16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

ANALYSIS OF LOCALIZATION CUE PRESERVATION BY MULTICHANNEL WIENER FILTERING BASED BINAURAL NOISE REDUCTION IN HEARING AIDS

Bram Cornelis[1] , *Simon Doclo*[1,3], *Tim Van den Bogaert*[2], *Marc Moonen*[1], *Jan Wouters*[2]

[1]ESAT-SCD
Dept. of electrical engineering, K.U.Leuven
Kasteelpark Arenberg 10, 3001 Heverlee, Belgium
phone: +32 16 321927
email: bram.cornelis@esat.kuleuven.be,
marc.moonen@esat.kuleuven.be
web: www.esat.kuleuven.be/scd

[2]ExpORL
Dept. of Neurosciences, K.U.Leuven
Herestraat 49/721, 3000 leuven, Belgium

[3]NXP Semiconductors, Corporate I&T-Research
Interleuvenlaan 80, 3001 Heverlee, Belgium
email: simon.doclo@nxp.com

ABSTRACT

In this paper the localization cue preservation of a noise reduction algorithm for binaural hearing aids is analyzed. In a binaural noise reduction procedure based on Multi-channel Wiener Filtering (MWF), the basic cost function is extended with terms related to the Interaural Transfer Functions (ITF) of a speech and noise source. To make the algorithm computationally feasible, these cost terms are simplified to quadratic terms. First it is shown that this approach cannot preserve the ITF of the speech and noise component simultaneously. However, the obtained output ITF is related to the output SNR. This will lead to a perceptually advantageous effect: a noisy output signal (low SNR) will be perceived in the direction of the noise source, while a clean output signal (high SNR) will be perceived in the direction of the speech source.

1. INTRODUCTION

Modern hearing aids make use of noise reduction algorithms to improve speech intelligibility in background noise. Hearing aids are mostly fitted with multiple microphones, which especially leads to an improvement in noise reduction performance because spatial information can be exploited in addition to spectral information [1, 2, 3]. In a binaural setup, the hearing impaired person has two hearing aids that can communicate over a wireless link. Microphone signals from both hearing aids can then be shared, which further improves the noise reduction performance over a monaural configuration or bilateral configuration (two hearing aids that work independently).

Current noise reduction algorithms in bilateral hearing aids are not designed to preserve the localization cues (also called *binaural cues*). Namely, the ITD (interaural time difference) and ILD (interaural level difference), which are used when localising sounds [4]. Incorrect sound localization can endanger the hearing aid user (e.g. in traffic situations), and also imposes a disadvantage in speech segregation in noisy environments. A normal hearing person suppresses unwanted signals, coming from different directions than a desired signal, by correctly localising the desired and undesired sounds. This effect of spatial unmasking can lead to a speech intelligibility improvement up to 10 dB [5].

In contrast to the monaural or bilateral setups, a *binaural noise reduction algorithm* can be designed so that the advantage of binaural hearing can be fully exploited. In [3], a binaural noise reduction algorithm based on *Multi-channel Wiener filtering (MWF)* was introduced. It was proven in [6] that this algorithm preserves the binaural cues of the speech component, but changes the noise cues to those of the speech component. In [6] an extension was also proposed to trade-off some noise reduction performance for better cue preservation. The MWF cost function was extended with terms related to the *interaural transfer functions (ITF)* of the speech and noise components. Simulations in [6] with this MWF-ITF algorithm showed that it is then indeed possible to preserve both speech and noise binaural cues. To reduce computational complexity, a simplification was introduced in the ITF cost function to obtain quadratic cost terms [7]. Remarkably, perceptual tests in [7] showed an improvement in localization performance, even though this simplified cost function was used.

In this paper, the performance (in noise reduction and cue preservation) of the MWF algorithm with quadratic ITF extension is analyzed. In section 2, the MWF-ITF procedure is summarized and some objective performance measures for noise reduction and cue preservation are defined. In section 3 closed form expressions for the optimal MWF-ITF filters are derived, assuming a single speech source. It is proven that the quadratic ITF approach cannot preserve speech and noise cues simultaneously, but that the speech and noise ITF are changed to one and the same value, which is a combination of the input speech and noise ITF's. Simulations in section 4 confirm these theoretical findings. However, it will be shown that there is an advantageous perceptual effect by evaluating the dependence of the obtained output ITF on the output SNR. This explains why the quadratic ITF extension indeed improved the perceptual localisation performance in [7].

2. BINAURAL MULTI-CHANNEL WIENER FILTER WITH ITF EXTENSION: MWF-ITF

2.1 Configuration and notation

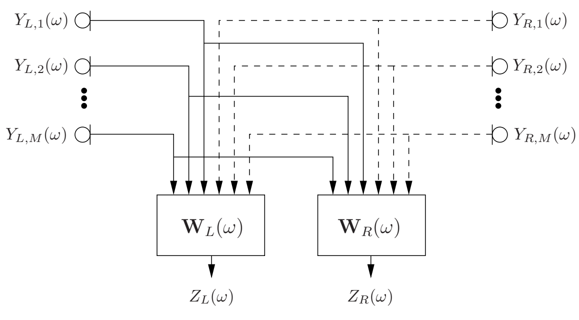

Figure 1: General binaural processing scheme

Bram Cornelis is funded by a Ph.D. grant of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). This research work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven in the frame of the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office IUAP P6/04 (DYSCO, 'Dynamical systems, control and optimization', 2007-2011), the Concerted Research Action GOA-AMBioRICS and the research project FWO nr. G.0600.08 ('Signal processing and network design for wireless acoustic sensor networks'). The scientific responsibility is assumed by its authors.

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

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

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 $Y_{R,m}(\omega) = X_{R,m}(\omega) + V_{R,m}(\omega)$. We define the $2M$-dimensional signal vector $\mathbf{Y}$, omitting the frequency-domain variable ω for conciseness, as

$$\mathbf{Y} = [Y_{L,1} \ \dots \ Y_{L,M} \ \ Y_{R,1} \ \dots \ Y_{R,M}]^T = [\mathbf{Y}_L^T \ \mathbf{Y}_R^T]^T. \tag{2}$$

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

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

where $\mathscr{E}$ denotes the expected value operator. Assuming that the speech and the noise components are uncorrelated, $\mathbf{R}_y = \mathbf{R}_x + \mathbf{R}_v$. It will be assumed that a single speech source is present, such that the speech component and speech correlation matrix are equal to:

$$\mathbf{X} = \mathbf{A}S, \qquad \mathbf{R}_x = P_s\mathbf{A}\mathbf{A}^H, \tag{4}$$

where vector $\mathbf{A}$ represents the transfer function between the speech source and the microphone array.
We will use the r_L-th microphone on the left hearing aid and the r_R-th microphone on the right hearing aid as the so-called reference microphones for the noise reduction algorithm (see section 2.2). The reference microphone signals are denoted 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

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

2.2 Binaural Multi-channel Wiener Filter (MWF)

The binaural MWF produces a minimum-mean-square-error (MMSE) estimate of the speech components in the reference microphone signals Y_L and Y_R. To provide a trade-off between speech distortion and noise reduction, the Speech Distortion Weighted Multi-channel Wiener filter (SDW-MWF) has been proposed which minimizes the weighted sum of the residual noise energy and the speech distortion energy [3]. The binaural SDW-MWF cost function is equal to

$$J_{SDW}(\mathbf{W}) = \mathscr{E}\left\{ \left\| \begin{bmatrix} X_L - \mathbf{W}_L^H\mathbf{X} \\ X_R - \mathbf{W}_R^H\mathbf{X} \end{bmatrix} \right\|^2 + \mu \left\| \begin{bmatrix} \mathbf{W}_L^H\mathbf{V} \\ \mathbf{W}_R^H\mathbf{V} \end{bmatrix} \right\|^2 \right\}, \tag{8}$$

where μ provides the mentioned trade-off. Hence, the filter minimizing $J_{SDW}(\mathbf{W})$ is equal to

$$\mathbf{W}_{SDW} = \mathbf{R}^{-1}\mathbf{r}_x, \tag{9}$$

with

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_x + \mu\mathbf{R}_v & \mathbf{0}_{2M} \\ \mathbf{0}_{2M} & \mathbf{R}_x + \mu\mathbf{R}_v \end{bmatrix}, \quad \mathbf{r}_x = \begin{bmatrix} \mathbf{R}_x\mathbf{e}_L \\ \mathbf{R}_x\mathbf{e}_R \end{bmatrix}. \tag{10}$$

where $\mathbf{e}_L$ and $\mathbf{e}_R$ are vectors of which only one element is equal to 1 and the other elements are equal to 0, i.e. $\mathbf{e}_L(r_L) = 1$ and $\mathbf{e}_R(r_R) = 1$.

2.3 Extension with interaural transfer function

To control the binaural cues of the speech and the noise component, it is possible to extend the SDW-MWF cost function with terms related to the *Interaural Transfer Function (ITF)* of the speech and the noise components, as has been proposed in [6, 7].
The input ITF of the speech and noise component are defined as

$$ITF_v^{in} = \frac{V_L}{V_R}, \qquad ITF_x^{in} = \frac{X_L}{X_R}, \tag{11}$$

while the output ITF's are equal to

$$ITF_v^{out} = \frac{Z_{vL}}{Z_{vR}} = \frac{\mathbf{W}_L^H\mathbf{V}}{\mathbf{W}_R^H\mathbf{V}}, \qquad ITF_x^{out} = \frac{Z_{xL}}{Z_{xR}} = \frac{\mathbf{W}_L^H\mathbf{X}}{\mathbf{W}_R^H\mathbf{X}}. \tag{12}$$

When the aim is to preserve the cues of the speech and the noise components, the desired output ITF's are equal to the input ITF's in (11). We assume the input ITF's to be constant (as is e.g. the case for a single source[1]), such that they can be estimated in a least-squares sense, e.g. for the noise ITF as

$$ITF_v^{des} = \frac{\mathscr{E}\{V_L V_R^*\}}{\mathscr{E}\{V_R V_R^*\}} = \frac{\mathbf{e}_L^H\mathbf{R}_v\mathbf{e}_R}{\mathbf{e}_R^H\mathbf{R}_v\mathbf{e}_R} \tag{13}$$

and similarly for the speech ITF. The ITF cost function for preserving the cues of the noise component is then defined as

$$\boxed{J_{ITF,1}^v(\mathbf{W}) = \mathscr{E}\left\{ \left| \frac{\mathbf{W}_L^H\mathbf{V}}{\mathbf{W}_R^H\mathbf{V}} - ITF_v^{des} \right|^2 \right\}} \tag{14}$$

Because (14) can only be minimized by means of iterative optimisation techniques, a simplified quadratic ITF cost function is also introduced in [7], i.e.

$$\boxed{J_{ITF,2}^v(\mathbf{W}) = \mathscr{E}\{|\mathbf{W}_L^H\mathbf{V} - ITF_v^{des}\,\mathbf{W}_R^H\mathbf{V}|^2\} = \mathbf{W}^H\mathbf{R}_{vt}\mathbf{W}} \tag{15}$$

with

$$\mathbf{R}_{vt} = \begin{bmatrix} \mathbf{R}_v & -ITF_v^{des,*}\,\mathbf{R}_v \\ -ITF_v^{des}\,\mathbf{R}_v & |ITF_v^{des}|^2\,\mathbf{R}_v \end{bmatrix}, \tag{16}$$

$$= \begin{bmatrix} \mathbf{I}_{2M} \\ -ITF_v^{des}\,\mathbf{I}_{2M} \end{bmatrix} \mathbf{R}_v \begin{bmatrix} \mathbf{I}_{2M} & -ITF_v^{des,*}\,\mathbf{I}_{2M} \end{bmatrix} \tag{17}$$

The simplified cost function (15) will be used throughout this paper. The ITF cost function for the speech component is defined similarly as the ITF cost function for the noise component, by replacing the noise correlation matrix with the speech correlation matrix and the desired noise ITF with the desired speech ITF.
The total cost function trading off noise reduction, speech distortion and cue preservation is defined as

$$\boxed{J_{SDW-ITF,2}(\mathbf{W}) = J_{SDW}(\mathbf{W}) + \gamma\, J_{ITF,2}^x(\mathbf{W}) + \delta\, J_{ITF,2}^v(\mathbf{W})} \tag{18}$$

[1] In this case, it can also be shown that preserving the ITF is equivalent to preserving the phase of the cross-correlation, i.e. the ITD, and preserving the power ratio, i.e. the ILD.

where the parameters γ and δ enable to put more emphasis on binaural cue preservation for the speech and the noise component. By combining the quadratic cost functions (15) and the SDW cost function (8), a closed-form expression for the optimal filter is obtained, i.e.

$$\boxed{\mathbf{W}_{SDW-ITF,2} = (\mathbf{R} + \gamma\, \mathbf{R}_{xt} + \delta\, \mathbf{R}_{vt})^{-1} \mathbf{r}_x} \tag{19}$$

2.4 Objective performance measures

In this section objective measures are defined to compute the performance (noise reduction and cue preservation) of the algorithm, namely the output SNR, the error on the ITD (interaural time difference) cues and the error on the ILD (interaural level difference) cues.

The *output SNR* is defined as the power ratio of speech and noise component in the output signal, i.e. for the left hearing aid

$$SNR_{out,L} = \frac{\mathscr{E}\{|Z_{xL}|^2\}}{\mathscr{E}\{|Z_{vL}|^2\}} = \frac{\mathbf{W}_L^H \mathbf{R}_x \mathbf{W}_L}{\mathbf{W}_L^H \mathbf{R}_v \mathbf{W}_L}\,. \tag{20}$$

The output SNR for the right hearing aid is defined in a similar fashion. The *ITD error* of the speech or noise component can be calculated as the difference of the phases of the cross-correlations at the input and the output. For example, the ITD error of the noise component is

$$\Delta ITD_v = \frac{|\angle\mathscr{E}\{Z_{vL}Z_{vR}^*\} - \angle\mathscr{E}\{V_L V_R^*\}|}{\pi}\,. \tag{21}$$

The *ILD error* is defined as the difference between the power level ratios's at the input and the output. For example, the ILD error of the noise component is

$$\Delta ILD_v = 10\log_{10}\frac{\mathscr{E}\{|Z_{vL}|^2\}}{\mathscr{E}\{|Z_{vR}|^2\}} - 10\log_{10}\frac{\mathscr{E}\{|V_L|^2\}}{\mathscr{E}\{|V_R|^2\}}\,. \tag{22}$$

The ITD and ILD error of the speech component are similarly defined.

3. THEORETICAL ANALYSIS OF MWF WITH QUADRATIC ITF EXTENSION

In this section, closed-form expressions are derived for the optimal filters and the output ITF corresponding to cost function (18). The ITF parameter γ will be fixed to 0 to reduce the complexity of the equations, however, the derivation can also be made for the case $\gamma \neq 0$ which will lead to similar conclusions as the results given here. As mentioned in section 2.1, it is assumed that a single speech source is present as in (4).

Equation (4) together with (17) can be plugged into formula (19) for the optimal SDW-ITF filters. The matrix inversion in (19) can then be worked out by applying the matrix inversion lemma. i.e.

$$(\mathbf{R} + \delta\mathbf{R}_{vt})^{-1} = \mathbf{R}^{-1} - \frac{\delta}{\mu^2}\begin{bmatrix}\mathbf{P}\\ \alpha_v\mathbf{P}\end{bmatrix}\left(\mathbf{I}_{2M} + \frac{\delta(1+|\alpha_v|^2)}{\mu}\mathbf{P}\right)^{-1}. \begin{bmatrix}\mathbf{P}\mathbf{R}_v^{-1} & \alpha_v^*\mathbf{P}\mathbf{R}_v^{-1}\end{bmatrix} \tag{23}$$

where

$$\rho = P_s\mathbf{A}^H\mathbf{R}_v^{-1}\mathbf{A}, \quad \alpha_v = -ITF_v^{des}, \quad \mathbf{P} = \mathbf{I}_{2M} - \frac{P_s\mathbf{R}_v^{-1}\mathbf{A}\mathbf{A}^H}{\mu+\rho} \tag{24}$$

The optimal filters are obtained by multiplying (23) by a factor $\mathbf{r}_x$, as in (19). Again, assuming a single speech source (4), this factor can be written as

$$\mathbf{r}_x = P_s\begin{bmatrix}\mathbf{A}A_L^*\\ \mathbf{A}A_R^*\end{bmatrix}, \tag{25}$$

so that it can be shown that the optimal filters reduce to

$$\mathbf{W}_{SDW-ITF,L} = \frac{P_s}{\mu+\rho}\left[A_L^* - \xi(A_L^* - ITF_v^{des,*}A_R^*)\right]\mathbf{R}_v^{-1}\mathbf{A}$$
$$\mathbf{W}_{SDW-ITF,R} = \frac{P_s}{\mu+\rho}\left[A_R^* + \xi ITF_v^{des}(A_L^* - ITF_v^{des,*}A_R^*)\right]\mathbf{R}_v^{-1}\mathbf{A}$$
$$\text{with} \qquad \xi = \frac{\delta}{\mu+\rho+\delta(1+|ITF_v^{des}|^2)}\,. \tag{26}$$

The expressions (26) show that the left and right filters (for every frequency ω) are parallel vectors. As a consequence, the output ITF's (12) of the speech and noise component will be the same and equal to

$$\boxed{ITF^{out} = \frac{ITF_x^{in} - \xi(ITF_x^{in} - ITF_v^{des})}{1 + \xi ITF_v^{des,*}(ITF_x^{in} - ITF_v^{des})}} \tag{27}$$

As the speech and noise ITF's are the same, it is not possible to preserve speech and noise binaural cues simultaneously with the quadratic ITF extension. The obtained ITF can be related to the input ITF's of the speech and noise as is apparent in (27).

For $\delta = 0$, the obtained filters (26) are the SDW-MWF optimal filters (9). The output ITF (27) becomes equal to ITF_x^{in}, which means speech and noise are both perceived in the speech direction. In [6], it was already shown that the standard SDW-MWF solution indeed preserves the speech cues, but distorts the noise cues.

The obtained output SNR can be derived by plugging (4) and (26) into (20). Again, because of the parallel filters, the output SNR at the left and right hearing aid are the same and equal to

$$SNR_{out} = P_s\mathbf{A}^H\mathbf{R}_v^{-1}\mathbf{A}, \tag{28}$$

which was defined as ρ in (24). Remarkably, the output SNR is independent of the ITF parameter δ. Furthermore, as the output ITF formula (27) contains a factor ξ defined in (26), and ξ in turn contains $\rho = SNR_{out}$, the obtained output ITF is related to the output SNR.

For $\gamma \neq 0$, it can be shown that the obtained filters are still parallel, which leads to the same conclusions as the case $\gamma = 0$: the speech and noise ITF's are the same so that speech and noise cues cannot be preserved simultaneously.

4. SIMULATIONS

In this section, the performance of the SDW-MWF algorithm with quadratic ITF extension will be tested in a scenario of one speech source and one noise source. First, the setup will be briefly discussed. Then the binaural cue preservation of the algorithm will be evaluated for different values of the ITF parameters γ and δ. Finally, the relation between the obtained output ITF and the output SNR will be analyzed.

4.1 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, with an elevation $\phi = 0$. The speech and noise components of the microphone signals can thus be written as

$$\mathbf{X}(\omega) = \mathbf{d}(\omega,\theta_x)S(\omega), \quad \mathbf{V}(\omega) = \mathbf{d}(\omega,\theta_v)V(\omega), \tag{29}$$

where $\mathbf{d}(\omega,\theta)$ is the steering vector for angle θ. The (omnidirectional) microphones are located on a head, so the head shadow effect will be taken into account. To achieve this, HRTF's measured on a KEMAR dummy-head [8] are incorporated in the steering vectors. The speech and noise correlation matrices are constructed as

$$\mathbf{R}_x(\omega) = P_s\,\mathbf{d}(\omega,\theta_x)\mathbf{d}^H(\omega,\theta_x), \tag{30}$$
$$\mathbf{R}_v(\omega) = P_v\,\mathbf{d}(\omega,\theta_v)\mathbf{d}^H(\omega,\theta_v) + P_{vs}\,\mathbf{I}_{2M}\,. \tag{31}$$

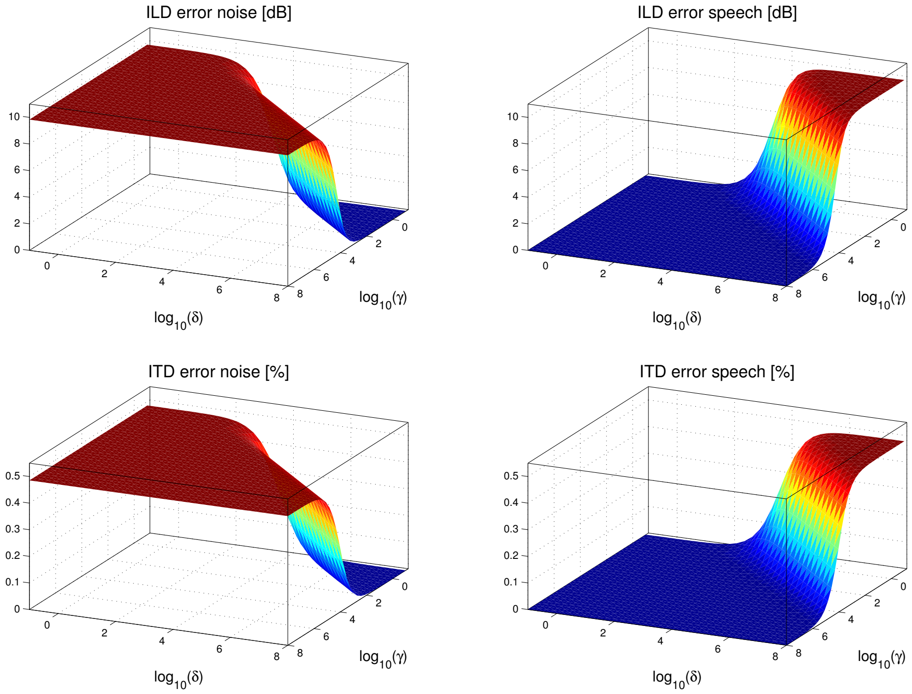

Figure 2: MWF-ITF with quadratic ITF extension; The ITD and ILD error for the noise and speech component are shown for different values of γ and δ.

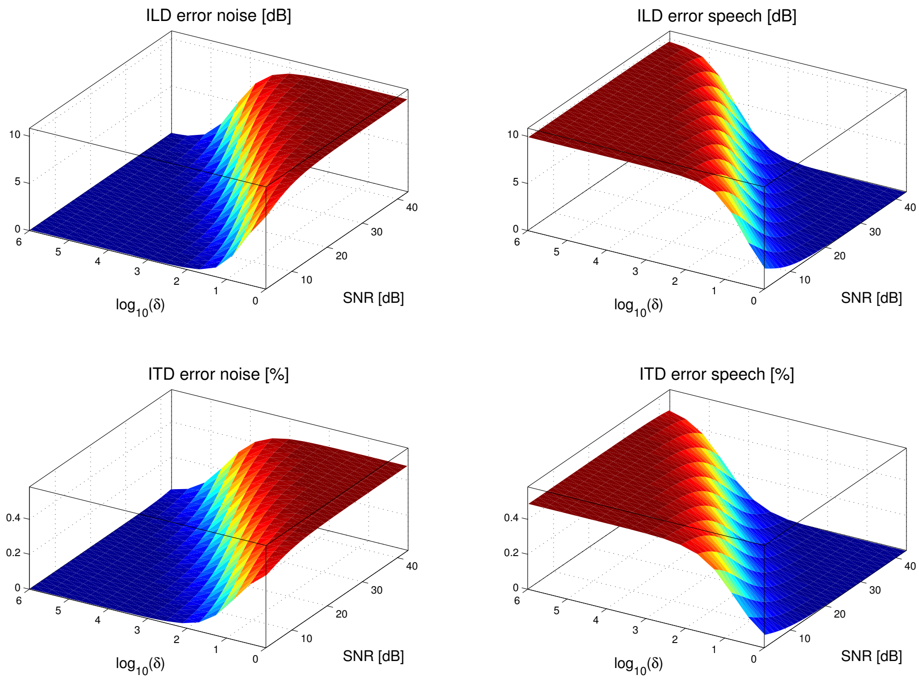

Figure 3: MWF-ITF with quadratic ITF extension; The ITD and ILD error for the noise and speech component are shown in function of the ITF parameter δ and in function of the output SNR.

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, modelled as spatially uncorrelated noise, is added in (31) to add a degree of realism and to make the noise correlation matrix invertible.

4.2 Parameters

The experiments are performed using a speech source at -5° and a noise source at 40°. A 2-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 1.5cm. The algorithm is tested at a frequency of $\omega = 2\pi 2000$ rad$/s$, the HRTF data is sampled at 44100 Hz. The signal powers are chosen as $P_s = P_v = 1$ and $P_{vs} = 0.01$. The parameter μ in the SDW-MWF cost function (8) is equal to 1.

4.3 ITD and ILD error for quadratic ITF extension

In figure 2 the ILD and ITD errors (defined in (21) and (22)) of speech and noise for the SDW-MWF with quadratic ITF extension (15) are shown. The ILD and ITD error for the noise component are shown in the left column, the ILD and ITD error for the speech component are shown on the right. For some choices of the ITF parameters γ and δ, the ITD/ILD errors on the noise component can be made arbitrarily small. However, the ITD/ILD errors on the speech component will then become large. On the other hand, when the ITD/ILD errors on the speech component are made small, the ITD/ILD errors on the noise component are large. An optimal choice of parameters, where the ITD/ILD errors of both noise and speech are small appears to be impossible. These results are in accordance with the theoretical discussion in the previous section, where it was shown that the output ITF's of speech and noise component are equal, so speech and noise cues cannot be preserved simultaneously.

4.4 MWF with quadratic ITF extension: 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 could seem to be inappropiate as a binaural noise reduction algorithm (if binaural cues should also be preserved), as it would be impossible to correctly localise both the speech and the noise source. In [7] the MWF-ITF algorithm (with quadratic ITF cost terms), was validaded perceptually. An improvement in the total localisation 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, which was also shown in (27), will be analyzed.

In figure 3, 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 (28). As in previous sections, the noise and speech ITD and ILD errors are shown. It can be seen that for certain values of the ITF parameter δ, the ITD/ILD errors of the noise component are small at low output SNR's, while the speech ITD/ILD errors are small at high output SNR's. This means 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 [7], the signals are processed with FFT's and in every frequency bin the optimal filters are calculated. The obtained output SNR's will vary in the different frequency bins, and similarly, the output ITF's 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.

5. CONCLUSION

In this paper the SDW-MWF with quadratic ITF extension was analyzed. For complexity reasons the ITF cost functions used here are simplified to quadratic cost terms. It was shown that the resulting procedure cannot preserve the speech and noise binaural cues simultaneously. However, it was also shown that there is an advantageous perceptual effect: in frequency bins where the output SNR is low, the ITF will be shifted towards the noise input ITF and vice versa for high output SNR and the speech ITF. As a result, the remaining noise in the output signals is perceived in the original direction of the noise component, and similarly, the speech component is perceived in the correct direction.

REFERENCES

[1] 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 and Audio Processing*, vol. 5, no. 6, pp. 543–551, Nov. 1997.

[2] T. Lotter, *Single and multimicrophone speech enhancement for hearing aids*. Ph.D. thesis, RWTH Aachen, Germany, Aug. 2004.

[3] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, *Speech Distortion Weighted Multichannel Wiener Filtering Techniques for Noise Reduction*. Springer-Verlag, 2005, ch. 9 in "Speech Enhancement" (J. Benesty, J. Chen, S. Makino, Eds.), pp. 199–228.

[4] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localisation*. MIT, Cambridge, MA, 1983.

[5] A. Bronkhorst and R. Plomp, "The effect of head-induced interaural time and level differences on speech intelligibility in noise," *Journal of the Acoustical Society of America*, vol. 83, no. 4, pp. 1508–1516, Apr. 1988.

[6] 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," *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Paris, France, Sep. 2006.

[7] 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," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 565–568, Honolulu HI, USA, Apr. 2007.

[8] B. Gardner and K. Martin, "HRTF measurements of a KEMAR dummy-head microphone," *MIT Media Lab Perceptual Computing, Tech. Rep. #280*, May 1994.