

ROBUSTNESS ANALYSIS OF BINAURAL HEARING AID BEAMFORMER ALGORITHMS BY MEANS OF OBJECTIVE PERCEPTUAL QUALITY MEASURES

Thomas Rohdenburg, Volker Hohmann, Birger Kollmeier

Universität Oldenburg
Medizinische Physik
26111 Oldenburg, Germany

thomas.rohdenburg@uni-oldenburg.de

ABSTRACT

In this contribution different microphone array-based noise reduction schemes for hearing aids are suggested and compared in terms of their performance, signal quality and robustness against model errors. The algorithms all have binaural output and are evaluated using objective perceptual quality measures [1, 2, 3]. It has been shown earlier that these measures are able to predict subjective data that is relevant for the assessment of noise reduction algorithms. The quality measures showed clearly that fixed beamformers designed with head models were relatively robust against steering errors whereas for the adaptive beamformers tested in this study the robustness was limited and the benefit due to higher noise reduction depended on the noise scenario and the reliability of a direction of arrival estimation. Furthermore, binaural cue distortions introduced by the different binaural output strategies could be identified by the binaural speech intelligibility measure [3] even in case monaural quality values were similar. Thus, this perceptual quality measure seems to be suitable to discover the benefit that the listener might have from the effect of spatial unmasking.

1. INTRODUCTION

In modern hearing aids multi-channel noise reduction schemes based on small microphone arrays are used for speech enhancement. These algorithms exploit the spatial configuration of the interfering signals and therefore generally lead to less signal distortion and higher noise reduction than single-channel envelope filters. The human ability to separate sound sources in a complex situation, namely the cocktail-party effect, partly arises from the use of binaural localization cues. If binaural information is lost or distorted by the processing, the hearing impaired listener may not make use of the effect of spatial unmasking as efficiently as in the undistorted binaural condition. The intelligibility improvement introduced by a spatial filter is counteracted by the decrease due to the deteriorated efficiency of the spatial unmasking in this case. Although bilateral supply with hearing aids is motivated by a better directional-hearing ability, it has been shown in [4] that binaural cues are distorted if the hearing aids at the left and right ears work independently. Therefore, researchers have suggested microphone array based binaural spatial filtering techniques [5, 6, 7, 8] that assume a connection between the left and right hearing aid. In this study we analyzed fixed and adaptive beamformer algorithms, that exploit *a priori* knowledge about array position, wave propagation and direction of arrival as these seem to be slowly varying parameters that can be estimated and used for the adaptation of the algorithms. Information about the voice activity which might also be helpful for noise estimation was not used here. The beam-

formers that were calculated using the constraint minimum variance distortionless response (MVDR) design [9] had single channel outputs that were extended by a binaural stage. Three different strategies for generating a binaural output have been applied and evaluated by perceptual measures. Furthermore, the robustness of fixed and adaptive beamformers using different propagation models have been analyzed against steering error, array position and head-size mismatch by appropriate perceptual quality measures.

2. SIGNAL MODEL

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. For condition 1) 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. For condition 2) we used only one directional signal (speaker from 30° (left)) mixed with an artificial diffuse noise. The artificial noise was generated by summing up a speech-colored random noise that was filtered with HRTFs from all directions to simulate a 2D-isotropic noise field. This abated the influence of the noise field characteristic on the signal quality which was helpful for the analysis of the steering mismatch. The 30° direction was chosen because it is asymmetric to the array and offers a more general assessment of the beamformers properties than a fixed 0° look direction.

3. ALGORITHMS

The multi-channel algorithms used here are designed using the well-known constraint Minimum Variance Distortionless Response (MVDR) solution [9], Eq. (1),

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

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

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

where f denotes the frequency, \mathbf{W} the beamformer coefficients, \mathbf{d} the propagation vector, a_m and τ_m the amplitude and the group delay at microphone m , \mathbf{X} the input vector, Y_f the output of the fixed beamformer (see Figure 1).

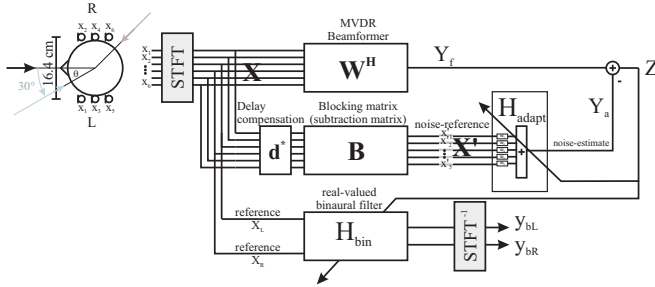


Figure 1: GSC beamformer and binaural post-filter

This solution allows to include different assumptions about the wave propagation of the target signal (included in the propagation vector \mathbf{d} , Eq. (2)), and the characteristics of the noise field as described by its cross power spectral density matrix Φ_{NN} .

The upper path of the signal diagram in Figure 1 shows the fixed beamformer which can be extended by an adaptive noise canceler path to form a Generalized Sidelobe Canceller (GSC) [10, 11]. Note that *fixed* beamformer means that a fixed noise field is assumed whereas a GSC can adapt to varying noise fields. However, for both beamformer types an adaptive steering to a moving target signal can be applied, e.g., if extended by a direction of arrival (DOA) estimation algorithm. Additionally, the beamformers are extended by a binaural stage with diverse methods to obtain a binaural output. All combinations of beamformer type (fixed/adaptive), binaural output method (bin1, bin2, bin3) and different assumptions about the wave propagation model (free-field (FF), head-model (HM1, HM2, HRTF)) are investigated in this study in terms of their performance and robustness.

Wave propagation models can be integrated into the beamformer design via the propagation vector \mathbf{d} and the noise field cross power density matrix Φ_{NN} . For the free-field (FF), \mathbf{d} has constant group-delay, τ_m , and unity amplitude, a_m , in the frequency domain. For head-models τ_m , a_m are frequency dependent accounting for head shadow and diffraction effects. The first head model (HM1) by [12] is a simple and effective parametric model that estimates the characteristic 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, the interaural level difference (ILD) cues, are covered by a single-pole, single-zero head-shadow filter which also accounts for an additional frequency dependent delay for low frequencies [12]. The second head model (HM2) by [13] additionally incorporates the distance of the source for modeling near-field effects and interference effects that introduce ripples in the response that are quite prominent on the shadowed side. It is calculated by a recursive algorithm given in [13]. The third head model (HRTF) uses the measured HRTF of the respective microphones directly as the propagation vector. The noise field matrix Φ_{NN} influences the amount of noise reduction achieved by the beamformer. In the free-field, a 3D-isotropic diffuse noise field matrix reduces to a coherence matrix with sinc-characteristic [9]. For the head-models the diffuse noise field is estimated by integrating the propagation vectors over all directions. Furthermore, Φ_{NN} needs to be constrained to reduce super directivity for feasible designs [9, 10].

Binaural Outputs are calculated using three different methods:

(i) (bin1) The binaural output is generated by a real-valued time-

varying post-filter based on [5] that is controlled by the monaural beamformer output Z :

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

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

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

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) (bin2) 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) (bin3) The array is split into a subarray of two parallel 3-channel beamformers $\mathbf{W}_L, \mathbf{W}_R$ 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 [4]:

$$Y_{bL}(t, f) = Z_L(t, f) = \mathbf{W}_L^H(f) \mathbf{X}_{135}(t, f) \quad (9)$$

$$Y_{bR}(t, f) = Z_R(t, f) = \mathbf{W}_R^H(f) \mathbf{X}_{246}(t, f) \quad (10)$$

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

4. QUALITY MEASURES

SNRE: The SNR-Enhancement (SNRE) is the difference of the signal-to-noise ratio (SNR) at the output of the beamformer and a reference input-SNR, both measured in dB. For a comparison of multi-channel algorithms the choice of the reference is crucial. Here, the SNRE is calculated between the left (right) output of the binaural stage and the left (right) input at the reference microphone, respectively.

PSM: The quality measure PSM from PEMO-Q [2] estimates the perceptual similarity between the processed signal and the clean speech source signal. For monaural noise reduction schemes this measure has shown a high correlation with subjective overall quality ratings according to [1, 14]. Here, the PSM is measured between the clean speech component at the left (right) reference microphone and the left (right) output of the binaural stage.

SRT: The speech reception threshold (SRT) is defined as the signal-to-noise ratio (SNR) at 50% speech intelligibility. In [3] a binaural model of speech intelligibility based on the equalization-cancellation (EC) processing by Durlach had been defined which is able to predict the SRT with high accuracy. For the objective quality assessment of binaural signals processed by noise reduction schemes, we are interested in the difference between the SRT of the input signal and the SRT of the output, namely the SRT Gain. Thus, the SRT Gain is the amount of SNR reduction achieved by

Algorithm	SNRE L dB	SNRE R dB	mean SNRE dB	PSM L	PSM R	SRT Gain dB	SNR L dB	SNR R dB	SRT dB
FF bin1	8.1	9.9	9.0	0.66	0.54	7.6	8.0	4.5	-15.4
FF bin2	4.9	10.2	7.6	0.53	0.55	3.3	4.8	4.8	-11.1
FF bin3	4.0	4.0	4.0	0.55	0.29	4.6	3.9	-1.4	-12.4
HM1 bin1	7.6	8.8	8.2	0.67	0.57	8.3	7.5	3.4	-16.1
HM1 bin2	4.0	9.7	6.9	0.55	0.58	4.3	3.9	4.3	-12.1
HM1 bin3	4.2	4.6	4.4	0.56	0.32	4.7	4.1	-0.8	-12.5
HM2 bin1	9.0	10.9	10.0	0.69	0.61	8.4	8.9	5.5	-16.2
HM2 bin2	6.5	13.0	9.8	0.59	0.62	5.1	6.4	7.6	-12.9
HM2 bin3	4.4	4.6	4.5	0.56	0.31	4.8	4.3	-0.8	-12.6
HRTF bin1	9.2	11.4	10.3	0.71	0.64	8.5	9.1	6.0	-16.3
HRTF bin2	7.2	13.8	10.5	0.61	0.65	5.6	7.1	8.4	-13.4
HRTF bin3	5.0	6.4	5.7	0.57	0.36	5.1	4.9	1.0	-12.9
input	-	-	-	0.38	0.14	-	-0.1	-5.4	-7.8

Table 1: Binaural output quality

the algorithm as estimated by intelligibility estimates including spatial unmasking. However, if the noise reduction algorithm is nonlinear the exact SRT Gain has to be calculated iteratively by reducing the SNR of the beamformer input signal until the predicted SRT has the same value as the original unprocessed reference signals.

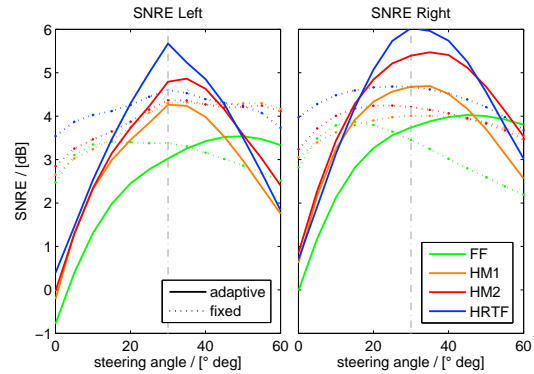
5. RESULTS

5.1. Binaural output quality

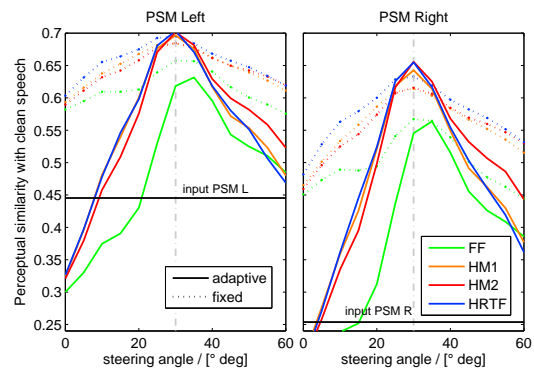
Table 1 shows the performance results for the three binaural strategies (bin1-3) which were evaluated for the fixed beamformers with different propagation models in signal condition 1). Although the mean SNRE values for bin1 and bin2 were in the same range, bin1 had a higher enhancement for the left channel and bin2 had a higher enhancement for the right channel. Interestingly, the SRT Gain of bin1 was significantly higher than for bin2. This behavior can be explained as follows: As the beamformer output Z is monaural and the multiplication with the left and right propagation vectors only turns the output into the target direction, all signals are perceptually still coming from one direction. In other words: the localization cues for the background noise are lost. The binaural SRT measure can identify the difference as it considers the spatial arrangement of speech and noise signals to calculate the SRT. For this, it does not need explicit knowledge about the interaural time and level difference (ITD, ILD). For bin3 the noise reduction performance was reduced compared to bin1 and bin2 as the bilateral beamformer uses a subarray of only three microphones. However, as the distortion of the binaural cues for bin3 is lower than for bin2, the values of the SRT are almost the same. In terms of the different propagation models, quality increases with the complexity and exactness of the model.

5.2. Robustness against steering errors

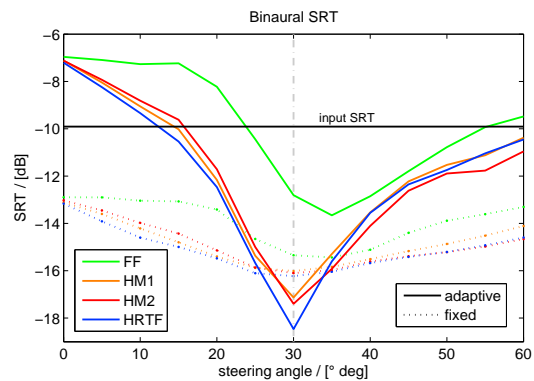
Figure 2 shows the three quality measures, (a) SNRE, (b) PSM and (c) SRT for different beamformers using the binaural post-filter (bin1) in signal condition 2) over the steering angle of the beamformer. The dotted lines refer to the fixed beamformers, the solid lines to the (adaptive) GSCs and the black lines show the quality values for the unprocessed input signals. The target speech signal came from the 30° direction, so the best quality values should have been expected if the beamformer was steered in this direction. However, depending on the underlying model, algorithm and noise field, this might not always be the case. It can be seen that the free-field coefficients (green curves) are suboptimal for the head-mounted array because the maximum values are not aligned with the steering direction of the beamformer. Among all beamformers, the free-field propagation model leads to the lowest SNRE and the lowest perceptual quality values (PSM, SRT), because it does not incorporate any head-shadow and diffraction effects. The HRTF



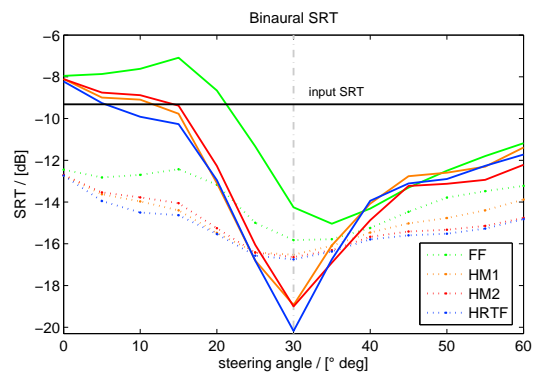
(a) SNRE in condition 2



(b) PSM in condition 2



(c) SRT in condition 2



(d) SRT in condition 1

Figure 2: Robustness evaluation against steering mismatch

coefficients led to the highest noise reduction performance but the head models (HM1, HM2) showed comparable results in terms of the predicted overall quality and SRT. The fixed head model beamformers could enhance the SNR in diffuse isotropic noise by about 4 dB. The flatness of the dotted curves shows that they are relative robust against steering errors. The GSCs (solid lines) had approximately 1 dB higher SNREs than the fixed beamformers, but in terms of the estimated overall quality the advantages were small. The SRT estimate was 2 dB lower but these values were only stable within a steering mismatch of $\pm 5^\circ$ degree which pointed out a lower robustness. However for condition 1) with a directional interfering noise source the adaptive beamformer could reduce the SRT by about 4dB more compared to the fixed beamformer that was optimized for suppressing isotropic noise (see Fig. 2 (d)). In summary it could be stated that the GSC was more susceptible to model errors and might only be beneficial in situations with directional interfering noise and small steering errors.

5.3. Robustness against model variation

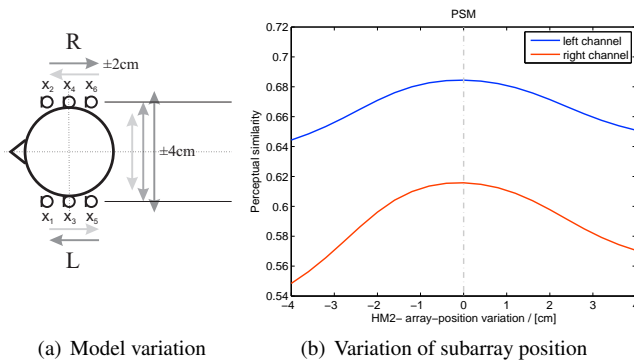


Figure 3: Robustness against variation of array position and model parameters for (HM2)

The second head model (HM2) had shown a good performance that was comparable to the measured HRTFs. However, the robustness of the beamformer designed with HM2 against variations of head-size and position is important for practical applications. Figure 3(b) shows that the HM2 is relative robust against the mismatch between the position of the left and right hearing aid and the true array positions (during the recording of the signals). The same applies to the variation of the head-model's parameter "sphere-size" which is not shown here. This results motivate the use of the head-model for hearing aid algorithms.

6. CONCLUSIONS

The robustness analysis has shown the importance of the incorporation of head-shadow and diffraction influences in the beamformer design for head-mounted arrays. The fixed beamformers designed with head models were relatively robust against steering errors whereas for adaptive beamformers the robustness was limited and a quality gain compared to fixed beamformers might only be reached in scenarios with directional noise sources and a reliable direction of arrival estimation. However, there are several approaches in literature to increase the robustness of the GSC [11] which have not been incorporated here.

The binaural speech intelligibility measure provides an integrative measure of binaural unmasking and could identify differences

in the estimated speech-reception threshold (SRT) if binaural information was distorted. Therefore, it seems to be an appropriate measure to evaluate the perceptual quality of noise reduction schemes with binaural output. In combination with different near-to-realistic sound-scenarios the quality measures showed encouraging results towards a robustness testbench for multichannel-hearing aid algorithms with binaural output. Further work should concentrate on a further empirical validation of the objective perceptual measures.

Work supported by the EC (DIRAC project IST-027787), HearCom-Project (IST-004171) and BMBF

7. REFERENCES

- [1] 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.
- [2] 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.
- [3] 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.
- [4] T. Van den Bogaert, T. J. Klaseen, 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.
- [5] 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.
- [6] T. Van den Bogaert, J. Wouters, S. Doclo, and M. Moonen, "Binaural cue preservation for hearing aids using an interaural transfer function multichannel wiener filter," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2007.
- [7] J. Desloge, W. Rabinowitz, and P. Zurek, "Microphone-array hearing aids with binaural output. i. fixed-processing systems," *IEEE Trans. on Speech and Audio Processing*, vol. 5, no. 6, pp. 529–542, Nov 1997.
- [8] D. Welker, J. Greenberg, J. Desloge, and P. Zurek, "Microphone-array hearing aids with binaural output. ii. a two-microphone adaptive system," *IEEE Trans. on Speech and Audio Processing*, vol. 5, no. 6, pp. 543–551, Nov. 1997.
- [9] J. Bitzer and K. U. Simmer, "Superdirective microphone arrays," in *Microphone Arrays*, Brandstein and Ward, Eds. Springer, 2001, ch. 2, pp. 19–38.
- [10] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. on Antennas Propagation*, vol. 30, pp. 27–34, 1982.
- [11] O. Hoshuyama and A. Sugiyama, "Robust adaptive beamforming," in *Microphone Arrays*, Brandstein and Ward, Eds. Springer, 2001, ch. 5, pp. 87–106.
- [12] 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.
- [13] 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.
- [14] ITU-T, "Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm," ITU, Series P: Telephone Transmission Quality Recommendation P.835, Nov. 2003.