# OBJECTIVE PERCEPTUAL QUALITY ASSESSMENT FOR SELF-STEERING BINAURAL HEARING AID MICROPHONE ARRAYS

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

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

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

## ABSTRACT

In this study a self-steering beamformer with binaural output for a head-worn microphone array is investigated in simulated and real-world conditions. The influence of the underlying sound propagation model on the estimation accuracy of the direction of arrival (DOA) estimation algorithm and the overall performance of the combined DOA-beamformer-system is evaluated. For this, technical performance measures as well as objective quality measures based on perceptual models of the auditory system are used. The self-steering beamformer showed better performance than a beamformer with fixed look-direction for SNR values above -2 dB if the propagation model includes at least a coarse head model.

***Index Terms***— Direction of arrival estimation, Array signal processing, Noise Reduction, Hearing aids, Perceptual audio quality estimation

## 1. INTRODUCTION

Multi-channel noise reduction schemes are promising solutions for hearing aids as they are capable to exploit the spatial distribution of the interfering signals. Thus, they lead generally to less signal distortion than single-channel noise reduction algorithms. For head-worn microphone arrays it is usually assumed that the look-direction is fixed at zero degrees, and that the user always turns his or her head towards the desired signal. This may become unsatisfying for the hearing aid user in particular for algorithms with a high spatial selectivity and if the signal of interest is moving. In this contribution a combination of a binaural beamformer [1, 2] and an automatic steering (electronic control of the look direction) based on the Generalized Cross Correlation (GCC) approach by Knapp and Carter [3] is applied. The importance of a proper model of wave propagation is investigated for a head-worn DOA-beamformer system. Furthermore, the performance of the system is evaluated in terms of estimation errors and signal-quality by means of objective perceptual measures that are based on models of the auditory system. With these measures the influences of inevitably occurring estimation errors can be quantified on a perceptual scale. Based on these results, the optimum compromise between algorithmic complexity and benefit can be derived.

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

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

## 2. SIGNAL MODEL AND BINAURAL MULTI-CHANNEL NOISE REDUCTION

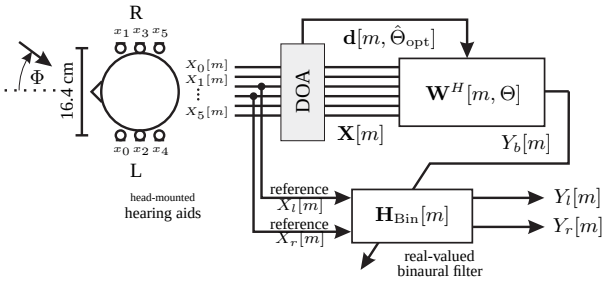



**Fig. 1**. Signal model and beamformer setup.

The noise reduction scheme used in this contribution is depicted in Fig. 1. With two 3-channel behind-the-ear (BTE) hearing aid shells mounted on a Brüel & Kjær (B&K) head and torso simulator (HATS), 6-channel head related transfer functions (HRTFs) were recorded in an anechoic room and in an office environment (reverberation time $\tau_{60} = 300$ ms) from different directions. A moving target signal was generated by filtering a speech signal with time-varying HRTFs that change due to a pre-defined virtual azimuth path (Fig. 2). Real-world environmental noise has also been recorded in a cafeteria and in an office room. Additionally, an artificial diffuse noise has been generated by summing up a speech-colored random noise that was filtered with HRTFs from all directions to simulate a cylindrical 2D-isotropic noise field. The moving speech signal was mixed with the noise signals at different signal-to-noise ratios (SNRs). In

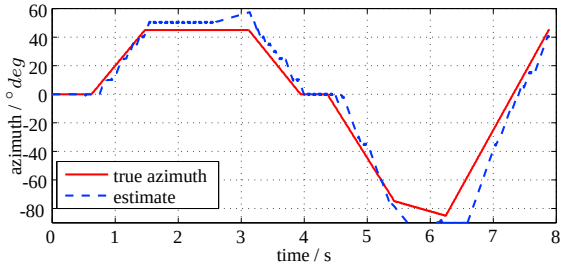



**Fig. 2**. Virtual azimuth path of moving speech source and its estimate for HM2 at 12 dB SNR.

Fig. 1, $X_i[m]$ denotes the audio-signal transformed into the fre-

quency domain by use of the short time Fourier transform (STFT), where $i = 0..5$ is the channel index. A DOA detection algorithm estimates the target signal's azimuth angle $\Theta$ which is used to steer the beamformer to this direction by means of the propagation vector $\mathbf{d}[m,\Theta]$. The beamformer $\mathbf{W}[m,\Theta]$ generates a single channel output $Y_b[m]$ via the well known Minimum Variance Distortionless Response (MVDR) approach [4]:

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

$$\mathbf{d}[m,\Theta] = [d_0[m,\Theta], d_1[m,\Theta], \ldots, d_{N-1}[m,\Theta]]^T \quad (2)$$

$$d_i[m,\Theta] = |d_i[m,\Theta]|e^{-j2\pi m\frac{f_s}{M}\tau_i[m,\Theta]},\ i = 0..N-1 \quad (3)$$

The fixed noise-field characteristic is coded in the coherence matrix $\mathbf{\Gamma_{NN}}[m]$ which additionally influences beamformer properties directivity and susceptibility to white noise, and therefore has to be constrained [4, 1]. Both, $\mathbf{d}[m,\Theta]$ and $\mathbf{\Gamma_{NN}}[m]$ depend on to the assumed wave propagation model which may differ from the true (and generally unknown) wave propagation from the source to the microphones. We distinguish four models, free-field (FF), two head models (HM1 [5], HM2 [6]) and the measured anechoic transfer functions from the source to the head-mounted hearing aid microphone array (HRTF). The simplest approach is to use a free-field / far-field assumption (FF), i.e., the sound propagation is modeled as a plane wave without interfering objects in the propagation path. For FF, $\mathbf{d}[m,\Theta]$ has unity magnitude, $|d_i[m,\Theta]| = 1\ \forall(i,m,\Theta)$ and constant group delay $\tau[m,\Theta] = \tau[\Theta]$ that can be calculated from the inter-microphone distance and the angle of incidence. For head-worn arrays it is beneficial to include knowledge about head shadow and diffraction effects [1, 11], especially for lateral target signal sources. Thus, head models by Duda et al. [5, 6] are applied which are effective parametric models that are based on the characteristics of a sphere. In HM1, 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 at low frequencies [5]. In HM2, near-field effects and interference effects that introduce ripples in the frequency response which are quite prominent on the shadowed side are incorporated as described in [6]. For both head models (HM1, HM2) the frequency dependent group delay $\tau[m,\Theta]$ and magnitude have to be calculated for each microphone and angle of incidence due to [5, 6]. For HRTF, the propagation vector $\mathbf{d}[m,\Theta]$ equals the measured anechoic 6-channel HRTF for the angle of incidence $\Theta$. $\Gamma_{NN}[m]$ can be estimated for a cylindrical isotropic diffuse noise field by integrating the propagation vectors over all directions $\Theta$. For FF, this solution can be calculated via the Bessel function of the first kind of order zero. For the white noise gain constraints and further details see [4].

The binaural output is calculated by a real-valued time-varying post-filter based on [2] that is controlled by the monaural beamformer output $Y_b$:

$$H_{\mathrm{Bin}}[m] = \frac{\left(|d_l[m,\Theta]|^2 + |d_r[m,\Theta]|^2\right)\Phi_{Y_bY_b}[m]}{\Phi_{X_lX_l}[m] + \Phi_{X_rX_r}[m]} \quad (4)$$

$$Y_l[m] = H_{\mathrm{Bin}}[m]X_l[m] \quad (5)$$

$$Y_r[m] = H_{\mathrm{Bin}}[m]X_r[m] \quad (6)$$

Here $X_l[m], X_r[m]$ (see Fig. 1) denote the reference input signals and $d_l[m], d_r[m]$ the propagation coefficients for the estimated signal direction $\Theta_{\mathrm{opt}}$, at the left and right reference microphone, respectively. $\Phi_{Y_bY_b}[m]$, $\Phi_{X_lX_l}[m]$ and $\Phi_{X_rX_r}[m]$ are the power spectral density estimates for the signals $Y_b[m], X_l[m], X_r[m]$, respectively. As depicted in Fig. 1 we chose channel 3 and 4 as reference channels for the left and right site. For a detailed analysis of the binaural output see [1].

## 3. DIRECTION OF ARRIVAL ESTIMATION

Direction of arrival estimation is done by estimating the signal delay between microphone pair $x_l[k], x_r[k]$ via the PHAT-GCC (Phase Transform Generalized Cross Correlation) [3] which has been proven to give reliable estimates for various environments:

$$\tau_d = \arg\max_k R_{x_lx_r}[k] \quad (7)$$

with the (PHAT) generalized cross correlation [3]

$$R_{x_lx_r}[k] = \frac{1}{L_{\mathrm{DFT}}}\sum_{m=0}^{L_{\mathrm{DFT}}-1}\frac{\Phi_{x_lx_r}[m]}{|\Phi_{x_lx_r}[m]|}e^{j\frac{2\pi}{M}mk},\ k = 0..L_{\mathrm{DFT}}-1 \quad (8)$$

Typical signal delays that occur between the left and right microphones are about $8.3\mu s/1^\circ$deg in the range of $\pm30^\circ$ deg. For a sampling rate of 16 kHz these are $7.5^\circ$ deg per sample. Thus, an appropriate oversampling of the generalized cross-correlation $R_{x_lx_r}[k]$ is suggested.

The time-delay of arrival due to diffraction is longer for lateral signals then expected in the free-field case. Therefore the time-delay corresponds to other angles of incidence for the head models than for the free-field. Fig. 3 depicts deviations that occur due to a wrong delay-to-azimuth mapping. Fig. 3(a) shows the time delay of arrival

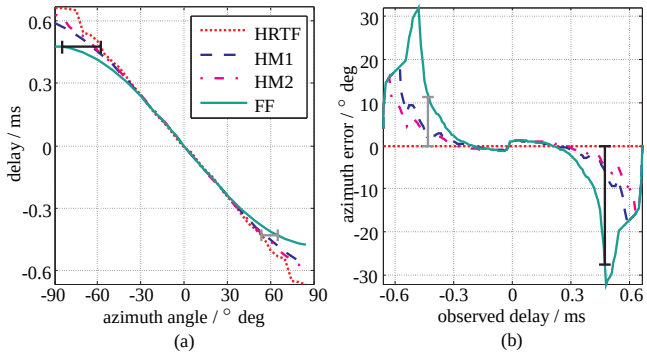



**Fig. 3**. Azimuth error for different time delays $\tau_d$ and propagation models.

between the microphones $x_l[k]$ and $x_r[k]$ against the azimuth angle for different propagation models. Between $\pm30^\circ$ the dependency is almost linear and only little deviations between the propagation models exist. For more lateral angles the differences increase due to the increased traveling time of the sound signals around the human head. In Fig. 3(b) the deviation of the estimated angle for the propagation model and true angle as determined from the measured HRTF is depicted. Note that for the free-field model (FF) delays beyond $\pm0.5$ ms are assigned to $\pm90^\circ$. Therefore, the azimuth error decreases for values beyond these maximum delays. The gray and black bars show the corresponding values in (a) and (b). It can be seen that the head models give a better approximation of the true time delay than FF assumptions. Although the group delays for the

head models are frequency dependent [5], these effects are omitted here as they only apply for low frequencies (< 200 Hz). A maximum tracking speed of the DOA estimator is limited to $125^\circ/s$ as described in [11] to avoid sudden peaks in the DOA estimate that lead to severe disturbances of the subsequent beamformer. A simple speech activity detector based on the magnitude of $R_{x_l x_r}[k]$ is applied by updating the DOA estimate only if $R_{x_l x_r}[k]$ is greater than a threshold $\xi$. During speech pauses a tracking algorithm based on the last estimates continues the update of the azimuth estimate. However for the application in a hearing aid it might be useful to apply more sophisticated tracking algorithms that increase the robustness of the estimate while at the same time allowing for a quick change of direction due to a moving speaker. Here, our main focus lies on understanding the principle problems due to imperfect propagation models.

## 4. QUALITY ASSESSMENT

It has been shown in Fig. 3 that the assumption of an imperfect propagation model leads to systematic errors in the estimation of the signal-source direction. As we are interested in the influence of these estimation errors on the performance and signal quality for realistic scenarios we propose three performance measures.

**SNRE:** The SNR-Enhancement (SNRE) is the difference of the SNR at the output of the beamformer and a reference input-SNR, both measured in dB. For binaural systems the SNRE is calculated between the left (right) output of the binaural post-filter and the left (right) input at the reference microphone, respectively; by simply taking the mean SNRE a better-ear effect would be ignored.

**PSM / ΔPSM:** The quality measure PSM from PEMO-Q [7] estimates the perceptual similarity between the processed signal and the clean speech source signal. It has shown high correlations between objective and subjective data and has been used for quality assessment of noise reduction schemes in [1, 8, 9]. PSM increases with increasing (input) SNR. As we are interested in the quality enhancement introduced by the algorithm, we use the deduced measure ΔPSM that is calculated as the difference between the Perceptual Similarity Measure (PSM) of the output and of the unprocessed input signal.

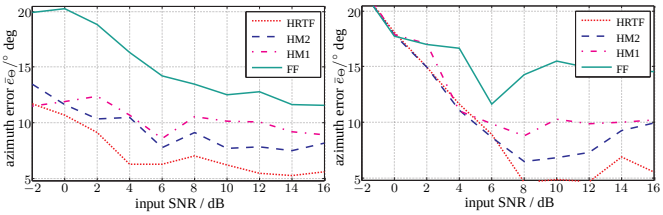
**Binaural SRT / ΔSRT:** The speech reception threshold (SRT) is defined as the signal-to-noise ratio (SNR) at 50% speech intelligibility. In [10] a binaural model of speech intelligibility based on the equalization-cancelation (EC) processing by Durlach had been defined which is able to predict the SRT with high accuracy. If the estimated SRT for the output of a noise reduction scheme is lower than for the input signal this means that the speech intelligibility has increased due to the algorithm. However, as the speech intelligibility is a nonlinear function of the SNR and other signal features such as the preservation of binaural cues, we use the difference between output and input SRT, namely the ΔSRT, as an indirect measure for the increase of intelligibility. The binaural SRT measure as described in [10, 1] assumes a spatially stationary source configuration. To be applicable to moving sources it had to be extended to a block-wise measure with subsequent averaging across blocks.

## 5. SIMULATION RESULTS

### 5.1. DOA Estimation Error

Fig. 4 shows the mean azimuth estimation error of the DOA algorithm $\bar{e}_\Theta = \frac{1}{|\mathcal{A}|}\sum_{\mathcal{A}} \Theta - \hat{\Theta}$ over the input SNR for the four propagation models. Here, $\Theta$ and $\hat{\Theta}$ 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. In artificial diffuse noise, Fig. 4(a), the mean azimuth error for the head models is below $15^\circ$ degree at an SNR of $-2$ dB and falls below $10^\circ$ for an SNR $> 2-4$ dB depending on the exactness of the model. The measured (in practice generally unknown) HRTF shows the best performance followed by HM2 which seems to be a feasible approximation. Assuming free-field, $\bar{e}_\Theta$ is persistently $3-7\,^\circ$ greater than for the head models.

The performance for this algorithm in a recorded real-world office environment with ambient noise, Fig.4(b), is worse at $-2$ dB SNR than for artificial diffuse noise, but $\bar{e}_\Theta$ also falls below $10^\circ$ for an input SNR $> 5$ dB for the head models. Compared to the results gained in [11] where a DOA estimator based on the dual delay line approach was evaluated, it can be can be stated that the GCC-PHAT algorithm performs much better, particularly in noisy conditions.




(a) Articial diffuse noise (b) Office noise

**Fig. 4**. Mean DOA error in different noise conditions.

### 5.2. Objective Perceptual Quality of the whole system

Fig. 5 shows the performance measures described in Section 4 over the SNR of the input signal ($\text{SNR}_{in}$). If not indicated otherwise, results are shown for the diffuse noise. The Signal to Noise Ratio Enhancement (SNRE) in Fig. 5(a) slightly decreases with increasing $\text{SNR}_{in}$ which is a fact common to all noise reduction systems as for infinite $\text{SNR}_{in}$ the SNRE converges to zero. The ideal system (solid black line) has *a priori* information about the direction of arrival and uses the measured HRTF as a propagation model. Therefore, it should set the upper performance limit. Also, it would be expected that the systems with the most exact propagation model (HRTF and HM2, before HM1 and FF) have the highest SNRE. However, this is not seen in the right channel where FF (solid green) crosses HM2 (dashed blue). This is an artifact of the broadband SNRE measure that is suboptimal for quality assessment, as it does not incorporate signal distortions. For PSM in Fig. 5(b) the ranking behaves as expected: The ideal system sets the upper limit and the system with the fixed look direction to $0^\circ$ shows the worst performance. The absolute PSM (not shown here) for the ideal system lies between 0.6 and 0.9 (where values close to 1 mean that the signal is perceptually undistinguishable from the clean speech [7]). A negative ΔPSM shows a signal degradation compared to the unprocessed signal, e.g., FF and $0^\circ$ fixed at SNR>12 dB. For the head models ΔPSM is consistently higher than for the fixed system, whereas for FF the quality enhancement is marginal. Fig. 5(c) shows the decrease of the Speech Reception Threshold (SRT) due to the noise reduction that also incorporates the speech intelligibility benefit due to the preservation of binaural cues. Again, the ranking is consistent with the exactness of the propagation model. For input SNR values where the DOA estimation has low errors, HM2 and HRTF have less than $0.5$ dB higher SRT than the ideal system. For FF, ΔSRT lies $1.5$ dB higher than for the ideal system. All self-steered systems with head models have a lower SRT than the system fixed to $0^\circ$ degree look-direction for all $\text{SNR}_{in}$ whereas for FF this is the case at an $\text{SNR}_{in}$ > 3dB. In those

cases steered systems are superior to fixed systems for the given input signals. Fig. 5(d) and 5(e) show the performance for real-world recordings in the office room mixed with (d) office ambient noise and (e) babble noise from a cafeteria. A $\Delta$SRT close to the ideal system indicates a good performance which is given for the head models at a $\mathrm{SNR}_{in} > 4$ dB for the ambient noise (d) and a $\mathrm{SNR}_{in} > 9$ dB for babble noise (e). For FF, $\Delta$SRT is significantly higher in (d) and it is close to the fixed system in (e). In summary it can be stated that for the difficult cafeteria noise condition where sudden correlated noise sources may occur, DOA estimation performance for a fast moving target signal source at low SNR is poor. However, for input $\mathrm{SNR}_{in}$ >9 dB automatic-steered systems are favorable, given an appropriate propagation model.

## 6. CONCLUSION

We presented a self-steering multi-channel noise reduction system with binaural output applicable to hearing aids. Estimation errors have been analyzed under the assumption of different wave propagation models. For a fast moving speech source under different simulated and real-world noise conditions, algorithm performance was evaluated using technically based measures and objective perceptual quality measures based on auditory models. The results show that for signal-to-noise ratios (SNRs) greater -2 dB self-steering systems are superior to fixed systems if a certain complexity of the propagation model is met. The DOA-beamformer system performs best in diffuse or ambient noise conditions. However, in difficult noise conditions such as cafeteria noise, the performance is lower than for a simulated system with a priori knowledge about the direction of arrival.

### 7. REFERENCES


[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] C. H. Knapp and G. C. Carter, "The Generalized Correlation Method for Estimation of Time Delay," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976.

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

[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, Sep 1998.

[6] 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.

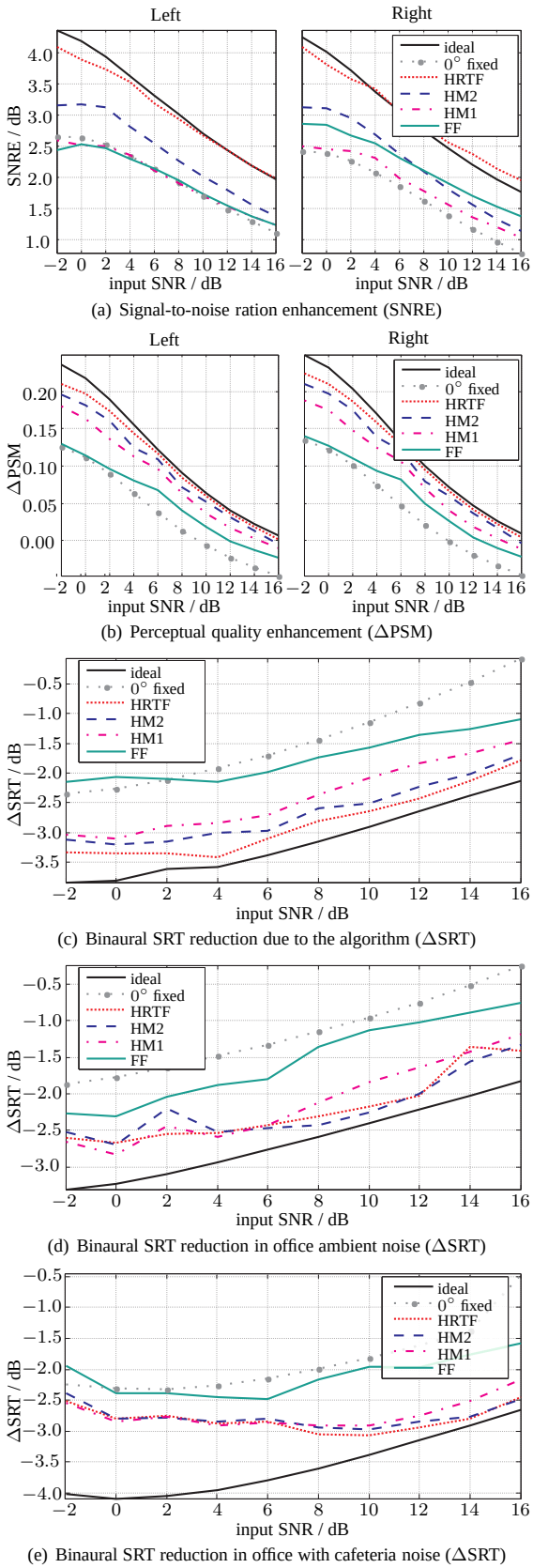
[7] R. Huber and B. Kollmeier, "Pemo-Q -A new Method for Objective Audio Quality Assessment using a Model of Auditory Perception." *IEEE Trans. on Audio, Speech and Language Processing*, 2006, special Issue on Objective Quality Assessment of Speech and Audio.

[8] 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.

[9] 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.

[10] 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.

[11] 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.


(a) Signal-to-noise ration enhancement (SNRE)

(b) Perceptual quality enhancement ($\Delta$PSM)

(c) Binaural SRT reduction due to the algorithm ($\Delta$SRT)

(d) Binaural SRT reduction in office ambient noise ($\Delta$SRT)

(e) Binaural SRT reduction in office with cafeteria noise ($\Delta$SRT)

**Fig. 5**. Objective quality assessment of DOA plus beamformer system with different wave propagation models