective Beamforming," *EURASIP Journal on Applied Signal Processing*, vol. 2006, no. Article ID 63297, pp. 1–14, 2006.

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

[8] L. Calmes, G. Lakemeyer, and H. Wagner, "Azimuthal Sound Localization using Coincidence of Timing Across Frequency on a Robotic Platform," *Journal of the Acoustical Society of America (JASA)*, vol. 121, no. 4, pp. 2034–2048, Apr. 2007.

[9] G. Doblinger, "Localization and Tracking of Acoustical Sources," in *Topics in Acoustic Echo and Noise Control*, chapter 6, pp. 91 – 122. Springer, Berlin - Heidelberg, 2006.

[10] R. Huber and B. Kollmeier, "PEMO-Q - A New Method for Objective Audio Quality Assessment using a Model of Auditory Perception," *IEEE Trans. on Audio, Speech and Language Processing*, vol. Special Issue on Objectiv Quality Assessment of Speech and Audio, 2006.

[11] T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Objective Measures for the Evaluation of Noise Reduction Schemes," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, 2005.

Work supported by the EC (DIRAC project IST-027787), BMBF and DFG