Speech intelligibility improvements with hearing aids using bilateral and binaural adaptive multichannel Wiener filtering based noise reduction[a)]

Bram Cornelis[b)] and Marc Moonen
Department of Electrical Engineering (ESAT-SCD)–Department of IBBT Future Health, KU Leuven, Kasteelpark Arenberg 10 B-3001 Heverlee, Belgium

Jan Wouters
Department of Neurosciences–ExpORL, KU Leuven, O & N 2, Herestraat 49 bus 721 B-3000 Leuven, Belgium

(Received 6 July 2011; revised 4 April 2012; accepted 6 April 2012)

This paper evaluates noise reduction techniques in bilateral and binaural hearing aids. Adaptive implementations (on a real-time test platform) of the bilateral and binaural speech distortion weighted multichannel Wiener filter (SDW-MWF) and a competing bilateral fixed beamformer are evaluated. As the SDW-MWF relies on a voice activity detector (VAD), a realistic binaural VAD is also included. The test subjects (both normal hearing subjects and hearing aid users) are tested by an adaptive speech reception threshold (SRT) test in different spatial scenarios, including a realistic cafeteria scenario with nonstationary noise. The main conclusions are: (a) The binaural SDW-MWF can further improve the SRT (up to 2 dB) over the improvements achieved by bilateral algorithms, although a significant difference is only achievable if the binaural SDW-MWF uses a perfect VAD. However, in the cafeteria scenario only the binaural SDW-MWF achieves a significant SRT improvement (2.6 dB with perfect VAD, 2.2 dB with real VAD), for the group of hearing aid users. (b) There is no significant degradation when using a real VAD at the input signal-to-noise ratio (SNR) levels where the hearing aid users reach their SRT. (c) The bilateral SDW-MWF achieves no SRT improvements compared to the bilateral fixed beamformer.
© *2012 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4707534]

PACS number(s): 43.66.Ts, 43.66.Pn, 43.60.Fg, 43.60.Mn [TD] Pages: 4743–4755

I. INTRODUCTION

Sensorineural hearing loss is most often accompanied by a loss of spectral and temporal resolution in the auditory processing (Plomp, 1978; Dillon, 2001). As a result, hearing aid users experience great difficulties in understanding speech in noisy environments. To reach the same amount of speech understanding as a normal hearing person, a hearing aid user requires a signal-to-noise ratio (SNR) increase of about 4–10 dB (Dillon, 2001; Hamacher *et al*., 2005). Therefore, as there is a great need for increasing the SNR, noise reduction in hearing aids has been an active research area for many years.

Noise reduction algorithms are commonly classified as either single-microphone algorithms or multi-microphone algorithms. Single-microphone noise reduction algorithms have been studied for over four decades [a recent overview can be found in Loizou (2007)]. Although these techniques may increase the SNR, this usually comes at the price of a high speech distortion (Chen *et al*., 2006). Current single-microphone noise reduction algorithms are then found not to be able to achieve a speech intelligibility improvement (Trine and Van Tasell, 2002; Hu and Loizou, 2007; Luts *et al*., 2010; Loizou and Kim, 2011). Therefore, hearing aids are usually fitted with multiple microphones, so that spatial information can be utilized in addition to temporal and spectral information in a multi-microphone noise reduction algorithm. Theoretically, the SNR can then be increased without introducing speech distortion (Chen *et al*., 2006).

Well-known multi-microphone noise reduction algorithms are based on fixed and adaptive beamforming (Van Veen and Buckley, 1988). Fixed beamformers use time-invariant filters and are thus computationally very cheap. Several techniques have been considered for hearing aids, e.g., directional microphones, delay-and-sum beamformers, and superdirective beamformers (Ricketts, 2001; Dillon, 2001; Hamacher *et al*., 2005). Adaptive beamformers are data-dependent techniques that can adapt to changing noise scenarios. The most commonly used and computationally cheapest adaptive beamforming technique is the so-called adaptive directional microphone (Elko and Pong, 1995), where two fixed spatial patterns are combined using an adaptive scalar gain, so that a null is formed in the direction of the strongest noise interferer. A more general class of techniques continuously minimizes the total output power of the beamformer under the constraint that the response towards one (or more) target speech directions is preserved. The so-called minimum variance distortionless response (MDVR) beamformer (Capon, 1969) constrains the response towards a single target direction (e.g., the facing direction of the hearing aid user), the so-called linearly

[a)] A preliminary version of this work was presented at the International Hearing Aid Research Conference (IHCON), Lake Tahoe, CA, August 2010.
[b)] Author to whom correspondence should be addressed. Electronic mail: bram.cornelis@gmail.com

Downloaded 29 Jun 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

constrained minimum variance (LCMV) beamformer (Frost, 1972; Van Veen and Buckley, 1988) extends this to a set of linear constraints. A popular implementation of these beamformers is the generalized sidelobe canceller (GSC) algorithm (Griffiths and Jim, 1982), which transforms the constrained optimization problem into an equivalent but easier unconstrained optimization problem. The GSC can be considered as the current state of the art solution adopted in hearing aids. A two-microphone implementation was indeed shown to be able to achieve a speech intelligibility improvement (Vanden Berghe and Wouters, 1998; Spriet *et al.*, 2007).

A more recent adaptive beamforming technique is based on multichannel Wiener filtering (MWF) (Doclo and Moonen, 2002; Chen *et al.*, 2006), where the goal is to obtain a minimum-mean-squared-error (MMSE) estimate of the speech component in a reference microphone signal. To provide an explicit tradeoff between speech distortion and noise reduction, the speech distortion weighted multichannel Wiener filter (SDW-MWF) was also proposed by Doclo and Moonen (2002). A benefit over the GSC is that, in principle, no *a priori* knowledge about the acoustic environment, target speech location, or microphone characteristics is required. As a result, the SDW-MWF is also less sensitive to imperfections such as microphone mismatch, as demonstrated by the physical evaluation of Spriet *et al.* (2005b). In the study by Luts *et al.* (2010), a three-microphone SDW-MWF implementation was evaluated at different test sites, together with other single- and multi-microphone noise reduction algorithms. Overall, the SDW-MWF achieved the largest speech intelligibility improvements even in highly reverberant environments. Even though in laboratory evaluations, it was demonstrated that current adaptive beamformers can achieve large speech intelligibility improvements compared to fixed beamformers (Hamacher *et al.*, 2005), there has remained some skepticism about the real-world benefit (Woods *et al.*, 2010). That is, relatively few situations are encountered in everyday life where current adaptive beamformers offer significant benefits over fixed beamformers (Woods *et al.*, 2010). Still, it should be noted that only so-called monaural implementations in a single device (with a small microphone array) were assessed in Woods *et al.* (2010) and similar studies.

In this paper, so-called binaural hearing aids (Hamacher *et al.*, 2005) are considered where a wireless link allows for exchanging one or more microphone signals between a left and a right device. As a higher number of microphone signals are then available in a binaural noise reduction procedure, an improved performance can be expected. However, as two separate microphone arrays are combined in the binaural procedure, small variations of intermicrophone distances (between microphones of the two different devices) can occur, for example if one of the devices is not properly placed. An adaptive beamforming algorithm such as the SDW-MWF, which does not require *a priori* knowledge and is not sensitive to such variations, is therefore suited for binaural noise reduction. Furthermore, the binaural noise reduction should also preserve the so-called binaural cues, i.e., the interaural time differences (ITDs) and interaural level differences (ILDs), which are used by the human brain to localize sounds (Blauert, 1996). Correct sound localization (of both speech and noise sources) is an important goal by itself, but can also further improve speech intelligibility (Zurek, 1992). It was shown that the SDW-MWF can be extended so that both the speech and the noise binaural cues (both ITDs and ILDs) can be preserved (Doclo *et al.*, 2006; Klasen *et al.*, 2007; Cornelis *et al.*, 2010). Hence, also from this perspective the SDW-MWF offers a valuable approach to binaural noise reduction.

In the study of Van den Bogaert *et al.* (2009), normal hearing subjects (wearing headphones) were tested by adaptive speech reception threshold (SRT) tests to evaluate the performances of bilateral (i.e., where a left and a right device are used but the two devices work independently) and binaural SDW-MWF algorithms. The binaural SDW-MWF algorithm indeed achieved significant SRT improvements compared to the bilateral SDW-MWF, even when only one microphone signal was transmitted to the contralateral device. However, the considered SDW-MWF algorithms were idealized in two ways. First, the noise reduction filters of the SDW-MWF were estimated in a batch procedure using the complete microphone signals, whereas in practice, the filters should be adaptively estimated in an online procedure. Second, perfect knowledge of the target speech activity was assumed. In practice, a voice activity detection (VAD) algorithm is required, which can indicate the periods where the target speech source is active. It was demonstrated by physical evaluations that the SDW-MWF outperforms the GSC for moderate VAD errors (Spriet *et al.*, 2005a). In Catic *et al.* (2010), it was similarly demonstrated by a physical evaluation that an offline implementation of the binaural SDW-MWF retains much of its benefit when using a real VAD, although the improvement degrades gradually at lower input SNRs.

In this study, a performance evaluation of an online adaptive binaural SDW-MWF implementation using a real VAD will be performed. The test subjects (both normal hearing subjects and hearing aid users) are seated inside a test room and wear two behind-the-ear hearing aids. The hearing aid microphone signals are processed by two external real-time test platforms. The tested spatial scenarios include scenarios similar to those in Van den Bogaert *et al.* (2009) and Luts *et al.* (2010), but a more challenging so-called cafeteria scenario with highly nonstationary noise will also be tested.

The following research questions are addressed and discussed in this paper: (i) Does an adaptive implementation of a bilateral and binaural SDW-MWF still achieve SRT improvements, even in challenging scenarios? (ii) Does the binaural SDW-MWF retain its benefit over bilateral algorithms? (iii) Does using a real VAD result in significant performance degradations for the bilateral and binaural SDW-MWF? (iv) Do the adaptive SDW-MWF algorithms yield significant improvements compared to a fixed beamformer? (v) Do the noise reduction algorithms yield different outcomes for the different subject groups?

II. NOISE REDUCTION ALGORITHMS

A. Configuration

A configuration is considered where the hearing aid user wears a device on each ear. Each device has a microphone

Downloaded 29 Jun 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

array consisting of M microphones. The mth microphone signal in the left device $Y_{L,m}(\omega)$ can then be specified in the frequency domain as

$$Y_{L,m}(\omega) = X_{L,m}(\omega) + V_{L,m}(\omega), \quad m = 0, \ldots, M-1, \quad (1)$$

where $X_{L,m}(\omega)$ represents the speech component and $V_{L,m}(\omega)$ represents the noise component. Similarly, the mth microphone signal in the right device is equal to $Y_{R,m}(\omega) = X_{R,m}(\omega) + V_{R,m}(\omega)$. Two microphone signals (one in each device) are selected as the so-called reference microphone signals for the noise reduction procedures, and are denoted as $Y_{L,\mathrm{ref}}(\omega)$ and $Y_{R,\mathrm{ref}}(\omega)$. Typically, the front microphone signals are used as reference microphone signals.

In a binaural configuration, microphone signals are exchanged between the two devices over a wireless link. In addition to its own M microphone signals, each device therefore also has access to N microphone signals of the contralateral device (with $N \leq M$). In a bilateral configuration, no signals are exchanged between the devices, so that each device works independently, and so that $N = 0$. In general, all available microphone signals in the left device can be stacked in the $(M+N)$-dimensional signal vector $\mathbf{y}_L(\omega)$ and similarly, all available microphone signals in the right device can be stacked in $\mathbf{y}_R(\omega)$. As in Eq. (1), these can also be written as

$$\mathbf{y}_L(\omega) = \mathbf{x}_L(\omega) + \mathbf{v}_L(\omega) \quad (2)$$

for the left device and similarly $\mathbf{y}_R(\omega) = \mathbf{x}_R(\omega) + \mathbf{v}_R(\omega)$ for the right device. For conciseness, we will omit the frequency-domain variable ω in the remainder of the paper.

The $(M+N) \times (M+N)$ correlation matrices, which contain the autocorrelation and cross correlation values of $\mathbf{y}_L$, $\mathbf{x}_L$, and $\mathbf{v}_L$ over the different microphones, are defined as

$$\mathbf{R}_{yL} = \mathcal{E}\{\mathbf{y}_L \mathbf{y}_L^H\}, \quad \mathbf{R}_{xL} = \mathcal{E}\{\mathbf{x}_L \mathbf{x}_L^H\}, \quad \mathbf{R}_{vL} = \mathcal{E}\{\mathbf{v}_L \mathbf{v}_L^H\}, \quad (3)$$

where $\mathcal{E}$ denotes the expected value operator. $\mathbf{R}_{yR}$, $\mathbf{R}_{xR}$, and $\mathbf{R}_{vR}$ are similarly defined. The estimation of the correlation matrices requires a voice activity detection (VAD) algorithm, which classifies frames as speech + noise or noise-only frames, so that the estimates $\hat{\mathbf{R}}_{yL}$,$\hat{\mathbf{R}}_{yR}$ and $\hat{\mathbf{R}}_{vL}$,$\hat{\mathbf{R}}_{vR}$ can be recursively updated as

- In speech + noise frames

$$\begin{aligned} \hat{\mathbf{R}}_{yL}[k] &= \lambda_y \hat{\mathbf{R}}_{yL}[k-1] + (1-\lambda_y)\mathbf{y}_L[k]\mathbf{y}_L^H[k] \,, \\ \hat{\mathbf{R}}_{vL}[k] &= \hat{\mathbf{R}}_{vL}[k-1] \,, \end{aligned} \quad (4)$$

- In noise-only frames

$$\begin{aligned} \hat{\mathbf{R}}_{yL}[k] &= \hat{\mathbf{R}}_{yL}[k-1] \,, \\ \hat{\mathbf{R}}_{vL}[k] &= \lambda_v \hat{\mathbf{R}}_{vL}[k-1] + (1-\lambda_v)\mathbf{v}_L[k]\mathbf{v}_L^H[k] \,, \end{aligned} \quad (5)$$

for the left device and similarly for the right device. k is the frame-index, λ_y and λ_v are exponential forgetting factors which are typically chosen close to one. Assuming the speech and noise are uncorrelated and that the noise is sufficiently stationary, $\hat{\mathbf{R}}_{xL}$ and $\hat{\mathbf{R}}_{xR}$ can then be found as $\hat{\mathbf{R}}_{xL} = \hat{\mathbf{R}}_{yL} - \hat{\mathbf{R}}_{vL}$ and $\hat{\mathbf{R}}_{xR} = \hat{\mathbf{R}}_{yR} - \hat{\mathbf{R}}_{vR}$.

The objective of the noise reduction procedure is to obtain output signals Z_L and Z_R at the left and the right device by filtering and summing the left and right signal vectors, i.e.,

$$Z_L = \mathbf{w}_L^H \mathbf{y}_L, \quad Z_R = \mathbf{w}_R^H \mathbf{y}_R, \quad (6)$$

where $\mathbf{w}_L$ and $\mathbf{w}_R$ are $(M+N)$-dimensional complex weight vectors.

B. Multichannel Wiener filter (MWF)

The multichannel Wiener filter (MWF) (Doclo and Moonen, 2002; Chen *et al.*, 2006) produces a minimum-mean-squared-error (MMSE) estimate of the speech component in the reference microphone signal of each device, hence simultaneously reducing noise and limiting speech distortion. To provide a more explicit tradeoff between speech distortion and noise reduction, the speech distortion weighted multichannel Wiener filter (SDW-MWF) has been proposed, which minimizes a weighted sum of the mean-squared residual noise and speech distortion (Doclo and Moonen, 2002). The SDW-MWF cost function for the filter $\mathbf{w}_L$ estimating the speech component $X_{L,\mathrm{ref}}$ in the reference microphone signal of the left device is equal to

$$J_{\mathrm{SDW-MWF}}(\mathbf{w}_L) = \mathcal{E}\left\{ \left| X_{L,\mathrm{ref}} - \mathbf{w}_L^H \mathbf{x}_L \right|^2 + \mu \left| \mathbf{w}_L^H \mathbf{v}_L \right|^2 \right\}, \quad (7)$$

where the speech distortion parameter μ provides a tradeoff between noise reduction and speech distortion. The SDW-MWF cost function for the right filter $\mathbf{w}_R$ can be similarly defined. The optimal SDW-MWF filters for the left and right device are then equal to

$$\begin{aligned} \mathbf{w}_L &= (\mathbf{R}_{xL} + \mu \mathbf{R}_{vL})^{-1} \mathbf{R}_{xL} \mathbf{e}_L, \\ \mathbf{w}_R &= (\mathbf{R}_{xR} + \mu \mathbf{R}_{vR})^{-1} \mathbf{R}_{xR} \mathbf{e}_R \,, \end{aligned} \quad (8)$$

where $\mathbf{e}_L$ and $\mathbf{e}_R$ are $(M+N)$-dimensional vectors with only one element equal to 1 and the other elements equal to 0, used to select the reference microphone signals out of the signal vectors $\mathbf{y}_L$ and $\mathbf{y}_R$, i.e., $Y_{L,\mathrm{ref}} = \mathbf{e}_L^H \mathbf{y}_L$ and $Y_{R,\mathrm{ref}} = \mathbf{e}_R^H \mathbf{y}_R$.

The SDW-MWF is a good choice for binaural noise reduction: if all microphone signals are exchanged between the devices in a binaural configuration, the SDW-MWF preserves the speech binaural cues (Doclo *et al.*, 2006) in addition to obtaining a noise reduction performance which is superior to the performance obtained in a bilateral configuration. The binaural cues of the residual noise can also be preserved by including extensions to the SDW-MWF cost function (Klasen *et al.*, 2007; Cornelis *et al.*, 2010). Perceptual evaluations have also illustrated that both a significant speech intelligibility improvement (Van den Bogaert *et al.*, 2009) and a good localization performance (Van den Bogaert *et al.*, 2008) can be achieved with the binaural SDW-MWF.

However, as previously mentioned, the evaluated binaural SDW-MWF algorithm was "idealized" in the work of Van den Bogaert *et al.* (2009). In this study, a realistic

Downloaded 29 Jun 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

implementation of the binaural SDW-MWF is therefore evaluated.

- The adaptive RLS-type SDW-MWF algorithm of Doclo *et al.* (2007) is extended for the binaural configuration (cf. Sec. III G) and used as an online adaptive procedure.
- The impact of using a realistic (binaural) VAD algorithm (cf. Sec. III H) is assessed.

Finally, it is noted that a so-called cafeteria scenario with highly nonstationary noise is also tested in this study (cf. Sec. III D). This scenario is particularly challenging for the SDW-MWF.

C. Fixed beamformer (FB)

A bilateral fixed beamformer (FB) is also considered in this study as a competing noise reduction algorithm to the SDW-MWF. The FB calculates output signals in each device as in Eq. (6), where the filters $\mathbf{w}_L$ and $\mathbf{w}_R$ are calibrated *a priori*, and where $\mathbf{y}_L$ and $\mathbf{y}_R$ only contain the two ipsilateral microphone signals, i.e., $M = 2, N = 0$.

FBs are well established in the market, and offer the advantages of very low computational demands and robustness to level and SNR changes. A major disadvantage of a FB is that an assumption has to be made about the target speech source location. Most often, it is assumed that the target speech source location is in the frontal (look) direction. Another FB disadvantage is the sensitivity to "model imperfections" such as microphone mismatch, caused by aging effects or environmental influences (Ricketts, 2001; Hamacher *et al.*, 2005). In addition to a degraded noise reduction performance, the model imperfections can also cause distortion of the binaural cues (Van den Bogaert *et al.*, 2005). Adaptive beamformers (such as the SDW-MWF) can offer a better noise reduction performance for coherent noise compared to a FB (Hamacher *et al.*, 2005), while the SDW-MWF in particular also offers increased robustness to model imperfections such as microphone mismatch, as demonstrated by the physical evaluation of Spriet *et al.* (2005b). As the SDW-MWF relies on a VAD, it can however be outperformed by a FB in very adverse scenarios (such as in low SNR or nonstationary noise scenarios), where the VAD fails to discriminate between speech + noise and noise-only frames.

The directivity pattern of the considered bilateral FB, measured in an anechoic room on a CORTEX MK2 manikin (01dB-Metravib, Limonest Cedex, France), is illustrated in Fig. 1. Only the directivity pattern for the left device is shown, as the pattern for the right device is an almost identical mirror image (i.e., mirrored around the vertical axis). The directivity from the specified angle is defined as the power ratio of the output signal (in dB) for sound arriving for a certain azimuthal angle (with 0° in front of the head, 90° to the right of the head) and for the diffuse case where sound arrives from all directions. In order to relate the performance to speech intelligibility, the directivity is averaged across frequencies with a weighting function according to the speech intelligibility index (SII) as in Dillon (2001) and Hamacher *et al.* (2005). The FB is designed so that a notch is formed at 180° while maximum directivity is obtained at 0° (cardioid-type pattern). It is however apparent from Fig. 1 that the direction of maximum directivity is shifted toward the side of the head in practice, when the device is worn behind the ear, as is also illustrated in Dillon (2001) and Hamacher *et al.* (2005). It should be noted that other patterns such as a hypercardioid or supercardioid could have been used instead of the considered cardioid pattern. The directivity indices of the hypercardioid and supercardioid indicate less attenuation to noise arriving from 180° (compared to the cardioid pattern), but they have the advantage that a better noise reduction performance is achieved in diffuse noise scenarios (Ricketts, 2001; Dillon, 2001; Hamacher *et al.*, 2005).

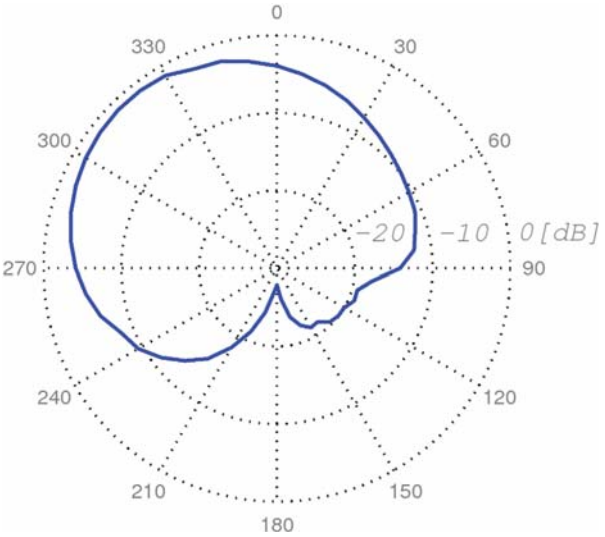


FIG. 1. (Color online) SII-weighted directivity pattern (10 dB grid) of FB at left device, for different azimuthal angles.

III. METHODS

A. General procedure

The speech reception threshold (SRT) is measured by the adaptive test procedure of Plomp and Mimpen (1979). In this procedure, the level of the target speech signal is adjusted in steps of 2 dB depending on the response of the test subject (−2 dB after a correct sentence recognition, +2 dB after an incorrect sentence recognition), while the noise level is fixed. The 50% SRT, i.e., the SNR level at which 50% of the presented speech is correctly identified by the listener, is then obtained as the average input SNR of the last 10 presentations, including a virtual SNR level based on the response to the last sentence. The input SNRs are measured at the center point of the setup, in absence of the subject.

Each subject is tested in either one or three different spatial scenarios (cf. Secs. III D and III F), with six noise reduction settings (including a reference condition with noise reduction switched off) per scenario (cf. Sec. III I). The resulting six or eighteen different conditions are evaluated in randomized order and conducted twice (test and retest) in one (normal hearing subjects) or two sessions (hearing aid users).

B. Setup

The test setup is depicted in Fig. 2. Eight Fostex 6310B loudspeakers (Fostex Company, Tokyo, Japan) are placed in

Downloaded 29 Jun 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

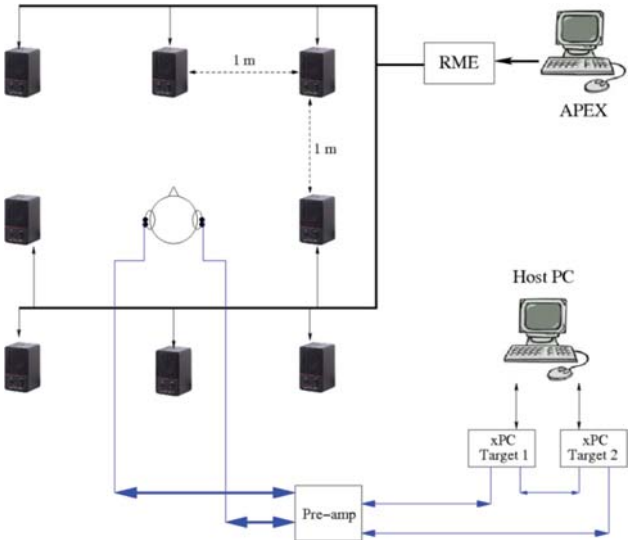


FIG. 2. (Color online) Test setup.

a lattice with an interspacing of 1 m, at a height of 1.3 m. The test subject is seated in the middle of this lattice, and is asked to look to the front loudspeaker during the test, where minor head movements are allowed. The experimenter and most of the equipment are placed in a separate control room. By means of a surveillance camera the experimenter is able to check if the test subject remains in the correct position during the test.

The adaptive SRT procedure is controlled by the APEX 3 psycho-acoustical test platform (Francart *et al.*, 2008), which generates the speech and noise stimuli. The generated stimuli are then presented through the loudspeakers using an RME Hammerfall DSP II soundcard (RME Audio, Haimhausen, Germany).

The test subjects wear behind-the-ear hearing aids (Phonak, Sonova Holding AG, Stäfa, Switzerland) in both ears, whereby each device has two omnidirectional microphones ($M = 2$). The hearing aids do not contain a signal processor, but allow access to the microphone signals. These microphone signals are pre-amplified and routed to two hardware platforms (where each platform represents one hearing aid), which perform real-time signal processing. The software was developed in the xPC Target real-time software environment (MathWorks, 2000), which compiles the complete signal processing scheme into C-code and transmits it to the two hardware platforms (the so-called Target machines).

The experimenter can change the parameter settings of the Target machines during the test (at runtime) by means of the so-called Host PC, which is connected to the Target machines via an Ethernet connection. A graphical user interface (GUI), running on the Host PC, was created so that the experimenter can easily change the audio volume, hearing loss thresholds, and noise reduction algorithm. As the subject's own voice could erroneously influence the filter adaptation during the test, a pushbutton was also implemented by which the updates of all internal variables (speech/noise statistics, noise floor thresholds used in the VAD,…) can be frozen. After each presented sentence, the button is pushed by the experimenter so that the filter adaptation is switched off during the subject's response, and pushed again after the subject's response so that the filter adaptation is again switched on.

Each Target machine has a total of four input audio signals: the two microphone signals of the associated hearing aid, the front microphone signal of the contralateral device (for binaural noise reduction), and a copy of the clean speech signal which is generated by APEX and is also sent to the appropriate loudspeaker. The copy of the clean speech signal is used for constructing a perfect VAD reference signal for some of the tested conditions. The (left and right) processed output signals are redirected to the receivers of the two hearing aids. Instead of earmolds, disposable foam earplugs with tubing are used, which prevent the direct sound from entering the ear canal (and can thus be viewed as a closed fitting). It is noted that if an open fitting with a large vent is applied in practice, the obtained noise reduction benefits may be lower than the values reported here due to low-frequency noise entering the ear canal through the vent (Dillon, 2001).

C. Room characteristics

Tests are performed in a soundproof room with dimensions 6 m × 5.5 m × 2.6 m (where 6 m corresponds to the horizontal direction in Fig. 2 and 2.6 m is the height of the room). Subjects are seated in the right half of the room, at 1.7 m from the right wall and 3.1 m from the front wall. The reverberation time was measured according to the ISO 3382 (1997) standard, where the equipment belonging to the setup was present, but in absence of a test subject. The room has an SII-weighted [averaged across frequencies with SII-weighting according to ANSI-SII (1997)] reverberation time of 0.62 s.

D. Spatial scenarios

Two single noise source scenarios (S_0N_{45} and $S_{90}N_{270}$) and a pseudo-diffuse (three noise sources) scenario ($S_0N_{90/180/270}$) are considered, which are similar to the scenarios tested by Van den Bogaert *et al.* (2009). The scenarios are denoted as S_xN_y throughout the paper, with x the azimuthal angle-of-arrival of the target speech (S) signal and y the azimuthal angle-of-arrival of the noise (N) signal(s). 0° is located in front of the subject, 90° to the right of the subject. In these scenarios, the loudspeakers are pointed toward the subject, seated at the center of the lattice.

A more challenging realistic scenario, representative of a cafeteria situation, is also considered and shown in Fig. 3. The target speech source is generated by the loudspeaker in front of the subject (denoted as loudspeaker 7 in Fig. 3), while the noise interferers are generated by loudspeakers 1, 2, 3, 4, and 8. For this scenario the loudspeakers are not pointed toward the subject but are pointed in a vertical direction, as illustrated in Fig. 3. This particular setup therefore mimics concurrent (interfering) conversations which are taking place around the hearing aid user, as in a cafeteria. For this scenario, the direct-to-reverberant energy ratio (DRR) also varies significantly depending on the loudspeaker from which the sound originates. For a sound originating from loudspeaker 7 (pointed toward the subject), the DRR (measured in the left front microphone) is equal to 4.5 dB, whereas

Downloaded 29 Jun 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

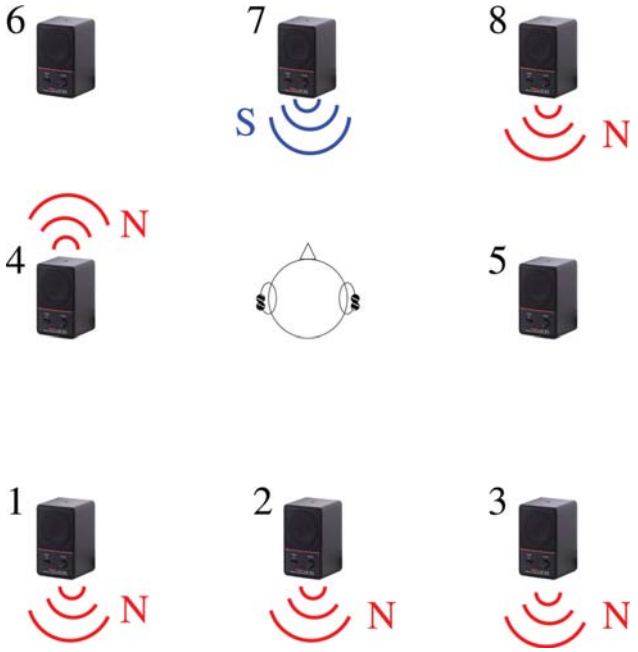


FIG. 3. (Color online) Cafeteria scenario.

a sound originating from loudspeaker 2 (pointed away from the subject) leads to a DRR equal to −2.0 dB.

E. Stimuli

Sentences spoken by a Dutch male talker (Versfeld *et al.*, 2000) are used as speech material. The database consists of 39 lists with 13 sentences per list. As six noise reduction algorithms (cf. Sec. III I) are tested (in a test/retest setup), where every algorithm is tested in each scenario, this implies that at most three different spatial scenarios can be tested per subject (as this results in a total of 36 test conditions).

In the single noise source and pseudo-diffuse (three noise sources) scenarios, multi-talker babble noise (Auditec, 1997) is used as noise signal(s). The signals are scaled so that the total input noise sound pressure level (SPL) at the center point of the lattice, measured in absence of the subject, is equal to 65 dBA.

For the cafeteria scenario, sentences of the English hearing-in-noise test (HINT) (Nilsson *et al.*, 1994) are used as noise signals. The cafeteria noise sentences are in a different language than the target sentences because of practical considerations: the test subjects will not be confused as to which sentences are the target speech (i.e., sentences in their own language) and which are unwanted noise (i.e., sentences in foreign language). The different cafeteria noise signals are time-shifted so that they are not all active at the same time, while at least one noise signal is active at any time instant. The signals are scaled so that the average SPL per noise signal is 60 dBA, measured at 1 m of the loudspeaker. For this calibration, speech-weighted noise with a spectrum that matches the average HINT spectrum, is used as calibration stimulus.

F. Subjects and fitting

Two groups of normal hearing subjects participated in the tests. The first group of ten subjects were tested in three spatial scenarios (S_0N_{45}, $S_{90}N_{270}$, and cafeteria). After these tests, it was decided to test a second group of ten normal hearing subjects in one additional spatial scenario ($S_0N_{90/180/270}$). All normal hearing subjects had average hearing thresholds better than or equal to 20 dB hearing level (HL) for octave frequencies between 250 and 8000 Hz.

Eight hearing impaired subjects, all experienced bilateral hearing aid users, participated in the tests. The subjects had a moderate symmetrical sensorineural hearing loss; the average hearing loss is depicted in Fig. 4. As discussed in Sec. III E, the available number of speech lists only allows testing a maximum of three different spatial scenarios per subject (as a list should only be presented once). For the hearing aid users (where only a small number of subjects is available), it was therefore decided to test the S_0N_{45}, $S_0N_{90/180/270}$, and cafeteria scenarios, and to omit the $S_{90}N_{270}$ scenario: for the $S_{90}N_{270}$ scenario, no significant improvements were found in the pilot tests on normal hearing subjects (cf. Sec. IV B), and it is assumed that for hearing aid users also no significant effects would be found.

A frequency-dependent gain is fitted based on the individual audiograms, in accordance to the NAL-RP prescription rule (Byrne *et al.*, 1991). Compression is not included, only a limiter set at 100 dB SPL per frequency band. Pilot tests are conducted prior to the actual SRT test procedures in order to check if the overall gain is experienced as too loud or too quiet for the hearing aid user. In a first pilot test, sentences are presented from the front loudspeaker, where the hearing aid user is asked if the sentences are understandable and at a comfortable volume level. If not, the audio volume is increased or decreased based on the hearing aid user's response, and the procedure is repeated until a comfortable (and fully audible) level is found. To check the audibility and the fitting of the amplification, the hearing aid users are also tested in two pilot adaptive SRT tests, where the SRT obtained with the subject's own devices is compared to the SRT obtained with the devices connected to the test platform. For the normal hearing subjects, the insertion gain was set to 0 dB.

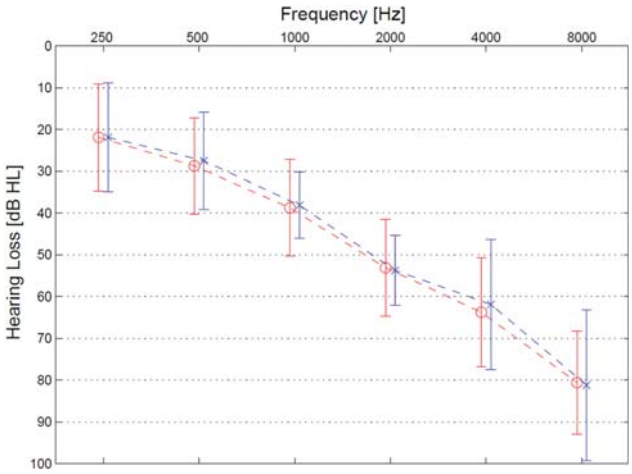


FIG. 4. (Color online) Average left (×) and right (○) hearing loss (HL) per octave frequency of eight hearing aid users. Standard deviations are indicated by error bars.

Downloaded 29 Jun 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

G. SDW-MWF implementation

In Doclo *et al.* (2007), several monaural adaptive frequency-domain SDW-MWF algorithms were proposed which perform frequency-domain filtering with an overlap-save (OLS) filterbank. The bilateral and binaural adaptive SDW-MWF algorithms considered here are based on the unconstrained algorithm with block-structured correlation matrices, denoted as Algo 1 in Doclo *et al.* (2007). Two major changes are applied to the algorithm.

- The fixed spatial preprocessing stage is omitted as this preprocessing can distort the binaural cues (Van den Bogaert *et al.*, 2005). The considered algorithm can therefore be viewed as an adaptive version of the algorithm evaluated by Van den Bogaert *et al.* (2009).
- A weighted overlap-add (WOLA) filterbank (cf. Sec. III H) is used instead of an OLS filterbank.

As in Van den Bogaert *et al.* (2009), the speech distortion parameter μ is chosen as $\mu = 5$ for all test conditions as this setting offers a good compromise between noise reduction and speech distortion.

H. Support algorithms

Besides the noise reduction algorithms, some support algorithms are also implemented in the real-time software environment.

- Filterbank: a weighted overlap-add (WOLA) filterbank is implemented as this is a flexible framework suitable for hearing aid applications (Crochiere, 1980; Brennan and Schneider, 1998). The microphone signals are sampled at 20 480 Hz, the DFT size is equal to 128, and the frames overlap by 75%. For these settings, the WOLA filterbank introduces an input–output delay of only 4.6 ms, which is well below the 20 ms limit reported by Stone and Moore (1999), where the hearing aid delay likely becomes disturbing to the hearing aid user.
- Voice activity detection (VAD): the VAD algorithm, denoted as fusion-2 VAD in Cornelis *et al.* (2011), is used in this evaluation. The fusion-2 VAD makes a decision fusion of log-energy based VAD algorithms operated at both sides of the head, together with a cross-correlation based VAD (which assumes the target speech signal azimuthal angle-of-arrival is 0°), where the decision fusion rule is based on local SNR estimates.
- Frequency dependent gain and subband power limiter: the gain and limiter use the same WOLA filterbank as the noise reduction algorithms so that no additional delay is introduced.

I. Overview evaluated algorithms

The following six noise reduction settings are evaluated:

- Unprocessed (REF): Bilateral reference condition where no noise reduction is applied. The front microphone signals are selected as output signals.
- Bilateral FB (BIL-FB): Bilateral 2-microphone FB (devices work independently; $M = 2, N = 0$).
- Bilateral MWF perfect VAD (BIL-MWF-P): Bilateral 2-microphone SDW-MWF (devices work independently; $M = 2,\ N = 0$), which makes use of an ideal VAD reference.
- Bilateral MWF real VAD (BIL-MWF-R): Bilateral 2-microphone SDW-MWF (devices work independently; $M = 2, N = 0$), which uses a realistic VAD (fusion-2 VAD).
- Binaural MWF perfect VAD (BIN-MWF-P): Binaural 3-microphone SDW-MWF (2 ipsilateral microphone signals + the contralateral front microphone signal; $M = 2$, $N = 1$), which makes use of an ideal VAD reference.
- Binaural MWF real VAD (BIN-MWF-R): Binaural 3-microphone SDW-MWF (2 ipsilateral microphone signals + the contralateral front microphone signal; $M = 2$, $N = 1$), which uses a realistic VAD (fusion-2 VAD).

For the binaural SDW-MWF algorithms, it is assumed that only one microphone signal ($N = 1$) can be exchanged between the devices (in full-duplex), as the transmission of (extra) microphone signals increases both the computational complexity and the power consumption. Moreover, it was also observed (for batch algorithms) that exchanging additional microphone signals does not result in significant performance improvements (Van den Bogaert *et al.*, 2009).

IV. RESULTS

The average SRT scores obtained in the four spatial scenarios are provided in Figs. 5–8. Statistical analyses were carried out on the SRT data using the SPSS 16.0 software. As it was illustrated by Van den Bogaert *et al.* (2009) that the performance of the algorithms is dependent on the position of the speech and noise source(s), every spatial scenario is analyzed separately. To analyze the SRT data, a mixed-model factorial repeated-measures analysis of variance (abbreviated to "ANOVA") is carried out. The mixed-model ANOVA includes two within-subjects factors, i.e., algorithm (6 levels) and test/retest (2 levels), and one between-subjects factor, i.e., subject group (normal hearing subject or hearing aid user). Pairwise comparisons between all algorithmic

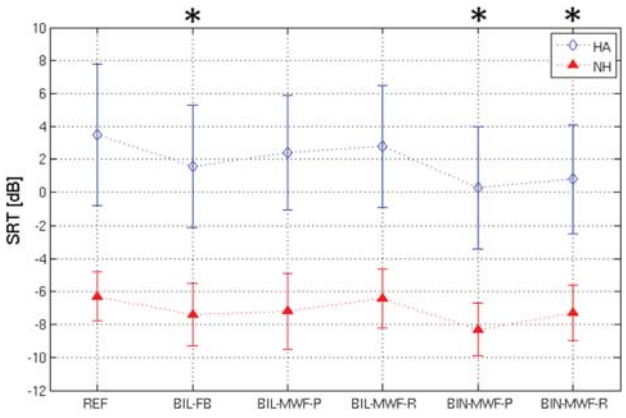


FIG. 5. (Color online) Average SRT results of ten normal hearing subjects (NH) and eight hearing aid users (HA) for the S_0N_{45} scenario. Standard deviations are indicated by error bars. Algorithms which significantly improve the SRT for both groups, compared to the unprocessed (REF) condition, are marked by an "*" above the graph.

Downloaded 29 Jun 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

conditions are carried out as post hoc tests. To guard against type I errors, Bonferroni corrections for multiple comparisons are applied. A p-value of $p = 0.05$ is used as the threshold for significance in all tests.

A. S_0N_{45} scenario

The SRT data of the normal hearing subjects and hearing aid users is provided in Fig. 5. It can be observed that the hearing aid users consistently reach their SRT at higher SNR levels than the normal hearing subjects. The mixed-model ANOVA however indicates that the between-subjects variable subject group does not interact with the within-subjects factors, i.e., p-values of $p = 0.118$ and $p = 0.267$ are measured for the interactions subject group-test/retest and subject group–algorithm, respectively. This finding motivates a single analysis for both subject groups.

A main effect is found for the factor algorithm ($p < 0.001$), but no significant effect of test/retest ($p = 0.866$) and no interaction effect between the two factors ($p = 0.072$). Post hoc tests show that three algorithms obtain a significant SRT improvement over the unprocessed condition: the largest SRT improvement of 2.6 ± 1.9 dB ($p < 0.001$) is obtained by BIN-MWF-P, followed by an improvement of 1.8 ± 1.9 dB ($p = 0.007$) for BIN-MWF-R, and finally an improvement of 1.5 ± 1.8 dB ($p = 0.035$) for the BIL-FB. The binaural SDW-MWF algorithms are significantly better than their bilateral counterparts ($p = 0.012$ for BIN-MWF-P versus BIL-MWF-P and $p = 0.037$ for BIN-MWF-R vs. BIL-MWF-R), but not significantly better than the FB ($p = 0.189$ and $p = 1.000$ for BIL-FB vs. BIN-MWF-P and BIN-MWF-R, respectively). BIN-MWF-R obtains a 0.8 dB lower SRT improvement than BIN-MWF-P, but this difference is statistically non-significant ($p = 0.911$).

B. $S_{90}N_{270}$ scenario

The average SRT data of ten normal hearing subjects is provided in Fig. 6. The ANOVA indicates a main effect for algorithm ($p = 0.002$), no significant effect for test/retest ($p = 0.526$), and no interaction effect ($p = 0.233$). As is apparent from the figure, none of the algorithms is however able to improve the intelligibility compared to the unprocessed (REF) condition. In fact, the post hoc tests reveal that BIL-MWF-R significantly degrades the performance, namely by 1.6 ± 1.3 dB ($p = 0.049$). For the other algorithms, no significant difference with the unprocessed (REF) condition can be demonstrated. Therefore, this scenario was not included in the evaluations with hearing aid users.

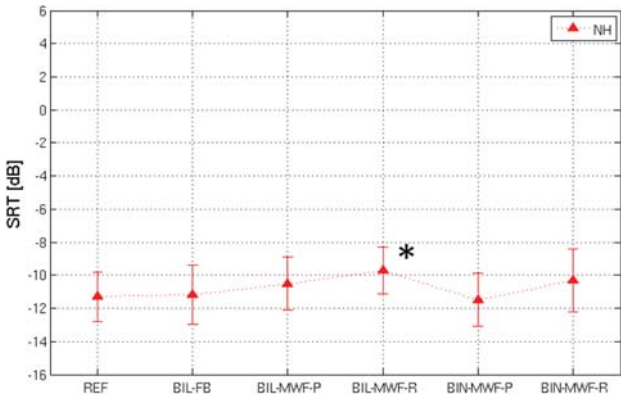


FIG. 6. (Color online) Average SRT results of ten normal hearing subjects (NH), for the $S_{90}N_{270}$ scenario. Standard deviations are indicated by error bars. Significant degradations compared to the unprocessed (REF) condition are marked by an "*."

C. $S_0N_{90/180/270}$ scenario

The SRT data of the normal hearing subjects and hearing aid users is provided in Fig. 7. It can again be observed that the hearing aid users consistently reach their SRT at higher SNR levels than the normal hearing subjects. The mixed-model ANOVA again indicates that the between-subjects variable subject group does not interact with the within-subjects factors, i.e., p-values of $p = 0.132$ and $p = 0.218$ are measured for the interactions subject group-test/retest and subject group–algorithm, respectively, so that a single analysis can be made for both subject groups.

A main effect is found for the factors algorithm ($p < 0.001$) and test/retest ($p = 0.026$). The significant effect for test/retest can most likely be attributed to the results of the hearing aid users, whose retest results are indeed 1.0 dB better on average, compared to the test results. This was also observed in the study of Luts *et al.* (2010), where it was hypothesized that the better retest results may be due to a learning effect. However, since no interaction between algorithm and test/retest is found ($p = 0.822$), the pairwise comparisons between algorithms are valid for both the test and retest session. All algorithms significantly improve the SRT over the unprocessed (REF) condition ($p < 0.001$). The largest improvements are found for the binaural SDW-MWF algorithms (5.4 ± 1.5 dB and 4.5 ± 1.6 dB for BIN-MWF-P and BIN-MWF-R, respectively), followed by the bilateral SDW-MWF algorithms (3.7 ± 1.1 dB and 3.7 ± 1.4 dB for BIL-MWF-P and BIL-MWF-R, respectively), and finally the smallest improvement for BIL-FB (3.3 ± 1.2 dB). While BIN-MWF-P significantly outperforms all bilateral

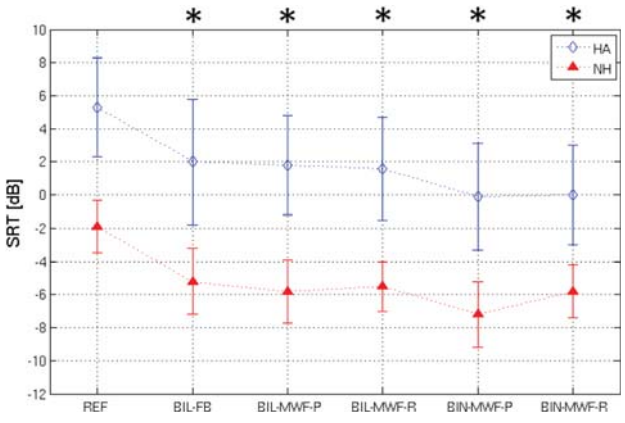


FIG. 7. (Color online) Average SRT results of ten normal hearing subjects (NH) and eight hearing aid users (HA) for the $S_0N_{90/180/270}$ scenario. Standard deviations are indicated by error bars. Algorithms which significantly improve the SRT for both groups, compared to the unprocessed (REF) condition, are marked by an "*" above the graph.

Downloaded 29 Jun 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

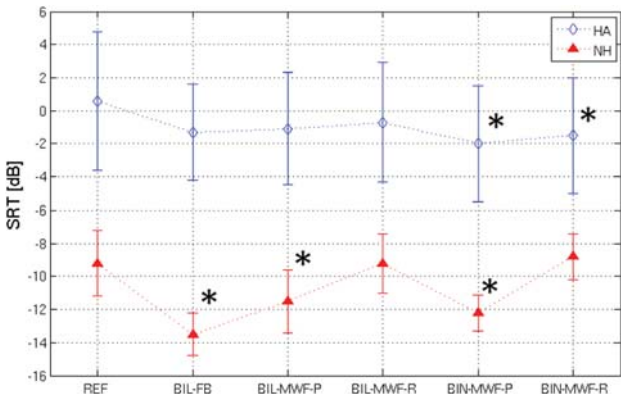


FIG. 8. (Color online) Average SRT results of ten normal hearing subjects (NH) and eight hearing aid users (HA), for the cafeteria scenario. Standard deviations are indicated by error bars. Significant SRT improvements compared to the unprocessed (REF) condition are marked by an "*."

algorithms ($p \leq 0.004$), when using a real VAD no significant improvement over the bilateral algorithms can be demonstrated ($p \geq 0.133$). BIN-MWF-R obtains a 0.9 dB lower SRT improvement than BIN-MWF-P, but this difference is statistically non-significant ($p = 0.331$).

D. Cafeteria scenario

The SRT data of the normal hearing subjects and hearing aid users is provided in Fig. 8. Again, the hearing aid users consistently reach their SRT at higher SNR levels than the normal hearing subjects. The mixed-model ANOVA indicates that there is a significant interaction effect between the factors subject group and algorithm ($p < 0.001$). It is also clearly visible in Fig. 8 that the SDW-MWF algorithms using a real VAD seem to lead to different SRT improvements in the two subject groups. A separate ANOVA is therefore carried out for each subject group.

For the normal hearing subjects, a main effect for the factor algorithm is found ($p < 0.001$), but no significant effect of test/retest ($p = 0.112$) and no interaction effect ($p = 0.411$). Post hoc tests reveal that three algorithms obtain a significant SRT improvement over the unprocessed (REF) condition: the BIL-FB obtains the largest SRT improvement of 4.2 ± 1.9 dB ($p = 0.002$), followed by the BIN-MWF-P with 3.0 ± 1.7 dB ($p = 0.006$) and the BIL-MWF-P with 2.2 ± 1.7 dB ($p = 0.033$). The SDW-MWF algorithms using a real VAD do not obtain SRT improvements, and are significantly outperformed by the other algorithms.

For the hearing aid users, a main effect for the factor algorithm is found ($p = 0.002$), but no significant effect of test/retest ($p = 0.437$) and no interaction effect ($p = 0.607$). Unlike for the normal hearing subjects, now only the binaural SDW-MWF obtains a significant SRT improvement which, moreover, does not seem affected by using a real VAD: the SRT improvements are 2.6 ± 1.1 dB ($p = 0.004$) for BIN-MWF-P and 2.2 ± 1.2 dB ($p = 0.025$) for BIN-MWF-R. No significant SRT improvements are found for the BIL-FB ($p = 0.529$) or BIL-MWF-P and BIL-MWF-R ($p = 0.435$ and $p = 1.000$, respectively), and no significant differences can be demonstrated for any other comparisons between algorithms.

V. DISCUSSION

A. Differences between subject groups

The absolute SRT scores show differences between subject groups. As commonly observed in perceptual evaluations, the hearing aid users have higher (=worse) SRT scores than the normal hearing subjects. This worse performance can be attributed to a combination of effects (Plomp, 1978; Dillon, 2001), but more particularly to supra-threshold auditory problems, such as worse spectral and temporal resolution.

As the SRT scores are determined at very different SNR levels for the normal hearing subjects and hearing aid users, it is possible that the algorithms performed differently for both subject groups. More in particular, a prior physical evaluation (not reported in this paper) indicated that the SNR improvement obtained by the SDW-MWF (using a perfect VAD) decreases if the input SNR is lower, while the speech distortion increases. Moreover, the performance likely degrades further when using a real VAD, as the VAD performs worse at low SNRs and thus introduces estimation errors.

The statistical analysis indicates that for the S_0N_{45} and $S_0N_{90/180/270}$ scenarios, the relative SRT improvements are not significantly different for both subject groups. This result suggests that on average, even for the normal hearing subjects, the SNR levels are sufficiently high for a good performance of both the SDW-MWF and the VAD. For the cafeteria scenario on the other hand, significantly different SRT improvements are observed for the two groups. First, as the normal hearing subjects are evaluated at very low SNR levels (also lower than for the S_0N_{45} and $S_0N_{90/180/270}$ scenarios), the VAD cannot perform well. As a result, the SDW-MWF algorithms using a real VAD are indeed not able to improve the speech intelligibility. Second, as this is actually a speech-in-speech scenario (thus with highly nonstationary noise, which violates the SDW-MWF noise-stationarity assumption, cf. Sec. II B), the SDW-MWF performance likely degrades, especially at low SNR levels. This may explain why the FB outperforms the SDW-MWF algorithms for the normal hearing subjects. At higher SNR levels, the VAD is less affected, and the noise component has less weight in the SDW-MWF cost function so that the SDW-MWF performance is also less affected. For the hearing aid users, who obtain their SRTs at higher SNR levels, the results indeed indicate that the binaural SDW-MWF (both BIN-MWF-P and BIN-MWF-R) generally achieves the best performance.

B. Algorithmic performance

1. Adaptive SDW-MWF performance

One of the general goals of this evaluation is to assess whether an adaptive bilateral or binaural SDW-MWF algorithm is able to improve speech intelligibility in a realistic listening environment. In previous work, a three-microphone adaptive bilateral SDW-MWF algorithm based on Doclo *et al.* (2007), was evaluated at multiple test sites and shown

Downloaded 29 Jun 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

to indeed offer significant intelligibility improvements (Luts *et al.*, 2010). Although the SDW-MWF algorithms in this paper are also based on Doclo *et al.* (2007), they are different in a few crucial points, most notably the omission of a fixed preprocessing stage and the use of a low-delay WOLA filterbank (cf. Sec. III G). The evaluated adaptive SDW-MWF algorithms are therefore better matched to the batch SDW-MWF algorithms evaluated by Van den Bogaert *et al.* (2009). The fixed preprocessing stage is omitted in order not to compromise the localization performance (an important aspect of binaural hearing aids). Moreover, if the target speech signal does not arrive from the assumed frontal direction, the performance can also degrade when using a fixed preprocessing stage.

The results indicate that the adaptive SDW-MWF algorithms retain their speech intelligibility improvements for scenarios with a single noise source (S_0N_{45}) and with multiple noise sources ($S_0N_{90/180/270}$), which are comparable to scenarios tested by Van den Bogaert *et al.* (2009). For the more realistic cafeteria scenario, it was anticipated that the nonstationarity of the noise would limit the SDW-MWF performance. However, for hearing aid users who obtain their SRTs at high SNR levels, the binaural SDW-MWF is still able to obtain significant SRT improvements.

As in Van den Bogaert *et al.* (2009), ten normal hearing subjects were also tested in a scenario with speech and noise arriving from different sides of the head ($S_{90}N_{270}$). Because of the headshadow effect, there is a large SNR difference between the two ears. In accordance with Van den Bogaert *et al.* (2009), prior physical evaluations (not reported in the paper) indicate that a large SNR improvement is obtained in the worst SNR ear, while only a small SNR improvement is obtained in the best SNR ear (which is mainly used for speech understanding). Moreover, as no effort is made here to preserve the residual noise cues, the listener cannot benefit from the binaural unmasking effect (Zurek, 1992). As a result, no SRT improvements are found for any of the tested SDW-MWF algorithms, which is in accordance with the results of Van den Bogaert *et al.* (2009). It should however be noted that it is possible to include so-called partial noise estimation, where part of the unprocessed microphone signal is mixed into the SDW-MWF output signal (Klasen *et al.*, 2007; Cornelis *et al.*, 2010). As the listener can then potentially benefit from the binaural unmasking effect, the SRT can be improved for this particular scenario (Van den Bogaert *et al.*, 2009). Due to limited testing time and limited sentence material, this algorithm is not included in the current evaluation.

2. Binaural versus bilateral SDW-MWF performance

The transmission of microphone signals between hearing aids clearly comes at a cost: both the computational complexity and the power consumption increase. A major research question is therefore whether the binaural benefit (in terms of SRT improvements) justifies the extra overhead. In array processing literature, it is well-known that if the size of an array increases and if more microphones are used, narrower beam patterns can be created. This helps in scenarios with a closely spaced speech and noise source as the noise can then be reduced without affecting the speech. Additional microphones also offer more degrees of freedom in the design, which particularly results in benefits in multiple noise sources scenarios. An extreme example based on these principles are the so-called hearing glasses where multiple microphones are built into the frame of eyeglasses (Soede *et al.*, 1993). The same effects are also observed for the binaural SDW-MWF (Srinivasan, 2008; Van den Bogaert *et al.*, 2009), although the analysis is only made for an idealized algorithm (with perfect knowledge of the speech and noise signal second order statistics).

The results of this work indicate that a binaural SDW-MWF indeed retains benefits in realistic circumstances (adaptive algorithm, real VAD, challenging listening environment). In the S_0N_{45} scenario, where a narrow beam width is required, the binaural SDW-MWF algorithm (both BIN-MWF-P and BIN-MWF-R) obtains the largest SRT improvements (average improvements of 2.6 dB and 1.8 dB, respectively), while the bilateral SDW-MWF algorithms (BIL-MWF-P and BIL-MWF-R) do not obtain a significant improvement. The bilateral FB (BIL-FB) is able to achieve a significant SRT improvement of 1.5 dB. This SRT improvement is slightly smaller than the improvement obtained with the binaural SDW-MWF, but not significantly different. In the multiple noise sources scenario ($S_0N_{90/180/270}$), all algorithms are able to improve the SRT, where the binaural SDW-MWF (both BIN-MWF-P and BIN-MWF-R) obtains the largest SRT improvement (about 2 dB additional benefit compared to the bilateral algorithms). However, a significant difference between the binaural SDW-MWF and the bilateral algorithms could only be demonstrated for BIN-MWF-P (i.e., using a perfect VAD). The cafeteria scenario again presents a listening environment with multiple noise sources, which are however not all active at the same time (so that the noise directions are constantly changing). Nevertheless, for the hearing aid users, the binaural SDW-MWF algorithms (BIN-MWF-P and BIN-MWF-R) offer significant SRT improvements, while no significant improvements can be demonstrated for any of the bilateral algorithms. Finally, for the $S_{90}N_{270}$ scenario, none of the algorithms is able to improve the SRT. As previously noted, it was demonstrated that including partial noise estimation can offer SRT improvements (Van den Bogaert *et al.*, 2009), but then only in combination with a binaural SDW-MWF algorithm.

In conclusion, it can be stated that in the tested scenarios, the binaural SDW-MWF can improve the SRT (up to 2 dB) versus the improvements obtained by the bilateral algorithms (for S_0N_{45} and $S_0N_{90/180/270}$), although a statistically significant difference is only achieved in the $S_0N_{90/180/270}$ scenario, if the binaural SDW-MWF uses a perfect VAD. However, for the cafeteria scenario only the binaural SDW-MWF achieves a significant SRT improvement versus the unprocessed condition (2.6 dB with perfect VAD, 2.2 dB with real VAD), for the group of hearing aid users.

As it can be assumed that the number of microphone signals that can be exchanged between the devices is limited, only the contralateral front microphone signals are exchanged in the evaluated binaural SDW-MWF algorithm.

Downloaded 29 Jun 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

Although exchanging additional microphone signals could result in a further improvement, it was observed (for batch algorithms) that this does not have a significant impact on the performance (Van den Bogaert *et al.*, 2009). Although this setup is not tested in the current study, the same effect would most likely be observed for the adaptive algorithms.

3. VAD performance

In the S_0N_{45} and $S_0N_{90/180/270}$ scenarios, the SRT improvement obtained with the binaural SDW-MWF degraded by 0.8 – 0.9 dB when using BIN-MWF-R instead of BIN-MWF-P, although this difference was non-significant. It should however be noted that the degradation is larger for the normal hearing subjects than for the hearing impaired subjects, as is visible in Figs. 5 and 7. Although statistically no differences can be demonstrated between both groups, the degradation of 0.9 dB may thus be an overestimation. Nevertheless, BIN-MWF-R still outperforms the other algorithms. For the bilateral SDW-MWF, no significant differences were observed between the performance of BIL-MWF-P and BIL-MWF-R.

As previously mentioned, significant differences between subjects groups are observed for the cafeteria scenario, in contrast to the other scenarios. The VAD does not perform well at the low SNR levels where the normal hearing subjects are measured. However, for the hearing aid users who obtain their SRTs at higher SNR levels, a non-significant difference of only 0.4 dB between BIN-MWF-P and BIN-MWF-R indicates that the VAD does not have a major impact. This example illustrates that, for some scenarios, it is indeed necessary to also test hearing aid users who are evidently the actual target users.

For the $S_{90}N_{270}$ scenario, where only normal hearing subjects were tested, a very low SRT is achieved as in this scenario the headshadow and binaural unmasking effects can be fully exploited. The VAD does not perform well which even results in a significant SRT degradation for the bilateral SDW-MWF. Again, as hearing aid users would obtain their SRT at higher SNR levels, the observed performance degradation could be smaller. For the binaural SDW-MWF algorithms, no significant degradation is observed.

In conclusion, it was demonstrated that the SDW-MWF retains its SRT improvements when using a real VAD at the SNR levels where the hearing aid users would actually experience intelligibility improvements (around the SRT). Moreover, the binaural SDW-MWF retains its benefit over bilateral SDW-MWF algorithms, so that it is indeed a valid candidate for binaural noise reduction. It should however be noted that degradations can be expected at lower input SNR levels (i.e., beneath the SRT level of the hearing aid users). This is reflected by the results of the normal hearing subjects (who obtain their SRT at lower SNR levels), and was also demonstrated by the physical evaluation of Catic *et al.* (2010).

4. SDW-MWF versus fixed beamformer performance

A bilateral FB is tested here as competing noise reduction algorithm to the bilateral and binaural SDW-MWF. The FB has the advantage that it does not rely on a VAD and is not affected in nonstationary noise conditions so that it was not clear *a priori* whether the SDW-MWF could outperform the FB in these scenarios.

The FB obtains a significant SRT improvement in the S_0N_{45} scenario, in contrast to the bilateral SDW-MWF. For the $S_0N_{90/180/270}$, both algorithms achieve comparable SRT improvements. The FB outperforms the bilateral SDW-MWF for the normal hearing subjects in the $S_{90}N_{270}$ and cafeteria scenarios, where very low SNRs are encountered. For hearing aid users, the algorithms obtain a similar performance. The results indicate that in these scenarios no SRT benefit can be demonstrated for the bilateral SDW-MWF, compared to the FB. These results may thus confirm the findings of Woods *et al.* (2010), i.e., that adaptive beamformers rarely add significant benefits over fixed beamformers in realistic situations. It should, however, be noted that larger benefits are achievable by the bilateral SDW-MWF when using an implementation as in Luts *et al.* (2010), where a fixed preprocessing stage is included in the scheme.

The binaural SDW-MWF usually leads to larger SRT improvements than the bilateral FB, as is most notably observed for the hearing aid users in the cafeteria scenario (cf. the discussion in Secs. V A and V B 3). The SDW-MWF can thus be viewed as a good choice for binaural noise reduction. In contrast, designing a binaural FB may be less straightforward: even small changes in the microphone placement (e.g., due to improper placement of the devices) could have an impact on the performance, similarly to the sensitivity to model imperfections as explained in Sec. II C. Binaural FBs have only recently been introduced in the market [e.g., the so-called StereoZoom algorithm (Nyffeler, 2010)], and to the author's knowledge there is no published study which thoroughly investigates the robustness of these new techniques. A bilateral fixed beamformer, which is already well-established in the market and extensively studied in literature, was therefore chosen as competing algorithm to the SDW-MWF in this study.

We note that the evaluated bilateral FB is optimally calibrated and so does not suffer from the imperfections due to the wear and tear of daily use (Hamacher *et al.*, 2005). The results obtained with the FB should thus be viewed as a performance upper bound, while real-life performance can be significantly worse. For the same reason, some of the advantages of the SDW-MWF algorithms, which are not sensitive to these imperfections (Spriet *et al.*, 2005b), can thus not be fully appreciated by this evaluation. Finally, it is also noted that localization performance was not considered in this work. A binaural SDW-MWF can preserve the binaural cues of the target speech and the residual noise (Van den Bogaert *et al.*, 2008; Cornelis *et al.*, 2010), whereas a bilateral FB introduces distortions (Van den Bogaert *et al.*, 2005) if model imperfections occur. Also from this perspective, the full benefit of the binaural SDW-MWF can not be appreciated by this evaluation.

VI. CONCLUSION

In the work of Van den Bogaert *et al.* (2009), it was demonstrated that a binaural SDW-MWF achieves

Downloaded 29 Jun 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

significant SRT improvements over a bilateral SDW-MWF or adaptive directional microphone. The evaluated algorithms were, however, idealized in a number of ways. The bilateral and binaural SDW-MWF filter coefficients were calculated in an off-line batch procedure, where the complete microphone signals are used. Moreover, perfect knowledge of the target speech activity, which is necessary to estimate the second order statistics, was assumed to be available. Finally, it should also be noted that the stimuli were presented through headphones, and only normal hearing subjects participated in the evaluation.

In this paper, online adaptive bilateral and binaural SDW-MWF algorithms were evaluated. The performance of the SDW-MWF algorithms, a bilateral fixed beamformer, and a reference unprocessed condition were assessed by an adaptive SRT procedure for a total of four spatial scenarios, including a challenging cafeteria scenario with highly nonstationary noise (i.e., interfering speakers). The impact of using a real VAD algorithm instead of a perfect VAD was also assessed. Finally, basic hearing aid functionality was included in the processing so that hearing aid users could also be evaluated in addition to normal hearing subjects.

The conclusions are as follows. First, it was observed that the binaural SDW-MWF can further improve the SRT (up to 2 dB) over the improvements achieved by bilateral algorithms, although a significant difference is only achievable if the binaural SDW-MWF uses a perfect VAD. However, in the cafeteria scenario only the binaural SDW-MWF achieves a significant SRT improvement (2.6 dB with perfect VAD, 2.2 dB with real VAD), for the group of hearing aid users. Second, no significant performance degradations were observed when using a real VAD instead of a perfect VAD, at the SNR levels where the hearing aid users reached their SRT. Third, it was observed that the bilateral adaptive SDW-MWF did not perform better than the bilateral FB. It was however noted that no model imperfections (such as microphone mismatch) were present. If these occur, it is known that the FB performance can degrade, while the SDW-MWF is robust against such effects. Fourth, a significant algorithmic performance difference between the two subject groups was only observed for the cafeteria scenario. For the normal hearing subjects, who reached their SRT at relatively low SNR levels, both the SDW-MWF and the VAD performance were degraded so that the FB achieved the best improvement. For the hearing aid users, who reached their SRT at higher SNR levels, the SDW-MWF and VAD were not degraded. Significant improvements could only be demonstrated for the binaural SDW-MWF. This example has also illustrated that for some scenarios evaluating hearing aid users who are obviously the actual target subjects indeed has great value.

In future research, the described setup will also be used to evaluate the influence of the discussed noise reduction algorithms on the localization performance. It can be assessed whether the online adaptive implementation of the binaural SDW-MWF indeed preserves the speech binaural cues and possibly also the noise binaural cues (by adding a so-called partial noise estimation parameter), as was the case for the batch implementation (Van den Bogaert *et al.*, 2008). It is also expected that the adaptive bilateral SDW-MWF and bilateral FB can degrade the localization performance. For a FB, it was indeed demonstrated by a physical evaluation (Van den Bogaert *et al.*, 2005) that the binaural cues are distorted if model imperfections (such as microphone mismatch) occur. So, also from this perspective, the binaural SDW-MWF can offer a valuable approach to noise reduction in hearing aids.

ACKNOWLEDGMENTS

The authors would like to thank all participants of this study. We also thank Kristien Kneepkens who carried out the tests as part of her M.Sc. thesis. The support of Dr. Tim Van den Bogaert and Dr. Heleen Luts, which enabled the start-up of the project, is highly appreciated. Finally, we thank Sofie Jansen who helped in contacting test subjects. B.C. was funded by a Ph.D. grant from the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). This research work was carried out at the Lab ExpORL and the Department of Electrical Engineering (ESAT/SCD) of the KU Leuven in the frame of the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office IUAP P6/04 (DYSCO, "Dynamical systems, control and optimization," 2007–2011), Concerted Research Action GOA-MaNet and research project FWO No. G.0600.08 ("Signal processing and network design for wireless acoustic sensor networks"). The scientific responsibility is assumed by its authors.

ANSI-SII. (**1997**). S3.5, *American National Standard Methods for Calculation of the Speech Intelligibility Index* (Acoustical Society of America, Melville, NY).

Auditec. (**1997**). "Auditory tests (revised)," Compact Disc, Auditec, St. Louis.

Blauert, J. (**1996**). *Spatial Hearing: The Psychophysics of Human Sound Localisation*, revised ed. (MIT Press, Cambridge, MA), pp. 21–508.

Brennan, R., and Schneider, T. (**1998**). "A flexible filterbank structure for extensive signal manipulations in digital hearing aids," *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*, Vol. 6, pp. 569–572.

Byrne, D., Parkinson, A., and Newall, P. (**1991**). "Modified hearing aid selection procedures for severe/profound hearing losses," in *The Vanderbilt Hearing Aid Report II*, edited by G. Studebaker, F. Bess, and L. Beck (York, Parkton, NC), pp. 295–300.

Capon, J. (**1969**). "High-resolution frequency-wavenumber spectrum analysis," Proc. IEEE **57**, 1408–1418.

Catic, J., Dau, T., Buchholz, J. M., and Gran, F. (**2010**). "The effect of a voice activity detector on the speech enhancement performance of the binaural multichannel Wiener filter," EURASIP J. Audio, Speech, Music Process. **2010**, 1–12.

Chen, J., Benesty, J., Huang, Y., and Doclo, S. (**2006**). "New insights into the noise reduction Wiener filter," IEEE Trans. Audio, Speech, Lang. Process. **14**, 1218–1234.

Cornelis, B., Doclo, S., Van den Bogaert, T., Moonen, M., and Wouters, J. (**2010**). "Theoretical analysis of binaural multimicrophone noise reduction techniques," IEEE Trans. Audio, Speech, Lang. Process. **18**, 342–355.

Cornelis, B., Moonen, M., and Wouters, J. (**2011**). "Binaural voice activity detection for MWF-based noise reduction in binaural hearing aids," in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, Barcelona, Spain, pp. 486–490.

Crochiere, R. (**1980**). "A weighted overlap-add method of short-time Fourier analysis/synthesis," IEEE Trans. Acoust., Speech, Signal Process. **28**, 99–102.

Dillon, H. (**2001**). *Hearing Aids* (Boomerang, Sidney, Australia), pp. 1–504.

Downloaded 29 Jun 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

Doclo, S., Klasen, T. J., Van den Bogaert, T., Wouters, J., and Moonen, M. (**2006**). "Theoretical analysis of binaural cue preservation using multi-channel Wiener filtering and interaural transfer functions," in *Proceedings of the International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Paris, France, pp. 1–4.

Doclo, S., and Moonen, M. (**2002**). "GSVD-based optimal filtering for single and multimicrophone speech enhancement," IEEE Trans. Signal Process. **50**, 2230–2244.

Doclo, S., Spriet, A., Wouters, J., and Moonen, M. (**2007**). "Frequency-domain criterion for speech distortion weighted multichannel Wiener filter for robust noise reduction," Speech Commun. **49**, 636–656.

Elko, G. W., and Pong, A. N. (**1995**). "A simple adaptive first-order differential microphone," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 169–172.

Francart, T., van Wieringen, A., and Wouters, J. (**2008**). "APEX 3: A multi-purpose test platform for auditory psychophysical experiments," J. Neurosci. Methods **172**, 283–293.

Frost, O. L. (**1972**). "An algorithm for linearly constrained adaptive array processing," Proc. IEEE **60**, 926–935.

Griffiths, L. J., and Jim, C. W. (**1982**). "An alternative approach to linearly constrained adaptive beamforming," IEEE Trans. Antennas Propagat. **30**, 27–34.

Hamacher, V., Chalupper, J., Eggers, J., Fischer, E., Kornagel, U., Puder, H., and Rass, U. (**2005**). "Signal processing in high-end hearing aids: State of the art, challenges, and future trends," EURASIP J. Appl. Signal Process. 2915–2929.

Hu, Y., and Loizou, P. C. (**2007**). "A comparative intelligibility study of single-microphone noise reduction algorithms," J. Acoust. Soc. Am. **122**, 1777–1786.

ISO 3382. (**1997**). "Acoustics—Measurements of the reverberation time of rooms with reference to other acoustical parameters" (International Organization for Standardization, Geneva, Switzerland).

Klasen, T. J., Van den Bogaert, T., Moonen, M., and Wouters, J. (**2007**). "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," IEEE Trans. Signal Process. **55**, 1579–1585.

Loizou, P. C. (**2007**). *Speech Enhancement: Theory and Practice* (CRC Press, New York), pp. 1–632.

Loizou, P. C., and Kim, G. (**2011**). "Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions," IEEE Trans. Audio, Speech, Lang. Process. **19**, 47–56.

Luts, H., Eneman, K., Wouters, J., Schulte, M., Vormann, M., Buechler, M., Dillier, N., Houben, R., Dreschler, W. A., Froehlich, M., Puder, H., Grimm, G., Hohmann, V., Leijon, A., Lombard, A., Mauler, D., and Spriet, A. (**2010**). "Multicenter evaluation of signal enhancement algorithms for hearing aids," J. Acoust. Soc. Am. **127**, 1491–1505.

MathWorks. (**2000**). *xPC Target* (The Mathworks Inc., Natick, MA).

Nilsson, M., Soli, S. D., and Sullivan, A. (**1994**). "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085–1099.

Nyffeler, M., (**2010**). "StereoZoom: Improvements with directional microphones," Field Study News, http://www.phonakpro.com (Last viewed April 4, 2012).

Plomp, R. (**1978**). "Auditory handicap of hearing impairment and the limited benefit of hearing aids," J. Acoust. Soc. Am. **63**, 533–549.

Plomp, R., and Mimpen, A. M. (**1979**). "Improving the reliability of testing the speech reception threshold for sentences," Audiology **18**, 43–52.

Ricketts, T. (**2001**). "Directional hearing aids," Trends Amplif. **5**, 139–176.

Soede, W., Berkhout, A. J., and Bilsen, F. A. (**1993**). "Development of a directional hearing instrument based on array technology," J. Acoust. Soc. Am. **94**, 785–798.

Spriet, A., Moonen, M., and Wouters, J. (**2005a**). "The impact of speech detection errors on the noise reduction performance of multi-channel Wiener filtering and generalized sidelobe cancellation," Signal Process. **85**, 1073–1088.

Spriet, A., Moonen, M., and Wouters, J. (**2005b**). "Robustness analysis of multi-channel Wiener filtering and generalized sidelobe cancellation for multi-microphone noise reduction in hearing aid applications," IEEE Trans. Speech Audio Process. **13**, 487–503.

Spriet, A., Van Deun, L., Eftaxiadis, K., Laneau, J., Moonen, M., van Dijk, B., van Wieringen, A., and Wouters, J. (**2007**). "Speech understanding in background noise with the two-microphone adaptive beamformer BEAM in the nucleus freedom cochlear implant system," Ear Hear. **28**, 62–72.

Srinivasan, S. (**2008**). "Noise reduction in binaural hearing aids: Analyzing the benefit over monaural systems," J. Acoust. Soc. Am. **124**, EL353–EL359.

Stone, M. A., and Moore, B. C. J. (**1999**). "Tolerable hearing aid delays. I. Estimation of limits imposed by the auditory path alone using simulated hearing losses," Ear Hear. **20**, 182–192.

Trine, T. D., and Van Tasell, D. (**2002**). "Digital hearing aid design: Fact vs. fantasy," Hear. J. **55**, 36–38, 40–42.

Van den Bogaert, T., Doclo, S., Moonen, M., and Wouters, J. (**2008**). "The effect of multi-microphone noise reduction systems on sound source localization in binaural hearing aids," J. Acoust. Soc. Am. **124**, 484–497.

Van den Bogaert, T., Doclo, S., Moonen, M., and Wouters, J. (**2009**). "Speech enhancement with multichannel Wiener filter techniques in multi-microphone binaural hearing aids," J. Acoust. Soc. Am. **125**, 360–371.

Van den Bogaert, T., Wouters, J., Klasen, T. J., and Moonen, M. (**2005**). "Distortion of interaural time cues by directional noise reduction systems in modern digital hearing aids," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, pp. 57–60.

Van Veen, B. D., and Buckley, K. M. (**1988**). "Beamforming: A versatile approach to spatial filtering," IEEE ASSP Mag. **5**, 4–24..

Vanden Berghe, J., and Wouters, J. (**1998**). "An adaptive noise canceller for hearing aids using two nearby microphones," J. Acoust. Soc. Am. **103**, 3621–3626.

Versfeld, N. J., Daalder, L., Festen, J. M., and Houtgast, T. (**2000**). "Method for the selection of sentence materials for efficient measurement of the speech reception threshold," J. Acoust. Soc. Am. **107**, 1671–1684.

Woods, W. S., Merks, I., Zhang, T., Fitz, K., and Edwards, B. (**2010**). "Assessing the benefit of adaptive null-steering using real-world signals," Int. J. Audiol. **49**, 434–443.

Zurek, P. (**1992**). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance*, 2nd ed. (Allyn and Bacon, Boston, MA), Chap. 15, pp. 255–276.

Downloaded 29 Jun 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp