Objective quality assessment of multi-channel noise reduction algorithms for hearing aids by means of psychoacoustic measures

Thomas Rohdenburg, Volker Hohmann, Birger Kollmeier

Universität Oldenburg, Medizinische Physik, 26111 Oldenburg, Deutschland, Email: thomas.rohdenburg@uni-oldenburg.de

Introduction

Modern hearing aids use microphone arrays for multi-channel noise reduction. A well-known algorithm type is the beamforming which uses the spatial distribution of natural audio signal sources by evaluating the correlation properties between the microphone signals. For this, information about the direction of arrival of the target signal and the relative microphone positions are needed. Compared to single-channel envelope filtering algorithms beamformers in general lead to less target signal distortion at a high amount of noise reduction. Likewise, the human auditory system uses correlation properties between the audio signals at the left and right ear for object segregation (cocktail party effect). Therefore it is desirable to preserve the binaural information even after the noise reduction. In this study, different methods are shown and compared that preserve or reconstruct the binaural signal properties which is particularly important for head-worn binaurally connected hearing aids. Additionally, head-shadow and diffraction effects have to be considered for the optimal design of the beamformer. The results show clearly, how the robustness of the noise reduction schemes is influenced under realistic signal conditions by using different parameters and assumptions about the sound wave propagation. The algorithms are evaluated using objective quality measures based on psychoacoustic models of the human auditory system.

Signal model and algorithms

The signals were recorded using two 3-channel behind-the-ear hearing aid shells mounted on a B&K dummy head. 6-channel head related transfer functions (HRTFs) in an anechoic room and real-world environmental noise in a cafeteria have been recorded. The input signal was composed from two directional signals filtered with HRTFs (target and interferer from 30° (front-left) and −135° (back-right) azimuth, respectively) and mixed with the recorded cafeteria noise to generate a near-to-realistic scenario. The 30° direction was chosen because it is asymmetric to the array and offers a more general assessment of the beamformer's properties than a fixed 0° look direction.
The multi-channel algorithms used here are designed using the well-known constraint Minimum Variance Distortionless Response (MVDR) solution [3], Eq. (1),

$$\boldsymbol{W}(f) = \frac{\boldsymbol{\Phi_{NN}}^{-1}(f)\boldsymbol{d}(f)}{\boldsymbol{d^H}(f)\boldsymbol{\Phi_{NN}}^{-1}(f)\boldsymbol{d}(f)} \quad (1)$$

$$\boldsymbol{d}(f) = \left[a_0 e^{j2\pi f\tau_0}, a_1 e^{j2\pi f\tau_1}, \ldots, a_{M-1} e^{j2\pi f\tau_{M-1}}\right]^T \quad (2)$$

$$Y_f(f) = \boldsymbol{W^H}(f)\boldsymbol{X}(f) \quad (3)$$

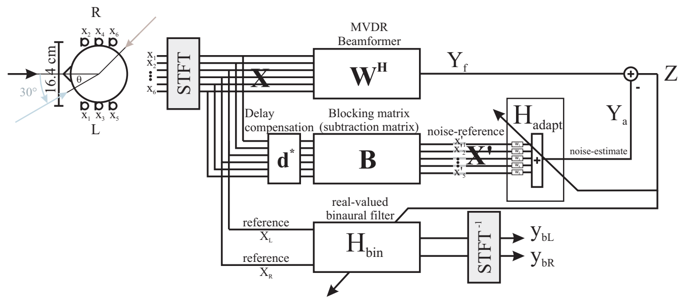

Abbildung 1: GSC beamformer and binaural post-filter

where f denotes the frequency, $\boldsymbol{W}$ the beamformer coefficients, $\boldsymbol{d}$ the propagation vector, a_m and τ_m the amplitude and the group delay at microphone m, $\boldsymbol{X}$ the input vector, Y_f the output of the fixed beamformer (see Fig. 1). The blocks $\boldsymbol{d}^*$ and $\boldsymbol{B}$ denote the delay and amplitude compensation followed by a blocking matrix that filters out the target signal component. An adaptive filter $\boldsymbol{H}_{\text{adapt}}$ calculates a noise reference Y_a that contains the residual noise components that Y_f and $\boldsymbol{X'}$ have in common.

Binaural Outputs

The binaural outputs are calculated using three different methods:
(**i**) (BIN_PF) The binaural output is generated by a real-valued time-varying post-filter adapted from [2] that is controlled by the monaural beamformer output Z:

$$H_{\text{Bin}}(t,f) = \frac{\left(|d_L(f)|^2 + |d_R(f)|^2\right)\Phi_{ZZ}(t,f)}{\Phi_{X_L X_L}(t,f) + \Phi_{X_R X_R}(t,f)} \quad (4)$$

$$Y_{bL}(t,f) = H_{\text{Bin}}(t,f)X_L(t,f) \quad (5)$$

$$Y_{bR}(t,f) = H_{\text{Bin}}(t,f)X_R(t,f) \quad (6)$$

where X_L,X_R (see Fig. 1) denote the input signals and d_L, d_R the propagation vectors for the expected signal direction θ_S, at the left and right reference microphone, respectively. Φ_{ZZ}, $\Phi_{X_L X_L}$ and $\Phi_{X_R X_R}$ are the power spectral density estimates for the signals Z, X_L, X_R, respectively. As the filter is real-valued, the phase of signal and noise are kept and therefore also most of the binaural cues. However, the envelope filter might introduce additional signal distortions.
(**ii**) (BIN_PR) The monaural beamformer output Z is multiplied by the propagation vectors of the reference microphones which reconstructs only the interaural phase

of the signal and may degrade spatial unmasking effects:

$$Y_{bL}(t,f) = d_L(f)Z(t,f) \quad (7)$$
$$Y_{bR}(t,f) = d_R(f)Z(t,f) \quad (8)$$

(iii) (BIN_BL) The array is split into a subarray of two parallel 3-channel beamformers $\boldsymbol{W_L}$, $\boldsymbol{W_R}$ which use common information about the target direction and the noise field. This simulates the behavior of independent bilateral hearing devices and binaural cues may be distorted as described in [11]:

$$Y_{bL}(t,f) = Z_L(t,f) = \boldsymbol{W_L^H}(f)\boldsymbol{X_{135}}(t,f) \quad (9)$$
$$Y_{bR}(t,f) = Z_R(t,f) = \boldsymbol{W_L^H}(f)\boldsymbol{X_{246}}(t,f) \quad (10)$$

where the numbers (1,3,5 and 2,4,6) refer to the microphones of the subarray, respectively.

Results

The performance evaluation was based on three objective measures: The broadband Signal-to-Noise Ratio Enhancement (SNRE), the Perceptual Similarity Measure (PSM) from PEMO-Q [6] and the estimated Speech Reception Threshold (SRT) by [9]. A description of these measures can be found in [1]. Figure 2 shows the robustness of adaptive and fixed beamformers against steering errors assuming wave propagation models of different accuracy. In case of a perfect steering towards the target signal from 30° and a correct propagation model (at least a simple head model) the adaptive beamformers have a slightly higher performance. However, this performance gain might break down in case i) the target direction is not exactly known, ii) the target is moving too fast, iii) the hearing aid user moves his/her head. The performance for the head model based beamformers is only slightly lower than for the exact HRTF, therefore using HM2 seems to be sufficient. Table 1 shows the results of

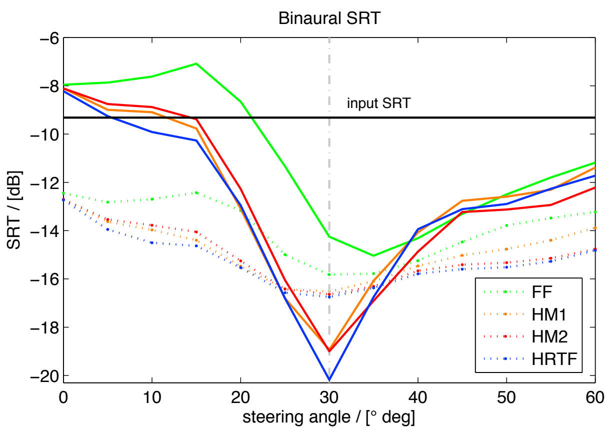

Abbildung 2: Estimated SRT used as a performance and robustness measure for different algorithms and propagation models

the three different binaural stages for HM2. Obviously, the maximum performance is reached by BIN_PF. If we look at the mean SNRE and the PSM of BIN_PR the performance seems to be in the same range as for BIN_PF, but the estimated SRT Gain is much lower. The binaural measure is able to identify the binaural information loss of BIN_PR which may result in a degradation in speech intelligibility. Additionally, the SNRE and PSM

Algorithm	SNRE L dB	SNRE R dB	mean SNRE dB	PSM L	PSM R	SRT Gain dB
HM2_BIN_PF	9,0	10,9	10,0	0,69	0,61	8,4
HM2_BIN_PR	6,5	13,0	9,8	0,59	0,62	5,1
HM2_BIN_BL	4,4	4,6	4,5	0,56	0,31	4,8

Tabelle 1: Performance of different binaural stages for the fixed beamformer using propagation model HM2

of BIN_BL are much lower (mainly due the lower number of microphones per array), but in terms of the SRT Gain it is comparable to BIN_PR which actually means that the binaural information loss is lower.
In conclusion, the SRT Gain seems to be an appropriate measure to identify binaural information loss. A head-worn beamformer system with binaural output can benefit from the integration of head models even if the exact HRTFs are unknown.
This work is supported by the EC (DIRAC project IST-027787), HearCom-Project (IST-004171), BMBF and DFG

Literatur

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

[2] T. Lotter and P. Vary, "Dual-channel speech enhancement by superdirective beamforming," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. Article ID 63 297, 14 pages, 2006.

[3] J. Bitzer and K. U. Simmer, "Superdirective microphone arrays," in *Microphone Arrays*, Brandstein and Ward, Eds. Springer, 2001, ch. 2, pp. 19–38.

[4] 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, Sep 1998.

[5] 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, 1998.

[6] 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*, 2006, special Issue on Objective Quality Assessment of Speech and Audio.

[7] T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Objective perceptual quality measures for the evaluation of noise reduction schemes," in *9th International Workshop on Acoustic Echo and Noise Control*, Eindhoven, 2005, pp. 169–172.

[8] T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Subband-based parameter optimization in noise reduction schemes by means of objective perceptual quality measures," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Paris, France, September 12-14 2006.

[9] R. Beutelmann and T. Brand, "Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners," *Journal of the Acoustical Society of America*, vol. 120, no. 1, pp. 331–342, 2006.

[10] S. Goetze, T. Rohdenburg, V. Hohmann, B. Kollmeier, and K.-D. Kammeyer, "Direction of Arrival Estimation based on the Dual Delay Line Approach for Binaural Hearing Aid Microphone Arrays," in *Proc. Int. Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, Xiamen, China, Nov. 2007.

[11] T. Van den Bogaert, T. J. Klasen, M. Moonen, and J. Wouters, "Distortion of interaural time cues by directional noise reduction systems in modern digital hearing," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2005.