The effect of multimicrophone noise reduction systems on sound source localization by users of binaural hearing aids

Tim Van den Bogaert[a)]
ExpORL, K.U.Leuven, O & N 2-Herestraat 49 bus 721, B-3000 Leuven, Belgium

Simon Doclo[b)]
ESAT-SCD, K.U.Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium

Jan Wouters
ExpORL, K.U.Leuven, O & N 2-Herestraat 49 bus 721, B-3000 Leuven, Belgium

Marc Moonen
ESAT-SCD, K.U.Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium

(Received 27 July 2007; revised 22 April 2008; accepted 24 April 2008)

This paper evaluates the influence of three multimicrophone noise reduction algorithms on the ability to localize sound sources. Two recently developed noise reduction techniques for binaural hearing aids were evaluated, namely, the binaural multichannel Wiener filter (MWF) and the binaural multichannel Wiener filter with partial noise estimate (MWF-N), together with a dual-monaural adaptive directional microphone (ADM), which is a widely used noise reduction approach in commercial hearing aids. The influence of the different algorithms on perceived sound source localization and their noise reduction performance was evaluated. It is shown that noise reduction algorithms can have a large influence on localization and that (a) the ADM only preserves localization in the forward direction over azimuths where limited or no noise reduction is obtained; (b) the MWF preserves localization of the target speech component but may distort localization of the noise component. The latter is dependent on signal-to-noise ratio and masking effects; (c) the MWF-N enables correct localization of both the speech and the noise components; (d) the statistical Wiener filter approach introduces a better combination of sound source localization and noise reduction performance than the ADM approach.
[DOI: 10.1121/1.2931962]

PACS number(s): 43.66.Ts, 43.66.Pn, 43.60.Fg, 43.66.Qp [BCM] Pages: 484–497

I. INTRODUCTION

Noise reduction algorithms in hearing aids are important for hearing-impaired persons to improve speech intelligibility in background noise. Multimicrophone noise reduction systems are able to exploit spatial in addition to spectral information and are hence typically preferred to single-microphone systems (Welker *et al.*, 1997; Lotter, 2004). However, the multimicrophone, typically adaptive, noise reduction algorithms currently used in hearing aids are designed to optimize the signal-to-noise ratio (SNR) in a monaural way and not to preserve binaural or interaural cues. Therefore, hearing aid users often localize sounds better when switching off the adaptive directional noise reduction in their hearing aids (Keidser *et al.*, 2006; Van den Bogaert *et al.*, 2006). This puts the hearing aid user at a disadvantage. In certain situations, such as traffic, incorrect localization of sounds may even endanger the user. In addition, interaural localization cues and spatial awareness are important for speech segregation in noisy environments due to spatial release from masking (Bronkhorst and Plomp, 1988; 1989). Changing from a bilateral, i.e., a dual-monaural, hearing aid configuration, to a binaural noise reduction algorithm, i.e., generating an output signal for both ears by using all available microphone signals, may enhance the amount of noise reduction and may increase the ability to control the adaptive processes to preserve the interaural cues between left and right hearing aids. An important limitation of most noise reduction array systems studied thus far is that they are designed to produce a single, i.e., a monaural, output. Extending these to a binaural output is not trivial.

Recently, several techniques to combine binaural noise reduction and preservation of spatial awareness have been studied. The first class of techniques is based on computational auditory scene analysis. Wittkop and Hohman (2003) proposed a method in which the incoming signal is split into different frequency bands. The estimated binaural properties, e.g., the coherence, of each frequency band are compared to the expected properties of the signal component (typically it is assumed that the signal component arrives from the frontal area with interaural time differences (ITDs) and interaural level differences (ILDs) close to 0 μs and 0 dB). This comparison determines whether these frequencies should be enhanced or attenuated. By applying identical gains at the left and the right hearing aids, interaural cues are preserved.

[a)]Electronic mail: tim.vandenbogaert@med.kuleuven.be
[b)]Presently at NXP Semiconductors, Interleuvenlaan 80 (C034), 3001 Leuven, Belgium.

 0001-4966/2008/124(1)/484/14/$23.00

However, the noise reduction performance of these methods is relatively limited and typically spectral enhancement problems such as "musical noise" occur.

The second class of techniques is based on fixed or adaptive beamforming. In the studies of Desloge *et al.* (1997), Welker *et al.* (1997), and Zurek and Greenberg (2000), fixed and adaptive multimicrophone beamforming systems were studied, designed to optimize their directional response and to faithfully preserve the interaural cues. In Desloge *et al.* (1997), six different fixed beamforming systems were tested and compared to a reference system which consisted of two independent cardioid microphones. Two of these systems used all microphone inputs from both hearing aids to calculate the output. The first system was a fixed processing scheme designed to limit the amount of ITD distortion at the output to 40 μs. The second system used a low/high pass filtering system and performed a fixed noise reduction on the higher frequencies ($f>800$ Hz) of the signal. The frequency band below 800 Hz remained unprocessed. This approach is inspired by the observation that the ITD information, which is mainly useful at low frequencies, is a dominant localization cue compared to the ILD information, present at the higher frequencies (Wightman and Kistler, 1992). Tests were performed with speech arriving from the front in a diffuse noise source scenario. Both systems showed a significant SNR gain of 2.7–4.4 dB in comparison to the reference system. In general, both systems provided the subjects with moderate localization capabilities using a test setup with a resolution of 30°.

In Welker *et al.* (1997), the low/high pass scheme described above was used in an adaptive noise reduction algorithm with two microphones, one at each ear. The high-frequency part ($f>f_c$) of the signal was now processed in an adaptive way. The algorithm was evaluated by normal hearing subjects. It was shown that f_c determined a trade-off between noise reduction and localization performance. An optimal setting of $f_c=500$ Hz was proposed which led to an effective noise reduction of 3 dB and a localization accuracy of 70%. Tests of Zurek and Greenberg (2000), with hearing-impaired subjects and $f_c=1000$ Hz, showed a SNR improvement of 2 dB when using the same algorithm.

The third class of techniques are based on blind source separation (BSS). Very recently, Aichner *et al.* (2007) proposed two methods for incorporating interaural cue preservation in BSS. The first method is based on using adaptive filters as a postprocessing stage after BSS. These filters remove the noise components, estimated by the BSS, from the reference microphone. By doing this at both sides of the head, the interaural cues of the speech component are preserved. Due to the fact that not all noise can be removed from the reference signal, it was claimed that the interaural cues of the remaining noise component are also preserved. The second method is based on constraining the BSS filters themselves, thereby avoiding distortion of the separated signals produced by the BSS. However, localization results were described very briefly using a quality rating on the output of the algorithm, and so far no results have been published on the source separation performance of these methods.

The last class, on which this paper will focus, is based on multichannel Wiener filtering (MWF). Recently, Doclo and Moonen (2002) mathematically described a MWF approach performing noise reduction in hearing aids. This approach, unlike an adaptive directional microphone (ADM), is based on using second-order statistics of the speech and the noise components to estimate the speech component in a noisy (reference) microphone signal. In Doclo *et al.* (2006), it was mathematically proven that a binaural version of the MWF generates filters which, in theory, perfectly preserve the interaural cues of the speech component but change the interaural cues of the noise component into those of the speech component. To optimally benefit from spatial release from masking and to optimize spatial awareness of the hearing aid user, it would be beneficial to also preserve the interaural cues of the noise component. Hence, two extensions of the MWF have been proposed. In the first extension, proposed by Klasen *et al.* (2006), an estimate of the interaural transfer function (ITF) was introduced into the cost function which was used to calculate the Wiener filters. This enabled putting more or less emphasis on preserving the interaural cues at the cost of some loss of noise reduction. However, if the ITF extension is emphasized too strongly, the interaural cues of the speech component will be distorted into those of the noise component. A perceptual validation of the MWF-ITF by Van den Bogaert *et al.* (2007) in a low reverberant environment showed that an optimal parameter setting could be found which improved localization performance compared to a binaural MWF without a large loss in noise reduction performance. However, this ITF extension is only valid for single noise source scenarios. The second extension is a MWF with partial noise estimate (MWF-N), first described by Klasen *et al.* (2007), which aims at eliminating only part of the noise component. The remaining, unprocessed, part of the noise signal then restores the spatial cues of the noise component of the signal at the output of the algorithm. This is similar to the work of Noble *et al.* (1998) and Byrne *et al.* (1998), in which improvements in localization were found when using open instead of closed earmolds. The open earmolds enabled the usage of the direct, unprocessed, sound at frequencies with low hearing loss to improve localization performance. In Klasen *et al.* (2007), the MWF and MWF-N approaches were compared to the approach of Welker *et al.* (1997), described earlier. This was done using objective performance measures based on anechoic data for a single noise source, fixed at 90°. To quantify localization performance, an ITD-error measure was defined, being the difference in ITD between the input and the output of the algorithms. ITD was calculated as the delay generating the maximum value in the cross correlation between the left and right ear signals. A maximum noise reduction of 27 dB was obtained and simulations showed that the ITD error of the speech component was close to zero for the MWF and the MWF-N. It was also shown that for the MWF, the ITD error of the noise component could exceed 500 μs. For the MWF-N, this error dropped below 50 μs. The work of Klasen *et al.* (2007) summarized the possible benefits and trade-offs of the MWF and the MWF-N compared to the approach of Welker *et al.* (1997). However, it remains hard to predict real-life performance, since an anechoic environment was used and since the ITD error measure, used to predict localization performance, is based on a very simple localization model.

The main purpose of this paper was to study the effect of noise reduction algorithms on the ability to localize sound sources when hearing aid users wear a hearing aid at both sides of the head. It evaluates two recently described binaural noise reduction algorithms, namely, the MWF and the MWF-N, as well as a widely used noise reduction approach, namely, an ADM. An unprocessed condition was used as a reference. The evaluation was performed in a room with a realistic reverberation time ($T60=0.61$ s) at two different SNRs, mainly using perceptual evaluations with normal hearing subjects. The focus of the manuscript is on localization performance in the horizontal plane, for which the ITD and ILD are the main cues (Hartmann, 1999; Blauert, 1997). Since the manuscript evaluates noise reduction algorithms, noise reduction data will also be presented.

The main research questions answered in this study are the following. (a) What is the influence of a commonly used noise reduction algorithm, namely, an ADM in a dual-monaural hearing aid configuration on the ability to localize sound sources in a realistic environment? (b) What is the influence of the binaural MWF in a binaural hearing aid configuration on the ability to localize sound sources? (c) Does the MWF-N improve localization performance in comparison to the MWF? (d) How do the MWF and MWF-N perform in terms of combining noise reduction and localization performance in comparison to the ADM configuration?

II. ALGORITHMS

A. ADM

An ADM is a commonly used noise reduction technique for hearing aids (Luo *et al.*, 2002; Maj *et al.*, 2004). Unlike the MWF-based algorithms, the ADM is based on the assumption that the target signal arrives from the frontal direction and that jammer signals arrive from the back hemisphere. The ADM uses the physical differences in time of arrival between the microphones to improve the SNR by steering a null in the direction of the jammer signals. The ADM uses the microphones of one hearing aid at a given ear and consists of two stages. The first stage generates two software directional microphone signals corresponding to front- and back-oriented cardioid patterns. In the second stage, these signals are combined by an adaptive, frequency dependent, scalar β that minimizes the energy arriving from the back hemisphere at the output of the algorithm. Typically, the value of β is constrained between 0 and 0.5 to avoid distortion in the frontal hemisphere.

B. Binaural MWF

In general, the goal of a Wiener filter is to filter out noise corrupting a desired signal. Using the second-order statistical properties of the desired signal and the noise, the optimal filter or Wiener filter can be calculated. It generates an output signal which approaches the desired signal as closely as possible in a mean-square error (MSE) sense. It is based on minimizing a cost function corresponding to the difference between the desired signal (the speech component which has to be estimated) and the output of the filter. In contrast with a single channel approach, a MWF uses multiple input signals to compute a set of filters generating this output signal. See Haykin (2002) for an overview on Wiener filtering.

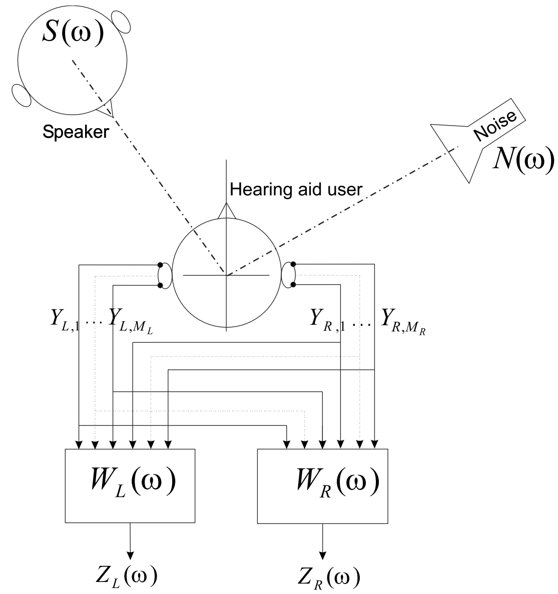

FIG. 1. Layout of a binaural noise reduction system.

Consider the binaural hearing aid configuration in Fig. 1, where the left and the right hearing aids have a microphone array consisting of, respectively, M_L and M_R microphones. The mth microphone signal $Y_{L,m}(\omega)$ of the left ear can be written in the frequency domain as

$$Y_{L,m}(\omega) = X_{L,m}(\omega) + V_{L,m}(\omega), \quad m = 1 \cdots M_L, \tag{1}$$

where $X_{L,m}(\omega)$ and $V_{L,m}(\omega)$ represent the speech and the noise components at the mth microphone input of the left hearing aid. $Y_{R,m}$, $X_{R,m}(\omega)$, and $V_{R,m}(\omega)$ are defined similarly for the right hearing aid. Assuming a link between the two hearing aids, microphone signals from a given ear (M_I) and contralateral ear (M_C) can be used to generate an output signal for each of the two hearing aids. The total number of microphones used at each ear is defined as $M=M_I+M_C$.[1] For the left and right ears, the M-dimensional input signal vectors $\mathbf{Y}_L$ and $\mathbf{Y}_R$ can be written as

$$\mathbf{Y}_L(\omega) = [Y_{L,1}(\omega), \ldots, Y_{L,M_I}(\omega), \quad Y_{R,1}(\omega), \ldots, Y_{R,M_C}(\omega)]^T, \tag{2}$$

$$\mathbf{Y}_R(\omega) = [Y_{L,1}(\omega), \ldots, Y_{L,M_C}(\omega), \quad Y_{R,1}(\omega), \ldots, Y_{R,M_I}(\omega)]^T, \tag{3}$$

with ^{T} the transpose operator. The vectors defining the speech component and the noise component, e.g., for the left ear $\mathbf{X}_L(\omega)$ and $\mathbf{V}_L(\omega)$, are defined in a similar way to the signal vectors. The filters which combine the microphone signals to optimally estimate the speech component are calculated using a Wiener filter procedure and are defined as $\mathbf{W}_L(\omega)$ and $\mathbf{W}_R(\omega)$ for the left and the right hearing aids, respectively. The output signals for the left and the right ears are equal to

$$\mathbf{Z}_L(\omega) = \mathbf{W}_L^H(\omega)\mathbf{Y}_L(\omega), \quad \mathbf{Z}_R(\omega) = \mathbf{W}_R^H(\omega)\mathbf{Y}_R(\omega), \tag{4}$$

with $\mathbf{W}_L(\omega)$ and $\mathbf{W}_R(\omega)$ M-dimensional complex vectors and ^{H} the Hermitian transpose operator. The $2M$-dimensional stacked weight vector $\mathbf{W}(\omega)$ is defined as

$$\mathbf{W}(\omega) = \begin{bmatrix} \mathbf{W}_L(\omega) \\ \mathbf{W}_R(\omega) \end{bmatrix}. \tag{5}$$

For conciseness, we will omit the frequency-domain variable ω in the remainder of the paper.

The binaural MWF produces a minimum MSE estimate of the speech component for each hearing aid. The MSE cost function J_{MSE} which should be minimized to calculate the filters $\mathbf{W}_L$ estimating the unknown speech component in the front microphone of the left hearing aid, i.e., $X_{L,1}$ from Eq. (1), and the filters $\mathbf{W}_R$ estimating the unknown speech component in the front microphone of the right hearing aid, i.e., $X_{R,1}$, equals

$$J_{\mathrm{MSE}}(\mathbf{W}) = \mathcal{E}\left\{\left\|\begin{bmatrix} X_{L,1} - \mathbf{W}_L^H\mathbf{Y}_L \\ X_{R,1} - \mathbf{W}_R^H\mathbf{Y}_R \end{bmatrix}\right\|^2\right\}, \tag{6}$$

with $\mathcal{E}$ the expected value operator. Minimizing $J_{\mathrm{MSE}}(\mathbf{W})$ leads to the optimal filters $\mathbf{W}$ producing the best minimum MSE estimate of the speech component $\mathbf{X}$ present in the reference microphones.

This cost function was, for a monaural hearing aid configuration, extended by Doclo and Moonen (2002) and Spriet *et al.* (2004) by using Eq. (1) and introducing an extra trade-off parameter μ. To enable a trade-off between speech distortion and noise reduction, they introduced the monaural speech distortion weighted MWF (SDW-MWF), which minimizes the weighted sum of the residual noise energy and the speech distortion energy. The binaural SDW-MWF cost function equals

$$J_{\mathrm{MWF}}(\mathbf{W}) = \mathcal{E}\left\{\left\|\begin{bmatrix} X_{L,1} - \mathbf{W}_L^H\mathbf{X}_L \\ X_{R,1} - \mathbf{W}_R^H\mathbf{X}_R \end{bmatrix}\right\|^2 + \mu\left\|\begin{bmatrix} \mathbf{W}_L^H\mathbf{V}_L \\ \mathbf{W}_R^H\mathbf{V}_R \end{bmatrix}\right\|^2\right\}, \tag{7}$$

where the first term represents speech distortion and the second term represents the residual noise. Note that when the trade-off parameter μ is set to 1, the SDW-MWF cost function (7) reduces to cost function (6). In the remainder of the paper the SDW-MWF algorithm will be used and evaluated. For conciseness the SDW-MWF algorithm will be referred to as MWF.

The Wiener filter solution minimizing the cost function $J_{\mathrm{MWF}}(\mathbf{W})$ equals

$$\mathbf{W}_{\mathrm{MWF}} = \begin{bmatrix} \mathbf{R}_{x,L} + \mu\mathbf{R}_{v,L} & 0_M \\ 0_M & \mathbf{R}_{x,R} + \mu\mathbf{R}_{v,R} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{R}_{x,L}e_L \\ \mathbf{R}_{x,R}e_R \end{bmatrix}, \tag{8}$$

with e_L and e_R being vectors with one element equal to 1 and the other elements equal to zero, defining the reference microphones used at both hearing aids, i.e., in the case of the front omnidirectional microphone $e_L(1)=1$ and $e_R(1)=1$. $\mathbf{R}_x$ and $\mathbf{R}_v$, which are at present still unknown, are defined as the $M \times M$-dimensional speech and noise correlation matrices, containing the autocorrelations and cross correlations (or the statistical information) of, respectively, the speech and noise components $\mathbf{X}$ and $\mathbf{V}$ over the different input channels, e.g., $\mathbf{R}_{x,L} = \mathcal{E}\{\mathbf{X}_L\mathbf{X}_L^H\}$. To find $\mathbf{W}_{\mathrm{MWF}}$ using Eq. (8), a voice activity detector is used to discriminate between "speech and noise periods" and "noise only periods." The noise correlation matrix $\mathbf{R}_v$ can be calculated during the noise only periods. By assuming a sufficient stationary noise signal, the speech correlation matrix $\mathbf{R}_x$ can be estimated during speech and noise periods by subtracting $\mathbf{R}_v$ from the correlation matrix $\mathbf{R}_y$ for the noisy signal Y. By using these correlation matrices, the filters $\mathbf{W}$ can be found [see Eq. (8)].

Since the binaural MWF is designed to produce two outputs, $Z_L(\omega)$ and $Z_R(\omega)$, respectively, estimating the speech components at the front omnidirectional microphones of the left and the right hearing aids, the interaural cues of the speech component are inherently preserved.

C. Binaural MWF-N

The rationale of the MWF-N is not to completely remove the noise component from the microphone signals but to remove only part of the noise component. The interaural cues of the unprocessed part can then be used to correctly localize the noise component. The MWF-N corresponds to estimating the desired speech component summed with a scaled version of the noise component (Klasen *et al.*, 2007). Consequently, Eq. (6) changes into

$$J_{\mathrm{MSE}\,\eta}(\mathbf{W}) = \mathcal{E}\left\{\left\|\begin{bmatrix} X_{L,1} + \eta V_{L,1} - \mathbf{W}_L^H\mathbf{Y}_L \\ X_{R,1} + \eta V_{R,1} - \mathbf{W}_R^H\mathbf{Y}_R \end{bmatrix}\right\|^2\right\}, \tag{9}$$

with η between 0 and 1. By using a small η more emphasis is put on noise reduction and less emphasis is put on preserving the interaural cues of the noise component. When $\eta=0$, the MWF-N reduces to the standard MWF. Similar to the MWF, a trade-off parameter can be introduced by weighting the amount of speech distortion with the residual noise energy in the partial noise estimate. In other words, the amount of speech distortion is limited at the cost of noise reduction on part $(1-\eta)$ of the noise signal. The cost function then becomes

$$J_{\mathrm{MWF}\,\eta}(\mathbf{W}) = \mathcal{E}\left\{\left\|\begin{bmatrix} X_{L,1} - \mathbf{W}_L^H\mathbf{X}_L \\ X_{R,1} - \mathbf{W}_R^H\mathbf{X}_R \end{bmatrix}\right\|^2 + \eta\left\|\begin{bmatrix} \eta V_{L,1} - \mathbf{W}_L^H\mathbf{V}_L \\ \eta V_{R,1} - \mathbf{W}_R^H\mathbf{V}_R \end{bmatrix}\right\|^2\right\}. \tag{10}$$

A simple relationship holds between the filter output of the MWF and the MWF-N,

$$Z_{\mathrm{MWF}_\eta L}(\eta,\mu) = \eta Y_{L,1} + (1-\eta)Z_{\mathrm{MWF},L}(\mu), \tag{11}$$

$$Z_{\mathrm{MWF}_\eta R}(\eta,\mu) = \eta Y_{R,1} + (1-\eta)Z_{\mathrm{MWF},R}(\mu). \tag{12}$$

In other words, the MWF-N solution is obtained by adding a portion of the unprocessed signals of the reference microphones (ηY) to the original MWF solution. This can be used to restore the spatial cues of the noise component in the processed signal. A similar theory is demonstrated in the work of Noble *et al.* (1998) and Byrne *et al.* (1998), in which localization performance was improved by using open instead of closed earmolds by dual-monaural hearing aid users.

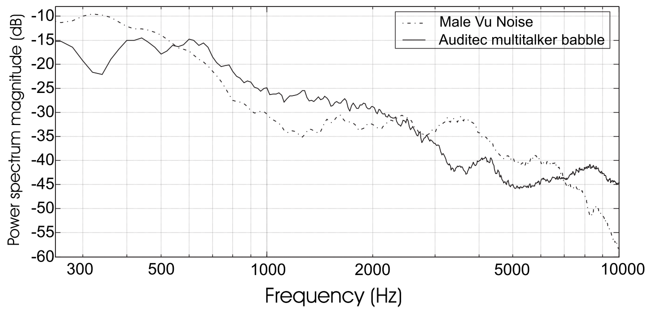

FIG. 2. Average power spectrum of the speech weighted noise signal (VU material) and the multitalker babble (Auditec). The overall SNR was 0 dB A.

Obviously, it is expected that noise reduction performance will decrease when increasing η.

III. LOCALIZATION PERFORMANCE

A. Test setup

Experiments were carried out in a reverberant room with dimensions $5.20\times 4.50\times 3.10$ m^3 (length $\times$ width $\times$ height) and a reverberation time $T60$, averaged over one/third octave frequencies from 100 to 8000 Hz, of 0.61 ± 0.08 s. Subjects were located at 1.90 m from the right wall and 2.05 m from the front wall. Stimuli were generated off line (see Sec. III B) and presented through headphones (Sennheiser HD650) using an RME Hamerfall DSPII soundcard. Subjects were placed inside an array of 13 Fostex 6301B single-cone speakers. The speakers were located in the frontal horizontal plane at angles ranging from $-90°$ to $+90°$ relative to the subject with a spacing of 15°. The speakers were placed at a distance of 1 m from the subject and were labeled 1–13. Since the stimuli were presented through headphones, loudspeakers were used only for visualization purposes. The task was to identify the loudspeaker where the target sound was heard.

B. Stimuli

The algorithms were evaluated using a steady speech weighted noise signal from the VU test material (Versfeld *et al.*, 2000) arriving from angle $x°$ as speech component (S). A multitalker babble (Auditec) was used as the jammer sound (N) arriving from angle $y°$, defining the spatial scenario S_xN_y. The spectra of the speech and noise source are depicted in Fig. 2. Three different spatial scenarios were evaluated: S_0N_{60}, $S_{90}N_{-90}$, and $S_{45}N_{-45}$.

To generate the input signals for all algorithms, stimuli were convolved with the appropriate impulse responses measured between the loudspeakers of the loudspeaker array and the microphones on two behind the ear (BTE) hearing aids worn by a Cortex MK2 manikin. The manikin was placed at the position of the test subjects. The BTE devices were two dual-microphone shells with direct microphone outputs from two omnidirectional microphones on each hearing aid. The intermicrophone distance was approximately 1 cm.

Three different noise reduction algorithms were evaluated. The first two algorithms were the binaural MWF with partial noise estimate using $\eta=0.2$ (MWF-N$_{0.2}$) and the standard binaural MWF, which corresponds to the MWF-N with $\eta=0$. Both of these algorithms were implemented using for each ear two omnidirectional microphones present at that ear and the front microphone of the contralateral hearing aid to generate an output for the given hearing aid. Simulations suggested that $\mu=5$ was an appropriate value for the trade-off parameter in Eqs. (7) and (10). The third algorithm was an ADM. The ADM configuration is a commonly used dual-monaural configuration which used for each ear both microphone signals of the given hearing aid to generate the output signal for that particular hearing aid. When testing performance in the unprocessed condition (unproc), the front omnidirectional microphone signals from the left and right hearing aids were presented to the subject. The outputs of all algorithms were calculated off line. ADM and MWF filters were trained on the specific spatial scenario and were fixed after convergence. For the MWF, a perfect voice activity detector was used to calculate the filters. Pilot testing suggested that the MWF filters behaved differently at different SNRs. Therefore, stimuli were generated at two different input SNRs (0 and -12 dB A), with the input SNR being calculated in the absence of the head.

C. Protocol

In the first test condition (S,N), the speech and the noise components were filtered by the fixed filters and presented separately to the subjects. By presenting the two components separately, interactions between components were avoided (masking effects, localizing two sounds is different from localizing one sound source). In the second condition (S+N) the speech and noise components were presented simultaneously and the subject was asked to localize both components. This resembled a steady-state real-life situation.

Subjects were instructed to keep their head fixed and pointed toward the 0° direction during stimulus playback and were supervised by the test leader. The task was to identify the loudspeaker where the target sound was perceived. Although only the locations of $-90°$, $-45°$, 0°, 45°, 60°, and 90° were used to generate the stimuli, subjects were free to use all given loudspeaker positions in the frontal horizontal hemisphere ($-90°$ to $+90°$ in steps of 15°) to identify where the sound was perceived. Tests were restricted to the frontal hemisphere to avoid front-back confusions which would complicate the analysis of the results and which are more related to spectral cues than to interaural cues. None of the subjects experienced major problems with this restriction. Subjects were clearly instructed that the test could be unbalanced. The five subjects were all normal hearing subjects working in the Department of Exp.ORL and were used to performing listening tests.

Pilot testing showed that the presented stimuli might sound diffuse or even arriving from two different angles instead of one clear direction. Therefore, subjects were asked to give comments on how the sound was perceived using the following classification: the sound arrives from a point source with one clear direction in space (point), the sound arrives from a wider area (wide), the sound arrives from everywhere (diffuse), or more than one sound source is perceived (dual). If they perceived multiple components at different locations, subjects were asked to report both locations and to report to which direction they would look when hearing this stimulus. This direction was then used as the response to the presented stimulus. Only for the condition S+N were the subjects explicitly asked to report two angles of arrival, one for the speech and one for the noise component.

The two different sound conditions (S,N and S+N) were presented in different test sessions with the angle of arrival, input SNR, and type of algorithm randomized throughout the test. Each stimulus was repeated three times, and an overall roving level of 6 dB was used (ranging from 0 to −6 dB). The presented stimuli were equalized in dB A level by adjusting the sound level, averaged over the left and right channels, to the same level for all generated stimuli. The stimuli were then presented at a comfortable level chosen by the subject. Because the task was quite hard, the subject had the possibility to repeat the same stimulus over and over again until a clear answer could be given to the test leader, who entered all responses and comments. The test leader had no information on the location of the stimulus nor the type of algorithm that was used and no feedback was given to the subjects. Typically one session took somewhat more than 1 h and several hours elapsed between different sessions. If fatigue or low concentration were observed, breaks were taken during the test.

D. Performance measures

Different error measures have been used in previous localization studies (Noble and Byrne, 1990; Lorenzi *et al.*, 1999; Van Hoesel *et al.*, 2002). Two commonly used error measures are the root-mean-square (rms) error and the mean absolute error (MAE). We focused on the MAE which is defined as

$$\mathrm{MAE}(^\circ) = \frac{\sum_{i=1}^{n} |(\text{stimulus azimuth} - \text{response azimuth})|}{n}, \tag{13}$$

with n the number of presented stimuli. For the MAE, all errors are weighted equally, while for the rms error, large errors have a larger impact than small errors. The smallest nonzero error a subject could make for one stimulus equaled 5° MAE (one error of 15° made during the three repetitions of the stimulus, $n=3$). In Sec. III E, the statistical analysis will show that this resolution was sufficient to illustrate effects of, and large differences between, the algorithms in the different spatial scenarios, which was the goal of this study.

E. Results and analysis

First the data and analysis for the condition S,N are presented, followed by the data and analysis for the condition S+N. All statistical analysis was done using SPSS 15.0. For conciseness, the term factorial repeated-measure ANOVA is abbreviated by ANOVA and pairwise comparisons discussed throughout the document were always Bonferroni corrected for multiple comparisons.

1. Condition S,N

Localization data for the condition with the speech and the noise component presented separately to the listener are given in Table I. Table I indicates where the stimulus was perceived by each subject, averaged over the three stimulus repetitions, together with the minimum, maximum, and averaged MAE values across subjects.

To compare the different algorithms, an ANOVA was carried out on the recorded MAE data. The factor algorithms (ADM, MWF, MWF-$N_{0.2}$), target (speech or noise component), SNRs (0 and −12 dB), and angles (S_0N_{60}, $S_{90}N_{-90}$, $S_{45}N_{-45}$) were used. As expected, many interactions were found between these factors, e.g., algorithm*target $p=0.004$. To disentangle these interactions, separate ANOVAs were carried out for the speech and noise components.

Speech component. An interaction was found between the factor angle and algorithm ($p=0.019$, $F=14.647$). Hence, separate ANOVAs were carried out for each spatial scenario.

For S_0N_{60} and $S_{45}N_{-45}$ no main effects were found ($p=0.470$ and $p=1.000$, respectively for the factor algorithm). For the scenario $S_{90}N_{-90}$, a main effect of the factor algorithm was found ($p=0.009$, $F=22.359$). Pairwise comparisons showed significantly lower performance for the ADM than for the MWF (difference averaged over the two SNRs =58° MAE, $p=0.039$) and the MWF-$N_{0.2}$ (difference averaged over the two SNRs=65° MAE, $p=0.019$). Table I shows that, for scenario $S_{90}N_{-90}$, none of the subjects was capable of localizing the speech component correctly when using the ADM, and sounds were most commonly localized around 0° (four out of five subjects). The MWF-$N_{0.2}$ scheme just failed to give significantly better performance than the MWF scheme (difference of 7° MAE, $p=0.057$).

When comparing the algorithms with the unprocessed condition, no main effects were found for scenarios S_0N_{60} and $S_{45}N_{-45}$ ($p>0.252$). For the scenario $S_{90}N_{-90}$, a main effect was found ($p=0.008$). Pairwise comparisons showed that only the ADM performed significantly more poorly than the unprocessed condition (a difference of 67° MAE, $p=0.038$, for SNR=0 dB and a difference of 65° MAE, $p=0.035$, for SNR=−12 dB).

Table II shows the percentage of reports of a clear directional sound image during the subjective classification of the stimuli. For the speech component, the combination of ADM and $S_{90}N_{-90}$ led to severely degraded performance compared to all other combinations. Interestingly, these stimuli were often perceived as being diffuse (53% for 0 dB and 60% for −12 dB). Subjects reported that, when perceiving a diffuse sound, 0° was often picked as the direction from where the sound was heard, since it is the neutral position in the middle of the sound array. Therefore, these 0° responses should be interpreted carefully.

Noise component. Due to an interaction with SNR ($p=0.050$), separate ANOVAs were carried out for each SNR. Since the speech and noise components were presented separately and since the presentation level for both components was calibrated to a comfortable level, the obtained results for the unprocessed stimuli are independent of SNR. Therefore, the data for the unprocessed condition were incorporated in the ANOVA for each SNR.

TABLE I. Response location (deg), averaged over three repetitions, together with the average, minimum, and maximum MAE across subjects for the three different spatial scenarios (S_0N_{60}, $S_{90}N_{-90}$, and $S_{45}N_{-45}$), and the different processing schemes (unprocessed, ADM, MWF, and MWF-$N_{0.2}$) at two different SNRs (0 and −12 dB). The speech and the noise sources were presented separately through headphones (S,N). The rows labeled "effect" show whether a significant difference from the unprocessed condition was found. *P*-values of pairwise comparisons are shown. If no main effects were found the term "nm" is used.

S_0N_{60}	S_0 unproc	ADM		MWF		MWF-$N_{0.2}$		N_{60} unproc	ADM		MWF		MWF-$N_{0.2}$	
		SNR0	SNR-12	SNR0	SNR-12	SNR0	SNR-12		SNR0	SNR-12	SNR0	SNR-12	SNR0	SNR-12
T	0	0	−5	−15	0	−15	0	90	90	75	0	50	0	85
J	0	0	0	0	0	0	0	90	90	90	0	90	80	90
H	0	0	0	−5	0	0	0	65	60	70	0	45	35	65
L	−5	−5	−5	−15	−10	−15	−10	90	80	85	−5	45	75	85
O	0	20	25	5	35	−0	0	85	80	90	10	80	65	80
Loc (av)	−1	3	3	−6	5	−8	−2	84	80	82	1	62	51	81
MAE (av)	1	5	7	8	9	8	2	24	20	22	59	28	25	21
Min-max MAE	0-5	0-20	0-25	0-15	0-35	0-15	0-10	5-30	0-30	10-30	50-65	20-35	5-60	5-30
Effect		nm	nm	nm	nm	nm	nm		$p=0.687$	nm	$p=0.027$	nm	$p=1.000$	nm
$S_{90}N_{-90}$	S_{90} unproc	ADM		MWF		MWF-$N_{0.2}$		N_{-90} unproc	ADM		MWF		MWF-$N_{0.2}$	
		SNR0	SNR-12	SNR0	SNR-12	SNR0	SNR-12		SNR0	SNR-12	SNR0	SNR-12	SNR0	SNR-12
T	90	0	0	85	80	85	90	−80	−15	0	80	−55	−90	−85
J	90	0	0	90	80	90	90	−90	0	0	45	−90	−90	−90
H	70	20	15	65	55	70	75	−85	−75	−60	−25	−75	−80	−90
L	80	0	15	80	75	85	75	−70	−35	−35	80	−60	−75	−80
O	75	50	50	70	55	70	75	−75	−20	−10	80	55	30	−65
Average	81	14	16	78	69	80	81	−80	−29	−21	52	−45	−61	−82
MAE (av)	9	76	74	12	21	10	9	10	61	69	142	45	29	8
Min-max MAE	0-20	40-90	40-90	0-25	10-35	0-20	0-15	0-20	15-90	30-90	65-170	0-145	0-120	0-25
Effect		$p=0.038$	$p=0.035$	$p=0.423$	$p=0.056$	$p=1.000$	$p=1.000$		$p=0.687$	nm	$p=0.027$	nm	$p=1.000$	nm
$S_{45}N_{-45}$	S_{45} unproc	ADM		MWF		MWF-$N_{0.2}$		N_{-45} unproc	ADM		MWF		MWF-$N_{0.2}$	
		SNR0	SNR-12	SNR0	SNR-12	SNR0	SNR-12		SNR0	SNR-12	SNR0	SNR-12	SNR0	SNR-12
T	50	60	50	80	65	65	80	−50	−60	−85	75	−90	−70	−75
J	60	90	90	90	45	90	50	−45	−90	−90	75	−70	−75	−90
H	45	40	30	45	45	45	45	−75	−70	−70	−50	−75	−70	−75
L	75	70	50	75	75	85	60	−60	−45	−50	−25	−50	−60	−60
O	75	70	75	70	75	75	70	−80	−60	−75	80	75	25	−65
Average	61	66	59	72	61	72	61	−62	−65	−74	31	−42	−50	−73
MAE (av)	16	23	20	27	16	27	16	17	20	29	78	45	37	28
Min-max MAE	0-30	5-45	5-45	0-45	0-30	0-45	0-35	0-35	0-45	5-45	5-125	5-120	15-90	15-45
Effect		nm	nm	nm	nm	nm	nm		$p=0.687$	nm	$p=0.027$	nm	$p=1.000$	nm

TABLE II. Percentage of stimuli perceptually classified as being a sound arriving from a point source with one clear direction in space, averaged over five subjects, for the three different spatial scenarios and the different processing schemes. The speech and the noise sources were presented separately through headphones (S,N). In the conditions in which most sounds were not categorized as arriving from one clear direction, the percentage of diffuse sounds (di), dual sounds (du), or very broad source (br) is added.

	Level (dB)	Unproc	ADM	MWF	MWF-$N_{0.2}$		Level (dB)	Unproc	ADM	MWF	MWF-$N_{0.2}$
S_0	0	67	87	93	100	N_{60}	0	89	80	80+7du	27+53du
	−12		53	87	93		−12		93	47+40du	67+27du
S_{90}	0	75	13+53di	100	100	N_{-90}	0	89	7+27di+40br	27+73du	7+93du
	−12		27+60di	60	100		−12		53+20di+20br	7+87du	13+87du
S_{45}	0	58	87	100	100	N_{-45}	0	89	93	20+67du	7+87du
	−12		87	87	73		−12		100	27+67du	13+80du

For SNR=0 dB, a main effect of algorithm was observed ($p=0.012$). Pairwise comparisons showed significantly lower performance for the MWF than for all other strategies (versus unprocessed $p=0.027$, versus ADM $p=0.017$, versus MWF-$N_{0.2}$ $p=0.049$). This can also be observed in Table I, which shows that the noise component at the output of the MWF was generally localized at the same location as the speech component. No significant differences were found between the unprocessed condition, the ADM, and the MWF-$N_{0.2}$ ($p \geqslant 0.687$). For SNR=−12 dB, no interactions or main effects were found (angle*algorithm $p=0.115$, angle $p=0.443$, algorithm $p=0.156$), implying that all algorithms, including the MWF, performed equally well at this SNR.

Interestingly, no interaction was found at either SNR between the factor algorithm and angle, although the results in Table I suggest that the ADM distorted the localization of the noise component in the scenario $S_{90}N_{-90}$ (which was also observed when analyzing the data of the speech component). Table I shows that only one out of five subjects, subject H, localized the noise component with the ADM equally well as in the unprocessed condition.

The subjective classification, shown in Table II, showed a clear drop in performance for almost all spatial scenarios for the MWF and the MWF-$N_{0.2}$ compared to the unprocessed condition. This was quite surprising for the MWF-$N_{0.2}$ and the MWF at SNR=−12 dB since their MAE values were relatively modest in these conditions and not statistically different from those for the unprocessed condition. Interestingly, the outputs of these algorithms were often classified as being a "dual sound." Averaged over the three spatial scenarios, there were 49% and 65% of such cases for the MWF and 78% and 65% of such cases for the MWF-$N_{0.2}$ at 0 and −12 dB, respectively. When dual sounds were reported, the sound was perceived as having two components, each arriving from a different angle. Subjects reported that one part arrived approximately from the position of the original noise component, whereas the other part arrived from around the position of the speech component. When using the MWF at a SNR of 0 dB, the sound arriving from the original noise position was typically described as being softer, lower in frequency and less distorted than the other part. For the SNR=−12 dB condition, the part arriving from the original noise position was reported as being louder than the distorted part arriving from the speech position.

2. Condition S+N

Whereas in the first experiment the goal was to gain understanding of how the filtering operations perceptually affect the localization cues, the second experiment was more related to real-life performance. In this experiment, speech and noise components were presented simultaneously which resembled more a steady-state real-life listening situation. Subjects were asked to localize both the speech and noise components. Table III shows the individual data indicating where the stimuli were perceived, averaged over three repetitions, together with the minimal, maximal, and averaged MAE values for the tested subjects.

In most conditions no differences were found between the data for condition S+N and condition S,N, leading to the same differences between algorithms as discussed for condition S,N. This was assessed for the unprocessed data, the ADM data, the MWF-$N_{0.2}$ data, and for the speech component data of the MWF by an ANOVA on all MAE data (S,N and S+N). For the noise component data of the MWF, a significant effect of the factor stimulus presentation (S,N versus S+N) was found for the 0 dB data ($p=0.006$) but not for the −12 dB data ($p=0.233$). An ANOVA comparing the 0 dB data of condition S+N demonstrated, in contrast with the S,N data, no significant difference between the MWF and all other conditions (factor algorithm, $p=0.322$). The data in Table III show that, for both SNRs, the performance of the MWF approaches that for the unprocessed condition for the noise component for all three spatial scenarios. The 0 dB data of the MWF contrast with the results obtained when speech and noise components were presented separately (Table I).

F. Discussion of reference condition

Since the unprocessed condition was used as a reference condition in the Results and analysis section, a short discussion of the results for this condition is in order. For the condition S,N, the average localization responses in the unprocessed condition were relatively accurate (Table I), with average MAE values between 1° and 24°, depending on the spatial scenario. Although localization was not perfect, these values are in reasonable agreement with those found by Van den Bogaert *et al.* (2006) who used similar procedures and stimuli in their tests. In their study, when testing subjects using their own ears to localize a broadband stimulus, the MAE, averaged over all angles, was about 8°, with large errors, up to 30°, occurring at the sides of the head. Poorer performance was expected here, since localization experiments were done using headphones and since the unprocessed stimuli were generated using signals at the front omnidirectional microphone of both hearing aids. Therefore, the signals could have sounded somewhat unnatural with slightly different ITDs and ILDs than normally occurring at the eardrums and with no relevant information about height and no externalization (pinnae effect). However, this condition was taken as the reference since an evaluation was made of the influence of noise reduction algorithms for hearing aids on the localization of sound sources. Since the allowed responses were limited to the frontal hemisphere, localization at the sides of the head might have been slightly biased toward the front. However, this was true for all conditions and does not explain the differences found between algorithms.

For the unprocessed condition, a similar localization performance was observed in conditions S,N and S+N. Since the data presented here were limited to only three repetitions for each spatial scenario with a limited number of subjects, one should be careful about generalizing this observation. Other researchers have demonstrated that localizing one sound source can be affected by the absence or presence of other sound signals (Lorenzi *et al.*, 1999).

TABLE III. Response location (deg), averaged over three repetitions, together with average, minimum, and maximum MAE data over the different subjects for the three different spatial scenarios (S_0N_{60}, $S_{90}N_{-90}$, and $S_{45}N_{-45}$), and the different processing schemes (unprocessed, ADM, MWF, and MWF-$N_{0.2}$) at two different SNRs (0 and −12 dB). The speech and the noise sources were presented simultaneously through headphones (S+N).

S_0N_{60}	S_0 unproc	ADM		MWF		MWF-$N_{0.2}$		N_{60} unproc	ADM		MWF		MWF-$N_{0.2}$	
		SNR0	SNR-12	SNR0	SNR-12	SNR0	SNR-12		SNR0	SNR-12	SNR0	SNR-12	SNR0	SNR-12
T	−10	0	−40	−10	−20	−5	−10	90	90	90	80	75	90	70
J	0	0	−25	0	0	0	0	90	90	90	80	75	90	90
H	0	0	0	0	0	0	−5	75	75	75	75	75	80	70
L	−15	−10	−15	−15	−15	−15	−20	90	90	80	90	90	85	90
O	0	50	−5	0	−5	0	−20	70	85	85	85	75	85	70
Average (°)	−5	8	−17	−5	−8	−4	−11	83	86	84	82	78	86	78
MAE av (°)	5	12	17	5	8	4	11	23	26	24	22	18	26	18
Min-max MAE	0-15	0-50	0-40	0-15	0-20	0-15	0-20	10-30	15-30	15-30	15-30	15-30	20-30	10-30
$S_{90}N_{-90}$	S_{90} unproc	ADM		MWF		MWF-$N_{0.2}$		N_{-90} unproc	ADM		MWF		MWF-$N_{0.2}$	
		SNR0	SNR-12	SNR0	SNR-12	SNR0	SNR-12		SNR0	SNR-12	SNR0	SNR-12	SNR0	SNR-12
T	75	0	−60	90	90	85	70	−60	−60	0	−75	−65	−85	−70
J	90	15	75	90	85	90	90	−90	−70	−80	−85	−85	−90	−90
H	65	20	20	65	55	75	45	−85	−60	−65	−65	−75	−75	−90
L	90	−5	85	90	80	90	75	−75	−55	−25	−65	−60	−70	−80
O	70	50	75	85	60	75	65	−90	−70	−35	−65	−40	−60	−60
Average (°)	78	16	39	84	74	83	69	−80	−63	−41	−71	−65	−76	−78
MAE av (°)	8	74	51	6	16	7	21	10	27	49	19	25	14	12
Min-max MAE	0-25	40-95	5-150	0-25	0-35	0-15	0-45	0-30	20-35	10-90	5-25	5-50	0-30	0-30
$S_{45}N_{-45}$	S_{45} unproc	ADM		MWF		MWF-$N_{0.2}$		N_{-45} unproc	ADM		MWF		MWF-$N_{0.2}$	
		SNR0	SNR-12	SNR0	SNR-12	SNR0	SNR-12		SNR0	SNR-12	SNR0	SNR-12	SNR0	SNR-12
T	75	75	55	75	80	60	75	−80	−60	−60	−90	−60	−90	−75
J	85	70	90	90	90	90	90	−80	−90	−80	−80	−65	−90	−90
H	45	45	45	45	45	40	50	−60	−60	−70	−75	−50	−70	−80
L	80	75	90	75	65	70	80	−65	−45	−45	−65	−55	−65	−60
O	85	90	80	75	70	75	80	−65	−60	−75	5	−10	−50	−40
Average (°)	74	71	72	72	70	67	75	−70	−63	−66	−61	−48	−73	−69
MAE av (°)	29	26	27	27	25	24	30	25	20	21	36	19	30	26
Min-max MAE	0-40	0-45	0-45	0-45	0-45	5-45	5-45	15-35	0-45	0-35	20-50	5-45	15-45	5-45

IV. NOISE REDUCTION PERFORMANCE

Besides the evaluation of the localization performance of the noise reduction algorithms, which was the main focus of this study, all tested algorithms were evaluated with respect to the suppression of noise, since a trade-off may exist between localization and noise reduction performance. The noise reduction performance of the different algorithms was measured using two out of the three spatial scenarios presented earlier.

A. Test setup

Speech reception thresholds (SRTs) were measured using an adaptive test procedure (Plomp and Mimpen, 1979). The level of the speech signals was adjusted to determine the 50% speech recognition level, i.e., the SRT. The VU sentences were used as speech material (Versfeld *et al.*, 2000) and a multitalker babble, the same as the one used in the localization experiment, was used as jammer signal. The performance of the three algorithms was evaluated for spatial scenarios S_0N_{60} and $S_{90}N_{-90}$. Tests were performed in a sound attenuating booth. Stimuli were presented under headphones (TDH-39) using a RME Hamerfall DSPII soundcard and a Tucker Davis HB7 headphone driver. The setup was calibrated so that the sound pressure level of the noise signal averaged over the left and right ears was constant and equal to 65 dB A. The level of the speech signal was adjusted with a step size of 2 dB during the adaptive procedure. The group of five normal hearing subjects tested in the localization experiment was expanded to nine normal hearing subjects, since the noise reduction data of five normal hearing subjects only showed close to significant trends.

B. Results and analysis

Table IV shows the individual SRT values (decibel SNR) of the nine normal hearing subjects for the unprocessed condition, together with the SRT gain obtained using the different noise reduction algorithms ($=SRT_{algo}-SRT_{unproc}$). To compare performance between algorithms, Bonferroni corrected pairwise comparisons were performed on the SRT data for each spatial scenario.

TABLE IV. SRT data, in decibel SNR, for the unprocessed condition, and SRT gain, of the different noise reduction algorithms relative to the unprocessed condition. Nine normal hearing subjects were tested using three different noise reduction algorithms in the spatial scenarios S_0N_{60} and $S_{90}N_{-90}$. A lower SNR score and a higher gain are better. The row labeled effect shows whether a significant difference with the unprocessed condition was observed.

	SRT (decibel SNR) unproc	Noise reduction gain MWF*	MWF-$N^*_{0.2}$	(decibel SRT) ADM*
		S_0N_{60}		
T	−5,4	6,0	2,8	2,8
J	−5,4	2,0	2,4	1,6
H	−9,8	1,6	0,0	−0,4
L	−8,2	4,4	2,8	2,0
O	−6,6	2,8	2,0	0,0
N	−5,4	4,0	2,8	2,0
A	−6,2	3,2	3.2	2.4
B	−4,2	4,4	5,6	2,4
TB	−3,8	6,4	4,8	5,6
Average		3,9	2,9	2,0
Stdev		1,6	1,6	1,7
Effect		$p=0.001$	$p=0.003$	$p=0.045$
		$S_{90}N_{-90}$		
T	−6,6	4,0	2,8	−2,8
J	−8,2	0,4	4,0	−4,0
H	−11,4	2,0	1,6	−5,2
L	−12,6	−1,6	0,4	−4,8
O	−8,6	1,2	2,0	−4,4
N	−9,0	−0,8	0,0	−2,4
A	−9,8	0,0	2,4	−6,0
B	−7,4	0,0	1,6	−5,6
TB	−9,0	−1,6	1,6	−4,4
Average		0,4	1,8	−4,4
Stdev		1,8	1,2	1,2
Effect		$p=1.000$	$p=0.011$	$p<0.001$

For the scenario S_0N_{60}, all three noise reduction algorithms gave a significant gain in SRT. The gains were 2.0 dB for the ADM ($p=0.045$), 3.9 dB for the MWF ($p=0.001$), and 2.9 dB for the MWF-$N_{0.2}$ ($p=0.003$). The MWF significantly outperformed the ADM by 1.2 dB ($p=0.003$). No significant difference was observed between the MWF and the MWF-$N_{0.2}$ ($p=0.392$), although six out of nine subjects performed more poorly with the MWF-$N_{0.2}$.

For the scenario $S_{90}N_{-90}$, a significant loss of 4.4 dB in SNR of the ADM was found ($p<0.001$). Moreover, the ADM performed significantly more poorly than all other algorithms (all comparisons $p<0.001$). The MWF gave no clear advantage over the unprocessed condition, with an average gain of 0.4 dB ($p=1.000$). The MWF-$N_{0.2}$ was the only algorithm that gave a significant SRT gain with an average gain of 1.8 dB ($p=0.011$). No significant difference was observed between the MWF and the MWF-$N_{0.2}$ ($p=0.166$), although seven out of nine subjects showed better performance with the MWF-$N_{0.2}$ than with the MWF.

V. DISCUSSION

Four research questions were raised related to the combined goals of improving speech perception in noise while preserving sound source localization using multimicrophone noise reduction algorithms. The results and analyses from the previous sections are used to answer these questions.

A. The influence of a dual-monaural ADM on the localization of sound sources

As a reference noise reduction algorithm for evaluating two recently introduced MWF-based noise reduction strategies for hearing aids, an ADM was used. Such a system is commonly implemented in current hearing aids to enhance speech perception in noise. In Secs. III E 1 and III E 2 it was observed that localization performance using the ADM was comparable to that for the unprocessed condition for spatial scenarios S_0N_{60} and $S_{45}N_{-45}$. However, a large degradation was found for scenario $S_{90}N_{-90}$ (Tables I and III), which was statistically verified for the speech component (Sec. III E 1). Perceptual evaluation showed that in spatial scenario $S_{90}N_{-90}$, the signals generated by the ADM were often described as being diffuse with no directional information present in the signal. Neither the perceptual data nor the statistical analysis showed a significant impact of SNR on localization performance.

The negative impact of adaptive and fixed directional microphones on localization performance was also observed in the work of Van den Bogaert *et al.* (2006) and Keidser *et al.* (2006). In Van den Bogaert *et al.* (2006), hearing-impaired users showed a significant decrease in localization performance when using their ADM systems compared to using omnidirectional microphones. This was observed when localizing a broadband stimulus in a noisy environment with the noise sources positioned at ±90°. A separate analysis showed that this was due to localization errors made when stimuli were presented from the sides, between ±60° and ±90°, of the head. When testing the ADM in silence with a broadband stimulus, no significant decrease in localization performance was observed. Keidser *et al.* (2006) tested the influence of multichannel wide dynamic range compression (WDRC), single channel noise reduction and directional microphones on localization performance. They observed that directional microphone settings had the largest influence on localization performance. The aspect of different directional microphone characteristics for the left and right hearing aids was assessed, using an omnidirectional pattern in both hearing aids as a reference condition. Combining a cardioid pattern at one ear with a figure-8 pattern at the other ear produced the largest decrease in localization performance. It was suggested that this could be an extreme hearing aid setting when using an ADM at both sides of the head.

In hearing aids, an ADM is typically constrained to avoid noise reduction and distortion of signals arriving from the front. For the scenarios S_0N_{60} and $S_{45}N_{-45}$, both the speech and the noise source were within or close to this area. Therefore, the interaural cues of the speech and noise components remained unchanged. However, due to this constraint, noise reduction will typically be limited in these areas. This was illustrated by the limited noise reduction performance of the ADM scheme in scenario S_0N_{60} (Sec. IV B, Table IV). Outside this area, sounds are suppressed. Therefore, both speech and noise sources were suppressed in the spatial scenario $S_{90}N_{-90}$ which led to the negative noise reduction performance (−4.4 dB) of the ADM.

Van den Bogaert *et al.* (2005) showed that distortion of ITD cues was proportional to the amount of noise reduction for a fixed and an ADM. This explains the drop in localization accuracy for scenario $S_{90}N_{-90}$ compared to the other spatial scenarios and compared to the unprocessed condition (Sec. III E 1). This is also illustrated by the work of Keidser *et al.* (2006), in which ITD and ILD measurements on directional microphones showed large ITD and ILD distortion at angles around 90° and much lower distortion between +50° and −50°.

Localization performance for the ADM was independent of SNR (Secs. III E 1 and III E 2). This can be explained by the fact that the ADM is based on exploiting physical differences in time of arrival between the microphones in the hearing aid, which are independent of SNR. Since the coherence between microphone signals was used to attenuate the strongest source in the back hemisphere, the most coherent part of the noise signal was removed. This would explain the classification of the output as sounding "diffuse."

B. The influence of the binaural MWF on the localization of sound sources

Doclo *et al.* (2006) mathematically proven that a binaural version of the MWF perfectly preserves the interaural cues of the speech component but changes the cues of the noise component into those of the speech component. This was also observed in ITD-error simulations, used to predict localization performance in the work of Klasen *et al.* (2007). As a consequence, large localization errors of the noise component were expected in the subjective evaluation discussed in this manuscript. These errors were indeed observed and statistically confirmed for the SNR=0 dB condition when the filtered speech and noise component were presented separately to the subjects (S,N). However, they were not observed when SNR=−12 dB (Sec. III E 1) nor when the speech and noise sources were presented simultaneously (S +N) (Sec. III E 2).

This can be explained using the subjective classification in Table II. Despite the good localization performance for the SNR=−12 dB condition, Table II suggests a decrease in sound quality for both SNRs. Subjects reported that the noise component at the output of the MWF sounded as if it was produced by sound sources at two different positions, one at the original noise position which sounded relatively clear and one at the speech position which sounded more distorted. In the SNR=−12 dB condition, subjects preferred the sound arriving from the original noise location, often resulting in a correct localization of the noise component. In the SNR=0 dB condition, subjects preferred the sound arriving from the original speech location. However, individual subjects did not follow this general trend, e.g., for spatial scenarios $S_{90}N_{-90}$ and $S_{45}N_{-45}$, subject O preferred the sound arriving from the original speech location when using a MWF at a SNR=−12 dB. This demonstrates that the variability between subjects in the MWF conditions can be explained by the dual-sound phenomenon.

The reason for the dual sounds can be found in the filter generation of the MWF. Since the speech correlation matrix is estimated as $\mathbf{R}_x = \mathbf{R}_y - \mathbf{R}_n$ (Sec. II B), where $\mathbf{R}_y$ and $\mathbf{R}_n$ were computed during different time periods, this estimate will be poor at a low SNR. Hence, in the frequency region with high SNR (in our case between 3000 and 5500 Hz, see Fig. 2), a good estimate was obtained, such that the interaural cues of the noise component were changed into those of the speech component, as illustrated by Fig. 3. On the other hand, in the frequency region with low SNR (in our case between 500 and 3000 Hz, see Fig. 2), a poor estimate was obtained, such that the output contained interaural cues corresponding to the original position of the noise source. Because of these different behaviors for different frequency regions, a dual sound was created. For the low overall SNR, i.e., SNR=−12 dB, a large proportion of the noise component contained the interaural cues of the original noise angle, which resulted in a correct localization of the noise component (Table I).

Figure 3 shows the cross-correlation function and the ILD between the left and right ear signals of the unprocessed speech and noise components and of the noise component processed by the MWF and the MWF-N. These are given for the spatial scenario $S_{90}N_{-90}$ at SNR=−12 and 0 dB. The ILDs were calculated using third-order butterworth filters with cutoff frequencies based on the Bark scale (Zwicker, 1961). The cross-correlation functions, used to interpret ITD information, were calculated for the low-pass filtered left and right ear signals and were normalized to a maximum value of 1 for identical signals. A cutoff frequency of 1000 Hz was used, since the most relevant ITD information for the human auditory system is present at frequencies below 1000 Hz, e.g., Hartmann (1999). The ITD is approximated by the delay for which the cross-correlation function reaches its maximum.

For SNR=0 dB, the ITD of the MWF noise component was shifted toward the ITD of the original speech component. Also, the amplitude of the cross correlation, the amount of coherence between the left and right signals, and the width of the curve totally agree with the curve for the original speech component. Also, the ILDs of the MWF processed noise component were shifted toward those of the speech component for SNR=0 dB, except for a small region around 1000 Hz, which could be due to the low input SNR in this region (Fig. 2).

For SNR=−12 dB, the cross-correlation function of the processed noise component was shifted toward that for the speech component. However, a second peak was present around −500 μs. Also the curve was somewhat more flat than the curve for SNR=0 dB, meaning that the ITD information was less coherent than for the SNR=0 dB condition. The ILD plots show that only ILDs for frequencies between 3000 and 5500 Hz (the region with a high input SNR) were shifted toward the ILDs for the unprocessed speech component. These observations, especially at SNR=−12 dB, illustrate the dual-sound phenomenon and explain the improved localization performance when using the MWF at SNR=−12 dB compared to SNR=0 dB.

The dual-sound phenomenon also explains the good lo-

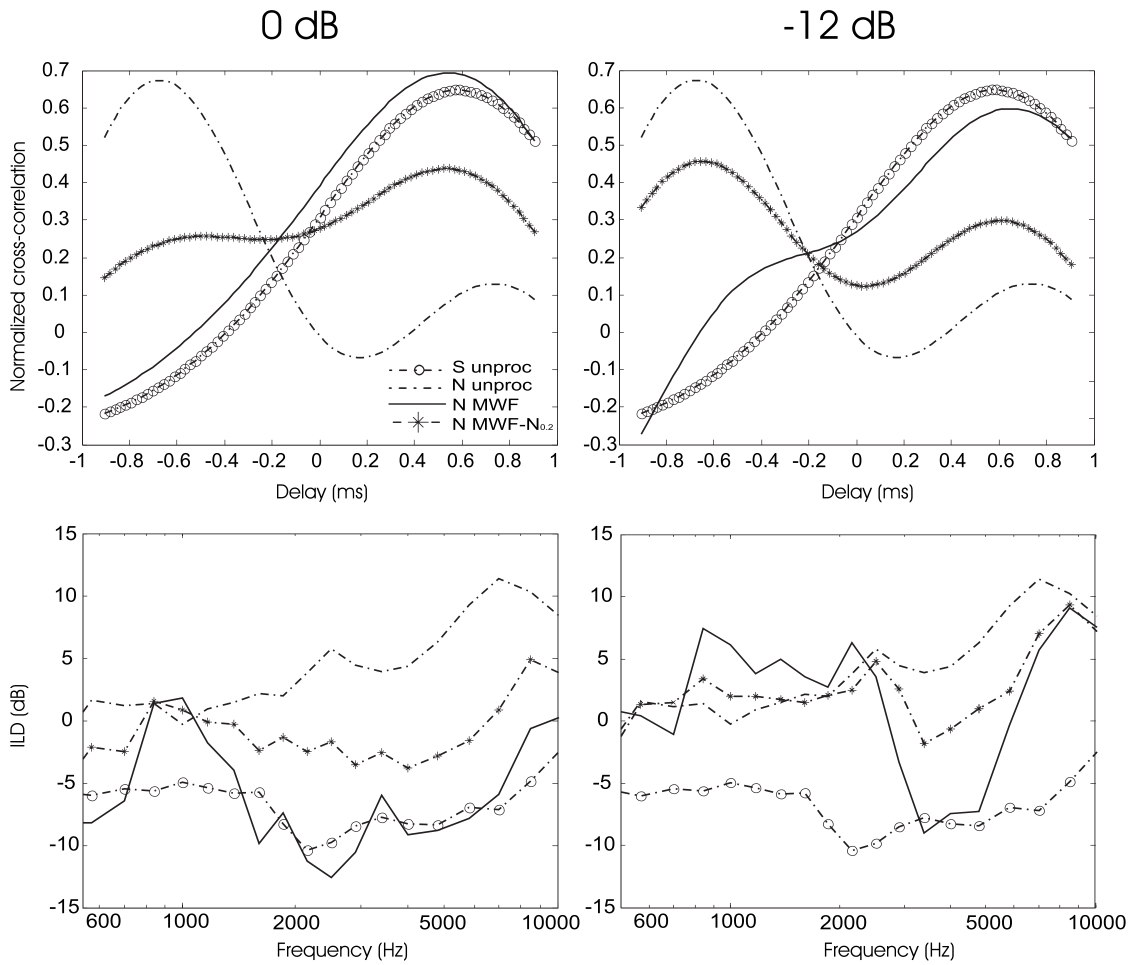

FIG. 3. ILD and cross-correlation functions between the left and right ear signals for the unprocessed speech and unprocessed noise components together with the MWF and MWF-$N_{0.2}$ processed noise components. This is shown for spatial scenario $S_{90}N_{-90}$. ILDs were calculated using a critical band analysis (bark bands).

calization performance when the speech and noise sources were played simultaneously (S+N). In this condition, the speech component masked parts of the frequency spectrum of the noise component at the output of the algorithm. The noise component was masked mostly in frequency regions with a good noise reduction performance. This is exactly the region where the interaural cues of the noise component were shifted toward the interaural cues of the speech component. When the sounds were played simultaneously, the part of the noise component with the incorrect cues was masked by the speech component. Due to this masking, the noise source could be correctly localized when using the MWF.

The significant effect of SNR and presentation format (S,N or S+N) illustrates that testing algorithms on localization performance in laboratory conditions is not straightforward and results should be interpreted carefully when generalizing to real world situations. Both presentation formats (S,N and S+N) could be relevant to real-life situations. Speech and noise presented simultaneously could be relevant to situations with converged filters and speech and noise sources playing continuously. Presenting speech and noise component separately could be relevant for the gaps present in the speech or noise components, e.g., when pauses are present in sentences.

C. The improvement in localization performance for MWF-N relative to MWF

Klasen *et al.* (2007) showed that the ITD error of the noise component generated by the MWF could be decreased by extending this algorithm to the MWF-N. It was suggested that this could result in improved localization performance. The perceptual relevance of the MWF-N was proven in Sec. III E. Large improvements were observed for all spatial scenarios when the speech and noise components were presented separately (S,N) for an input SNR=0 dB. In the other conditions, less or no room for improvement was available due to the reasons explained in the evaluation of the MWF (masking, errors in estimating the speech correlation matrix at low SNR). Hence, no statistical evidence of improvement was found for these conditions. However, nonsignificant trends were sometimes observed and a subset of the data, i.e., the data for the noise component at S+N at both SNRs in the spatial scenario $S_{90}N_{-90}$ did show significantly better performance for the MWF-N than for the MWF. Although the MWF-$N_{0.2}$ improved localization performance of the MWF to that for the unprocessed condition, a difference in perceptual evaluation remained. When presenting speech and noise components separately, the output signals of the MWF-$N_{0.2}$ were still described as arriving from two different directions

(Table II). Adding more of the unprocessed signal (e.g., MWF-$N_{0.3}$) would probably improve the sound quality but would further decrease noise reduction performance.

Figure 3 illustrates the interaural information present in the MWF and MWF-N processed noise components. When comparing the MWF curves with those for the MWF-N, it is observed that the distorted ILD and ITD cues at the output of the MWF were corrected toward the values for the unprocessed condition when using the MWF-$N_{0.2}$. This is true for both SNR=−12 and SNR=0 dB. Still, both the cross correlation and ILD graphs illustrate that not all cues were corrected. The cross-correlation curves for the signals at the output of the MWF-N still show a local maximum around the peak generated by the original speech component and the ILD cues for some frequency regions remain close to the ILD cues for the original speech component. This was observed more for SNR=0 dB than for SNR=−12 dB, since the MWF introduced larger distortions at high SNRs, meaning that a larger correction factor η was needed in this condition. This is consistent with the dual-sound phenomenon which was observed, despite the good localization performance, when using the MWF-N in the S,N condition.

D. Overall comparison of the ADM, MWF, and MWF-N

For the spatial scenario S_0N_{60}, Table IV shows that both the binaural MWF and the binaural MWF-N outperformed the dual-monaural ADM in terms of noise reduction. This is logical since the MWF is not constrained to suppressing sound sources only in the rear hemisphere. No significant difference in speech perception was found between the MWF and MWF-N. This occured despite the introduction of the unprocessed component in the MWF-N scheme [Eqs. (11) and (12)]. In scenario S_0N_{60}, the ADM perfectly preserved localization performance for both the speech and noise components, since almost no processing was done on the speech and noise components. The MWF preserved the ability to localize the speech component but not always the ability to localize the noise component, especially when speech and noise sources were presented separately at a high SNR (Table I). The MWF-N seems to enable the user to localize both the speech and the noise sources correctly at the cost of some noise reduction, e.g., 1.0 dB compared to the MWF (Sec. IV B).

For the spatial scenario $S_{90}N_{-90}$, the ADM gave a large drop in noise reduction performance, since it is designed to remove sounds not arriving from the front. Therefore, in scenario $S_{90}N_{-90}$ both the speech and the noise components were suppressed. The processing of these signals was accompanied by a large drop in localization performance (Tables I and III) which was discussed in Sec. V A. Again, the MWF outperformed the ADM in terms of noise reduction but not in terms of localization of the noise component at high SNRs (see Table I, Sec. V B). The MWF-$N_{0.2}$ seems to enable the user to combine correct sound source localization (Table I) with good noise reduction performance (see Table IV). Interestingly, the MWF-$N_{0.2}$, which adds an unprocessed component to the MWF output, outperformed the MWF in terms of noise reduction in this spatial scenario (Table IV). This might be explained by the improved localization performance when using the MWF-$N_{0.2}$ compared to the MWF (Sec. III E 1), which might have led to better speech segregation due to spatial unmasking.

VI. CONCLUSIONS

In this paper, four research questions were addressed which are related to the influence of noise reduction techniques for hearing aids on the localization of sound sources. First, the localization performance of normal hearing subjects was quantified using a dual-monaural noise reduction system, namely, an ADM which is commonly used in current high-end hearing aids. The ADM led to a significant drop in localization performance when sounds were presented from outside the frontal direction. As second and third research questions, two newly proposed binaural noise reduction algorithms were evaluated in terms of localization performance. The binaural MWF led to good localization performance for the speech component. The noise component on the other hand could be perceived as arriving from the location of the speech component when the speech and noise components were presented separately to the subjects. However, localization performance when using the MWF was in many cases better than expected due to errors in the estimation of the speech correlation matrix and due to masking effects when the speech and noise components were presented simultaneously. Results for the binaural MWF-N showed that, by adding part of the unprocessed signal (η =0.2) to the output of the MWF, localization of the noise component improved. Hence, no significant difference in localization performance was found in all scenarios when comparing the MWF-N to the unprocessed condition for both the speech and the noise components. Fourth, the combination of noise reduction and localization performance was studied, leading to the conclusion that the dual-monaural ADM configuration was not able to provide both good localization of the speech and the noise components and good noise reduction performance. On the other hand, the MWF-N enables correct sound localization of both the speech and the noise components, together with good noise reduction performance. Both MWF and MWF-N were based on a statistical Wiener filter approach. This is different from the ADM, which used the physical delay between microphones to improve the SNR. We suggest that MWF-based, binaural noise reduction techniques might introduce a better combination of sound source localization and noise reduction performance compared to a traditional ADM.

The full data set consisted of many conditions: two different ways of presenting stimuli, three spatial scenarios, four different algorithms, and two SNRs. Therefore, the subset of data for each condition became relatively small and due to this limitation, small differences between algorithms may have been undetected. However, even these limited subsets of data were sufficient to illustrate effects of, and differences between, noise reduction algorithms. Moreover, it was shown that interpreting results of localization experiments with noise reduction systems is not straightforward since these results are dependent on spatial scenario, SNR, and masking effects. Further research, involving larger data sets for each condition, might reveal smaller differences between algorithms.

ACKNOWLEDGMENTS

T.V.d.B was funded by a Ph.D. grant of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). This research was partly carried out in the framework of the K.U.Leuven Concerted Research Action GOA-AMBioRICS. The authors would like to thank Steffen Fieuws of the Biostatistical center of the K.U.Leuven. The authors would also like to thank both reviewers and Brian Moore for their detailed comments, which significantly improved the manuscript.

[1]Typically M_I equals the number of microphones on the hearing aid at a given ear (for the left hearing aid: $M_I = M_L$). M_C can be constrained due to power or bandwidth limitations of the binaural link (for the left hearing aid: $M_C \leq M_R$). In theory M_I and M_C can be different for the left and the right hearing aids. For reasons of clarity, we assume this is not the case.

Aichner, R., Buchner, H., Zourub, M., and Kellerman, W. (**2007**). "Multichannel source separation preserving spatial information," in *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. I5–I8.

Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization*, revised ed. (The MIT Press, Cambridge).

Bronkhorst, A. W., and Plomp, R. (**1988**). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," J. Acoust. Soc. Am. **83**, 1508–1516.

Bronkhorst, A. W., and Plomp, R. (**1989**). "Binaural speech intelligibility in noise for hearing impaired listeners," J. Acoust. Soc. Am. **86**, 1374–1383.

Byrne, D., Sinclair, S., and Noble, W. (**1998**). "Open earmold fittings for improving aided auditory localization for sensorineural hearing losses with good high-frequency hearing," Ear Hear. **19**, 62–71.

Desloge, J. G., Rabinowitz, W. M., and Zurek, P. M. (**1997**). "Microphone-array hearing aids with binaural output part I: Fixed processing systems," IEEE Trans. Speech Audio Process. **5**, 529–542.

Doclo, S., Klasen, T. J., Van den Bogaert, T., Wouters, J., and Moonen, M. (**2006**). "Theoretical analysis of binaural cue preservation using multichannel Wiener filtering and interaural transfer functions," in *Proceedings 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 multi-microphone speech enhancement," IEEE Trans. Signal Process. **50**, 2230–2244.

Hartmann, W. M. (**1999**). "How we localize sound," Phys. Today **52**(11), 24–29.

Haykin, S. (**2002**). *Adaptive Filter Theory*, 4th ed. (Prentice Hall, New Jersey), Chap. 2, pp. 94–135.

Keidser, G., Rohrseitz, K., Dillon, H., Hamacher, V., Carter, L., Rass, U., and Convery, E. (**2006**). "The effect of multi-channel wide dynamic range compression, noise reduction, and the directional microphone on horizontal localization performance in hearing aid wearers," Int. J. Audiol. **45**, 563–579.

Klasen, T. J., Doclo, S., Van den Bogaert, T., Moonen, M., and Wouters, J. (**2006**). "Binaural multi-channel Wiener filtering for hearing aids: preserving interaural time and level differences," in *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toulouse, France.

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.

Lorenzi, C. S., Gatehouse, S., and Lever, C. (**1999**). "Sound localization in noise in normal hearing listeners," J. Acoust. Soc. Am. **105**, 1810–1820.

Lotter, T. (**2004**), "Single and multimicrophone speech enhancement for hearing aids," Ph.D. thesis, RWTH Aachen.

Luo, F. L., Yang, J., Pavlovic, C., and Nehorai, A. (**2002**). "Adaptive null-forming scheme in digital hearing aids," IEEE Trans. Signal Process. **50**, 1583–1590.

Maj, J. B., Wouters, J., and Moonen, M. (**2004**). "Noise reduction results of an adaptive filtering technique for dual microphone behind the ear hearing aids," Ear Hear. **25**, 215–229.

Noble, W., and Byrne, D. (**1990**). "A comparison of different binaural hearing aid systems for sound localization in the horizontal and vertical planes," Br. J. Audiol. **24**, 335–346.

Noble, W., Sinclair, S., and Byrne, D. (**1998**). "Improvements in aided sound localization with open earmolds: observations in people with high-frequency hearing loss," J. Am. Acad. Audiol **9**, 25–34.

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

Spriet, A., Moonen, M., and Wouters, J. (**2004**). "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction," Signal Process. **84**, 2367–2387.

Van den Bogaert, T., Doclo, S., Moonen, M., and Wouters, J. (**2007**). "Binaural cue preservation for hearing aids using an interaural transfer function multichannel Wiener filter," in *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vol. IV, pp. 565–568.

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

Van den Bogaert, T., Klasen, T. J., Van Deun, L., Wouters, J., and Moonen, M. (**2006**). "Localization with bilateral hearing aids: without is better than with," J. Acoust. Soc. Am. **119**, 515–526.

Van Hoesel, R., Ramsden, R., and O'Driscoll, M. (**2002**). "Sound direction identification, interaural time delay discrimination, and speech intelligibility advantages in noise for a bilateral cochlear implant user," Ear Hear. **23**, 137–149.

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.

Welker, D. P., Greenberg, J. E., Desloge, J. G., and Zurek, P. M. (**1997**). "Microphone-array hearing aids with binaural output part II: A two-microphone adaptive system," IEEE Trans. Speech Audio Process. **5**, 543–551.

Wightman, F. L., and Kistler, D. J. (**1992**). "The dominant role of low-frequency interaural time differences in sound localization," J. Acoust. Soc. Am. **91**, 1648–1661.

Wittkop, T., and Hohman, V. (**2003**). "Strategy selective noise reduction for binaural digital hearing aids," Speech Commun. **39**, 111–138.

Zurek, P. M., and Greenberg, J. E. (**2000**). "Two-microphone adaptive array hearing aids with monaural and binaural outputs," in *Proceedings of the Ninth IEEE DSP Workshop*, Hunt, TX.

Zwicker, E. (**1961**). "Subdivision of the audible frequency range into critical bands," J. Acoust. Soc. Am. **33**, 248–249.