Combined Source Tracking and Noise Reduction for Application in Hearing Aids

*Thomas Rohdenburg, †Stefan Goetze, *Volker Hohmann, †Karl-Dirk Kammeyer, and *Birger Kollmeier

*Medical Physics Group, University of Oldenburg, 26111 Oldenburg, Germany
E-Mail: thomas.rohdenburg@uni-oldenburg.de
Web: http://medi.uni-oldenburg.de

†Dept. of Communications Engineering, University of Bremen, 28359 Bremen, Germany
E-Mail: goetze@ant.uni-bremen.de
Web: http://www.ant.uni-bremen.de

Abstract

Multi-channel self-steering microphone arrays for hearing aid applications enable the hearing impaired user to follow conversations coming from other directions than the front direction. Binaural connections between left and right hearing aids allow for (i) the estimation of the direction of arrival (DOA) and thus for automatic beam-steering, (ii) a higher noise reduction due to a greater number of sensors, and (iii) the preservation or reconstruction of binaural cues that can be used by the hearing impaired listener for object segregation. Since the microphones are placed near the human head, diffraction and shadowing effects have to be incorporated into the beamformer design as well as into the DOA estimator.
In this contribution a combined system of DOA estimation and a head-worn 6-channel binaural noise reduction scheme is presented and evaluated using perceptual measures that are based on auditory models.

1 Introduction

Multi-channel noise reduction schemes are frequently used for application in hearing aids as they exploit spatial information about the signals and therefore, in general, lead to a higher noise reduction and lower signal distortion than single channel techniques.
Binaural connections between left and right hearing aids have been approved in recent publications [1, 2] and the first hearing aids with wireless links are available on the market that transfer program and algorithm settings. It can be expected that in near future also full-band audio information will be transmitted provided that a significant performance gain can be archieved. For binaural head-worn beamformer systems head shadow and diffraction effects become important, in particular for algorithms with a high spatial selectivity [1]. Up to now, beamformer systems for hearing aids made the assumption that the relative target direction is at the front direction. However, this assumption might become unsatisfying for the hearing aid user if the signal of interest is coming from the side or is even moving. In this contribution a direction of arrival (DOA) estimator in combination with a beamformer that automatically steers to the most prominent source is suggested. For head-worn binaurally linked microphone arrays it has been shown in [1, 3] that for a moving sound source, self-steering beamformers lead to better performance than beamfomers with fixed look-direction for SNR values above -2 dB if the propagation model at least included a coarse head model. In this contribution the head related DOA estimation technique is reviewed and different approaches are compared in terms of their performance and robustness under adverse noise conditions. The remainder of this paper is organized as follows: In Section 2 the binaural noise reduction scheme is briefly introduced. Head shadowing and diffraction effects on the noise reduction scheme are described in Section 3 and DOA estimation is discussed in Section 4. Section 5 presents simulations results and Section 6 concludes the paper.

2 Binaural Noise Reduction

Figure 1 shows the combined system for noise reduction and DOA estimation. 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 have been measured, both in an anechoic and an office environment (room reverberation time $\tau_{60} \approx 300$ms). Additionally, 6-channel recordings have been made under realistic noise conditions in an office and in a cafeteria.

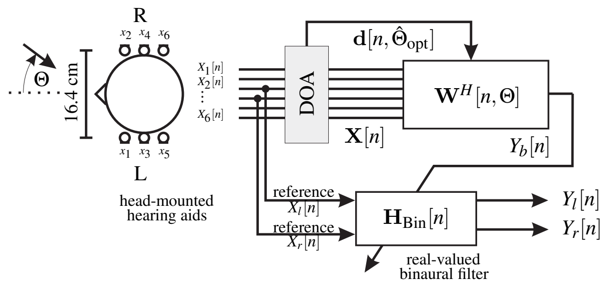

Figure 1: Signal model and beamformer setup.

The well-known minimum variance distortionless response (MVDR) beamformer is applied in the frequency domain to generate a single-channel estimate of the desired signal from the 6-channel input signals. We used a constrained superdirective beamformer design [4].

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

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

$$d_m[n,\Theta] = |d_m[n,\Theta]| e^{-j2\pi n \frac{f_s}{N} \tau_m[n,\Theta]}, \; m = 0..M-1 \quad (3)$$

In (1) $\mathbf{\Gamma}_{NN}[n]$ and $\mathbf{d}[n,\Theta]$ are the noise coherence matrix and the propagation vector in discrete frequency domain, respectively. k, n, N, Θ, m, M, and τ are the discrete time index, the discrete frequency index, the DFT length, the angle of incidence, the channel index, the number of microphones, and the traveling delay between the microphones, respectively. The desired direction of arrival and the propagation model are included in the propagation vector (see Section 3 and 4). A post filter [5, 6] generates a binaural audio signal from the beamformer's output and two reference signals to provide the user with binaural information.

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

$$Y_l[n] = H_{\text{Bin}}[n] X_l[n] \quad (5)$$

$$Y_r[n] = H_{\text{Bin}}[n] X_r[n] \quad (6)$$

For a comparison of different binaural techniques see [6].

3 Incorporation of Head Influences

Humans make use of binaural information for signal unmasking and object segregation. Two important binaural cues are, amongst others, the interaural time difference (ITD) and the interaural level difference (ILD), depicted in Figure 2 and 4. These effects are due to diffraction and shadowing of sound waves at the human head and also need to be considered in algorithms for head-worn microphone arrays, particularly in the DOA estimator and the beamformer [1, 6, 3]. Since head related transfer functions (to the BTE hearing aids) are different for each user, it is evident to use parametric head models [7, 8] instead of individual HRTFs to incorporate the coarse head characteristics into the algorithm design. Figure 2 shows the interaural time differences (ITD) for free-field assumptions (FF), measured HRTFs, head models HM1 [7], and HM2 [8]. HM2 uses a somewhat more sophisticated model and has proven to lead to better results for noise reduction and DOA estimation schemes (see [1, 6, 3] and Section 5). While hardly any difference exists at angles at about 0° (front direction) delay differences of about 0.3ms at ±90° correspond to an angle estimation errors of up to 30° [1]. Thus a mapping to the correct angles is an essential step if DOA estimators shall work with head mounted devices.

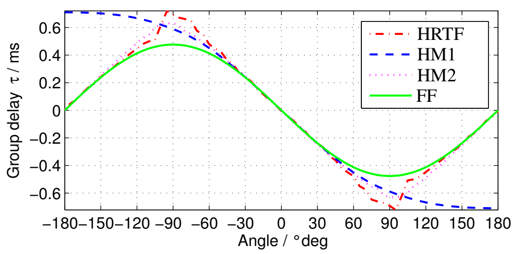

Figure 2: Group delay for different propagation models.

For the incoming sound wave the head is a dispersive spatial filter which leads to diffraction (and thus to longer traveling times around the head) for lower frequencies and shadowing (i.e.: amplitude reduction) for higher frequencies. Figure 3 visualizes the differences of the traveling distances of a sound wave under free-field conditions and for a spherical head model. The differences become particularly large for angles around 90° as depicted in Figure 2.

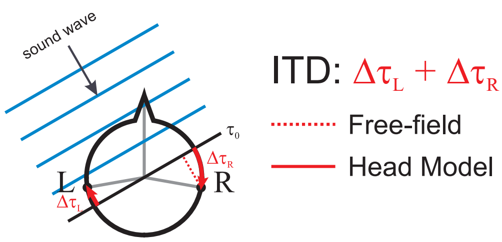

Figure 3: Diffraction at a sphere.

Figure 4 shows the influence of ILD for measured HRTFs (left) and for the head model (HM2). Although head models do not give an exact estimate of the user-dependent HRTF, the microphone and frequency dependent ILDs are roughly estimated by HM2.

4 Direction of Arrival Estimation

In [3] and [1] DOA estimation methods known from the literature (e.g. based on the dual delay line approach [9] or the generalized cross correlation (GCC) approach [10]) have been extended to work for head mounted microphone arrays. Since the GCC approach turned out to be more robust in terms of environmental noise and room reverberation [1, 3] we focus on this approach

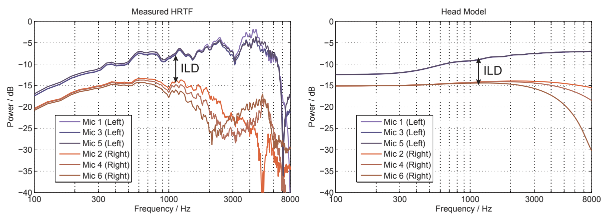

Figure 4: Interaural level differences for measured HRTFs and head models.

and its extensions here. The time delay corresponding to the estimated direction of arrival can be determined by

$$\hat{\tau} = \arg\max_{k} R_{x_l x_r}[k] \tag{7}$$

with the generalized cross correlation [10]

$$R_{x_l x_r}[k] = \frac{1}{N}\sum_{n=0}^{N-1} \Psi[n] X_l[n] X_r^*[n] e^{j\frac{2\pi}{N}nk}. \tag{8}$$

For the weighting function $\Psi[n]$ several proposals exist which aim to emphasize the GCC value at the true time delay of arrival over the undesired local maxima. One well-known weighting function which has proven to be robust in noisy and reverberant conditions is the PHAse Transform (PHAT) weighting function $\Psi_{\mathrm{PHAT}}[n] = 1/|\Phi_{X_l X_r}[n]|$, aka the crosspower spectrum phase (CSP).

In practice, the direction of arrival for the GCC-PHAT method is determined in three steps. First, $R_{x_l x_r}[k]$ is calculated at equidistant time samples k. Since in practice, time differences of arrival (TDOA) between two microphones are short and the interesting area is covered by only a few samples, the crosscorrelation $R_{x_l x_r}[k]$ is interpolated by an oversampled IFFT. Afterwards, the time-delay which corresponds to the highest correlation value is found by a maximum search $\arg\max_k$. As we are interested in the correlation function on a equidistant azimuth angle scale, namely, the direction of arrival, we finally have to re-map the time delay $\hat{\tau}$ to the azimuth angle with a non-linear mapping function which can become quite complex for the head-related case. Figure 2 depicts this mapping for free-field, head-models and measured HRTFs. This three-step estimation method is suboptimal, because for a satisfying resolution of lateral azimuth angles between $|\Theta| = 30° \ldots 90°$, the oversampling needs to be high whereas for angles between $[-30° \ldots 30°]$ the DFT resolution is sufficient (for a microphone pair in broadside direction). Thus, by directly applying time-delays that are equidistant on the azimuth-scale we can save computational costs for the IFFT and the mapping function.

$$R_{x_l x_r}[\Theta] = \frac{1}{N}\sum_{n=0}^{N-1} \Psi[n] X_l[n] X_r^*[n] e^{j\frac{2\pi}{N}n\tau_{lr}[\Theta,n]} \tag{9}$$

Here $e^{j\frac{2\pi}{N}n\tau_{lr}[\Theta,n]}$ is the phase component of the inter-microphone transfer function between microphone l and r for a source signal impinging from the direction Θ. Note, that τ_{lr} may also be frequency dependent accounting for dispersion effects observed for head-worn arrays.

The DOA estimation method in (9) has been described by DiBiase in [11] as the spatial response phase transform (SRP-PHAT). DiBiase analyzed the performance of the SRP-PHAT using all microphone combinations l and r.

$$\hat{\Theta} = \arg\max_{\Theta} \sum_{i=1}^{M}\sum_{j=1}^{M} R_{x_i x_j}[\Theta] \tag{10}$$

Although redundancies of the correlation between microphone pairs $[l,r]$,$[r,l]$ and autocorrelations $[l,l]$ where included in the estimate, DiBiase found no detrimental effect using all

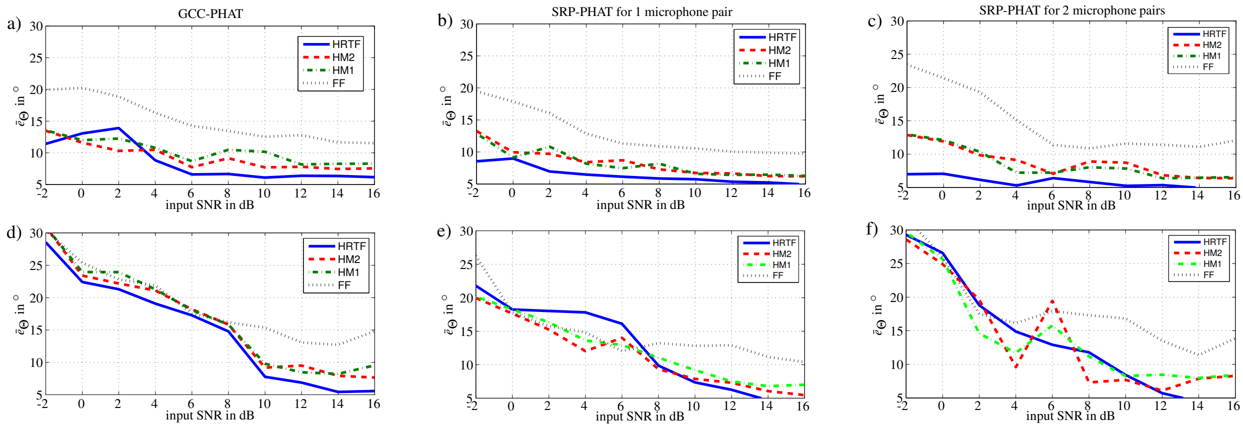

Figure 5: Mean DOA estimation error over input SNR for diffuse noise conditions without reverberation (upper part, a)-c)) and in reverberant environment ($\tau_{60} \approx 300$ms) and babble noise conditions (lower part, d)-f)) for GCC-PHAT (a), d)) and SRP-PHAT for 1 microphone pair (b), e)) and 2 microphone pairs (c), f)).

combinations. However, for the microphone array used here, it was expected that some microphone pairs might not yield any information about the direction of arrival because of a very low inter-microphone distance.

	Position		
Mic. no.	x in mm	y in mm	z in mm
(1) LF	14.9	0	4.7
(2) RF	14.9	-164	4.7
(3) LM	7.3	0	2.6
(4) RM	7.3	-164	2.6
(5) LB	0	0	0
(6) RB	0	-164	0

Table 1: Microphone positions in mm (compare Figure 1).

In Table 1 and 2 the absolute positions of the microphones used in the hearing aid setup and the inter-microphone spacings are given, respectively.

	Distance in mm to microphone no.					
Mic. no.	(1) LF	(2) RF	(3) LM	(4) RM	(5) LB	(6) RB
(1) LF	-	164.0	7.9	164.2	15.6	164.7
(2) RF	164.0	-	164.2	7.9	164.7	15.6
(3) LM	7.9	164.2	-	164.0	7.7	164.2
(4) RM	164.2	7.9	164.0	-	164.2	7.7
(5) LB	15.6	164.7	7.7	164.2	-	164.0
(6) RB	164.7	15.6	164.1	7.7	164.0	-

Table 2: Distances between microphones in mm. (L:left, R:right, F:front, M:middle, B:back).

As it can be seen from Table 2 the distances between some of the microphone pairs (e.g. microphone 1 and 3) are very small. Thus, they may be too small for a DOA estimation by means or GCC-PHAT or SRP-PHAT since real-world noise fields often are diffuse. Diffuse noise fields are highly correlated up to a certain frequency which raises for smaller inter-microphone distances. Furthermore, the spatial positions of the microphone pair consisting of microphone 1 and 2 and the microphone pair consisting of microphone 3 and 4, e.g., are quite similar. Thus a combination of these microphone pairs may provide only little more information about the desired signal. For the experiments a subset of all possible microphone pairs was used and different combinations were evaluated compared to a single microphone pair.

$$\hat{\Theta} = \arg\max_{\Theta} \sum_{p=1}^{P} R_{x_{p,1}x_{p,2}}[\Theta] \qquad (11)$$

Here p is number of the actual microphone pair and P is the number of microphone pairs used. For the investigated hearing aid system, the DOA could theoretically be estimated from $P_{\max} = \frac{(M-1)\cdot M}{2}$ pairs where $M = 6$ is the number of microphones.

5 Simulation Results

In Figure 5 the mean DOA estimation error $\bar{e}_\Theta = \frac{1}{|\mathscr{A}|}\sum_{\mathscr{A}} \Theta - \hat{\Theta}$ is shown depending on the input SNR for the GCC-Phat method (subplots a) and d)), the SRP-PHAT method using one microphone pair (subplots b) and e)) and the SRP-PHAT method using two microphone pairs (subplots c) and f)). Here, Θ and $\hat{\Theta}$ are the true and the estimated direction of arrival, respectively. $\mathscr{A}$ is the set of frames where speech is present and $|\mathscr{A}|$ its cardinality. The upper subplots (a)-c)) show the performance of the algorithms for an anechoic situation (no reverberation) in diffuse noise conditions while the lower subplots (d)-f)) show the performance of the DOA estimators in a reverberant environment ($\tau_{60} \approx 300$ms) and babble noise conditions. Different propagation models (free-field (FF), head models (HM1, HM2) and measured HRTFs) were evaluated. It can be seen that, in general, the algorithms perform best for measured HRTFs and worst if no diffraction and shadowing effects are incorporated into the design (FF). Using head models is a good approximation for the measured HRTF which is unknown in practical systems. Comparing the GCC-PHAT and SRP-PHAT curves reveals that the SRP-PHAT algorithm performs slightly better. However, the averaging over multiple microphone pairs did not lead to the expected performance improvement (particularly not for real-world conditions including reverberation and high environmental noise) that was reported in literature. It was found that a small variance decrease of the correlation matrix $R_{x_{p,1}x_{p,2}}$ could be seen for an ideal diffuse noise field. Looking at real-world recorded babble noise this small effect disappeared because it had a stronger correlation which was seen by all microphone pairs simultaneously. In summary it can be stated that the SRP-PHAT algorithm using only one microphone pair (9) showed the best performance for the given microphone setup.

Figure 6 shows the performance of the combined SRP-PHAT-steered noise reduction system evaluated by three objective measures: the signal-to-noise ratio enhancement SNRE, the perceptual similarity measure ΔPSM from PEMO-Q [12, 1] and the binaural speech intelligibility measure ΔBSIM [14, 1]. All three measures show the relative enhancement of the input signal. The broadband SNRE shows the amount of noise reduction in a technical sense. Although it is an established measure which is correlated with the perceived amount of noise reduction [13] it has

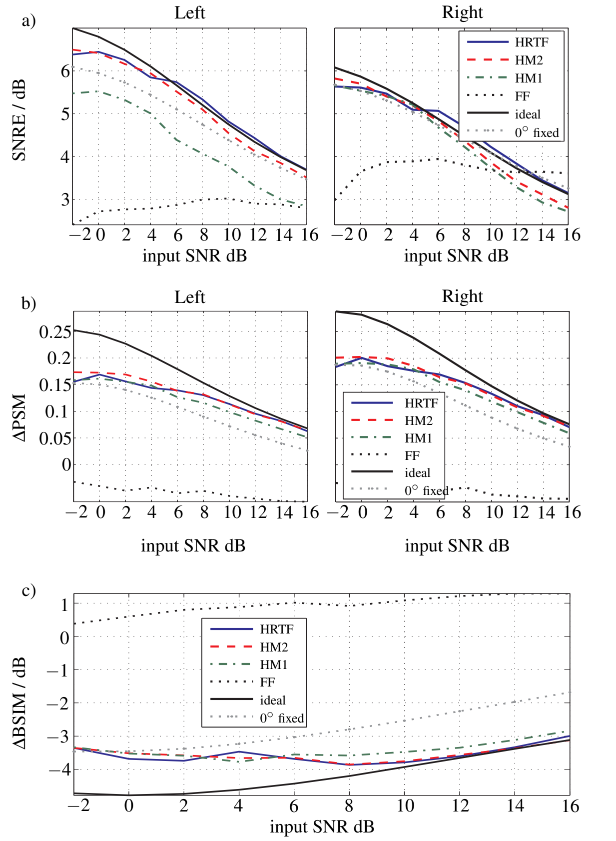

Figure 6: Performance evaluation with the objective measures SNRE, ΔPSM and ΔBSIM for a speech signal in a reverberant office environment ($\tau_{60} = 300$ ms) mixed with babble noise at different input SNR's and processed by the SRP-PHAT-steered binaural noise reduction scheme.

some severe deficiencies as signal distortions may not be seen properly. Thus inconsistencies may occur, e.g., in Figure 6 a) the estimated system is better than the optimal system. In subplot b), ΔPSM shows the increase of the estimated perceptual quality compared to the input signal. Both, SNRE and PSM are monaural measures and therefore are evaluated for left and right output signal, respectively. The difference between the input and output speech reception threshold is estimated by ΔBSIM in Figure 6 c). This measure integrates binaural information that might be used by the listener for localization and object segregation. Obviously, the performances of the self-steered systems converge to the ideal system at about 8 dB. A ΔBSIM value of -4 dB means, e.g., that the expected speech reception threshold (i.e., 50% speech intelligibility) of the binaural output is 4 dB lower than the speech reception threshold estimated for the input signal. Thus ΔBSIM can be interpreted as the amount of additional *head-room* of speech intelligibility achieved by the binaural noise reduction scheme. For a further evaluation of ΔBSIM (which was formerly named ΔSRT), and other measures the reader is referred to [1, 6, 14].
The performance of the beamformer designed for free-field was much lower than for the head-model based designs. This is partly due to the white noise gain constraint which is an important factor that influences the amount of noise reduction and that had been optimized for the head-worn array. However, it has already been shown in [6, 1] that free-field beamformers are suboptimal for head-worn arrays.

6 Conclusions

The results show that the SRP-PHAT method using only a subset of microphone pairs leads to a slightly higher performance than the GCC-PHAT method and simultaneously reduces computational load which is due to a better sampling of the parameter space with equidistant azimuth angles and the dispensable TDOA-mapping. For the microphone array used here, a combination of multiple microphone pairs did not lead to a consistent improvement compared to a single microphone pair. This can be explained by the low inter-microphone distances and the similarity of the correlation patterns for the microphone pairs that are applicable to DOA estimation. The combination of SRP-PHAT based direction of arrival estimation and a constrained superdirective beamformer showed that the objectively estimated signal quality and the estimated speech intelligibility were improved compared to traditional non-steered systems if at least a course head model was included in the DOA estimation algorithm and the binaural noise reduction scheme for head-worn microphone arrays.

References

[1] T. Rohdenburg, S. Goetze, V. Hohmann, K.-D. Kammeyer, and B. Kollmeier, "Objective Perceptual Quality Assessment for Self-Steering Binaural Hearing Aid Microphone Arrays," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Las Vegas, USA, pp. 2449–2452, March 30 - April 4, 2008.

[2] S. Doclo, T. van den Bogaert, J. Wouters, and M. Moonen, "Comparison of Reduced-Bandwidth MWF-Based Noise Reduction Algorithms for Binaural Hearing Aids," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'07)*, New Paltz, USA, pp. 223–226, October 21 - 24, 2007.

[3] 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.

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

[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. 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, USA, Oct. 2007.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] J.H. DiBiase. *A High-Accuracy, Low-Latency Technique for Talker Localization in Reverberant Environments Using Microphone Arrays*. PhD thesis, Brown University, Providence, Rhode Island, USA, May 2000.

[12] 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.

[13] 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, pp. 169–172, 2005.

[14] 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.