Trends in Hearing
2015, Vol. 19: 1–16
© The Author(s) 2015
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2331216515617916
tia.sagepub.com
SAGE

Comparing Binaural Pre-processing Strategies I: Instrumental Evaluation

Regina M. Baumgärtel[1,2], Martin Krawczyk-Becker[2,3], Daniel Marquardt[2,4], Christoph Völker[1,2], Hongmei Hu[1,2], Tobias Herzke[2,5], Graham Coleman[2,5], Kamil Adiloğlu[2,5], Stephan M. A. Ernst[1,2], Timo Gerkmann[2,3], Simon Doclo[2,4], Birger Kollmeier[1,2], Volker Hohmann[1,2,5], and Mathias Dietz[1,2]

Abstract

In a collaborative research project, several monaural and binaural noise reduction algorithms have been comprehensively evaluated. In this article, eight selected noise reduction algorithms were assessed using instrumental measures, with a focus on the instrumental evaluation of speech intelligibility. Four distinct, reverberant scenarios were created to reflect everyday listening situations: a stationary speech-shaped noise, a multitalker babble noise, a single interfering talker, and a realistic cafeteria noise. Three instrumental measures were employed to assess predicted speech intelligibility and predicted sound quality: the intelligibility-weighted signal-to-noise ratio, the short-time objective intelligibility measure, and the perceptual evaluation of speech quality. The results show substantial improvements in predicted speech intelligibility as well as sound quality for the proposed algorithms. The evaluated coherence-based noise reduction algorithm was able to provide improvements in predicted audio signal quality. For the tested single-channel noise reduction algorithm, improvements in intelligibility-weighted signal-to-noise ratio were observed in all but the nonstationary cafeteria ambient noise scenario. Binaural minimum variance distortionless response beamforming algorithms performed particularly well in all noise scenarios.

Keywords

binaural, signal enhancement, instrumental evaluation, speech intelligibility, audio quality

Introduction

Many conversations today take place in rather noisy environments. For normal-hearing (NH) listeners, this degraded speech does not pose a major challenge and is typically intelligible. Hearing aid (HA) or cochlear implant (CI) users, on the other hand, are much more impacted in their speech intelligibility by interfering noise sources (Festen & Plomp, 1990; Peters, Moore, & Baer, 1998; Qin & Oxenham, 2003; Stickney, Zeng, Litovsky, & Assmann, 2004).

Considerable effort has been made to develop and investigate single- as well as multichannel noise reduction algorithms for HAs and CIs (for comprehensive reviews, see e.g., Bentler, 2005; Doclo, Gannot, Moonen, & Spriet, 2010; Doclo, Kellermann, Makino, & Nordholm, 2015; Hamacher, Kornagel, Lotter, & Puder, 2008; Levitt, 2001; Wouters, Doclo, Koning, & Francart, 2013). Spatial filtering (typically referred to as beamforming) has become a standard in modern hearing devices. By enhancing signals originating from one direction (usually the front) and suppressing signals originating from other locations, these algorithms are able to achieve large improvements in signal-to-noise ratio (SNR). A wireless link between hearing devices on the left and right side is already available in commercial HAs. These binaural HAs also feature binaural noise reduction algorithms. With the prevalence of bilateral cochlear

[1]Medical Physics Group, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany
[2]Cluster of Excellence "Hearing4all," Oldenburg, Germany
[3]Speech Signal Processing Group, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany
[4]Signal Processing Group, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany
[5]HörTech gGmbH, Oldenburg, Germany

Corresponding author:
Regina M. Baumgärtel, Carl von Ossietzky Universität Oldenburg, Medizinische Physik and Cluster of Excellence "Hearing4all," Küpkersweg 74, D-26129 Oldenburg, Germany.
Email: Regina.Baumgaertel@uni-oldenburg.de

implantation increasing, the possibility of providing such algorithms to bilateral CI users is emerging. These multichannel, binaural algorithms use the microphone signals from both hearing devices and result in larger SNR improvements compared to monaural beamforming algorithms (Cornelis, Moonen, & Wouters, 2012; Van den Bogaert, Doclo, Wouters, & Moonen, 2009), providing improved speech intelligibility in noise. Algorithms operating on single channel signals, on the other hand, do not usually result in speech intelligibility improvements but have been shown offer improved signal quality and listening comfort (e.g., Luts et al., 2010).

The objective of signal enhancement strategies is to enhance two fundamental perceptual aspects of noisy speech signals: speech intelligibility and sound quality. However, these two objectives cannot always be achieved simultaneously. When assessing the merit of signal enhancement algorithms, both aspects should be taken into consideration, although improved speech intelligibility typically is considered to be more important. In general, there is a trade-off between noise reduction and speech distortion. An increase in speech intelligibility can, for example, be achieved at the cost of lower signal quality (e.g., due to distortions). Especially for algorithms operating on single channel signals, an improved signal quality does not necessarily entail improved speech intelligibility at the same time (Hu & Loizou, 2007a, 2007b). Multichannel noise reduction algorithms are often able to achieve both, increased speech intelligibility as well as increased signal quality.

Instrumental measures are commonly used to evaluate an algorithm's capabilities in speech intelligibility enhancement and quality improvement (e.g., Hendriks & Gerkmann, 2012). Perceptual speech intelligibility measurements in NH listeners (Fink, Furst, & Muchnik, 2012; Healy, Yoho, Wang, & Wang, 2013; Kim, Lu, Hu, & Loizou, 2009; Yousefian & Loizou, 2012) have also been used regularly to evaluate and characterize signal enhancement algorithms, often in combination with other measures. Yousefian and Loizou (2012), for example, supplemented their speech intelligibility evaluation with an instrumental evaluation of signal quality, while Healy et al. (2013) and Fink et al. (2012) additionally reported speech intelligibility improvements in hearing-impaired (HI) subjects. Large-scale evaluation studies in HI listeners (e.g., Cornelis et al., 2012; Luts et al., 2010) or CI users (e.g., Brockmeyer & Potts, 2011) have been geared toward comparing the value of different signal enhancement algorithms for the respective listener groups.

Although a large number of studies have evaluated signal enhancement schemes perceptually as well as with the help of instrumental measures, most studies focus on the evaluation of only a small number of signal processing schemes. Differences between studies in measurement design, speech and noise material, as well as subject groups in the case of perceptual evaluations, or choice of measures in the case of instrumental evaluations, limit the comparability across studies.

This article is the first in a series of three articles in this issue originating from a collaborative project of several research groups within the Cluster of Excellence "Hearing4all" in Oldenburg. The goal of this collaborative research project was to comprehensively evaluate state-of-the-art signal enhancement algorithms, with emphasis on binaural algorithms. We tested (a) different listening situations, (b) different instrumental measures, (c) subjects with a very different hearing status, and (d) a variety of different algorithms. These four aspects taken together provide an overview of the benefits obtainable by monaural and binaural signal enhancement algorithms. A coherent study design was maintained across all evaluations to ensure high comparability of the results. Several state-of-the-art noise reduction algorithms were selected, with a focus on binaural algorithms but also including two monaural algorithms as references. The selected algorithms consisted of established algorithmic building blocks, such as (fixed and adaptive) minimum variance distortionless response (MVDR) beamforming and spectral post-filtering, which were combined in innovative ways. All algorithms were implemented in real time on a common signal processing platform, namely the Master Hearing Aid (MHA; Grimm, Herzke, Berg, & Hohmann, 2006), making the setup ideal for perceptual listening evaluations.

Four different, synthetic but highly realistic scenarios were designed to reflect real-world listening situations. All scenarios included a significant amount of reverberation ($T_{60} \approx 1.25s$), further challenging the algorithms. The noise scenarios were created in a three-dimensional listening environment using head-related impulse responses (HRIRs; Kayser et al., 2009).

Starting from these common algorithms and test scenarios, which are described in detail in the following Methods section, we have branched out into specific studies reported in the three articles. The current article presents the common framework and the instrumental evaluation of speech intelligibility and quality. Three measures were employed: the speech intelligibility-weighted signal-to-noise ratio (iSNR; Greenberg, Peterson, & Zurek, 1993), the short-time objective intelligibility (STOI; Taal, Hendriks, Heusdens, & Jensen, 2011) measure, and the perceptual evaluation of speech quality (PESQ; ITU-T, 2001). The second article aims at evaluating the same signal enhancement strategies through perceptual evaluations in bilateral CI users (Baumgärtel et al., 2015). In a third article, perceptual evaluations in NH listeners and HI subjects, as well as an evaluation using a binaural speech intelligibility model, are presented (Völker, Warzybok, & Ernst, 2015).

Methods

Noise Reduction Algorithms

The signal enhancement strategies evaluated in this study were implemented on a common processing platform. All output files had a sampling rate of 16 kHz.

Adaptive differential microphone (ADM). The adaptive differential microphone algorithm was implemented according to the description in Elko and Anh-Tho Nguyen (1995). The two omnidirectional microphones present in each hearing device were combined adaptively so that the sound energy from the rear hemisphere is minimized in the output of the algorithm. This is achieved by steering a spatial zero to suppress sound originating from the loudest source in the rear hemisphere. The ADM algorithm first computed front-facing and back-facing differential microphones with a spatial zero pointing to 180° and 0°, respectively. These signals were then weighted and combined, with the weight parameter determining the direction of the spatial zero. The weight parameter was adapted using a gradient-descent procedure to ensure the above energy criterion. The combination of two closely spaced omnidirectional microphones resulted in a comb-filter effect present in the output signal of the ADM. Therefore, a low-pass filter was used to counter the effect of the first minimum of the comb filter. The ADMs worked on the left and right side independently and were included here as a second reference condition alongside the unprocessed signals.[1]

Coherence filter (COH). The coherence-based noise reduction algorithm (Grimm, Hohmann, & Kollmeier, 2009; Luts et al., 2010) computes a spectral gain based on the concept of coherence to separate the desired speech signal from undesired noisy components. Coherent signal components were assumed to belong to the desired target signal, for example, a single speaker talking to the listener. Incoherent signal components are assumed to belong to the undesired noisy part.

The processing algorithm works in the short-time Fourier transform (STFT) domain, where STFT bins were grouped into 15 non-overlapping third-octave frequency bands with center frequencies ranging from 250 Hz to 8 kHz. The interaural phase difference (IPD) was used as an estimate for the coherence. The coherence $C(k,\ l)$ in each frequency band k and time segment l is estimated from the vector strength of the complex IPD $c_{\mathrm{IPD}}(k,l)$, as defined in Grimm et al. (2009):

$$C(k,l) = \left| \left\langle c_{\mathrm{IPD}}(k,l) \right\rangle_{\tau(k)} \right|. \tag{1}$$

The coherence value was estimated using a running average $\langle \cdot \rangle_{\tau(k)}$ with time constant $\tau(k)$. Since, for short-time constants, the estimate $C(k,\ l)$ may be larger than the actual coherence, a linear mapping of the coherence was introduced. The coherence interval $[C_1, C_2]$ was mapped linearly to the interval [0,1]:

$$\hat{C}(k,l) = \begin{cases} \frac{C(k,l)-C_1}{C_2-C_1} & C_1 < C(k,l) < C_2 \\ 0 & C(k,l) \leqslant C_1 \\ 1 & C(k,l) \geqslant C_2. \end{cases} \tag{2}$$

An identical mapping interval was used for all frequency bands. The gain in each frequency band was then computed by applying an efficiency exponent $\alpha(k)$, i.e.:

$$G(k,l) = \hat{C}(k,l)^{\alpha(k)}. \tag{3}$$

By applying the same gain to both channels, the binaural cues were preserved.

The two main parameters for the algorithm are the time constant $\tau(k)$ and the efficiency exponent $\alpha(k)$. Both frequency-dependent parameters were optimized manually. The efficiency exponent $\alpha(k)$ roughly followed the band importance function for the calculation of the speech intelligibility index (SII; American National Standard Institute, 1997). The values for the time constant $\tau(k)$ were approximated by $\frac{1}{f_k} \cdot 100$, where f_k denotes the center frequency of the k^{th} frequency band Hz. In this study, the coherence-based noise reduction algorithm was used in serial processing after the ADM algorithm, that is, the ADM supplied a binaural input signal for the coherence-based noise reduction algorithm (see Figure 1).

Single-channel noise reduction (SCNR). In this processing scheme, the frontal microphone signals of the left and the right hearing device were enhanced separately using an STFT-based single-channel noise reduction setup as outlined in Figure 2. For the STFT, we used a segment length of 32 ms with 50% overlap, and a square root Hanning window for analysis and overlap-add synthesis.

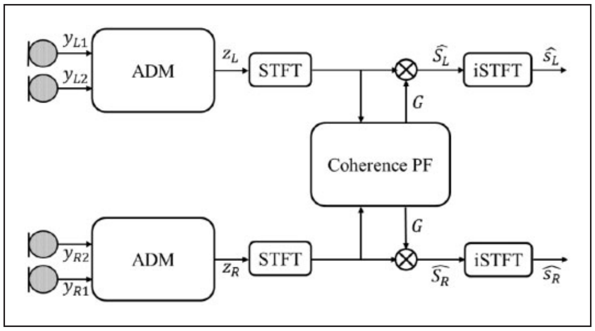

Figure 1. Block diagram illustrating the coherence filter setup.

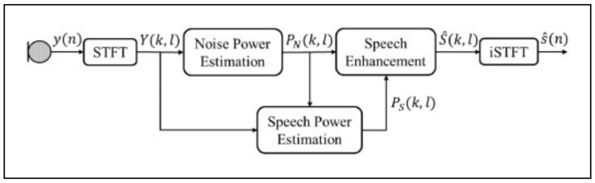

Figure 2. Block diagram illustrating the single-channel noise reduction setup.

In the STFT domain, the noise power spectral density $P_N(k,l)$ was estimated from the noisy input STFT $Y(k, l)$, using the speech presence probability-based estimator (Gerkmann & Hendriks, 2012). An estimate of the speech power $P_S(k,l)$ was obtained by temporal cepstrum smoothing as proposed in Gerkmann, Breithaupt, and Martin (2008). In the next step, the speech power $P_S(k,l)$ and the noise power $P_N(k,l)$ were used to estimate the clean speech spectral amplitude $|\hat{S}(k,l)|$ according to Breithaupt, Krawczyk, and Martin (2008), which is parameterized with a compression parameter β and a form parameter μ. As in Breithaupt et al. (2008), here we used $\mu = \beta = 0.5$, that is, the so-called super-Gaussian amplitude root (SuGAR) estimator. While $\mu = 0.5$ modeled the clean speech STFT coefficients as being complex super-Gaussian distributed, $\beta = 0.5$ corresponded to minimizing the mean square error between the square roots of the true and the estimated amplitudes. This choice has been reported to yield a good noise reduction performance with only little audible speech distortions (Breithaupt et al., 2008). The clean speech spectral amplitude is estimated by multiplying the input STFT amplitude with a real-valued gain function, i.e., $|\hat{S}(k,l)| = G(k,l)|Y(k,l)|$. To minimize speech distortions, we applied a lower limit of -9 dB to the gain function $G(k, l)$. Finally, the estimated spectral amplitude was combined with the noisy spectral phase of the input signal, i.e., $\hat{S}(k,l) = |\hat{S}(k,l)| \exp(\mathrm{i}\angle Y(k,l))$ and the enhanced time domain signal $\hat{s}(n)$, with time index n, was synthesized via overlap-add, which is denoted as iSTFT in Figure 2. The employed monaural enhancement scheme is used due to its generality. With more knowledge about the specific acoustic scenario, such as the noise type, alternative methods, for example, based on supervised-learning techniques (Kim et al., 2009), might lead to further improvements at the cost of a loss in generality.

Fixed MVDR beamformer (fixed MVDR). The binaural MVDR beamformer aimed at minimizing the overall noise output power, subject to the constraint of preserving the desired speech component in the frontal microphone signals of the left and the right hearing device. The frequency-domain binaural MVDR filters for the left and the right hearing devices $\mathbf{W}_L(k)$ and $\mathbf{W}_R(k)$ were equal to Van Veen and Buckley (1988):

$$\mathbf{W}_L(k) = \frac{\mathbf{\Gamma}^{-1}(k)\mathbf{A}(k)}{\mathbf{A}^H(k)\mathbf{\Gamma}^{-1}(k)\mathbf{A}(k)} A_L^*(k), \tag{4}$$

$$\mathbf{W}_R(k) = \frac{\mathbf{\Gamma}^{-1}(k)\mathbf{A}(k)}{\mathbf{A}^H(k)\mathbf{\Gamma}^{-1}(k)\mathbf{A}(k)} A_R^*(k), \tag{5}$$

where $\mathbf{\Gamma}(k)$ denotes the spatial coherence matrix of the noise field (assumed to be diffuse), $\mathbf{A}(k)$ denotes the anechoic head-related transfer function (HRTF) vector between the speech source and the microphones of the left and the right hearing device, and $A_L(k)$ and $A_R(k)$ denote the anechoic HRTFs of the frontal microphones in the left and the right hearing device, respectively. A detailed description of the beamforming scheme employed here can be found in Doclo et al. (2015). Assuming the speech source to be fixed in front of the listener, the filters $\mathbf{W}_L(k)$ and $\mathbf{W}_R(k)$ can be precalculated. The output signal at the left hearing device $z_L(n)$ was obtained by filtering and summing all microphone signals using the time-domain representation of the filter $\mathbf{W}_L(k)$. The output signal at the right hearing device $z_R(n)$ was obtained similarly.

Adaptive MVDR beamformer (adapt MVDR). Since in practice the noise field is generally not known and changes over time, fixed beamformers such as the described binaural MVDR are only able to achieve a limited amount of noise reduction. To adapt to changing noise environments, the noise coherence matrix $\mathbf{\Gamma}(k)$ needs to be updated, or alternatively, the generalized sidelobe canceler (GSC; Gannot, Burshtein, & Weinstein, 2001; Griffiths & Jim, 1982) structure has been proposed, consisting of a fixed beamformer, a blocking matrix, and an adaptive filtering stage, as depicted in Figure 3. The fixed beamformer generated a speech reference signal, the blocking matrix generated so-called noise reference signals by steering spatial zeros in the direction of the speech source, and the adaptive filtering stage used a multichannel adaptive filter aiming to remove the remaining correlation between the residual noise component in the speech reference signal and the noise reference. For the fixed beamformer, the binaural MVDR beamformer was used. The spatial zero toward the speech source (assumed to be in front of the listener) in the blocking matrix was realized by subtracting the microphone signals of the right hearing device from the microphone signals of the left hearing device, such that two noise reference signals, one for each side, are available at the input of the adaptive filter. The adaptive filtering stage was realized using a frequency-domain normalized least mean squares (NLMS) algorithm according to Shynk (1992).

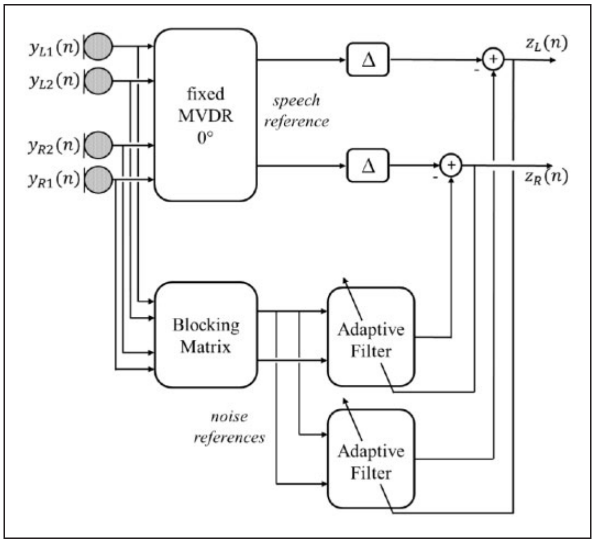

Figure 3. Block diagram illustrating the adaptive binaural MVDR beamformer setup.

Combination of beamformer and postfiltering. Both the fixed and the adaptive beamformers only consider spatial characteristics of the microphone signals. To additionally exploit spectro-temporal characteristics, we also considered the combination of the beamformers described earlier with a postfiltering based on single-channel speech enhancement, as presented above. The basic block diagram, encompassing all combinations considered in this article, is illustrated in Figure 4. First, a binaural MVDR beamformer was applied as presented in the preceding paragraphs. The binaural output signals of the beamformer were then transformed into the STFT domain, followed by the same SCNR that has been outlined earlier. Based on the signals at the output of the SCNR processing, gain functions for the left and for the right ear were computed, i.e., $G_{\mathrm{L}}(k,l)$ and $G_{\mathrm{R}}(k,l)$, and applied to the left and right frontal microphone signals. Finally, the enhanced signals were synthesized via overlap-add. For more details on spectral post-processing for binaural speech enhancement, see Lotter (2004), Rohdenburg (2008), Simmer, Bitzer, and Marro (2001), and Gannot and Cohen (2007). The three combinations under investigation differed only in the choice of the beamformer and the postfiltering scheme, either using a common postfiltering, i.e., $G_{\mathrm{L}}(k,l) = G_{\mathrm{R}}(k,l)$, or an individual postfiltering, i.e., $G_{\mathrm{L}}(k,l) \neq G_{\mathrm{R}}(k,l)$.

Common postfilter based on fixed binaural MVDR beamformer (com Pf (Fixed MVDR)). In this setup, the fixed MVDR beamformer was combined with a common postfilter, defined as:

$$G_{\mathrm{L}}(k,l) = G_{\mathrm{R}}(k,l) = \sqrt{\frac{|\tilde{Z}_{\mathrm{L}}(k,l)|^2 + |\tilde{Z}_{\mathrm{R}}(k,l)|^2}{|Y_{\mathrm{L1}}(k,l)|^2 + |Y_{\mathrm{R1}}(k,l)|^2}}. \qquad (6)$$

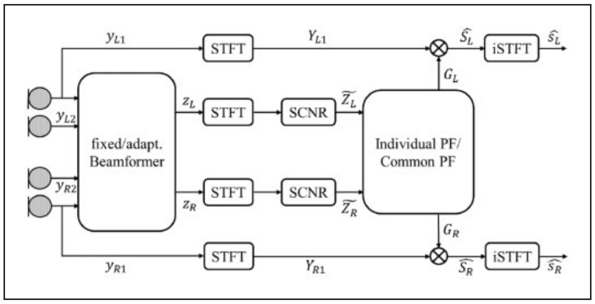

Figure 4. Block diagram illustrating the combination of a binaural MVDR beamformer with single-channel postfiltering. The indices *n*, *k*, and *l* were omitted for the sake of clarity.

By using the same real-valued gain on both signals, the interaural level differences (ILDs) and interaural time differences (ITDs) of both the speech and noise components were maintained, as the signals at both ears were scaled by the same factor. This was not necessarily the case for the output of the binaural MVDR beamformer without postfilter.

Common postfilter based on adaptive MVDR beamformer (com PF (adapt MVDR)). In this setup, the adaptive MVDR beamformer was combined with the same common postfilter as introduced above.

Individual postfilter based on adaptive MVDR beamformer (ind PF (adapt MVDR)). In this setup, the adaptive binaural MVDR beamformer is combined with a different postfilter, which works individually on the left and the right hearing device. The gain functions were defined as:

$$G_{\mathrm{L}}(k,l) = \frac{|\tilde{Z}_{\mathrm{L}}(k,l)|}{|Y_{\mathrm{L1}}(k,l)|}, \qquad (7)$$

$$G_{\mathrm{R}}(k,l) = \frac{|\tilde{Z}_{\mathrm{R}}(k,l)|}{|Y_{\mathrm{R1}}(k,l)|}. \qquad (8)$$

On the one hand, since the SCNR scheme itself is minimum mean-square error (MMSE) optimal (Breithaupt et al., 2008), using individual postfilters potentially achieved an increased SNR improvement compared with using the common postfilter described earlier. On the other hand, the input ILDs were not maintained anymore.

Real-time implementation on the MHA platform. The MHA (Grimm et al., 2006) is a real-time signal processing platform designed for implementation and evaluation of hearing device algorithms. It runs on multiple operating systems and processor architectures. The MHA framework as well as the existing algorithms were implemented in C++. Using the MHA configuration language, the implemented algorithms could be easily configured by

Table 1. List of Signal Enhancement Strategies.

No.	Abbreviation	Algorithm
1	NoPre	no pre-processing
2	ADM*	adaptive differential microphones
3	ADM + coh	adaptive differential micro-phones in combination with coherence filter
4	SCNR*	single-channel noise reduction
5	fixed MVDR	fixed binaural MVDR beamformer
6	adapt MVDR	adaptive binaural MVDR beamformer
7	com PF (fixed MVDR)	common postfilter based on fixed binaural MVDR beamformer
8	com PF (adapt MVDR)	common postfilter based on adaptive binaural MVDR beamformer
9	ind PF (adapt MVDR)	individual postfilter based on adaptive binaural MVDR beamformer

Note. Two algorithms marked with asterisks are established monaural strategies, which were included as reference (ADM) and because they have been used as processing blocks in some of the binaural algorithms (ADM and SCNR). MVDR = minimum variance distortionless response.

setting their parameters and could be combined with each other. Once a configuration had been loaded, all corresponding algorithms were loaded with their current settings as plug-ins at runtime into the MHA. The MHA supports re-configuration of algorithms at runtime. For this, a network connection can be established using network tools (e.g., telnet) as well as using MATLAB tools, which are part of the MHA distribution. All algorithms presented here (see Table 1 for overview) were implemented on the MHA platform. Although the save-to-file function was used for the instrumental evaluations of algorithm performance, all algorithms ran in real time, making the system an ideal platform for subjective listening tests. Such tests have been conducted with bilaterally implanted CI users, HI, and NH listeners. The results from these evaluations are presented in the two accompanying studies (Baumgärtel et al., 2015; Völker et al., 2015).

Speech and Noise Material

All scenarios were created in a highly reverberant, cafeteria-style room (see Figure 5 and Kayser et al., 2009). The reverberation time of this cafeteria of $T_{60} \approx 1.25s$ is larger than one would expect in typical conversation environments, yet listeners will at times be faced with environments exhibiting such long reverberation. The scenarios created here can, in terms of reverberation time, be understood as worst-case scenarios. With the exception of the cafeteria ambient noise (see below for details), all scenarios were created by convolving target speech and background noise signals with HRIRs recorded using behind-the-ear (BTE) HA shells on a dummy head in a reverberant cafeteria (Kayser et al., 2009). In the work described here, only front and rear BTE microphone channels were used, mimicking two-microphone HA or CI devices. The distance between the two microphones on each side was approximately 1.6 cm.

Speech material. The Oldenburg sentence test (OLSA) (Wagener, Brand, & Kollmeier, 1999) was used as speech material. The OLSA speech material shows a phoneme distribution that is equivalent to the mean phoneme distribution of the German language and is spoken at medium speed. Dry recordings of the OLSA sentences were convolved with HRIRs as described above in order to create the four-channel target input signals. The target speech source was located at 0° (front) at a distance of 102 cm in all test conditions (Position A in Figure 5). A total of 120 sentences were used in this instrumental evaluation.

Noise material. Instrumental evaluations of all algorithms were performed in four distinct acoustic scenarios described in detail below. To create interfering speech signals, speech material from the German few-talker corpus of the EUROM1 speech corpus (Chan et al., 1995) was used, where we chose only the five male talkers. To create the speech signals, 35 randomly selected passages by one talker were concatenated and the resulting signal was then cropped to ten minutes length. The four scenarios were as follows:

1. *olnoise (OLN)*. To create a stationary yet spatial noise scenario, the speech-shaped noise file provided with the OLSA sentence material (olnoise) was used. A 10-minute long version of the noise file was created. Five different (uncorrelated) sections of this noise were chosen by delaying the starting point of the signal by 0, 2, 4, 6, and 8 seconds. Each noise signal was then assigned to one of five locations (positions B, C, D, E, and F, see Figure 5) in the cafeteria environment to create incoherent stationary background noise.
2. *20-talker babble (20T)*. A multitalker babble noise was created by placing 20 talkers at five different

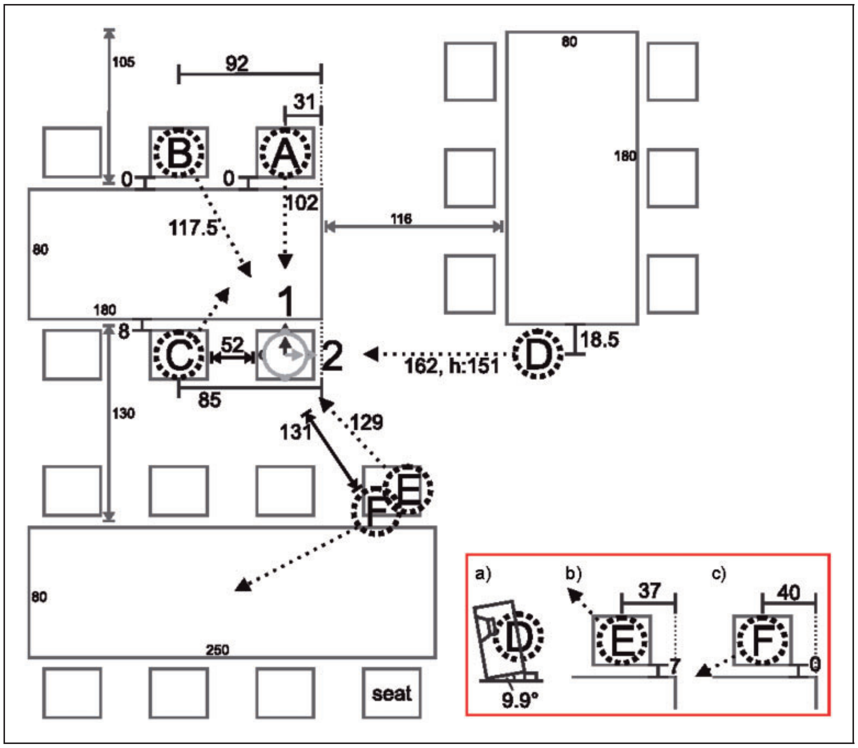

Figure 5. Layout of the cafeteria-type room used to create the target speech and noise signals. Position and head orientation 1 was chosen for the listener. Target speech originated from Position A, interfering talkers were located at either Position D or Positions B, C, D, E, and F. The inset (marked by red box) shows the detailed position and orientation of speakers located at Position D (a), Position E (b), and Position F (c).

locations (four talkers at each location) around the listener. Speech signals were created using the speech material taken from the EUROM1 corpus. Each of the five talkers was used four times (at four different locations). Therefore, for each of the five talkers, four different 10-minute signals were created as described above. The talkers were located to the left and right of the listener, as well as front left and back right (positions B, C, D, E, and F, see Figure 5).

3. *Single competing talker (SCT)*. A single, male interfering talker was placed at +90° (right-hand side, Position D in Figure 5) of the subject at a distance of 162 cm and an elevation of 40 cm above ear level (tilted to be pointed directly at the ear). The speech material was taken from the EUROM1 speech corpus as described above.
4. *Cafeteria ambient noise (CAN)*. As the most realistic scenario, the CAN signal was recorded alongside the HRIRs in the same cafeteria-type room (Kayser et al., 2009) during lunch hour. The noise signal includes periods of two-person conversations being carried out next to the recording dummy, periods of more diffuse talking in the background as well as typical cafeteria sounds such as dishes and cutlery being used and chairs being pushed across the floor.

Signal generation. Clean speech and noise signals were mixed at a broadband, long-term SNRs between -10 dB and $+10$ dB. The SNR was determined from the reverberant speech and noise signals, averaged across left and right ears, front and back microphones. Three seconds of noise-only in the beginning of each signal were provided to allow enough time for all algorithms to converge before target speech onset. For each test scenario, each of the 120 OLSA sentences used in this evaluation was mixed with one noise segment randomly cut from the longer, original noise files. The same noise segment was used for all SNRs. Signals were then processed by the signal pre-processing strategies using the MHA platform. The processed speech and noise signals were computed following the protocol introduced by Hagerman and Olofsson (2004). In short, two different signals were produced and processed by the algorithms: Speech mixed with the original noise signal ($S+N$) and speech mixed with a phase-inverted version of the noise signal ($S-N$). Under the assumption that both signals are processed equally by the algorithms, the processed speech and noise signals were calculated as follows:

$$S_{proc} = \frac{1}{2} \cdot ((S+N)_{proc} + (S-N)_{proc}), \tag{9}$$

and

$$N_{proc} = \frac{1}{2} \cdot ((S+N)_{proc} - (S-N)_{proc}). \tag{10}$$

Subsequently, all signals were time-aligned to compensate for different processing delays introduced by the algorithms. In this step, the 3 seconds of noise added at the beginning of each signal were also eliminated.

Reference signals. To compute the STOI and PESQ measures (see later for measure descriptions), a clean speech reference signal is required. All algorithms aim at estimating the anechoic speech component at the BTE microphones; therefore, clean speech convolved with anechoic HRIRs (Kayser et al., 2009) rather than dry clean speech was used as a reference.

Instrumental Measures

Instrumental evaluations of the considered algorithms were performed using instrumental measures of speech intelligibility as well as speech quality. Here the STOI measure as well as the iSNR were used as the instrumental speech intelligibility measures, while PESQ was used to evaluate speech quality.

Intelligibility-weighted SNR. The iSNR (compare Greenberg et al., 1993) calculates the long-term SNR in 18 frequency bands and weighs the obtained SNRs with the band-importance function according to the SII standard (American National Standard Institute, 1997) to obtain an overall iSNR measure. Since this measure does not require a reference signal, it was computed based on the processed speech (S_{proc}) and noise (N_{proc}) signals obtained from equations (9) and (10).

Short-time objective intelligibility (STOI) measure. The STOI measure (Taal et al., 2011) determines the correlation between time-frequency segments of a clean speech reference signal and a (noisy) processed speech test signal. Both signals are divided into 25.6-ms, Hanning-windowed segments with 50% overlap. After decomposition into 15 third-octave bands with center frequencies ranging from 150 Hz to 4.3 kHz, the correlation between the clean speech reference signal and the processed signal is determined for temporal envelope segments of 384 ms length. Before calculation of the correlation coefficient, the processed signal is normalized to compensate for global level differences. Additionally, the signal is clipped, resulting in an upper bound for the sensitivity of the measure toward severely degraded time-frequency units (Taal et al., 2011). The obtained intermediate intelligibility measures are averaged across all time frames and all frequency bands to obtain one value, the STOI score. STOI scores are mapped to an absolute intelligibility prediction (Taal et al., 2011) where a score of 1 corresponds to 100% speech intelligibility. For NH listeners, the measure shows a high correlation with subjective speech intelligibility in different noise types for speech processed with different noise reduction schemes. In Hu et al. (2012), it has also been shown that STOI was able to predict speech intelligibility for noise-vocoded speech.

Perceptual evaluation of speech quality (PESQ). The PESQ measure is more complex than the other two measures used. It was developed and introduced by Rix, Beerends, Hollier, and Hekstra (2001) and is recommended by ITU-T for speech quality assessment of telephone networks (ITU-T, 2001). PESQ compares a clean speech reference signal with a processed speech signal by means of a perceptual model. PESQ was found to be in good agreement with subjective quality measures for NH listeners (Hu & Loizou, 2008). In short, the test and reference signals are time- and level-aligned and filtered to model a standard telephone handset. Subsequently, both the reference and the test signal are passed through an auditory transform. Two parameters are calculated from differences between the two transformed signals and are aggregated in time and frequency. These differences are then mapped to a mean opinion score (MOS), covering a range from 0.5 (highly degraded test signal) to 4.5 (no difference between reference and test signal). PESQ results will be reported here in terms of MOS. For reference, a decrease in SNR from 0 dB to -5 dB in the unprocessed signal results in a reduction in MOS of 0.3. The choice of an anechoic, clean-speech reference file resulted in the evaluation of dereverberation and SNR improvements as quality improvements.

Results and Discussion

In this section, we compare the performance of the considered noise reduction schemes by means of three different instrumental measures. The same algorithms have been evaluated in the same noise conditions with bilaterally implanted CI subjects by Baumgärtel et al. (2015) and in NH and HI subjects by Völker et al. (2015). Absolute values obtained from the instrumental evaluation at an input SNR of 0 dB are presented in Figure 6.

Additionally, for each measure, the improvements provided by each algorithm in each scenario were determined as

$$\Delta = \max(Score_{Algo,L}, Score_{Algo,R}) - \max(Score_{Ref,L}, Score_{Ref,R}), \tag{11}$$

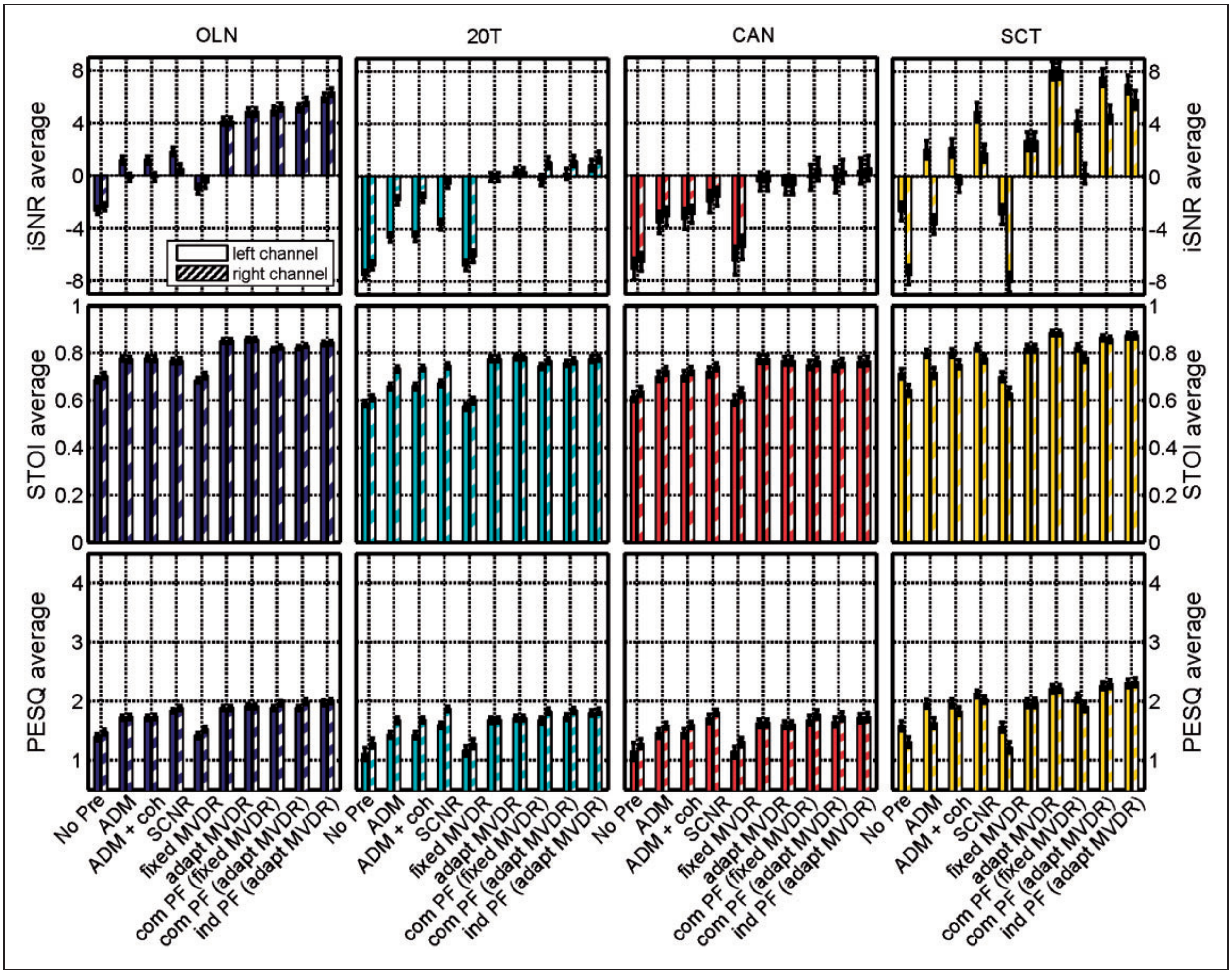

Figure 6. Instrumental evaluation results at 0 dB input SNR. Top panels show the results for the iSNR measure, middle panels show the results for the STOI measure, and bottom panels show the results for the PESQ measure. Columns from left to right show results for OLN (Navy), 20T (Turquoise), CAN (Red), and SCT (Yellow) noise scenarios. Left channel results are indicated by bar graphs with solid filling, and right channel results by bar graphs with hashed filling. Error bars denote the standard deviation.

where either the unprocessed condition (NoPre) or the signals processed with ADMs were chosen as the reference condition. We refer to these improvements as better-channel-improvements, and they are plotted for an input SNR of 0 dB in Figure 7(a) with respect to the unprocessed signal and Figure 7(b) with respect to the ADM processed signals. In Figure 8, better-channel-improvements with respect to the unprocessed reference condition are depicted for each algorithm as a function of the input SNR.

All results presented here are averaged across 120 sentences and, consequently, across 120 different noise segments for each test scenario. The error bars in Figures 6 to 8 (standard deviation) therefore provide an estimate of the variation in algorithmic performance for each algorithm in each test scenario. In the scenarios tested here, the fluctuations are rather small, suggesting all algorithms work robustly in each of the tested scenarios. The fluctuations in the highly nonstationary CAN and SCT scenarios are larger than fluctuations in the more stationary OLN and 20T babble scenarios as can be expected. For PESQ, the variation decreases with increasing input SNR. The same is true for STOI, albeit to a lesser extent. The variation seems to be caused almost exclusively by the noise characteristics; the standard deviations for all algorithms within one noise scenario are very similar.

The results from the three instrumental measures differed slightly, as each measure sheds light on certain signal characteristics (see Instrumental Measures section for details). It should be noted that the absolute mean opinion scores obtained from the PESQ evaluation are comparatively low. The full scale of the PESQ scores ranges from 0.5 to 4.5, whereas the results here only

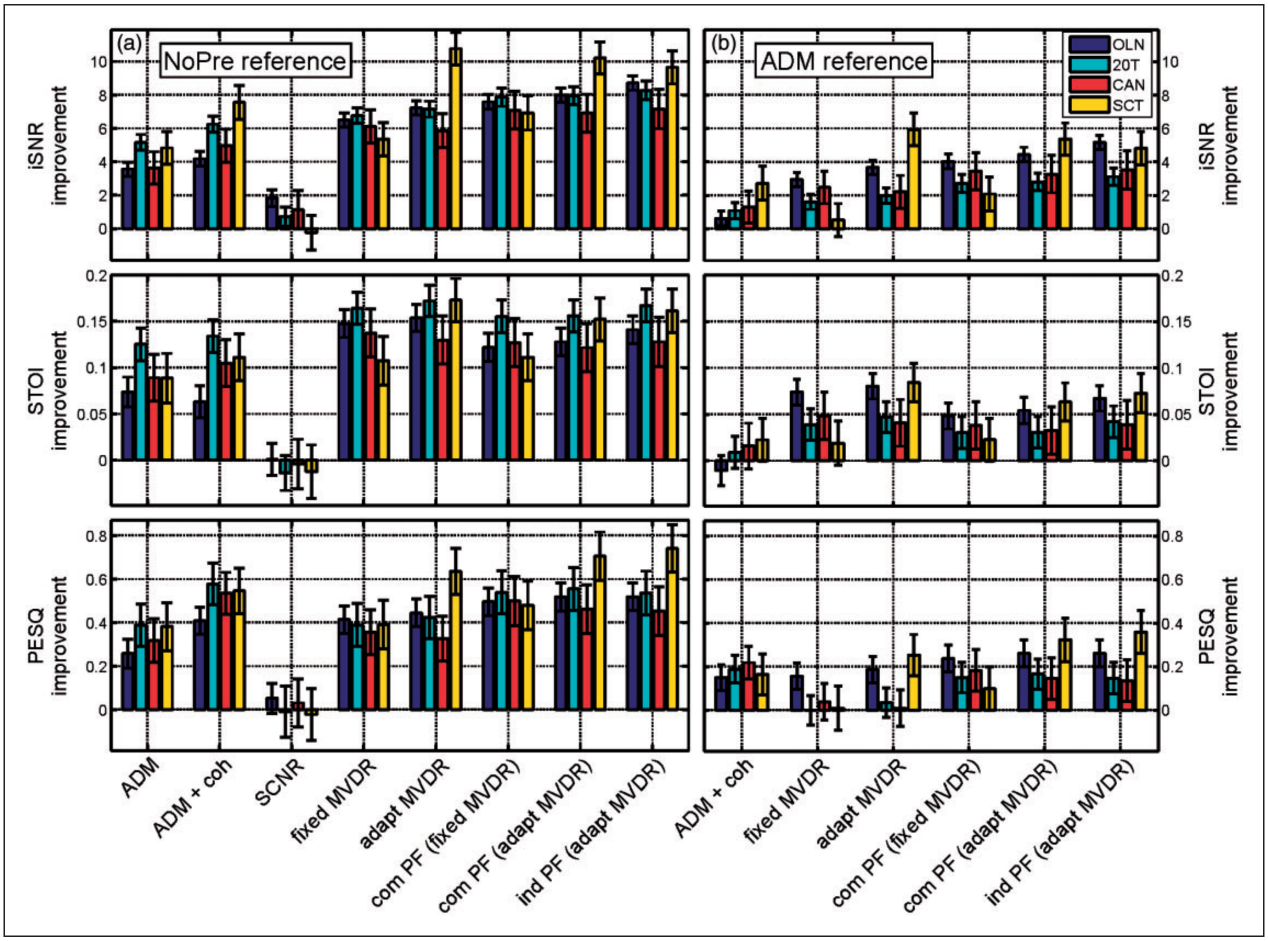

Figure 7. Better-channel-improvements obtained for each instrumental measure score at an input SNR of 0 dB. (a) The better channel for each algorithm condition is compared with the better channel in the corresponding no pre-processing condition. (b) The better channel for each algorithm condition is compared with the better channel in the corresponding ADM processed condition. Color codes are used for the test scenario, error bars denote the standard deviation.

covered a range up to 2.3. The PESQ measure was originally developed to evaluate telephone transmission by comparing a clean speech signal to a transmitted (and presumably degraded in quality), yet still clean signal. Here, however, we have used PESQ as an instrumental quality measure by comparing a noisy speech signal (output of the signal enhancement algorithms) with a clean reference speech signal (compare Reference Signals section). Residual noise, not accounted for in the original model, was therefore treated as a quality impairment.

For the unprocessed reference condition (NoPre) in the SCT scenario, a large difference between the left and right channels was found (Figure 6, rightmost column), with the left channel showing better values in all measures. This finding was expected considering the highly asymmetric setup of this noise scenario: One competing talker was located to the right of the listener, while no noise sources are present to the listener's left. In the other three conditions, only small differences were observed between the left and right channels, with the right channel being evaluated as slightly better than the left. This difference could again be attributed to a slight asymmetry in the measurement scenario setup (see Figure 5 for geometric layout of the measurement environment). In the OLN and 20T scenarios, noise sources were located at positions B–F. The left side sources were located closer to the hearing devices and therefore produced higher noise power than the sources located at the right side. Additionally, the listener was seated in close proximity to a wall on the left side resulting in left-biased reflections, while the listener's right side faced an open room.

Signal processing with the ADM algorithm enhanced the differences between left and right channels, especially in the 20T condition. This is due to the ADM acting on the right microphone channels being able to steer a spatial zero toward two interfering noise sources located toward the back (positions E and F, see Figure 5), resulting in high noise suppression, while the ADM acting on

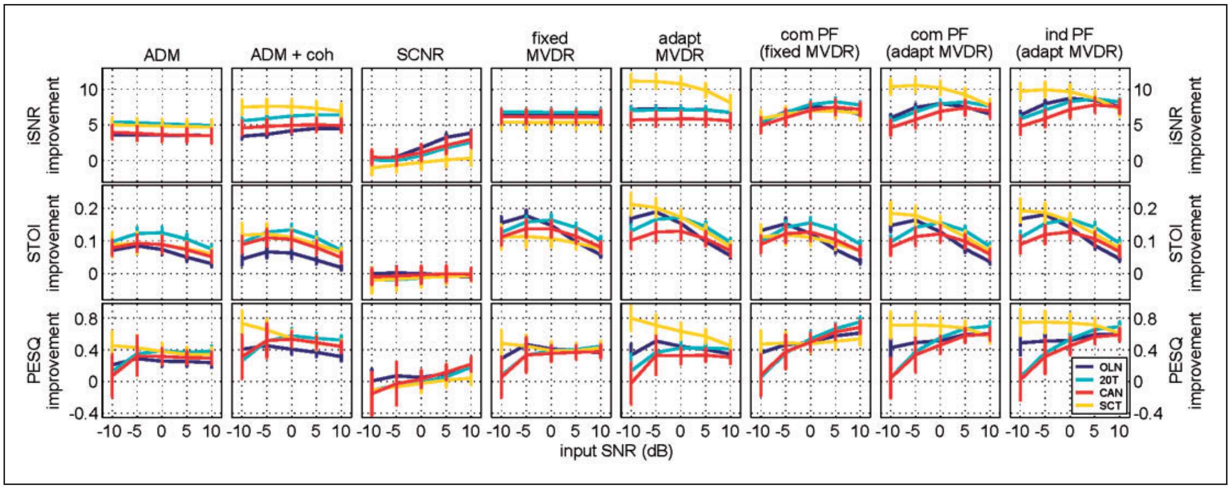

Figure 8. Better-channel-improvements for all tested algorithms are plotted at varying input SNRs between −10 dB and +10 dB. Top panels show results for iSNR, middle panels STOI, and bottom panels PESQ results. Error bars denote the standard deviation. Color codes are used for the different test scenarios.

the left signal channels were more influenced by the interfering noise source at a rather frontal position (B), which cannot sufficiently be suppressed due to the close proximity to the target direction (A). Better-channel-improvements obtained by processing with ADMs were seen with all measures in all scenarios. Considering that the target and interfering sound source are spatially separated in the SCT scenario, the ADMs were expected to yield the largest improvements in this scenario. None of the measures, however, matched this expectation. As a function of the input SNR, the iSNR measure shows a minimal decrease in better-channel-improvement with increasing input SNR. The reason for this behavior is that at low input SNRs, the noise sources are more prominent than the target source which allows for an efficient adaptation of the algorithm to the interfering sound. Both the STOI and PESQ measures predict the best performance at input SNRs of −5 dB or 0 dB. While the decrease in performance at higher input SNRs can be attributed to the loss in algorithm efficiency as discussed above, the decrease at lower input SNRs is likely influenced by distortions introduced by the processing algorithm which are evaluated negatively by the STOI and PESQ measures, but not by the iSNR measure.

The combination of ADMs with coherence-based postfiltering (ADM + coh) resulted in further increases in better-channel-improvements for all measures in all conditions (see Figure 7(b)), with the only exception being STOI which showed a slight decrease in performance due to the addition of the coherence-based postfilter in the stationary OLN scenario. In both, the iSNR and STOI results, the same trend is apparent: ADM + coh provides larger benefits with increasing nonstationarity of the interfering noise. We speculate that this is due to the temporal variation of the interaural coherence decreasing with increasing stationarity of the interfering noise. Since the coherence-based postfilter derives the gain from the interaural coherence, it provides less benefit with decreasing temporal variation of the interaural coherence. Unlike the iSNR measure, the STOI measure takes into account signal distortions to a certain extent. In all test scenarios but OLN, the improved SNR (as apparent from the iSNR scores) outweighs the negative impact of the distortions. In the stationary OLN, however, the reduced STOI score for ADM + coh with respect to the ADM reference is likely due to distortions introduced by the processing that could not be offset by SNR improvements. As a function of input SNR, each measure will be discussed individually. For the iSNR measure and nonstationary noises (SCT and CAN), a similar behavior as ADM alone was found but with overall larger benefits. In the stationary noises (OLN and 20T), an increase in benefit with increasing input SNR can be seen. The STOI measure shows similar behavior to the ADM algorithm alone except for two findings: in the OLN scenario, overall benefits are smaller and, in the SCT scenario, the benefits at low input SNRs are larger. Overall larger benefits than ADM alone were seen in the PESQ measure, especially in the SCT scenario at low input SNRs.

The SCNR algorithm yielded the smallest improvements (sometimes degradations) in all scenarios, using all measures. Since all other algorithms evaluated in this study are multichannel processing schemes, this finding was expected. Multichannel algorithms are well known to provide larger benefits in both speech intelligibility and signal quality than single-channel algorithms. For all measures, the best performance of the SCNR algorithm was observed, as expected, in the stationary OLN condition. The worst performance was seen in either the 20T or SCT scenarios. The SCNR scheme employed here relied on speech and noise power estimates based on the speech presence probability. For the stationary OLN, these estimates and therefore the separation of a noisy signal into speech and noise components worked quite well, giving rise to the observed improvement in all measures. For nonstationary noise scenarios, where the noise contains more speech-like signal parts (SCT being the extreme case), estimation errors occur and consequently no improvements were found. It is notable, however, that even in the extreme case of speech-on-speech masking (SCT scenario), where the SCNR scheme was expected to fail in its ability to correctly estimate speech and noise powers, only rather small degradations were observed. With respect to the iSNR and PESQ measures, the algorithm performance increased with the input signal's SNR for all interfering noise conditions. This behavior was anticipated as lower interfering noise power reduces errors in the speech probability estimates. The STOI measure, however, predicts no change in performance with input SNR or even a small decrease.

Two versions of the binaural MVDR beamformer were tested here, a fixed MVDR and an adaptive MVDR beamformer. In the OLN, 20T and CAN scenarios both beamformers performed similarly, whereas in SCT scenario, the performance of the adaptive MVDR beamformer algorithms was substantially better than the fixed MVDR beamformer. While both beamformers were designed to enhance signals originating from directly in front of the listener, the adaptive beamforming algorithm had the additional ability to selectively suppress an interfering noise source originating from a different direction. This additional noise suppression did not yield much advantage in environments containing many noise sources located at a number of locations; however, in the SCT environment, the suppression of the single noise source in combination with the enhancement of the target speech source resulted in a much more favorable SNR than the target source enhancement alone (fixed MVDR).

For the fixed binaural MVDR, the best performance determined by each of the measures was seen in one of the diffuse-like noise scenarios (iSNR: 20T, STOI: 20T, and PESQ: OLN). For the iSNR and the STOI measure, the lowest performance of the fixed MVDR was seen in the highly directional SCT scenario. Since this binaural beamforming algorithm utilizes the assumption of a diffuse noise field in the calculation of the filters $\mathbf{W}_L(k)$ and $\mathbf{W}_R(k)$, this trend was anticipated. Compared with the ADM baseline, the fixed binaural MVDR showed the largest improvements in the stationary OLN scenario. The iSNR and STOI measure revealed smaller, yet noticeable improvements also for the CAN, 20T, and SCT scenarios. The PESQ measure shows further improvements only for the CAN condition and no difference to ADM for the remaining two (20T and SCT). The similarities and differences between ADM and fixed MVDR are also apparent when comparing the algorithms' performance across input SNRs for the iSNR and PESQ measures. The STOI measure, however, while showing similar trends for three of the noise scenarios (20T, CAN, and SCT), shows a larger improvement for the fixed binaural MVDR at negative input SNRs. For the fixed binaural MVDR beamformer, the addition of a common postfilter resulted in improved PESQ and iSNR scores in all scenarios. The STOI measure, however, showed decreased performance in all scenarios, except for the SCT scenario. As a function of input SNR, the behavior of the common postfilter based on the fixed MVDR beamformer is similar to the sum of the fixed binaural MVDR alone and the SCNR algorithm.

The adaptive binaural MVDR beamformer yielded the largest improvements overall for all measures in the SCT scenario. Since this scenario consists of spatially separated target and interfering sources, it is an ideal match for the adaptive MVDR algorithm. For the adaptive MVDR without postfilters, all measures revealed the largest better-channel-improvements in the SCT scenario. When regarding the iSNR and STOI measures, the adaptive MVDR in the SCT scenario also yielded the largest improvements across all algorithms and noise scenarios. For iSNR, both combinations of the adaptive MVDR with postfilters yielded the second- and third-largest overall better-channel-improvements. For STOI, however, these two algorithms achieved better results in the 20T condition. PESQ showed the highest improvements for each of the adaptive MVDR algorithms in the SCT scenario. For this measure, the overall (across all scenarios and all algorithms) best performance was achieved by the adaptive MVDR in combination with the individual postfilter in the SCT scenario. Compared to the ADM, the adaptive binaural MVDR showed the largest improvements in the stationary OLN scenario and the nonstationary SCT scenario. For the 20T and CAN scenarios, iSNR and STOI predict an improvement while the PESQ measure shows no difference to ADMs. These improvements are caused by the enhanced directivity of the adaptive MVDR beamformer compared to the ADMs. The same trend held true when comparing the common and individual postfilters based on the adaptive MVDR beamformer (with respect to the ADM baseline): The best performance was seen in the OLN and SCT scenarios. The amount of improvement provided by each of the postfilters, however, differs from measure to measure. STOI showed a decrease in all algorithm benefits cause by the addition of postfilters. The iSNR measure, on the other hand, showed increases in performance in all but the SCT scenario and PESQ revealed increases for both types of postfilters in all test scenarios. It can be assumed that the decrease in STOI score caused by the addition of postfiltering is due to the introduction of distortions, while the improvements in iSNR and PESQ are caused by an increase in SNR achieved by improved noise reduction. The decrease in iSNR in the SCT scenario can be attributed to errors in the speech-presence probability estimation of the postfilter, when confronted with two single speech sources. As a function of the input SNR, three general trends can be identified when comparing the adaptive binaural MVDR to the previously discussed ADMs: In the 20T and CAN scenarios, all measures show very a similar behavior between the ADMs and adaptive MVDR, with slightly larger improvements by the adaptive MVDR. In the OLN scenario, all measures show notably larger benefits for the adaptive MVDR than the ADM, yet the input SNR-dependence is similar. In the SCT scenario, we see drastically larger benefits across all measures provided by the adaptive MVDR that also shows a very different SNR-dependence. The benefits provided by the adaptive MVDR algorithm decrease with increasing SNR. This behavior can again be explained by the nature of the noise scenario and the algorithm itself: At low input SNRs, the speech power of the interfering talker dominates the acoustic scene and the algorithm can efficiently adapt to this interfering sound source. The direction of enhancement is set and therefore not impacted by the low speech power of the target speaker source at low SNRs. As with the common postfilter based on the fixed binaural MVDR, both postfiltering schemes based on the adaptive binaural MVDR reveal SNR-dependencies that can be understood as the sum of the SCNR algorithm and the adaptive binaural MVDR alone.

It can be observed that the individual postfilter (ind PF) performs slightly better than the common postfilter (com PF) for most scenarios and most measures. Exceptions to this finding were the iSNR in the SCT scenario, which showed a slight decrease in performance for the individual postfilter compared to the common postfilter and PESQ, which revealed the exact opposite: a slight increase in performance for the individual postfilter only for the SCT scenario and slight decreases for the three other scenarios.

The common postfilter was motivated to cause no distortions to the binaural cues (most importantly ILD) by applying the same (real-valued) gains to the left and the right channels. In contrast, for the individual postfilter, the gains were calculated for the left and right channels individually. While this approach produced distortions in the interaural level difference, the SNR improvement for each channel was maximized. NH listeners can benefit from a spatial separation between a target speech source and a noise source by exploiting the interaural cues resulting from this separation (Plomp & Mimpen, 1981). Consequently, it has previously been shown that they can benefit from a signal processing scheme preserving binaural cues (Van den Bogaert et al., 2009). In the instrumental evaluation presented here, however, no binaural instrumental measures were used. The left and the right channels were always regarded separately and such a binaural interaction benefit could not be assessed. Accordingly, the individual postfilter scheme yielded, with few exceptions, the expected better performance compared to the common postfilter scheme.

General Discussion

The SCNR scheme included here for reference performed similarly to what had previously been reported for subjective speech intelligibility measurements in noise, using (single-channel) noise reduced signals. Luts et al. (2010), in a large, multicenter study of signal enhancement algorithms, included two algorithms operating on single channel signals: noise suppression based on perceptually optimized spectral subtraction as well as Wiener filter-based noise suppression. Both of those algorithms showed no change in speech reception threshold compared to unprocessed signals, neither improving speech intelligibility in noise, nor impairing it.

The coherence-based noise reduction scheme investigated here had also previously been investigated in the aforementioned study by Luts et al. (2010). The algorithm was evaluated using speech intelligibility tests with a total of 109 subjects (NH and HI) across four countries. In all test sites, the algorithm showed no improvement in speech intelligibility, contrary to what our instrumental evaluation predicted. In a subjective preference test, however, Luts' subjects preferred the coherence-filtered signals over the unprocessed signals, rating them as *slightly better*. These findings are in line with the PESQ results presented here.

The instrumental evaluation performed here revealed differences between the common and the individual postfilter scheme, that despite being rather small (less than 1 dB iSNR and less than 3% predicted speech intelligibility (STOI)), were highly consistent across measures. The current instrumental evaluation suggested a benefit in speech intelligibility from using individual postfilters, providing maximal noise reduction for each channel.

Overall, all three instrumental measures considered here predicted good performance of the noise reduction algorithms with respect to speech intelligibility as well as speech quality. The optimal working point for most algorithms is around 0 dB input SNR, according to the STOI measure. Algorithms including SCNR can benefit from higher SNRs and provide more benefit in these more favorable conditions. Algorithms including the adaptive binaural MVDR yield the best performance in OLN and the SCT scenario at negative SNRs. The best results were obtained with the binaural MVDR beamforming algorithms (fixed and adaptive MVDR with and without postfilter). It should be noted, however, that these beamformers assume the direction of the target speaker to be known and would hence be unable to cope with nonfrontal talker locations or moving speech targets. For fixed sources located around 0°, however, these algorithms were able to provide improvements in all three instrumental measures.

Summary

In this article, an extensive instrumental evaluation of six binaural and two monaural signal enhancement schemes was presented. Evaluations were performed in four distinct reverberant scenarios that were designed to reflect real-world listening situations. All algorithms were implemented on a common real-time signal processing platform, making the setup ideal for perceptual evaluations.

The following findings emerged:

1. The adaptive differential microphones (ADMs) showed good results in the perceptual evaluation of speech quality (PESQ) evaluation.
2. The predicted speech quality was even more improved when using a coherence-based postfilter in combination with the ADMs.
3. The single-channel noise reduction (SCNR) algorithm tested here showed intelligibility-weighted signal-to-noise-ratio (iSNR) improvements in the stationary speech-shaped noise.
4. The adaptive binaural minimum variance distortionless response (MVDR) beamformers showed larger improvements in predicted speech intelligibility and predicted speech quality than the fixed binaural MVDR beamformer.
5. Processing with the binaural adaptive MVDR beamformer resulted in larger improvements than the monaural ADM in all measures.
6. Postfiltering schemes deriving gains for the left and right channels individually resulted in larger improvements than postfiltering schemes deriving a common gain for the left and the right channels when employed based on the adaptive MVDR beamformer.
7. The best overall performance was seen with the adaptive binaural MVDR beamformer in the single competing talker (SCT) scenario, resulting in a better-ear iSNR improvement of 10.8 dB.

These results are encouraging for perceptual listening tests to assess speech intelligibility and signal quality in NH listeners, HI listeners, and CI users. These tests have been performed and are reported in two subsequent studies in this issue (Baumgärtel et al., 2015; Völker et al., 2015).

Acknowledgements

The authors thank Rainer Huber for his input and for providing code used in this instrumental evaluation, Hendrik Kayser for help in using the HRTF database and providing Figure 5, Giso Grimm for the parameter optimization of the coherence filter, and two anonymous reviewers for their helpful comments.

Author Note

Portions of this work were presented at the *12th Congress of the European Federation of Audiology Societies (EFAS)*, Istanbul, Turkey, May 2015.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under ABCIT grant agreement no. 304912, the DFG (SFB/TRR31 "The Active Auditory System"), and DFG Cluster of Excellence "Hearing4all."

Note

1. Fixed differential microphones (FDMs) working on the left and right side independently are another possible choice of reference algorithm condition. FDMs implemented analogously to the ADMs presented here have been tested using the iSNR measure. Performance differed by less than .1 dB, except for one condition (right side channels, SCT noise), where the ADMs outperformed the FDMs by 2.9 dB. We therefore decided to include the technically more refined ADMs as reference algorithms in this study rather than FDMs.

References

American National Standard Institute. (1997). *Methods for the calculation of the speech intelligibility index*. Washington, DC: Author(ANSI 20S3.5-1997).

Baumgärtel, R., Hu, H., Krawczyk-Becker, M., Marquardt, D., Herzke, T., Coleman, G., . . . ,Dietz, M. (2015). Comparing binaural pre-processing strategies II: Speech intelligibility of bilateral cochlear implant users. *Trends in Hearing*, *19*, 1–18.

Bentler, R. A. (2005). Effectiveness of directional microphones and noise reduction schemes in hearing aids: A systematic review of the evidence. *Journal of the American Academy of Audiology*, *16*, 473–484.

Breithaupt, C., Krawczyk, M., & Martin, R. (2008). Parameterized MMSE spectral magnitude estimation for the enhancement of noisy speech. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 4037–4040. doi:10.1109/ICASSP.2008.4518540.

Brockmeyer, A. M., & Potts, L. G. (2011). Evaluation of different signal processing options in unilateral and bilateral cochlear freedom implant recipients using R-space background noise. *Journal of the American Academy of Audiology*, *22*(2), 65–80.

Chan, D., Fourcin, A., Gibbon, D., Granstrom, B., Huckvale, M., Kokkinakis, G., . . . ,Zeiliger, J. (1995). EUROM—A spoken language resource for the EU. *Proceedings of the 4th European Conference on Speech Communication and Speech Technology*, *1*, 867–870.

Cornelis, B., Moonen, M., & Wouters, J. (2012). Speech intelligibility improvements with hearing aids using bilateral and binaural adaptive multichannel Wiener filtering based noise reduction. *The Journal of the Acoustical Society of America*, *131*(6), 4743–4755.

Doclo, S., Gannot, S., Moonen, M., & Spriet, A. (2010). Acoustic beamforming for hearing aid applications. In S. Haykin, & K. Ray Liu (Eds.), *Handbook on array processing and sensor networks* (pp. 269–302). Hoboken, NJ, USA: Wiley Chapter 9.

Doclo, S., Kellermann, W., Makino, S., & Nordholm, S. (2015). Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones. *IEEE Signal Processing Magazine*, *32*(2), 18–30.

Elko, G. W., & Anh-Tho Nguyen, P. (1995). A simple adaptive first-order differential microphone. *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, 169–172. doi:10.1109/ASPAA.1995.482983.

Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America*, *88*(4), 1725–1736.

Fink, N., Furst, M., & Muchnik, C. (2012). Improving word recognition in noise among hearing-impaired subjects with a single-channel cochlear noise-reduction algorithm. *The Journal of the Acoustical Society of America*, *132*(3), 1718–1731.

Gannot, S., Burshtein, D., & Weinstein, E. (2001). Signal enhancement using beamforming and nonstationarity with applications to speech. *IEEE Transactions on Signal Processing*, *49*(8), 1614–1626.

Gannot, S., & Cohen, I. (2007). Adaptive beamforming and postfiltering. In J. Benesty, Y. Huang, & M. M. Sondhi (Eds.), *Springer handbook of speech processing* (pp. 199–228). Springer: New York, NY, USA.

Gerkmann, T., Breithaupt, C., & Martin, R. (2008). Improved a posteriori speech presence probability estimation based on a likelihood ratio with fixed priors. *IEEE Transactions on Audio, Speech, and Language Processing*, *16*(5), 910–919.

Gerkmann, T., & Hendriks, R. C. (2012). Unbiased MMSE-based noise power estimation with low complexity and low tracking delay. *IEEE Transactions on Audio, Speech, and Language Processing*, *20*(4), 1383–1393.

Greenberg, J. E., Peterson, P. M., & Zurek, P. M. (1993). Intelligibility-weighted measures of speech-to-interference ratio and speech system performance. *The Journal of the Acoustical Society of America*, *94*(5), 3009–3010.

Griffiths, L. J., & Jim, C. (1982). An alternative approach to linearly constrained adaptive beamforming. *IEEE Transactions on Antennas and Propagation*, *30*(1), 27–34.

Grimm, G., Herzke, T., Berg, D., & Hohmann, V. (2006). The master hearing aid: A PC-based platform for algorithm development and evaluation. *Acta Acustica United with Acustica*, *92*(4), 618–628.

Grimm, G., Hohmann, V., & Kollmeier, B. (2009). Increase and subjective evaluation of feedback stability in hearing aids by a binaural coherence-based noise reduction scheme. *IEEE Transactions on Audio, Speech, and Language Processing*, *17*(7), 1408–1419.

Hagerman, B., & Olofsson, A. (2004). A method to measure the effect of noise reduction algorithms using simultaneous speech and noise. *Acta Acustica United with Acustica*, *90*(2), 356–361.

Hamacher, V., Kornagel, U., Lotter, T., & Puder, H. (2008). Binaural signal processing in hearing aids: Technologies and algorithms. In R. Martin, U. Heute, C. Antweiler, & H. Puder (Eds.), *Advances in digital speech transmission* (pp. 401–429). New York, NY, USA: Wiley.

Healy, E. W., Yoho, S. E., Wang, Y., & Wang, D. (2013). An algorithm to improve speech recognition in noise for hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *134*(4), 3029–3038.

Hendriks, R., & Gerkmann, T. (2012). Noise correlation matrix estimation for multi-microphone speech enhancement. *IEEE Transactions on Audio, Speech, and Language Processing*, *20*(1), 223–233.

Hu, Y., & Loizou, P. (2008). Evaluation of objective quality measures for speech enhancement. *IEEE Transactions on Audio, Speech, and Language Processing*, *16*(1), 229–238.

Hu, Y., & Loizou, P. C. (2007a). A comparative intelligibility study of single-microphone noise reduction algorithms. *Journal of the Acoustical Society of America*, *122*(3), 1777–1786.

Hu, Y., & Loizou, P. C. (2007b). Subjective comparison and evaluation of speech enhancement algorithms. *Speech Communication*, *49*(7), 588–601.

Hu, H., Mohammadiha, N., Taghia, J., Leijon, A., Lutman, M. E., & Wang, S. (2012). Sparsity level in a non-negative matrix factorization based speech strategy in cochlear implants. In: *Proceedings of the 20th European Signal Processing Conference (EUSIPCO 2012)* (pp. 2432–2436).

ITU-T. (2001). ITU-T. *Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs* (ITU-T recommendation P.862). Geneva, Switzerland.

Kayser, H., Ewert, S. D., Anemüller, J., Rohdenburg, T., Hohmann, V.Kollmeier, B. (2009). Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses. *EURASIP Journal on Advances in Signal Processing*, *2009*(1), 298605.

Kim, G., Lu, Y., Hu, Y., & Loizou, P. C. (2009). An algorithm that improves speech intelligibility in noise for normal-hearing listeners. *The Journal of the Acoustical Society of America*, *126*(3), 1486–1494.

Levitt, H. (2001). Noise reduction in hearing aids: A review. *Journal of Rehabilitation Research and Development*, *38*, 111–121.

Lotter, T. (2004). *Single and Multimicrophone Speech Enhancement for Hearing Aids* (PhD thesis). RWTH Aachen, Aachen, Germany.

Luts, H., Eneman, K., Wouters, J., Schulte, M., Vormann, M., Buechler, M., . . . ,Spriet, A. (2010). Multicenter evaluation of signal enhancement algorithms for hearing aids. *Journal of the Acoustical Society of America*, *127*(3), 1491–1505.

Peters, R. W., Moore, B. C. J., & Baer, T. (1998). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *The Journal of the Acoustical Society of America*, *103*(1), 577–587.

Plomp, R., & Mimpen, A. M. (1981). Effect of the orientation of the speaker's head and the azimuth of a noise source on the speech-reception threshold for sentences. *Acta Acustica United with Acustica*, *48*(5), 325–328.

Qin, M. K., & Oxenham, A. J. (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *Journal of the Acoustical Society of America*, *114*(1), 446–454.

Rix, A., Beerends, J., Hollier, M., & Hekstra, A. (2001). Perceptual evaluation of speech quality (PESQ)—A new method for speech quality assessment of telephone networks and codecs. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, *2*, 749–752.

Rohdenburg, T. (2008). *Development and objective perceptual quality assessment of monaural and binaural noise reduction schemes for hearing aids* (PhD dissertation). University of Oldenburg, Germany.

Shynk, J. (1992). Frequency-domain and multirate adaptive filtering. *IEEE Signal Processing Magazine*, *9*(1), 14–37.

Simmer, K., Bitzer, J., & Marro, C. (2001). Post-filtering techniques. In M. Brandstein, & D. Ward (Eds.), *Microphone arrays: Signal processing techniques and applications* (pp. 39–57). Berlin, Germany: Springer, Chapter 3.

Stickney, G. S., Zeng, F. G., Litovsky, R., & Assmann, P. (2004). Cochlear implant speech recognition with speech maskers. *Journal of the Acoustical Society of America*, *116*(2), 1081–1091.

Taal, C. H., Hendriks, R. C., Heusdens, R., & Jensen, J. (2011). An evaluation of objective measures for intelligibility prediction of time-frequency weighted noisy speech. *Journal of the Acoustical Society of America*, *130*(5), 3013–3027.

Van den Bogaert, T., Doclo, S., Wouters, J., & Moonen, M. (2009). Speech enhancement with multichannel Wiener filter techniques in multimicrophone binaural hearing aids. *The Journal of the Acoustical Society of America*, *125*(1), 360–371.

Van Veen, B., & Buckley, K. (1988). Beamforming: A versatile approach to spatial filtering. *IEEE ASSP Magazine*, *5*(2), 4–24.

Völker, C., Warzybok, A., & Ernst, S. M. A. (2015). Comparing binaural pre-processing strategies III: Speech intelligibility of normal-hearing and hearing-impaired listeners. *Trends in Hearing*, *19*, 1–18.

Wagener, K., Brand, T., & Kollmeier, B. (1999). Entwicklung und Evaluation eines Satztests für die deutsche Sprache III: Evaluation des Oldenburger Satztests [Development and Evaluation of a Sentence Test for the German Language III: Evaluation of the Oldenburg Sentence Test]. *Zeitschrift für Audiologie*, *38*(3), 86–95.

Wouters, J., Doclo, S., Koning, R., & Francart, T. (2013). Sound processing for better coding of monaural and binaural cues in auditory prostheses. *Proceedings of the IEEE*, *101*(9), 1986–1997.

Yousefian, N., & Loizou, P. (2012). A dual-microphone speech enhancement algorithm based on the coherence function. *IEEE Transactions on Audio, Speech, and Language Processing*, *20*(2), 599–609.