Theoretical Analysis of Binaural Multimicrophone Noise Reduction Techniques

Bram Cornelis, *Member, IEEE*, Simon Doclo, *Member, IEEE*, Tim Van dan Bogaert, *Member, IEEE*, Marc Moonen, *Fellow, IEEE*, and Jan Wouters

***Abstract*—Binaural hearing aids use microphone signals from both left and right hearing aid to generate an output signal for each ear. The microphone signals can be processed by a procedure based on speech distortion weighted multichannel wiener filtering (SDW-MWF) to achieve significant noise reduction in a speech + noise scenario. In binaural procedures, it is also desirable to preserve binaural cues, in particular the interaural time difference (ITD) and interaural level difference (ILD), which are used to localize sounds. It has been shown in previous work that the binaural SDW-MWF procedure only preserves these binaural cues for the desired speech source, but distorts the noise binaural cues. Two extensions of the binaural SDW-MWF have therefore been proposed to improve the binaural cue preservation, namely the MWF with partial noise estimation (MWF-η) and MWF with interaural transfer function extension (MWF-ITF). In this paper, the binaural cue preservation of these extensions is analyzed theoretically and tested based on objective performance measures. Both extensions are able to preserve binaural cues for the speech and noise sources, while still achieving significant noise reduction performance.**

***Index Terms*—Binaural cues, binaural hearing aid, localization, multichannel Wiener filtering, noise reduction.**

I. Introduction

MODERN hearing aids make use of noise reduction algorithms to improve speech intelligibility in background noise. Hearing aids are usually fitted with multiple microphones, which generally leads to an improvement in noise reduction performance because spatial sound information can then be exploited in addition to spectral information. In a binaural setup, the hearing impaired person has two hearing aids that communicate over a wireless link [1]–[10]. In principle, microphone signals from both hearing aids could be shared, leading to further noise reduction performance compared to a monaural configuration or bilateral configuration in which two hearing aids work independently. Current noise reduction algorithms in bilateral and binaural configurations are not designed to preserve the *binaural cues*, in particular the interaural time difference (ITD) and interaural level difference (ILD), which are used to localize sounds [11]. Preservation of binaural cues is crucial as incorrect sound localization can endanger the hearing aid user. In addition to sound localization, binaural cues also improve speech perception in noisy environments. The so-called spatial release from masking effect leads to a speech intelligibility improvement of up to 10 dB [12]. Part of this improvement is caused by the spatial separation of speech and noise sources which generally improves the signal-to-noise ratio (SNR) at one of the ears. Another part is however purely caused by the binaural processing, and denoted as binaural unmasking. Other studies confirm this and report speech reception threshold (SRT) improvements of 2–3 dB due to the binaural processing [13]. In contrast to the monaural or bilateral setups, a *binaural noise reduction algorithm* could fully exploit this binaural unmasking effect to further improve speech intelligibility.

In [14], a binaural noise reduction algorithm based on *Multichannel Wiener Filtering (MWF)* has been introduced, referred to as *Speech Distortion Weighted MWF (SDW-MWF)*. It has been proven in [15] that this algorithm preserves the binaural cues for the speech component, but changes noise binaural cues to equal those of the speech component. The binaural SDW-MWF approach is reviewed in Section III.

In [16], an extension of the binaural SDW-MWF algorithm has been proposed, introducing a new tradeoff parameter. This parameter allows a tradeoff between noise reduction performance and the preservation of noise binaural cues. Perceptual tests in [17] have shown that this approach, referred to as *MWF with partial noise estimation (MWF-η)*, enables correct localization of both speech and noise components. Although noise reduction performance decreases in this approach, this does not necessarily lead to a loss in speech intelligibility due to the compensatory effect of binaural unmasking [18]. In Section IV, MWF-η is reviewed and its binaural cue preservation is analyzed theoretically.

Manuscript received October 02, 2008; revised June 26, 2009. First published July 24, 2009; current version published November 20, 2009. This research work was carried out at the ESAT Laboratory of Katholieke Universiteit Leuven in the frame of the Belgian Program 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-AMBioRICS and research project FWO nr. G.0600.08 ("Signal processing and network design for wireless acoustic sensor networks"). The work of B. Cornelis was supported by a Ph.D. grant from the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). The scientific responsibility is assumed by its authors. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Malcolm Slaney.

B. Cornelis and M. Moonen are with the Department of Electrical Engineering (ESAT–SCD), Katholieke Universiteit Leuven, B-3001 Leuven, Belgium (e-mail: bram.cornelis@esat.kuleuven.be; marc.moonen@esat.kuleuven.be).

S. Doclo was with the Department of Electrical Engineering (ESAT–SCD), Katholieke Universiteit Leuven, B-3001 Leuven, Belgium. He is now with the Sounds and Acoustics Group, NXP Semiconductors, B-3001 Leuven, Belgium (e-mail: simon.doclo@esat.kuleuven.be).

T. Van den Bogaert and J.Wouters are with the Department of Neurosciences, ExpORL, Katholieke Universiteit Leuven, 3000 Leuven, Belgium (e-mail: tim.vandenbogaert@med.kuleuven.be; jan.wouters@med.kuleuven.be).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASL.2009.2028374

1558-7916/$26.00 © 2009 IEEE

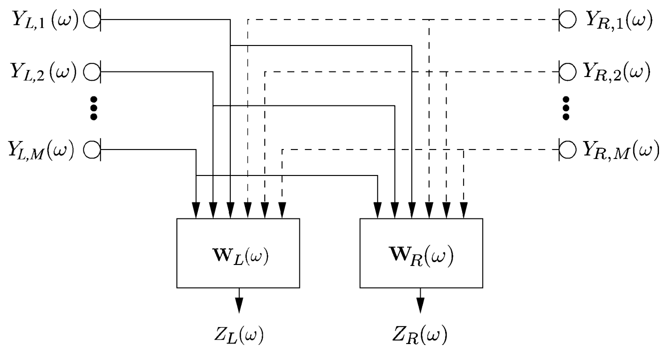


Fig. 1. General binaural processing scheme.

In [15], a different extension was proposed to tradeoff noise reduction performance for better cue preservation. Here, the MWF cost function is extended with extra terms related to the *interaural transfer functions (ITF)* of the speech and noise components. Simulations in [15] with this *MWF-ITF* algorithm have shown that it is then indeed possible to preserve both speech and noise binaural cues. To reduce computational complexity, a simplification was introduced in the MWF-ITF cost function to obtain quadratic cost terms [19]. Remarkably, perceptual tests in [19] showed an improvement in localization performance, even though a simplified cost function was used. In Section V closed form expressions for the optimal MWF-ITF filters are derived. It is proven that this MWF-ITF approach cannot preserve speech and noise binaural cues simultaneously, but that speech and noise ITFs are changed into one and the same value, which is a combination of the input speech and noise ITFs. However, to explain the improvement in the perceptual tests, Section VI illustrates that the obtained output ITF is related to the output SNR.

Both extended MWF algorithms are validated using objective performance measures in Section VI. A direct comparison is made, which evaluates both approaches. Finally, overall conclusions are drawn in Section VII.

II. Configuration and Notation

A. Microphone Signals and Output Signals

We consider the binaural hearing aid configuration depicted in Fig. 1, where both hearing aids have a microphone array consisting of M microphones.[1] The mth microphone signal in the left hearing aid $Y_{L,m}(\omega)$ can be specified in the frequency-domain as

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

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 hearing aid is equal to $Y_{R,m}(\omega) = X_{R,m}(\omega) + V_{R,m}(\omega)$. For conciseness, we will omit the frequency-domain variable ω from now on.

We define the M-dimensional stacked vectors $\mathbf{Y}_L$ and $\mathbf{Y}_R$ and the $2M$-dimensional signal vector $\mathbf{Y}$ as

$$\mathbf{Y}_L = \begin{bmatrix} Y_{L,1} \\ \vdots \\ Y_{L,M} \end{bmatrix}, \ \mathbf{Y}_R = \begin{bmatrix} Y_{R,1} \\ \vdots \\ Y_{R,M} \end{bmatrix}, \ \mathbf{Y} = \begin{bmatrix} \mathbf{Y}_L \\ \mathbf{Y}_R \end{bmatrix}. \tag{2}$$

The signal vector can be written as $\mathbf{Y} = \mathbf{X} + \mathbf{V}$, where $\mathbf{X}$ and $\mathbf{V}$ are defined similarly as $\mathbf{Y}$. The correlation matrix $\mathbf{R}_y$, the speech correlation matrix $\mathbf{R}_x$, and the noise correlation matrix $\mathbf{R}_v$ are defined as[2]

$$\mathbf{R}_y = \mathcal{E}\{\mathbf{Y}\mathbf{Y}^H\}, \ \mathbf{R}_x = \mathcal{E}\{\mathbf{X}\mathbf{X}^H\}, \ \mathbf{R}_v = \mathcal{E}\{\mathbf{V}\mathbf{V}^H\} \tag{3}$$

where $\mathcal{E}$ denotes the expected value operator. Assuming that the speech and the noise components are uncorrelated, $\mathbf{R}_y = \mathbf{R}_x + \mathbf{R}_v$.

We will use the r_Lth microphone on the left hearing aid and the r_Rth microphone on the right hearing aid as the so-called reference microphones for the speech enhancement algorithms. Typically, the front microphones are used as reference microphones. For conciseness, the reference microphone signals Y_{L,r_L} and Y_{R,r_R} at the left and the right hearing aid are denoted as Y_L and Y_R, which are then equal to

$$Y_L = \mathbf{e}_L^H \mathbf{Y}, \quad Y_R = \mathbf{e}_R^H \mathbf{Y} \tag{4}$$

where $\mathbf{e}_L$ and $\mathbf{e}_R$ are $2M$-dimensional vectors with only one element equal to 1 and the other elements equal to 0, i.e., $\mathbf{e}_L(r_L) = 1$ and $\mathbf{e}_R(M + r_R) = 1$. The reference microphone signals can be written as $Y_L = X_L + V_L$ and $Y_R = X_R + V_R$.

The output signals Z_L and Z_R at the left and the right hearing aid are obtained by filtering and summing all microphone signals from both hearing aids, i.e.,

$$Z_L = \mathbf{W}_L^H \mathbf{Y}, \quad Z_R = \mathbf{W}_R^H \mathbf{Y} \tag{5}$$

where $\mathbf{W}_L$ and $\mathbf{W}_R$ are $2M$-dimensional complex weight vectors. The output signal at the left hearing aid can be written as

$$Z_L = Z_{xL} + Z_{vL} = \mathbf{W}_L^H \mathbf{X} + \mathbf{W}_L^H \mathbf{V} \tag{6}$$

where Z_{xL} represents the speech component and Z_{vL} represents the noise component of the output signal. Similarly, the output signal at the right hearing aid can be written as $Z_R = Z_{xR} + Z_{vR} = \mathbf{W}_R^H \mathbf{X} + \mathbf{W}_R^H \mathbf{V}$. We define the $4M$-dimensional complex stacked weight vector $\mathbf{W}$ as

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_L \\ \mathbf{W}_R \end{bmatrix}. \tag{7}$$

B. Special Case: Single Speech Source

In the case of a single speech source, the speech signal vector can be modeled as

$$\mathbf{X} = \mathbf{A}S \tag{8}$$

where the $2M$-dimensional steering vector $\mathbf{A}$ contains the acoustic transfer functions from the speech source to the micro-

[1] It is possible to use different array sizes as in [15], but here both arrays are fixed at M microphones for the sake of simplicity.

[2] In [17] a distinction is made between $\mathbf{R}_{y,L}$ and $\mathbf{R}_{y,R}$, but this distinction is omitted here for notational convenience. Similarly, the definition (2) has been altered compared to the definition found in [17].

phones (including room acoustics, microphone characteristics, and head shadow effect) and S denotes the speech signal. The vector $\mathbf{A}$ is defined similarly as $\mathbf{Y}$ in (2), i.e.,

$$\mathbf{A}_L = \begin{bmatrix} A_{L,1} \\ \vdots \\ A_{L,M} \end{bmatrix}, \ \mathbf{A}_R = \begin{bmatrix} A_{R,1} \\ \vdots \\ A_{R,M} \end{bmatrix}, \ \mathbf{A} = \begin{bmatrix} \mathbf{A}_L \\ \mathbf{A}_R \end{bmatrix}. \quad (9)$$

The speech correlation matrix is then a rank-1 matrix, i.e.,

$$\mathbf{R}_x = P_s \mathbf{A}\mathbf{A}^H \quad (10)$$

with $P_s = \mathcal{E}\{|S|^2\}$ the power of the speech signal. The reference microphone signals at the left and the right hearing aid can be written as

$$Y_L = A_L S + V_L, \quad Y_R = A_R S + V_R \quad (11)$$

with A_L the r_Lth element of $\mathbf{A}_L$ and A_R the r_Rth element of $\mathbf{A}_R$.

C. Performance Measures

The *input SNR* is defined as the power ratio of speech and noise component in the reference microphones, i.e.,

$$\text{SNR}_L^{\text{in}} = \frac{\mathcal{E}\{|X_L|^2\}}{\mathcal{E}\{|V_L|^2\}} = \frac{\mathbf{e}_L^H \mathbf{R}_x \mathbf{e}_L}{\mathbf{e}_L^H \mathbf{R}_v \mathbf{e}_L} \quad (12)$$

$$\text{SNR}_R^{\text{in}} = \frac{\mathcal{E}\{|X_R|^2\}}{\mathcal{E}\{|V_R|^2\}} = \frac{\mathbf{e}_R^H \mathbf{R}_x \mathbf{e}_R}{\mathbf{e}_R^H \mathbf{R}_v \mathbf{e}_R} \quad (13)$$

and the *output SNR* is defined as the power ratio of speech and noise component in the output signals, i.e.,

$$\text{SNR}_L^{\text{out}} = \frac{\mathcal{E}\{|Z_{xL}|^2\}}{\mathcal{E}\{|Z_{vL}|^2\}} = \frac{\mathbf{W}_L^H \mathbf{R}_x \mathbf{W}_L}{\mathbf{W}_L^H \mathbf{R}_v \mathbf{W}_L} \quad (14)$$

$$\text{SNR}_R^{\text{out}} = \frac{\mathcal{E}\{|Z_{xR}|^2\}}{\mathcal{E}\{|Z_{xR}|^2\}} = \frac{\mathbf{W}_R^H \mathbf{R}_x \mathbf{W}_R}{\mathbf{W}_R^H \mathbf{R}_v \mathbf{W}_R}. \quad (15)$$

The *SNR improvement* at the left and the right hearing aid is defined as

$$\Delta\text{SNR}_L = \frac{\text{SNR}_L^{\text{out}}}{\text{SNR}_L^{\text{in}}}, \quad \Delta\text{SNR}_R = \frac{\text{SNR}_R^{\text{out}}}{\text{SNR}_R^{\text{in}}}. \quad (16)$$

The input and output *interaural transfer function (ITF)* of the speech and noise components are defined as the ratio of the components at the left and right hearing aid, i.e.,

$$\text{ITF}_x^{\text{in}} = \frac{X_L}{X_R} = \frac{\mathbf{e}_L^H \mathbf{X}}{\mathbf{e}_R^H \mathbf{X}}, \ \text{ITF}_x^{\text{out}} = \frac{Z_{xL}}{Z_{xR}} = \frac{\mathbf{W}_L^H \mathbf{X}}{\mathbf{W}_R^H \mathbf{X}} \quad (17)$$

$$\text{ITF}_v^{\text{in}} = \frac{V_L}{V_R} = \frac{\mathbf{e}_L^H \mathbf{V}}{\mathbf{e}_R^H \mathbf{V}}, \ \text{ITF}_v^{\text{out}} = \frac{Z_{vL}}{Z_{vR}} = \frac{\mathbf{W}_L^H \mathbf{V}}{\mathbf{W}_R^H \mathbf{V}}. \quad (18)$$

In the case of a *single speech source*, the ITF_x^{in} and $\text{ITF}_x^{\text{out}}$ are independent of the actual input signal S, namely,

$$\text{ITF}_x^{\text{in}} = \frac{\mathbf{e}_L^H \mathbf{A}}{\mathbf{e}_R^H \mathbf{A}} = \frac{A_L}{A_R}, \quad \text{ITF}_x^{\text{out}} = \frac{\mathbf{W}_L^H \mathbf{A}}{\mathbf{W}_R^H \mathbf{A}}. \quad (19)$$

Hence, it is also possible to use alternative formulas for these ITFs, namely,

$$\text{ITF}_x^{\text{in}} = \frac{\mathbf{e}_L^H \mathbf{R}_x \mathbf{e}_R}{\mathbf{e}_R^H \mathbf{R}_x \mathbf{e}_R} = \frac{\mathbf{e}_L^H \mathbf{R}_x \mathbf{e}_L}{\mathbf{e}_R^H \mathbf{R}_x \mathbf{e}_L} \quad (20)$$

$$\text{ITF}_x^{\text{out}} = \frac{\mathbf{W}_L^H \mathbf{R}_x \mathbf{W}_R}{\mathbf{W}_R^H \mathbf{R}_x \mathbf{W}_R} = \frac{\mathbf{W}_L^H \mathbf{R}_x \mathbf{W}_L}{\mathbf{W}_R^H \mathbf{R}_x \mathbf{W}_L}. \quad (21)$$

These input and output ITFs are complex valued scalars, of which the amplitude and phase can be defined as the (square root of the) *interaural level differences (ILDs)* and *interaural time differences (ITDs)*, namely,

$$\text{ILD}_x^{\text{in}} = \left[\left(\frac{\mathbf{e}_L^H \mathbf{R}_x \mathbf{e}_R}{\mathbf{e}_R^H \mathbf{R}_x \mathbf{e}_R}\right)\left(\frac{\mathbf{e}_L^H \mathbf{R}_x \mathbf{e}_L}{\mathbf{e}_R^H \mathbf{R}_x \mathbf{e}_L}\right)^*\right] = \frac{\mathbf{e}_L^H \mathbf{R}_x \mathbf{e}_L}{\mathbf{e}_R^H \mathbf{R}_x \mathbf{e}_R} \quad (22)$$

$$\text{ILD}_x^{\text{out}} = \frac{\mathbf{W}_L^H \mathbf{R}_x \mathbf{W}_L}{\mathbf{W}_R^H \mathbf{R}_x \mathbf{W}_R} \quad (23)$$

and

$$\text{ITD}_x^{\text{in}} = \angle\left(\frac{\mathbf{e}_L^H \mathbf{R}_x \mathbf{e}_R}{\mathbf{e}_R^H \mathbf{R}_x \mathbf{e}_R}\right) = \angle\left(\mathbf{e}_L^H \mathbf{R}_x \mathbf{e}_R\right) \quad (24)$$

$$\text{ITD}_x^{\text{out}} = \angle\left(\mathbf{W}_L^H \mathbf{R}_x \mathbf{W}_R\right). \quad (25)$$

Formulas (20)–(25) will be used as ITF, ILD, and ITD definitions, also in the more general case with more than one speech source. The input and output ITF, ILD, and ITD for the noise component are then also similarly defined as

$$\text{ITF}_v^{\text{in}} = \frac{\mathbf{e}_L^H \mathbf{R}_v \mathbf{e}_R}{\mathbf{e}_R^H \mathbf{R}_v \mathbf{e}_R} = \frac{\mathbf{e}_L^H \mathbf{R}_v \mathbf{e}_L}{\mathbf{e}_R^H \mathbf{R}_v \mathbf{e}_L} \quad (26)$$

$$\text{ITF}_v^{\text{out}} = \frac{\mathbf{W}_L^H \mathbf{R}_v \mathbf{W}_R}{\mathbf{W}_R^H \mathbf{R}_v \mathbf{W}_R} = \frac{\mathbf{W}_L^H \mathbf{R}_v \mathbf{W}_L}{\mathbf{W}_R^H \mathbf{R}_v \mathbf{W}_L} \quad (27)$$

$$\text{ILD}_v^{\text{in}} = \frac{\mathbf{e}_L^H \mathbf{R}_v \mathbf{e}_L}{\mathbf{e}_R^H \mathbf{R}_v \mathbf{e}_R} \quad (28)$$

$$\text{ILD}_v^{\text{out}} = \frac{\mathbf{W}_L^H \mathbf{R}_v \mathbf{W}_L}{\mathbf{W}_R^H \mathbf{R}_v \mathbf{W}_R} \quad (29)$$

$$\text{ITD}_v^{\text{in}} = \angle\left(\mathbf{e}_L^H \mathbf{R}_v \mathbf{e}_R\right) \quad (30)$$

$$\text{ITD}_v^{\text{out}} = \angle\left(\mathbf{W}_L^H \mathbf{R}_v \mathbf{W}_R\right). \quad (31)$$

The algorithms discussed in subsequent sections require the noise covariance matrix $\mathbf{R}_v$ to be of full rank, apparently excluding the interesting case with one single noise source. However, the above definitions will prove useful in practical cases with, for instance, a single dominant noise source in additive background noise such as sensor noise. Furthermore, in a practical single noise source scenario, the $\mathbf{R}_v$ matrices will generally be full-rank because of the finite dimensional discrete Fourier transforms used for the frequency domain processing.

The *ILD errors* are calculated as

$$\Delta\text{ILD}_x = 10\log_{10}(\text{ILD}_x^{\text{out}}) - 10\log_{10}(\text{ILD}_x^{\text{in}}) \quad (32)$$

$$\Delta\text{ILD}_v = 10\log_{10}(\text{ILD}_v^{\text{out}}) - 10\log_{10}(\text{ILD}_v^{\text{in}}). \quad (33)$$

The *ITD errors* are calculated as

$$\Delta\text{ITD}_x = \frac{|\text{ITD}_x^{\text{out}} - \text{ITD}_x^{\text{in}}|}{\pi} \quad (34)$$

$$\Delta\text{ITD}_v = \frac{|\text{ITD}_v^{\text{out}} - \text{ITD}_v^{\text{in}}|}{\pi}. \quad (35)$$

By this definition, ΔITD_x and ΔITD_v are relative errors which always lie between 0 and 1.

III. Multichannel Wiener Filter (MWF and SDW-MWF)

A. General Case

The binaural Multichannel Wiener Filter (**MWF**) produces a minimum-mean-square-error (MMSE) estimate of the speech component in the reference microphone of each hearing aid, hence simultaneously reducing noise and limiting speech distortion [20].

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 residual noise energy and the speech distortion energy [14]. The binaural SDW-MWF[3] cost function for the filter $\mathbf{W}_L$ estimating the speech component X_L in the reference microphone of the left hearing aid and for the filter $\mathbf{W}_R$ estimating the speech component X_R in the reference microphone of the right hearing aid is equal to

$$J_{\text{MWF}}(\mathbf{W}) = \mathcal{E}\left\{\left\|\begin{bmatrix} X_L - \mathbf{W}_L^H\mathbf{X} \\ X_R - \mathbf{W}_R^H\mathbf{X}\end{bmatrix}\right\|^2 + \mu\left\|\begin{bmatrix}\mathbf{W}_L^H\mathbf{V} \\ \mathbf{W}_R^H\mathbf{V}\end{bmatrix}\right\|^2\right\} \quad (36)$$

where μ provides a tradeoff between noise reduction and speech distortion. The optimal SDW-MWF filters for the left and right hearing aid are equal to

$$\boxed{\begin{aligned}\mathbf{W}_{\text{MWF},L} &= (\mathbf{R}_x + \mu\mathbf{R}_v)^{-1}\mathbf{R}_x\mathbf{e}_L \\ \mathbf{W}_{\text{MWF},R} &= (\mathbf{R}_x + \mu\mathbf{R}_v)^{-1}\mathbf{R}_x\mathbf{e}_R.\end{aligned}} \quad (37)$$

B. Single Speech Source

As shown in [15], in the case of a single speech source, the optimal filters are equal to

$$\boxed{\begin{aligned}\mathbf{W}_{\text{MWF},L} &= \frac{P_s\mathbf{R}_v^{-1}\mathbf{A}}{\mu+\rho}A_L^* \\ \mathbf{W}_{\text{MWF},R} &= \frac{P_s\mathbf{R}_v^{-1}\mathbf{A}}{\mu+\rho}A_R^*\end{aligned}} \quad (38)$$

with

$$\rho = P_s\mathbf{A}^H\mathbf{R}_v^{-1}\mathbf{A}. \quad (39)$$

[3]For conciseness, *SDW-MWF* is abbreviated to *MWF* in the formulas in this paper, following the same convention as in [17].

This implies that $\mathbf{W}_{\text{MWF},L}$ and $\mathbf{W}_{\text{MWF},R}$ are *parallel* [21], i.e.,

$$\boxed{\mathbf{W}_{\text{MWF},L} = \text{ITF}_x^{in,*}\,\mathbf{W}_{\text{MWF},R}} \quad (40)$$

with the interaural transfer function of the speech component ITF_x^{in} now equal to

$$\text{ITF}_x^{\text{in}} = \frac{X_L}{X_R} = \frac{A_L}{A_R}. \quad (41)$$

Using definition (14) and the fact that $\mathbf{W}_{\text{MWF},L}$ and $\mathbf{W}_{\text{MWF},R}$ are parallel (40), the output SNRs of the left and right hearing aid are the same and equal to

$$\boxed{\text{SNR}_L^{\text{out}} = \text{SNR}_R^{\text{out}} = \rho = P_s\mathbf{A}^H\mathbf{R}_v^{-1}\mathbf{A}.} \quad (42)$$

Using definition (16), the SNR improvement in the left and the right hearing aid is equal to

$$\boxed{\Delta\text{SNR}_L = \frac{\rho P_{vL}}{P_s|A_L|^2}, \quad \Delta\text{SNR}_R = \frac{\rho P_{vR}}{P_s|A_R|^2}} \quad (43)$$

with

$$P_{vL} = \mathbf{e}_L^H\mathbf{R}_v\mathbf{e}_L, \quad P_{vR} = \mathbf{e}_R^H\mathbf{R}_v\mathbf{e}_R. \quad (44)$$

As the SDW-MWF filter weight vectors for the left and the right hearing aid are parallel, the ITFs of the output speech and noise components are the same and equal to ITF_x^{in}

$$\boxed{\text{ITF}_x^{\text{out}} = \text{ITF}_v^{\text{out}} = \text{ITF}_x^{\text{in}}} \quad (45)$$

implying that all sounds (including the noise components) are perceived as coming from the speech direction.

IV. Partial Noise Estimation (MWF-η)

A. General Case

An extension of the binaural MWF that aims to partly preserve the binaural cues of the noise component has been proposed in [16] and validated through perceptual tests in [17] and [18]. The objective is to produce an MMSE estimate of the sum of the speech component and a scaled version of the noise component in the reference microphones. This partial noise estimation was also presented in [22] and [23] in single-channel noise reduction procedures. The cost function of the SDW-MWF with partial noise estimation (**MWF-η**) is equal to

$$J_{\text{MWF}\eta}(\mathbf{W}) = \mathcal{E}\left\{\left\|\begin{bmatrix} X_L - \mathbf{W}_L^H\mathbf{X} \\ X_R - \mathbf{W}_R^H\mathbf{X}\end{bmatrix}\right\|^2 + \mu\left\|\begin{bmatrix}\eta V_L - \mathbf{W}_L^H\mathbf{V} \\ \eta V_R - \mathbf{W}_R^H\mathbf{V}\end{bmatrix}\right\|^2\right\} \quad (46)$$

with the scaling parameter $0 \leq \eta \leq 1$. When $\eta = 0$, this cost function reduces to the standard SDW-MWF cost function in (36). When $\eta = 1$, the optimal filters are equal to $\mathbf{W}_{\text{MWF}\eta,L} = \mathbf{e}_L$ and $\mathbf{W}_{\text{MWF}\eta,R} = \mathbf{e}_R$, resulting in perfect preservation of the binaural cues for the speech and the noise component, but no noise reduction. The optimal filters for each hearing aid are equal to

$$\boxed{\begin{aligned}\mathbf{W}_{\mathrm{MWF}\eta,L} &= (\mathbf{R}_x + \mu\mathbf{R}_v)^{-1}(\mathbf{R}_x + \eta\mu\mathbf{R}_v)\mathbf{e}_L\\ &= (1-\eta)\mathbf{W}_{\mathrm{MWF},L} + \eta\mathbf{e}_L\\ \mathbf{W}_{\mathrm{MWF}\eta,R} &= (\mathbf{R}_x + \mu\mathbf{R}_v)^{-1}(\mathbf{R}_x + \eta\mu\mathbf{R}_v)\mathbf{e}_R\\ &= (1-\eta)\mathbf{W}_{\mathrm{MWF},R} + \eta\mathbf{e}_R.\end{aligned}} \tag{47}$$

The MWF-η procedure thus corresponds to a mixing of the output signal of the standard SDW-MWF [weighted with $(1-\eta)$] and the reference microphone signals (weighted with η).

B. Single Speech Source

Using (38), in the case of a single speech source, the optimal filters are equal to

$$\boxed{\begin{aligned}\mathbf{W}_{\mathrm{MWF}\eta,L} &= (1-\eta)\frac{P_s\mathbf{R}_v^{-1}\mathbf{A}}{\mu+\rho}A_L^* + \eta\,\mathbf{e}_L\\ \mathbf{W}_{\mathrm{MWF}\eta,R} &= (1-\eta)\frac{P_s\mathbf{R}_v^{-1}\mathbf{A}}{\mu+\rho}A_R^* + \eta\,\mathbf{e}_R.\end{aligned}} \tag{48}$$

Plugging the filters (48) in definition (14), the output SNRs for the left and right hearing aid are obtained as

$$\boxed{\begin{aligned}&\mathrm{SNR}_L^{\mathrm{out}}\\ &= \frac{(\eta\mu+\rho)^2\mathrm{SNR}_L^{\mathrm{in}}}{(1-\eta)(2\eta\mu+\rho+\eta\rho)\mathrm{SNR}_L^{\mathrm{in}} + \eta^2(\mu+\rho)^2}\\ &\mathrm{SNR}_R^{\mathrm{out}}\\ &= \frac{(\eta\mu+\rho)^2\mathrm{SNR}_R^{\mathrm{in}}}{(1-\eta)(2\eta\mu+\rho+\eta\rho)\mathrm{SNR}_R^{\mathrm{in}} + \eta^2(\mu+\rho)^2}\end{aligned}} \tag{49}$$

where

$$\mathrm{SNR}_L^{\mathrm{in}} = \frac{P_s|A_L|^2}{\mathbf{e}_L^H\mathbf{R}_v\mathbf{e}_L}, \quad \mathrm{SNR}_R^{\mathrm{in}} = \frac{P_s|A_R|^2}{\mathbf{e}_R^H\mathbf{R}_v\mathbf{e}_R}. \tag{50}$$

The output SNRs for the left and right hearing aid are now only the same if the input SNRs are the same, or if $\eta = 0$, in which case the output SNR of the SDW-MWF approach is obtained (i.e., $\mathrm{SNR}_L^{\mathrm{out}} = \mathrm{SNR}_R^{\mathrm{out}} = \rho$).

Using (49) and definition (16) and by defining $\Delta\mathrm{SNR}_L^o$ and $\Delta\mathrm{SNR}_R^o$ as the SNR improvements of the binaural SDW-MWF, cfr. (43)

$$\Delta\mathrm{SNR}_L^o = \frac{\rho P_{vL}}{P_s|A_L|^2}, \quad \Delta\mathrm{SNR}_R^o = \frac{\rho P_{vR}}{P_s|A_R|^2} \tag{51}$$

the SNR improvement at the left hearing aid is equal to

$$\Delta\mathrm{SNR}_L = \Delta\mathrm{SNR}_L^o \frac{\left(\frac{\eta\mu+\rho}{\mu+\rho}\right)^2}{\left(\frac{\eta\mu+\rho}{\mu+\rho}\right)^2 + (\Delta\mathrm{SNR}_L^o - 1)\eta^2}. \tag{52}$$

If $\eta = 0$, the SNR improvement is equal to $\Delta\mathrm{SNR}_L^o$, whereas if $\eta = 1$ no SNR improvement is obtained, i.e., $\Delta\mathrm{SNR}_L = 1$. Since the SNR improvement $\Delta\mathrm{SNR}_L^o$ of the binaural SDW-MWF is always larger than or equal to 1 [24], it can easily be shown that

$$\boxed{1 \le \Delta\mathrm{SNR}_L \le \Delta\mathrm{SNR}_L^o.} \tag{53}$$

Similar expressions can be derived for the SNR improvement at the right hearing aid, also leading to $1 \le \Delta\mathrm{SNR}_R \le \Delta\mathrm{SNR}_R^o$.

If ρ is sufficiently large[4], i.e., $\rho \gg \mu,\ \eta,\ \mathrm{SNR}_L^{\mathrm{in}}$, and $\eta \neq 0$, then the SNR improvement for the left hearing aid in (52) becomes approximately equal to

$$\lim_{\rho\to\infty} \Delta\mathrm{SNR}_L = \frac{1}{\eta^2}. \tag{54}$$

Similarly, if $\rho \gg \mu,\ \eta,\ \mathrm{SNR}_R^{\mathrm{in}}$, the SNR improvement in the right hearing aid will be approximately equal to $1/\eta^2$.

From (48) it can be seen that the MWF-η filters are generally *not parallel*, such that the ITF of the output speech and noise components are typically different. Using (17), the ITF of the output speech component is equal to

$$\begin{aligned}\mathrm{ITF}_x^{\mathrm{out}} &= \frac{\mathbf{W}_{\mathrm{MWF}\eta,L}^H\mathbf{A}}{\mathbf{W}_{\mathrm{MWF}\eta,R}^H\mathbf{A}}\\ &= \frac{(1-\eta)\frac{\rho}{\mu+\rho}A_L + \eta A_L}{(1-\eta)\frac{\rho}{\mu+\rho}A_R + \eta A_R} = \frac{A_L}{A_R}\end{aligned} \tag{55}$$

and so

$$\boxed{\mathrm{ITF}_x^{\mathrm{out}} = \mathrm{ITF}_x^{\mathrm{in}}.} \tag{56}$$

such that the ITF of the speech component is again preserved.

By plugging the filters (48) into definition (27), the output noise ITF is obtained as

$$\begin{aligned}\mathrm{ITF}_v^{\mathrm{out}} &= \frac{\mathbf{W}_L^H\mathbf{R}_v\mathbf{W}_R}{\mathbf{W}_R^H\mathbf{R}_v\mathbf{W}_R}\\ &= \frac{\psi(P_sA_LA_R^*) + \eta^2(\mathbf{e}_L^H\mathbf{R}_v\mathbf{e}_R)}{\psi(P_s|A_R|^2) + \eta^2(\mathbf{e}_R^H\mathbf{R}_v\mathbf{e}_R)}\end{aligned} \tag{57}$$

where

$$\psi = \left[(1-\eta)^2\frac{\rho}{(\mu+\rho)^2} + 2\eta(1-\eta)\frac{1}{(\mu+\rho)}\right]. \tag{58}$$

By using definitions (19), (26), and (50), it can be shown that (57) is equal to

$$\boxed{\mathrm{ITF}_v^{\mathrm{out}} = \frac{\psi\,\mathrm{SNR}_R^{\mathrm{in}}}{\psi\,\mathrm{SNR}_R^{\mathrm{in}}+\eta^2}\,\mathrm{ITF}_x^{\mathrm{in}} + \frac{\eta^2}{\psi\,\mathrm{SNR}_R^{\mathrm{in}}+\eta^2}\,\mathrm{ITF}_v^{\mathrm{in}}.} \tag{59}$$

Equation (59) shows that $\mathrm{ITF}_v^{\mathrm{out}}$ is a weighted sum of $\mathrm{ITF}_x^{\mathrm{in}}$ and $\mathrm{ITF}_v^{\mathrm{in}}$. If $\eta = 1$, the factor ψ is equal to 0 so that $\mathrm{ITF}_v^{\mathrm{out}} = \mathrm{ITF}_v^{\mathrm{in}}$, whereas for $\eta = 0$, $\mathrm{ITF}_v^{\mathrm{out}} = \mathrm{ITF}_x^{\mathrm{in}}$.

[4]This is the case when the spatial separation between the speech source and the noise sources is sufficiently large, in which case the product $\mathbf{A}^H\mathbf{R}_v^{-1}\mathbf{A}$ is large. A high speech power P_s also leads to a large ρ, but then the assumption $\rho \gg \mathrm{SNR}_L^{\mathrm{in}}$ may not be valid as the input SNR is also high.

By rearranging the terms in (59) and using (49), the output ITF can also be related to the SNR improvement

$$\mathrm{ITF}_v^{\mathrm{out}} = \mathrm{ITF}_x^{\mathrm{in}} + \frac{\eta^2(\mu+\rho)^2}{(\eta\mu+\rho)^2}\,\Delta\mathrm{SNR}_R\,(\mathrm{ITF}_v^{\mathrm{in}} - \mathrm{ITF}_x^{\mathrm{in}}). \quad (60)$$

Under the assumptions leading to (54), the obtained noise output ITF thus becomes

$$\mathrm{ITF}_v^{\mathrm{out}} \approx \mathrm{ITF}_x^{\mathrm{in}} + \frac{(\mu+\rho)^2}{(\eta\mu+\rho)^2}\,(\mathrm{ITF}_v^{\mathrm{in}} - \mathrm{ITF}_x^{\mathrm{in}}) \approx \mathrm{ITF}_v^{\mathrm{in}} \quad (61)$$

so it is seen that the error on the noise binaural cues will become small for any choice of $\eta \neq 0$ under these assumptions.

V. Interaural Transfer Function Extension (MWF-ITF)

A. General Case

To control the binaural cues of the speech and the noise component, it is also possible to extend the binaural SDW-MWF cost function with terms related to the ITF of the speech and the noise components, as has been proposed in [15], [19], and [25]. When the aim is to preserve the binaural cues of the speech and the noise components, the desired output ITFs are equal to the input ITFs in (20) and (26).

In [15], the ITF cost function for the noise component is defined as

$$J_{\mathrm{ITF},1}^{v}(\mathbf{W}) = \frac{\mathcal{E}\{|\mathbf{W}_L^H\mathbf{V} - \mathrm{ITF}_v^{\mathrm{in}}\,\mathbf{W}_R^H\mathbf{V}|^2\}}{\mathcal{E}\{|\mathbf{W}_R^H\mathbf{V}|^2\}} = \frac{\mathbf{W}^H\mathbf{R}_{vt}\mathbf{W}}{\mathbf{W}^H\mathbf{R}_{v1}\mathbf{W}} \quad (62)$$

with

$$\mathbf{R}_{vt} = \begin{bmatrix} \mathbf{R}_v & -\mathrm{ITF}_v^{in,*}\,\mathbf{R}_v \\ -\mathrm{ITF}_v^{\mathrm{in}}\,\mathbf{R}_v & |\mathrm{ITF}_v^{\mathrm{in}}|^2\,\mathbf{R}_v \end{bmatrix} \quad (63)$$

$$= \begin{bmatrix} \mathbf{I}_{2M} \\ -\mathrm{ITF}_v^{\mathrm{in}}\,\mathbf{I}_{2M} \end{bmatrix} \mathbf{R}_v \begin{bmatrix} \mathbf{I}_{2M} & -\mathrm{ITF}_v^{in,*}\,\mathbf{I}_{2M} \end{bmatrix} \quad (64)$$

$$\mathbf{R}_{v1} = \begin{bmatrix} \mathbf{0}_{2M} & \mathbf{0}_{2M} \\ \mathbf{0}_{2M} & \mathbf{R}_v \end{bmatrix} \quad (65)$$

where $\mathbf{0}_{2M}$ is a $2M \times 2M$ all-zero matrix and $\mathbf{I}_{2M}$ is the $2M \times 2M$ unity matrix. For mathematical convenience, we also introduce a *simplified quadratic ITF cost function* which is obtained by removing the denominator in (62), corresponding to the output noise power in the right hearing aid, i.e.,

$$J_{\mathrm{ITF},2}^{v}(\mathbf{W}) = \mathcal{E}\{|\mathbf{W}_L^H\mathbf{V} - \mathrm{ITF}_v^{\mathrm{in}}\,\mathbf{W}_R^H\mathbf{V}|^2\} = \mathbf{W}^H\mathbf{R}_{vt}\mathbf{W}. \quad (66)$$

The ITF cost function for the speech component is defined similarly as the ITF cost function for the noise component, i.e.,

$$J_{\mathrm{ITF},1}^{x}(\mathbf{W}) = \frac{\mathbf{W}^H\mathbf{R}_{xt}\mathbf{W}}{\mathbf{W}^H\mathbf{R}_{x1}\mathbf{W}} \quad (67)$$

$$J_{\mathrm{ITF},2}^{x}(\mathbf{W}) = \mathbf{W}^H\mathbf{R}_{xt}\mathbf{W}. \quad (68)$$

The SDW-MWF with interaural transfer function extension (**MWF-ITF**) thus minimizes the overall cost function which trades off noise reduction, speech distortion and binaural cue preservation, i.e.,

$$J_{\mathrm{MWF-ITF}}(\mathbf{W}) = J_{\mathrm{MWF}}(\mathbf{W}) + \gamma J_{\mathrm{ITF}}^{x}(\mathbf{W}) + \delta J_{\mathrm{ITF}}^{v}(\mathbf{W}) \quad (69)$$

where the parameters γ and δ allow to put more emphasis on binaural cue preservation for the speech and the noise component. The cost function $J_{\mathrm{MWF}}(\mathbf{W})$ is defined in (36) and for the ITF cost functions we can either use the cost functions defined in (62) and (67) or the simplified cost functions defined in (66) and (68).

When using the ITF cost functions $J_{\mathrm{ITF},1}^{x}(\mathbf{W})$ and $J_{\mathrm{ITF},1}^{v}(\mathbf{W})$, no closed-form expression is available for the filter minimizing $J_{\mathrm{MWF-ITF}}(\mathbf{W})$, such that we have to use iterative, e.g., quasi-Newton, optimization techniques [26], [27].

When using the simplified quadratic ITF cost functions $J_{\mathrm{ITF},2}^{x}(\mathbf{W})$ and $J_{\mathrm{ITF},2}^{v}(\mathbf{W})$, the filter minimizing $J_{\mathrm{MWF-ITF}}(\mathbf{W})$ is equal to

$$\boxed{\mathbf{W}_{\mathrm{MWF-ITF},2} = (\mathbf{R} + \gamma\mathbf{R}_{xt} + \delta\mathbf{R}_{vt})^{-1}\mathbf{r}_x} \quad (70)$$

with

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_x + \mu\mathbf{R}_v & \mathbf{0}_M \\ \mathbf{0}_M & \mathbf{R}_x + \mu\mathbf{R}_v \end{bmatrix}, \quad \mathbf{r}_x = \begin{bmatrix} \mathbf{R}_x\mathbf{e}_L \\ \mathbf{R}_x\mathbf{e}_R \end{bmatrix}. \quad (71)$$

B. Single Speech Source

In the case of a single speech source, the filter $\mathbf{W}_{\mathrm{MWF-ITF},2}$ in (70) can be reduced (cfr. Appendix) to (72), as shown at the bottom of the page, with

$$\xi = \frac{\delta}{\mu + \rho + \delta(1 + |\mathrm{ITF}_v^{\mathrm{in}}|^2)} \quad (73)$$

$$\alpha_t = 1 + \mathrm{ITF}_v^{in,*}\mathrm{ITF}_x^{\mathrm{in}} \quad (74)$$

$$\boxed{\begin{aligned} \mathbf{W}_{\mathrm{MWF-ITF},2,L} &= \frac{P_s}{\mu+\rho}\Big\{A_L^* - \xi(A_L^* - \mathrm{ITF}_v^{in,*}A_R^*)\big[1 - \nu\alpha_t^*(1-\xi\alpha_t)\big]\Big\}\mathbf{R}_v^{-1}\mathbf{A} \\ \mathbf{W}_{\mathrm{MWF-ITF},2,R} &= \frac{P_s}{\mu+\rho}\Big\{A_R^* + \xi\mathrm{ITF}_v^{\mathrm{in}}(A_L^* - \mathrm{ITF}_v^{in,*}A_R^*)\Big[1 - \nu\alpha_t^*\Big(\frac{\mathrm{ITF}_x^{\mathrm{in}}}{\mathrm{ITF}_v^{\mathrm{in}}} - \xi\alpha_t\Big)\Big]\Big\}\mathbf{R}_v^{-1}\mathbf{A} \end{aligned}} \quad (72)$$

$$\nu = \frac{\gamma\rho}{\mu + \rho + \gamma\rho(1 + |\mathrm{ITF}_x^{\mathrm{in}}|^2 - \xi|\alpha_t|^2)}. \quad (75)$$

Hence, the filter $\mathbf{W}_{\mathrm{MWF-ITF},2,L}$ is equal to a scaled version of $\mathbf{W}_{\mathrm{MWF},L}$ in (38) and likewise $\mathbf{W}_{\mathrm{MWF-ITF},2,R}$ is equal to a scaled version of $\mathbf{W}_{\mathrm{MWF},R}$. The scaling factors reflect the extensions with the ITF cost functions. As a consequence, the output SNR will be the same for the left and the right hearing aid, and will be equal to the output SNR of the SDW-MWF (cf. (42)), i.e.,

$$\boxed{\mathrm{SNR}_L^{\mathrm{out}} = \mathrm{SNR}_R^{\mathrm{out}} = \rho = P_s\mathbf{A}^H\mathbf{R}_v^{-1}\mathbf{A}} \quad (76)$$

and the SNR improvements are given by

$$\boxed{\Delta\mathrm{SNR}_L = \frac{\rho P_{vL}}{P_s|A_L|^2}, \quad \Delta\mathrm{SNR}_R = \frac{\rho P_{vR}}{P_s|A_R|^2}} \quad (77)$$

which is the same as (43).

As the filters $\mathbf{W}_{\mathrm{MWF-ITF},2,L}$ and $\mathbf{W}_{\mathrm{MWF-ITF},2,R}$ are still *parallel*, the ITFs of the output speech and noise components are the same and result in (78), as shown at the bottom of the page. It can be shown that for $\gamma \rightarrow \infty$ and $\delta \neq 0$, $\mathrm{ITF}_x^{\mathrm{out}} = \mathrm{ITF}_v^{\mathrm{out}} = \mathrm{ITF}_x^{\mathrm{in}}$ and that for $\delta \rightarrow \infty$ and $\gamma \neq 0$, $\mathrm{ITF}_x^{\mathrm{out}} = \mathrm{ITF}_v^{\mathrm{out}} = \mathrm{ITF}_v^{\mathrm{in}}$.

Hence, extending the SDW-MWF with the ITF cost functions gives rise to the same SNR improvement and output SNR as the binaural SDW-MWF, but changes the ITF of the output components. It is also possible to link the output SNR with the ITF of the output components. Consider two frequencies, ω_1 and ω_2, where the output SNR for ω_1 is larger than the output SNR for ω_2, i.e., $\rho_1 > \rho_2$ (ρ_1 and ρ_2 are given by (39) at frequencies ω_1 and ω_2, respectively). Hence, when using the same value for δ, $\xi_1 < \xi_2$ (where ξ_1 and ξ_2 are given by (73) at frequencies ω_1 and ω_2, respectively), such that for ω_1 the output ITF is closer to the ITF of the speech component than for ω_2. This is an advantageous perceptual effect that will be illustrated in simulations in Section VI.

VI. Experimental Results

In this section, the performance of the MWF-η and MWF-ITF algorithms will be tested in a scenario with one speech source and one noise source. First, the experimental setup will be discussed.

A. Data Model

The sources are located in the far-field of the microphone arrays in a non-reverberant environment. It is assumed that there is one speech and one noise source, and that they are located at angles θ_x and θ_v from the head ($\theta = 0°$: front, $\theta = 90°$: right), with an elevation $\phi = 0°$. The speech and noise components of the microphone signals can thus be written as

$$\mathbf{X}(\omega) = \mathbf{d}(\omega, \theta_x)S(\omega), \quad \mathbf{V}(\omega) = \mathbf{d}(\omega, \theta_v)V(\omega) \quad (79)$$

with the steering vector $\mathbf{d}(\omega, \theta)$ equal to (80), as shown at the bottom of the page. The (omnidirectional) microphones are located on a head, so the head shadow effect will be taken into account. To achieve this, head-related transfer functions (HRTFs) measured on a KEMAR dummy-head [28] are incorporated in the steering vectors. It is assumed that the same HRTF can be used for all the microphones at the left hearing aid, so that for example an entry $d_{L,m}(\omega, \theta)$, representing the mth microphone of the left hearing aid, at an angle θ, can be calculated as

$$d_{L,m}(\omega, \theta) = HRTF_L(\omega, \theta)e^{-j\omega\tau_{L,m}(\theta)} \quad (81)$$

where $\tau_{L,m}(\theta)$ represents the delay between the mth microphone and the reference point at the left hearing aid. The steering vector entries for the right hearing aid are obtained in a similar way (using $HRTF_R$ data and $\tau_{R,m}(\theta)$). The speech and noise correlation matrices are constructed as

$$\mathbf{R}_x(\omega) = P_s\ \mathbf{d}(\omega, \theta_x)\mathbf{d}^H(\omega, \theta_x) \quad (82)$$
$$\mathbf{R}_v(\omega) = P_v\ \mathbf{d}(\omega, \theta_v)\mathbf{d}^H(\omega, \theta_v) + P_{vs}\ \mathbf{I}_{2M}. \quad (83)$$

The parameters P_s, P_v, and P_{vs} represent the powers of the speech source, (located) noise source and (internal) sensor noise. Some sensor noise, modeled as spatially uncorrelated noise, is added in (83) to add a degree of realism and also to make the noise correlation matrix invertible.

The experiments are performed using a speech source at $-5°$ and a noise source at 40°. A two-microphone array is used on both the left and the right hearing aid. The microphone distance on the left hearing aid is 2 cm, whereas the right hearing aid has a microphone distance of 1.5 cm. The algorithms are tested at a frequency of $\omega = 2\pi \times 2000$ rad/s, the HRTF data is sampled at 44 100 Hz. The parameter μ in the SDW-MWF cost function (36) is set equal to 1.

$$\boxed{\mathrm{ITF}_x^{\mathrm{out}} = \mathrm{ITF}_v^{\mathrm{out}} = \frac{\mathrm{ITF}_x^{\mathrm{in}} - \xi(\mathrm{ITF}_x^{\mathrm{in}} - \mathrm{ITF}_v^{\mathrm{in}})\left[1 - \nu\alpha_t(1 - \xi\alpha_t^*)\right]}{1 + \xi\mathrm{ITF}_v^{in,*}(\mathrm{ITF}_x^{\mathrm{in}} - \mathrm{ITF}_v^{\mathrm{in}})\left[1 - \nu\alpha_t\left(\frac{\mathrm{ITF}_x^{in,*}}{\mathrm{ITF}_v^{in,*}} - \xi\alpha_t^*\right)\right]}} \quad (78)$$

$$\mathbf{d}(\omega, \theta) = \left[d_{L,1}(\omega, \theta)\ \ldots\ d_{L,M}(\omega, \theta)\ d_{R,1}(\omega, \theta)\ \ldots\ d_{R,M}(\omega, \theta)\right]^T \quad (80)$$

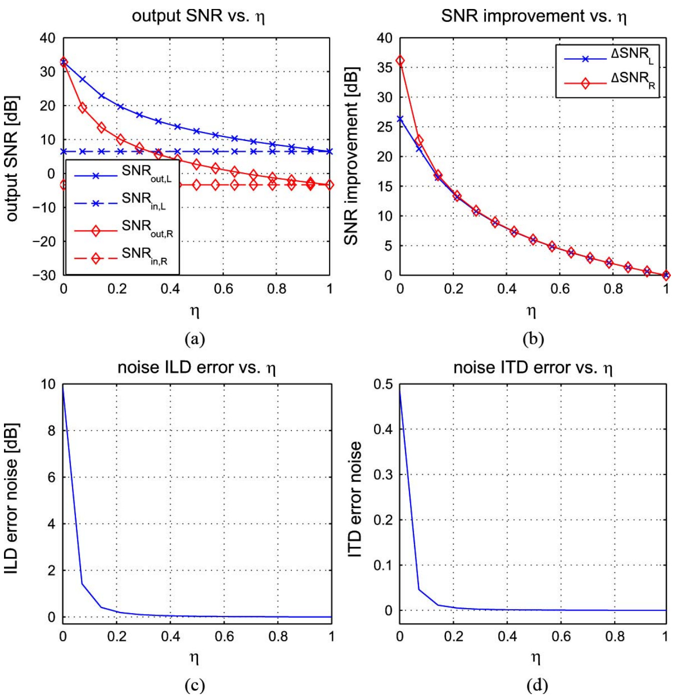


Fig. 2. Dependence of output SNR (a), SNR improvement (b), and noise binaural cues [ILD (c) and ITD (d)] on partial noise parameter η. Larger values of η lead to lower ILD/ITD errors, but also decrease the noise reduction performance.

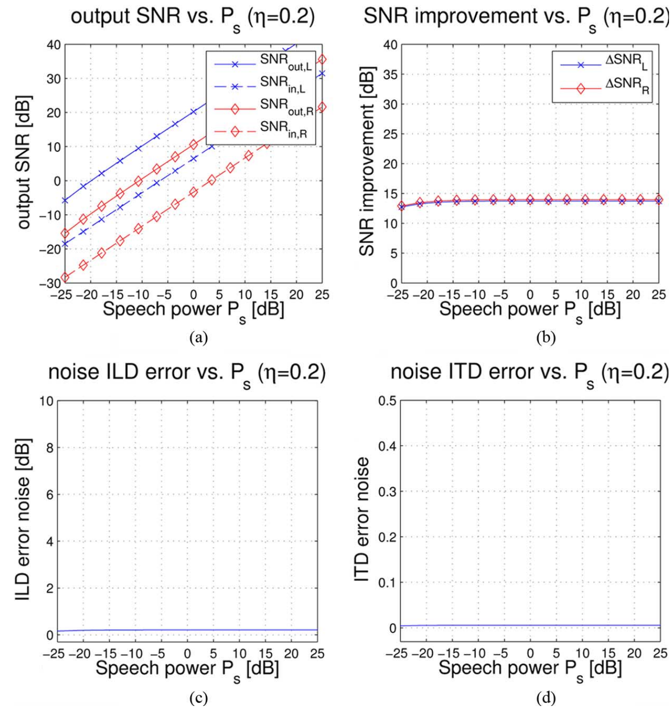


Fig. 3. Dependence of output SNR (a), SNR improvement (b), and noise binaural cues [ILD (c) and ITD (d)] on signal power P_s. The signal power has only a small influence on the noise reduction performance and on the binaural noise cue preservation.

B. Performance of MWF-η

The performance of the MWF-η algorithm is now tested with the data model of the previous section. As the ITF of the speech component is always preserved for any choice of η, as was shown in (56), only the noise ITD and ILD errors will be shown. The performance measures (32) and (34) are used for the ILD and ITD errors. The ILD error can thus be expressed in dB, while the ITD error is a relative error which lies between 0 and 1.

1) Dependence On Parameter η: The speech and noise powers in (82) and (83) will be fixed to $P_s = P_v = 1$, the sensor noise power is fixed to $P_{vs} = 0.01$. In Fig. 2, the output SNRs, SNR improvements (for left and right hearing aid) and the noise cue errors are shown for different values of the parameter η.

The case $\eta = 0$ corresponds to the binaural SDW-MWF solution. It can be seen that the ITD/ILD errors of the noise component are large. It was indeed shown in Section III-B and in [15] that the binaural SDW-MWF algorithm distorts the noise cues.

For larger values of η, the ILD/ITD errors of the noise component decrease. However, as more noise is mixed into the output signals, the obtained SNR improvement will also decrease, so that there is a tradeoff between SNR improvement and binaural cue preservation. In these simulations the factor ρ (39) is sufficiently large so that the assumptions in Section IV-B are valid and the SNR improvement (for both hearing aids) is approximately equal to $\Delta\text{SNR}_{L,R} = 1/\eta^2$, as in (54).

A large loss in SNR improvement can be seen for a choice of $\eta = 0.2$ compared to the SDW-MWF case $(\eta = 0)$. It does not seem possible for the binaural unmasking effect [13] to compensate for this loss. The SNR improvement obtained in a more realistic setup is however smaller than the theoretical limits shown here, so that it is expected that the real-life absolute SNR loss will be smaller than the loss shown here, and so that the compensatory effect of the binaural unmasking can be sufficient. Perceptual tests using this parameter setting [18] have indeed shown an improvement in speech intelligibility for some of the speech-noise configurations.

2) Dependence on Signal Power p_s: In Fig. 3 the signal power P_s is varied (dB values are relative to noise power P_v), while P_v and P_{vs} are fixed to 1 and 0.01, respectively. The partial noise parameter η is fixed to 0.2.

The results show that for the ILD/ITD errors and for the SNR improvement, there is no significant dependence on the signal power P_s. The obtained SNR improvement can be predicted by (54), which gives an SNR improvement of $10\log_{10}(1/0.2^2) = 13.98$ dB. The ITD/ILD errors are relatively small for all values of P_s, so that $\eta = 0.2$ seems to be sufficient for this scenario. Perceptual tests in [17] indeed show that for $\eta = 0.2$, both speech and noise sources are localized correctly.

3) Dependence on Noise Power p_v: In Fig. 4, the noise power P_v is varied (dB values are relative to signal power P_s), while P_s and P_{vs} are fixed to 1 and 0.01, respectively. The partial noise parameter η is fixed to 0.2.

Higher noise powers P_v lead to lower input and output SNRs whereas the SNR improvement is again given by (54). The ITD/ILD errors of the noise component are small when the obtained output SNR is low, which is in fact an advantageous perceptual effect: in frequency bins with a lot of residual noise, this noise will be perceived in the correct direction.

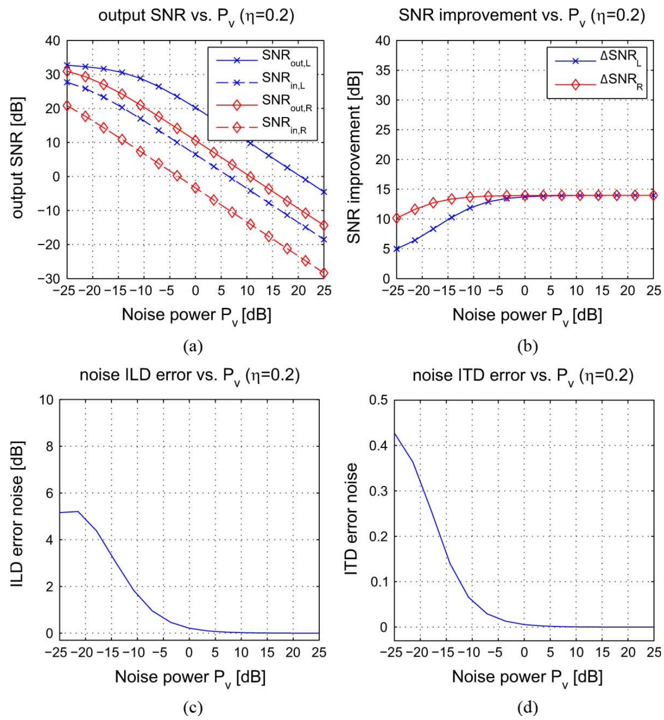


Fig. 4. Dependence of output SNR (a), SNR improvement (b), and noise binaural cues [ILD (c) and ITD (d)] on noise power P_v. When the output SNR is low, the binaural noise cues are better preserved, which is an advantageous perceptual effect.

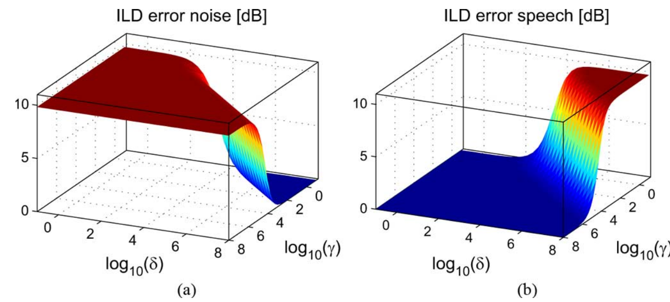


Fig. 5. MWF-ITF with quadratic ITF extension; The ILD errors for (a) the noise and (b) speech components are shown for different values of γ and δ.

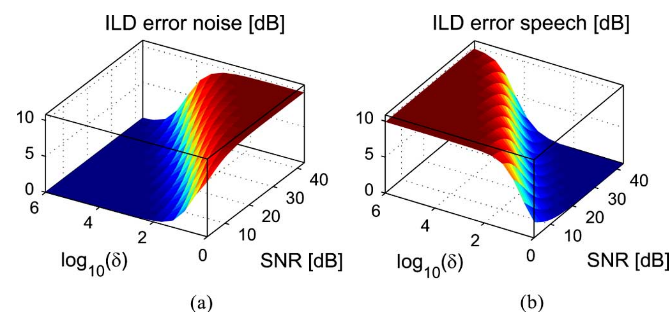


Fig. 6. MWF-ITF with quadratic ITF extension; The ILD errors for (a) the noise and (b) speech components are shown as a function of the ITF parameter δ and the output SNR.

At lower noise powers, the assumptions leading to (54) are no longer valid (the input SNRs are high) so that the SNR improvement is lower than predicted in (54). As seen in (60), $\text{ITF}_v^{\text{out}}$ will be shifted towards ITF_x^{in} if the SNR improvement is low, which leads to higher ILD/ITD errors. However, these higher errors are still acceptable because the output SNR is high. So, the residual noise will then be masked by the speech signal.

C. *Performance of MWF-ITF*

In this section, the performance of the MWF-ITF algorithm is tested. In [15], it was shown that the algorithm with non-quadratic ITF extension (62) allows preservation of both speech as noise binaural cues. However, as for the MWF-η algorithm, the obtained SNR improvement will decrease if more emphasis is put on cue preservation.

Because the MWF-ITF algorithm with non-quadratic extension is not computationally feasible for an hearing aid application, we will use the quadratic ITF extension in (66) from now on. The results obtained in [29] will be reviewed here.

1) ITD and ILD Error for Quadratic ITF Extension: In Fig. 5, the ILD errors of speech and noise components for the SDW-MWF with quadratic ITF extension (66) are shown. The ILD error for the noise component is shown on the left, the ILD error for the speech component is shown on the right. The ITD errors are omitted here, but lead to similar conclusions [29]. For some choices of the ITF parameters γ and δ, the ILD errors on the noise component can be made arbitrarily small. However, the ILD errors on the speech component will then become large. On the other hand, when the ILD errors on the speech component are made small, the ILD errors on the noise component are large. An optimal choice of parameters, where the ILD errors of both noise and speech are small appears to be impossible. These results are in accordance with the theoretical discussion in Section V, where it was shown that the output ITFs of speech and noise component are equal, so that speech and noise cues cannot be preserved simultaneously.

2) Output ITF Versus Output SNR: The theoretical discussion and the previous simulations showed that it is impossible to preserve speech and noise binaural cues simultaneously. As such, this approach may seem inappropriate as a binaural noise reduction algorithm, as it would be impossible to correctly localize both the speech and the noise source. However, in [19] the MWF-ITF algorithm (with quadratic ITF cost terms), was validated perceptually. An improvement in the total localization performance (speech+noise) was observed, which seems to contradict the previous discussion. To explain the improvement, the relationship between the output SNR and the output ITF will be analyzed.

In Fig. 6, the output SNR and the noise ITF parameter δ are varied. γ is fixed to 0 in this simulation. To vary the output SNR, the signal power P_s will be varied which changes the output SNR as seen in (39). As in the previous section, the noise and speech ILD errors are shown. It can be seen that for certain values of the ITF parameter δ, the ILD error of the noise component is small at low output SNRs, while the speech ILD error is small at high output SNRs. The ITD errors, which are omitted here, behave similarly [29]. These results show that the output ITF is shifted towards the input ITF of the noise component when the output SNR is low, while the output ITF is shifted towards the input ITF of the speech component in high SNR regions.

When the algorithm is applied on broadband signals, as in [19], the obtained output SNRs will vary in the different frequency bins, and similarly, the output ITFs will vary. This in fact represents an advantageous perceptual effect: in frequency bins with a low output SNR, the ITF is shifted towards the noise ITF, so that the residual noise in the output signals can still be heard in the noise direction, and vice versa for the speech component.

Although MWF-ITF thus allows for correct localization, a benefit in speech intelligibility due to binaural unmasking was not observed in the perceptual tests in [19]. This can be expected as the binaural unmasking effect relies on differences in the binaural cues of target and interferer in the same frequency bin, which is not achieved by MWF-ITF. Only when the interfering source is speech, perceiving target and interferer in different locations (for example by the advantageous perceptual effect of MWF-ITF) leads to speech intelligibility improvements [30], but not for speech-shaped noise as interfering source, as was the case in [19].

D. Comparison of SDW-MWF, MWF-ITF, and MWF-η in a Reverberant Room

The broadband performances of the SDW-MWF, MWF-η and MWF-ITF algorithms are now compared in a reverberant environment. HRTFs were measured in a reverberant room ($\mathrm{RT}_{60} = 500$ ms) on a binaural microphone array mounted on a dummy-head, so that the head-shadow effect is taken into account. To generate the microphone signals, the speech and noise signals are convolved with the measured HRTFs corresponding to their angles of arrival ($\theta_x = -15^\circ$, $\theta_v = 45^\circ$), and then added together. To further increase the degree of realism, the correlation matrices are estimated in a batch procedure on the speech and noise signals instead of being constructed with the steering vectors. The speech signal consists of four sentences of the Hearing In Noise Test (HINT) list [31], the noise signal is multitalker babble noise [32]. The microphone signals have a total duration of 26 s, and are sampled at 8000 Hz.

In practice, a voice activity detector (VAD) has to be implemented to distinguish between segments were speech and noise are both active ($\mathbf{R}_y$ is updated), and segments were only noise is active ($\mathbf{R}_v$ is updated), but here a perfect VAD is assumed. The correlation matrix estimates are plugged into (37), (47), and (70) to obtain the optimal filter coefficients of SDW-MWF, MWF-η and MWF-ITF, respectively. The speech distortion tradeoff parameter μ is set to 5, the MWF-ITF parameters are $\gamma = 100$ and $\delta = 300$, and the MWF-η parameter is $\eta = 0.2$. The signals are processed by 64-point fast Fourier transforms (FFTs), and the performance measures defined in Section II-C are then calculated for every frequency bin independently.

Fig. 7 shows the average output SNR per frequency bin of the SDW-MWF, MWF-ITF, and MWF-η. The SNRs of the left and right output signals were averaged to obtain the average SNRs shown in the figure. As a reference, the (unprocessed) input SNR of this setup is also shown. It can be observed that every procedure achieves an SNR improvement over the unprocessed case. As expected, the MWF-η algorithm obtains the lowest output SNR. The SDW-MWF and MWF-ITF procedures obtain similar output SNRs as predicted by the theory, except for a few

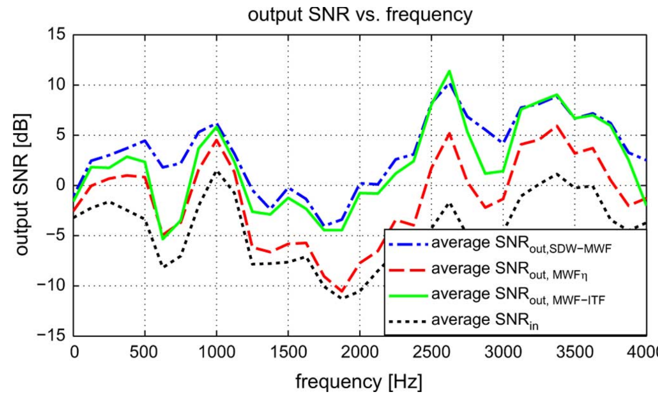
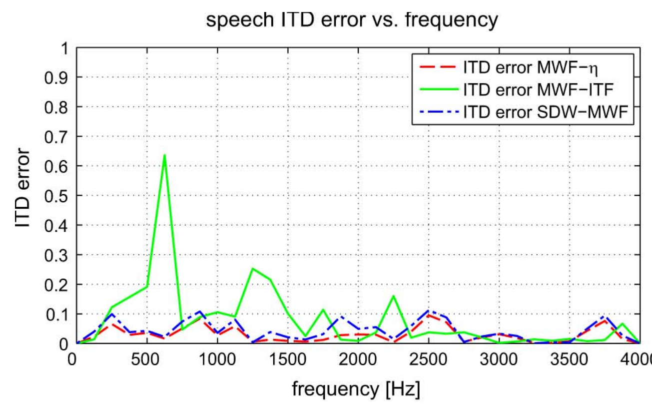
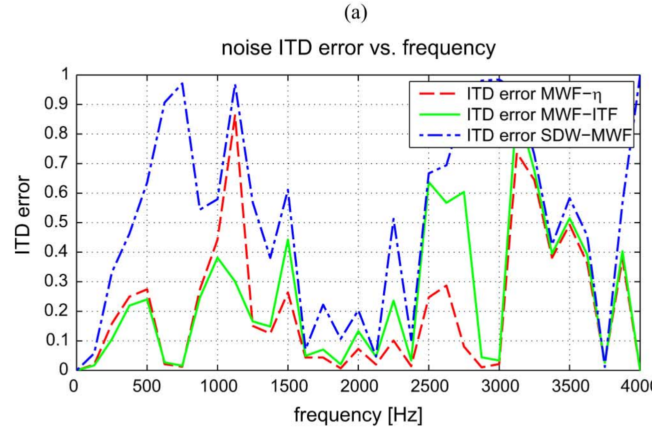


Fig. 7. Performance comparison of MWF-ITF ($\gamma = 100$ and $\delta = 300$), MWF-η ($\eta = 0.2$) and SDW-MWF in a reverberant room. The average output SNRs and average input SNR per frequency are shown.



(a)



(b)

Fig. 8. Performance comparison of MWF-ITF ($\gamma = 100$ and $\delta = 300$), MWF-η ($\eta = 0.2$) and SDW-MWF in a reverberant room. (a) The speech ITD errors and (b) noise ITD errors per frequency are shown.

bins where some dips in the output SNR curves occur (around 500, 1250, 1750, and 3000 Hz).

Fig. 8 shows the relative speech and noise ITD errors (34) for the three procedures, as a function of frequency. According to (45) and (56), both SDW-MWF as MWF-η do not distort the speech binaural cues. The speech ITD errors in Fig. 8(a) are indeed small for these procedures. In addition, MWF-η is expected to obtain the smallest noise ITD error, whereas SDW-MWF is expected to obtain a large noise ITD error, which is illustrated in Fig. 8(b). In the theoretical analysis, it was furthermore shown that the output ITF of the MWF-ITF procedure depends on the obtained output SNR. This advantageous perceptual effect is also illustrated by Fig. 8: at the frequency bins where dips in the obtained output SNR in Fig. 7 occur, a large speech ITD error is seen, whereas the noise ITD error is small. Thus, the output ITF is effectively shifted towards the noise ITF at frequencies with a low output SNR, which improves localization of the noise source.

VII. Conclusion

In this paper, a class of binaural noise reduction algorithms based on multichannel Wiener filtering (SDW-MWF) has been discussed. To fully exploit the advantage of binaural hearing, the preservation of binaural cues is an important objective, in addition to noise reduction. Perceptual tests in previous work have shown that two extensions of the SDW-MWF, namely MWF-ITF and MWF-η, can lead to localization and speech intelligibility improvements. In this paper, the binaural cue preservation performance of these procedures has been analyzed theoretically and validated with objective performance measures.

The binaural SDW-MWF algorithm was shown to be inadequate as a binaural noise reduction strategy, if binaural cues are to be preserved; both the speech and residual noise will be perceived in the direction of the speech signal, so that the hearing aid user cannot rely on binaural unmasking to improve speech understanding.

The binaural MWF-η algorithm is an extension that mixes a fraction of the original noise signal into the output signals. As a drawback, the SNR improvement decreases. However, this approach makes it possible to preserve both speech and noise binaural cues. These findings support the perceptual tests in [17], where an improvement in localization performance was observed. Furthermore, correct localization can lead to a benefit in speech intelligibility if the compensatory effect of binaural unmasking is larger than the SNR loss [18].

The MWF-ITF algorithm extends the binaural SDW-MWF cost function with terms related to the ITFs of the speech and noise component. These terms are simplified to quadratic terms to make the algorithm computationally feasible. It has been shown that the obtained output SNR remains equal to the output SNR of the binaural SDW-MWF algorithm. While it is impossible to preserve the speech and noise cues at the same time, there is an advantageous perceptual effect: if the output SNR is high, the obtained ITF will be shifted towards the ITF of the speech component, whereas for a low output SNR, the ITF will be shifted towards the ITF of the noise component. This effect makes a correct localization possible, and explains the localization performance improvement in prior perceptual tests [19]. As the binaural unmasking effect requires different speech and noise ITFs in a single frequency bin, a speech intelligibility improvement due to binaural unmasking is not possible however. Another drawback of MWF-ITF is the fact that it is not straightforward to find an optimal γ and δ parameter setting for different scenarios.

TABLE I
COMPARISON OF SDW-MWF, MWF-η, MWF-ITF

	SDW-MWF	**MWF-η**	**MWF-ITF**
Advantages	• High noise reduction performance • No parameter tuning • Preservation of speech cues	• Preservation of both speech and noise cues • Easy parameter tuning	• High noise reduction performance • Advantageous perceptual effect
Disadvantages	• Speech and noise components perceived in same direction • Noise cues cannot be preserved	• Lower noise reduction performance	• Speech and noise components with the same ITF's • Difficult parameter tuning

The advantages and disadvantages of the different algorithms are summarized in Table I.

Appendix

In this section, the MWF-ITF optimal filters are derived for the special case of a single speech source. First, the derivation will be made for the special case $\gamma = 0$. This result will then be reused in the derivation of the general case $\gamma \neq 0$.

A) ITF Cost Function for Only Noise Component ($\gamma = 0$)*:* The filter in (70) reduces to

$$\mathbf{W}_{\mathrm{MWF-ITF},2} = (\mathbf{R} + \delta\mathbf{R}_{vt})^{-1}\mathbf{r}_x. \tag{84}$$

For easier notation, we define $\alpha_v = -\mathrm{ITF}_v^{\mathrm{in}}$, such that (64) can be written as

$$\mathbf{R}_{vt} = \begin{bmatrix} \mathbf{I}_{2M} \\ \alpha_v \mathbf{I}_{2M} \end{bmatrix} \mathbf{R}_v \left[\mathbf{I}_{2M} \quad \alpha_v^* \mathbf{I}_{2M} \right]. \tag{85}$$

Using the matrix inversion lemma

$$(\mathbf{A} + \mathbf{BDC})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left[\mathbf{CA}^{-1}\mathbf{B} + \mathbf{D}^{-1}\right]^{-1}\mathbf{CA}^{-1} \tag{86}$$

it can be shown that $(\mathbf{R} + \delta\mathbf{R}_{vt})^{-1}$ is equal to expression

$$\begin{aligned}(\mathbf{R} + \delta\mathbf{R}_{vt})^{-1} &= \mathbf{R}^{-1} - \mathbf{R}^{-1}\begin{bmatrix} \mathbf{I}_{2M} \\ \alpha_v \mathbf{I}_{2M} \end{bmatrix} \\ &\times \left(\left[\mathbf{I}_{2M} \quad \alpha_v^* \mathbf{I}_{2M} \right] \mathbf{R}^{-1} \begin{bmatrix} \mathbf{I}_{2M} \\ \alpha_v \mathbf{I}_{2M} \end{bmatrix} + \frac{\mathbf{R}_v^{-1}}{\delta} \right)^{-1} \\ &\times \left[\mathbf{I}_{2M} \quad \alpha_v^* \mathbf{I}_{2M} \right] \mathbf{R}^{-1}. \end{aligned} \tag{87}$$

By defining

$$\mathbf{P} = \mathbf{I}_{2M} - \frac{P_s \mathbf{R}_v^{-1}\mathbf{A}\mathbf{A}^H}{\mu + \rho} \tag{88}$$

the matrix $\mathbf{R}^{-1}$ can be written as

$$\mathbf{R}^{-1} = \frac{1}{\mu}\begin{bmatrix} \mathbf{PR}_v^{-1} & \mathbf{0}_{2M} \\ \mathbf{0}_{2M} & \mathbf{PR}_v^{-1} \end{bmatrix} \tag{89}$$

and the following useful expression can be formulated:

$$\mathbf{P}\mathbf{R}_v^{-1}\mathbf{A} = \frac{\mu}{\mu+\rho}\mathbf{R}_v^{-1}\mathbf{A}. \tag{90}$$

Using (89), it can be shown that

$$\mathbf{R}^{-1}\begin{bmatrix}\mathbf{I}_{2M}\\ \alpha_v\,\mathbf{I}_{2M}\end{bmatrix} = \frac{1}{\mu}\begin{bmatrix}\mathbf{P}\mathbf{R}_v^{-1}\\ \alpha_v\mathbf{P}\mathbf{R}_v^{-1}\end{bmatrix} \tag{91}$$

$$[\,\mathbf{I}_{2M} \quad \alpha_v^*\,\mathbf{I}_{2M}\,]\,\mathbf{R}^{-1}\begin{bmatrix}\mathbf{I}_{2M}\\ \alpha_v\,\mathbf{I}_{2M}\end{bmatrix} = \frac{1+|\alpha_v|^2}{\mu}\mathbf{P}\mathbf{R}_v^{-1} \tag{92}$$

such that (87) reduces to

$$(\mathbf{R}+\delta\mathbf{R}_{vt})^{-1} = \mathbf{R}^{-1} - \frac{\delta}{\mu^2}\begin{bmatrix}\mathbf{P}\\ \alpha_v\mathbf{P}\end{bmatrix} \times\left(\mathbf{I}_{2M}+\frac{\delta(1+|\alpha_v|^2)}{\mu}\mathbf{P}\right)^{-1}[\,\mathbf{P}\mathbf{R}_v^{-1} \quad \alpha_v^*\mathbf{P}\mathbf{R}_v^{-1}\,]. \tag{93}$$

The filter in (84) is then equal to

$$\mathbf{W}_{\mathrm{MWF-ITF},2} = (\mathbf{R}+\delta\mathbf{R}_{vt})^{-1}P_s\begin{bmatrix}\mathbf{A}A_L^*\\ \mathbf{A}A_R^*\end{bmatrix} \tag{94}$$

$$= \mathbf{R}^{-1}P_s\begin{bmatrix}\mathbf{A}A_L^*\\ \mathbf{A}A_R^*\end{bmatrix} - \frac{\delta}{\mu^2}\begin{bmatrix}\mathbf{P}\\ \alpha_v\mathbf{P}\end{bmatrix} \times\left(\mathbf{I}_{2M}+\frac{\delta(1+|\alpha_v|^2)}{\mu}\mathbf{P}\right)^{-1} \times\mathbf{P}\mathbf{R}_v^{-1}\mathbf{A}P_s(A_L^*+\alpha_v^*A_R^*). \tag{95}$$

Using the matrix inversion lemma, it can be shown that

$$(\mathbf{I}_{2M}+c\mathbf{P})^{-1} = \frac{1}{c+1}\left[\mathbf{I}_{2M}+\frac{cP_s\mathbf{R}_v^{-1}\mathbf{A}\mathbf{A}^H}{(c+1)\mu+\rho}\right] \tag{96}$$

with c an arbitrary constant, such that

$$(\mathbf{I}_{2M}+c\mathbf{P})^{-1}\mathbf{R}_v^{-1}\mathbf{A} = \frac{\mu+\rho}{(c+1)\mu+\rho}\mathbf{R}_v^{-1}\mathbf{A}. \tag{97}$$

Using (38), (90), and (97), the filter in (95) is equal to

$$\mathbf{W}_{\mathrm{MWF-ITF},2} = \frac{P_s}{\mu+\rho}\begin{bmatrix}A_L^*\mathbf{R}_v^{-1}\mathbf{A}\\ A_R^*\mathbf{R}_v^{-1}\mathbf{A}\end{bmatrix} - \frac{\delta}{\mu}\frac{P_s(A_L^*+\alpha_v^*A_R^*)}{\mu+\rho+\delta(1+|\alpha_v|^2)}\begin{bmatrix}\mathbf{P}\\ \alpha_v\mathbf{P}\end{bmatrix}\mathbf{R}_v^{-1}\mathbf{A}$$

such that, using (90), and by defining

$$\xi = \frac{\delta}{\mu+\rho+\delta(1+|\alpha_v|^2)} \tag{98}$$

the optimal MWF-ITF filter reduces to expression

$$\boxed{\begin{aligned}\mathbf{W}_{\mathrm{MWF-ITF},2} &= (\mathbf{R}+\delta\mathbf{R}_{vt})^{-1}P_s\begin{bmatrix}\mathbf{A}A_L^*\\ \mathbf{A}A_R^*\end{bmatrix}\\ &= \frac{P_s}{\mu+\rho}\begin{bmatrix}[A_L^*-\xi(A_L^*+\alpha_v^*A_R^*)]\,\mathbf{R}_v^{-1}\mathbf{A}\\ [A_R^*-\xi\alpha_v(A_L^*+\alpha_v^*A_R^*)]\,\mathbf{R}_v^{-1}\mathbf{A}\end{bmatrix}.\end{aligned}} \tag{99}$$

B) ITF Cost Function for Speech and Noise Component ($\gamma \neq 0$)*:* For easier notation, we define $\alpha_x = -\mathrm{ITF}_x^{\mathrm{in}}$, such that $\mathbf{R}_{xt}$ is equal to

$$\mathbf{R}_{xt} = \begin{bmatrix}\mathbf{I}_{2M}\\ \alpha_x\,\mathbf{I}_{2M}\end{bmatrix}\mathbf{R}_x\,[\,\mathbf{I}_{2M} \quad \alpha_x^*\,\mathbf{I}_{2M}\,] = P_s\begin{bmatrix}\mathbf{A}\\ \alpha_x\,\mathbf{A}\end{bmatrix}[\,\mathbf{A}^H \quad \alpha_x^*\,\mathbf{A}^H\,]. \tag{100}$$

Hence, using the matrix inversion lemma, the filter in (70) is equal to expression (101), as shown at the bottom of the page. Similarly to (99), by setting $A_L^* = 1$ and $A_R^* = \alpha_x$, it can be shown that

$$P_s(\mathbf{R}+\delta\mathbf{R}_{vt})^{-1}\begin{bmatrix}\mathbf{A}\\ \alpha_x\,\mathbf{A}\end{bmatrix} = \frac{P_s}{\mu+\rho}\begin{bmatrix}(1-\xi\alpha_t)\mathbf{R}_v^{-1}\mathbf{A}\\ (\alpha_x-\alpha_v\xi\alpha_t)\mathbf{R}_v^{-1}\mathbf{A}\end{bmatrix} \tag{102}$$

with

$$\alpha_t = 1+\alpha_v^*\alpha_x. \tag{103}$$

Using (102), it can be shown that

$$P_s\,[\,\mathbf{A}^H \quad \alpha_x^*\,\mathbf{A}^H\,]\,(\mathbf{R}+\delta\mathbf{R}_{vt})^{-1}\begin{bmatrix}\mathbf{A}\\ \alpha_x\,\mathbf{A}\end{bmatrix} = \frac{\rho}{\mu+\rho}(1+|\alpha_x|^2-\xi|\alpha_t|^2). \tag{104}$$

Using (99) and $\alpha_x = -A_L/A_R$, it can be shown that

$$[\,\mathbf{A}^H \quad \alpha_x^*\,\mathbf{A}^H\,]\,(\mathbf{R}+\delta\mathbf{R}_{vt})^{-1}\mathbf{r}_x = -\frac{\rho\xi\alpha_t^*}{\mu+\rho}(A_L^*+\alpha_v^*A_R^*). \tag{105}$$

By plugging (99), (102), (104), and (105) into (101), and by defining

$$\nu = \frac{\gamma\rho}{\mu+\rho+\gamma\rho(1+|\alpha_x|^2-\xi|\alpha_t|^2)} \tag{106}$$

$$[(\mathbf{R}+\delta\mathbf{R}_{vt})+\gamma\mathbf{R}_{xt}]^{-1}\,\mathbf{r}_x = (\mathbf{R}+\delta\mathbf{R}_{vt})^{-1}\mathbf{r}_x - \frac{\gamma P_s(\mathbf{R}+\delta\mathbf{R}_{vt})^{-1}\begin{bmatrix}\mathbf{A}\\ \alpha_x\,\mathbf{A}\end{bmatrix}[\,\mathbf{A}^H \quad \alpha_x^*\,\mathbf{A}^H\,]\,(\mathbf{R}+\delta\mathbf{R}_{vt})^{-1}\mathbf{r}_x}{1+\gamma P_s\,[\,\mathbf{A}^H \quad \alpha_x^*\,\mathbf{A}^H\,]\,(\mathbf{R}+\delta\mathbf{R}_{vt})^{-1}\begin{bmatrix}\mathbf{A}\\ \alpha_x\,\mathbf{A}\end{bmatrix}}. \tag{101}$$

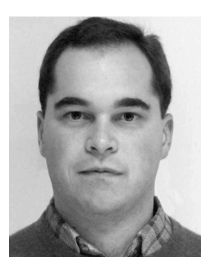
the filter $\mathbf{W}_{\mathrm{MWF-ITF},2}$ reduces to the previously shown expression (72).

References

[1] B. Kollmeier, J. Peissig, and V. Hohmann, "Real-time multiband dynamic compression and noise reduction for binaural hearing aids," *J. Rehabil. Res. Develop.*, vol. 30, no. 1, pp. 82–94, 1993.
[2] J. Desloge, W. Rabinowitz, and P. Zurek, "Microphone-array hearing aids with binaural output-Part I: Fixed-processing systems," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 6, pp. 529–542, Nov. 1997.
[3] D. Welker, J. Greenberg, J. Desloge, and P. Zurek, "Microphone-array hearing aids with binaural output-Part II: A two-microphone adaptive system," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 6, pp. 543–551, Nov. 1997.
[4] I. Merks, M. Boone, and A. Berkhout, "Design of a broadside array for a binaural hearing aid," in *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust. (WASPAA)*, New Paltz, NY, Oct. 1997.
[5] V. Hamacher, "Comparison of advanced monaural and binaural noise reduction algorithms for hearing aids," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Orlando, FL, May 2002, pp. 4008–4011.
[6] R. Nishimura, Y. Suzuki, and F. Asano, "A new adaptive binaural microphone array system using a weighted least squares algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Orlando, FL, May 2002, pp. 1925–1928.
[7] T. Wittkop and V. Hohmann, "Strategy-selective noise reduction for binaural digital hearing aids," *Speech Commun.*, vol. 39, no. 1–2, pp. 111–138, Jan. 2003.
[8] M. Lockwood, D. Jones, R. Bilger, C. Lansing, W. O'Brien, B. Wheeler, and A. Feng, "Performance of time- and frequency-domain binaural beamformers based on recorded signals from real rooms," *J. Acoust. Soc. Amer.*, vol. 115, no. 1, pp. 379–391, Jan. 2004.
[9] T. Lotter and P. Vary, "Dual-channel speech enhancement by superdirective beamforming," *EURASIP J. Appl. Signal Process.*, vol. 2006, 2006, Article ID 63297.
[10] O. Roy and M. Vetterli, "Rate-constrained beamforming for collaborating hearing aids," in *Proc. Int. Symp. Inf. Theory (ISIT)*, Seattle, WA, Jul. 2006, pp. 2809–2813.
[11] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localisation*. Cambridge, MA: MIT Press, 1983.
[12] A. Bronkhorst and R. Plomp, "The effect of head-induced interaural time and level differences on speech intelligibility in noise," *J. Acoust. Soc. Amer.*, vol. 83, no. 4, pp. 1508–1516, Apr. 1988.
[13] P. Zurek, "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing aid Performance*, 2 ed. Boston, MA: Allyn and Bacon, 1992, ch. 15, pp. 255–276.
[14] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Frequency-domain criterion for speech distortion weighted multichannel Wiener filter for robust noise reduction," *Speech Commun., Special Iss. Speech Enhancement*, vol. 49, no. 7–8, pp. 636–656, Jul.-Aug. 2007.
[15] S. Doclo, T. J. Klasen, T. Van den Bogaert, J. Wouters, and M. Moonen, "Theoretical analysis of binaural cue preservation using multi-channel Wiener filtering and interaural transfer functions," in *Proc. Int. Workshop Acoust. Echo Noise Control (IWAENC)*, Paris, France, Sep. 2006.
[16] T. Klasen, T. Van den Bogaert, M. Moonen, and J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1579–1585, Apr. 2007.
[17] T. Van den Bogaert, S. Doclo, M. Moonen, and J. Wouters, "The effect of multi-microphone noise reduction systems on sound source localization in binaural hearing aids," *J. Acoust. Soc. Amer.*, vol. 124, no. 1, pp. 484–497, 2008.
[18] T. Van den Bogaert, S. Doclo, M. Moonen, and J. Wouters, "Speech enhancement with multichannel Wiener filter techniques in multi-microphone binaural hearing aids," *J. Acoust. Soc. Amer.*, vol. 125, no. 1, pp. 360–371, 2009.
[19] T. Van den Bogaert, S. Doclo, M. Moonen, and J. Wouters, "Binaural cue preservation for hearing aids using an interaural transfer function multichannel Wiener filter," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Honolulu, HI, Apr. 2007, pp. 565–568.
[20] T. Klasen, T. Van den Bogaert, M. Moonen, and J. Wouters, "Preservation of interaural time delay for binaural hearing aids through multichannel Wiener filtering based noise reduction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Philadelphia, PA, Mar. 2005, vol. III, pp. 29–32.
[21] S. Doclo, T. Van den Bogaert, J. Wouters, and M. Moonen, "Reduced-bandwidth and distributed MWF-based noise reduction algorithms for binaural hearing aids," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 1, pp. 38–51, Jan. 2009.
[22] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1218–1234, Jul. 2006.
[23] B. de Vries and R. A. J. de Vries, "An integrated approach to hearing aid algorithm design for enhancement of audibility, intelligibility and comfort," in *Proc. IEEE Benelux Signal Process. Symp. (SPS2004)*, Hilvarenbeek, The Netherlands, Apr. 2004, pp. 65–68.
[24] S. Doclo and M. Moonen, "On the output SNR of the speech-distortion weighted multichannel Wiener filter," *IEEE Signal Process. Lett.*, vol. 12, pp. 809–811, Dec. 2005.
[25] T. Klasen, S. Doclo, T. Van den Bogaert, M. Moonen, and J. Wouters, "Binaural multi-channel Wiener filtering for hearing aids: Preserving interaural time and level differences," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Toulouse, France, May 2006, pp. 145–148.
[26] R. Fletcher, *Practical Methods of Optimization*. New York: Wiley, 1987.
[27] T. Coleman, M. A. Branch, and A. Grace, *MATLAB Optimization Toolbox User's Guide*. Natick, MA: The Mathworks, Inc., 1999.
[28] B. Gardner and K. Martin, "HRTF measurements of a KEMAR dummy-head microphone," MIT Media Lab Perceptual Computing, 1994, Tech. Rep. #280.
[29] B. Cornelis, S. Doclo, T. B. Van Den, M. Moonen, and J. Wouters, "Analysis of localization cue preservation by multichannel Wiener filtering based binaural noise reduction in hearing aids," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Lausanne, Switzerland, Aug. 2008.
[30] C. Darwin, "Contributions of binaural information to the separation of different sound sources," *Int. J. Audiol.*, vol. 45, pp. 20–24.
[31] M. Nilsson, S. D. Soli, and A. Sullivan, "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," *J. Acoust. Soc. Amer.*, vol. 95, no. 2, pp. 1085–1099, Feb. 1994.
[32] Auditec, Auditory Tests (Revised), Compact Disc, Auditec. St. Louis, MO, 1997.

Bram Cornelis (M'09) was born in Bornem, Belgium, in 1984. He received the M.Sc. degree in electrical engineering from the Katholieke Universiteit Leuven, Leuven, Belgium, in 2007. He is currently pursuing the Ph.D. degree at the Electrical Engineering Department, Katholieke Universiteit Leuven, and is supported by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen).

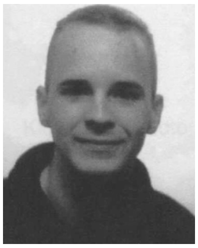
His research interest are in binaural signal processing for hearing aids, speech enhancement algorithms, and acoustical beamforming.



Simon Doclo (S'95–M'03) was born in Wilrijk, Belgium, in 1974. He received the M.Sc. degree in electrical engineering and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Leuven, Belgium, in 1997 and 2003, respectively.

From 2003 until 2007, he was a Postdoctoral fellow at the Electrical Engineering Department, Katholieke Universiteit Leuven, supported by the Research Foundation—Flanders. In 2005, he was a Visiting Postdoctoral Fellow at the Adaptive Systems Laboratory, McMaster University, Hamilton, ON, Canada. Currently, he is a Principal Scientist with NXP Semiconductors in the Sounds and Acoustics Group, Leuven, and still holds an honorary postdoctoral fellowship of the Research Foundation—Flanders. His research interests are in signal processing for acoustical applications, more specifically microphone array processing for speech enhancement and source localization, adaptive filtering, active noise control, and hearing aid processing.

Dr. Doclo received the Master Thesis Award of the Royal Flemish Society of Engineers in 1997 (with E. De Clippel), the Best Student Paper Award at the International Workshop on Acoustic Echo and Noise Control in 2001, the EURASIP Signal Processing Best Paper Award in 2003 (with M. Moonen), and the IEEE Signal Processing Society 2008 Best Paper Award (with J. Chen, J. Benesty, A. Huang). He is a member of the IEEE Signal Processing Society Technical Committee on Audio and Electroacoustics. He has been secretary of the IEEE Benelux Signal Processing Chapter (1998–2002), and has served as a Guest Editor for the *EURASIP Journal on Applied Signal Processing* (2005–2006).



Tim Van den Bogaert (M'05) was born in Kapellen, Belgium, in 1978. He received the M.Sc. degree in electrical engineering and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Leuven, Belgium, in 2002 and 2008, respectively.

Currently, he is a Postdoctoral Researcher at the Laboratory for Experimental ORL, Katholieke Universiteit Leuven. His research interests are in the area of binaural signal processing for hearing aids and cochlear implants.

Marc Moonen (M'94–SM'06–F'07) received the electrical engineering degree and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Leuven, Belgium, in 1986 and 1990, respectively.

Since 2004, he has been a Full Professor with the Electrical Engineering Department, Katholieke Universiteit Leuven, where he is heading a research team working in the area of numerical algorithms and signal processing for digital communications, wireless communications, DSL, and audio signal processing.

Dr. Moonen received the 1994 KU Leuven Research Council Award, the 1997 Alcatel Bell (Belgium) Award (with P. Vandaele), the 2004 Alcatel Bell (Belgium) Award (with R. Cendrillon), and was a 1997 "Laureate of the Belgium Royal Academy of Science." He received a journal best paper award from the IEEE TRANSACTIONS ON SIGNAL PROCESSING (with G. Leus) and from *Elsevier Signal Processing* (with S. Doclo). He was Chairman of the IEEE Benelux Signal Processing Chapter (1998-2002), past President of EURASIP (European Association for Signal, Speech, and Signal Processing) (2006-2008) and a member of the IEEE Signal Processing Society Technical Committee on Signal Processing for Communications. He served as Editor-in-Chief for the *EURASIP Journal on Applied Signal Processing* (2003-2005), and was a member of the editorial board of *Integration*, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II (2002-2003) and the *IEEE Signal Processing Magazine* (2003-2005). He is currently a member of the editorial board of *the VLSI Journal*, the *EURASIP Journal on Advances in Signal Processing*, the *EURASIP Journal on Wireless Communications and Networking*, and *Signal Processing*.

Jan Wouters was born in Leuven, Belgium, in 1960. He received the physics degree and the Ph.D. degree in sciences/physics from the Katholieke Universiteit Leuven, Leuven, Belgium, in 1982 and 1989, respectively.

From 1989 to 1992, he was a Research Fellow with the Belgian National Fund for Scientific Research (NFWO), Institute of Nuclear Physics (UCL Louvain-la-Neuve and KU Leuven) and at NASA Goddard Space Flight Centre, Greenbelt, MD. Since 1993, he has been a Professor at the Neurosciences Department, Katholieke Universiteit Leuven. His research activities center around audiology and the auditory system and signal processing for cochlear implants and hearing aids. He is the author of 110 articles in international peer-reviewed journals and is a reviewer for several international journals.

Dr. Wouters received an Award of the Flemish Ministry in 1989, a Fullbright Award and a NATO Research Fellowship in 1992, and the 1996 Flemish VVL Speech Therapy-Audiology Award. He is member of the International Collegium for Rehabilitative Audiology, a Board Member of the NAG (Dutch Acoustical Society), and is responsible for the Laboratory for Experimental ORL, Katholieke Universiteit Leuven.