BINAURAL MULTI-CHANNEL WIENER FILTERING FOR HEARING AIDS: PRESERVING INTERAURAL TIME AND LEVEL DIFFERENCES

Thomas J. Klasen, Simon Doclo, Tim Van den Bogaert, Marc Moonen, Jan Wouters

Department of Electrical Engineering
Katholieke Universiteit Leuven, Belgium
{tklasen,doclo,moonen}@esat.kuleuven.be

Laboratory of Exp. ORL
Katholieke Universiteit Leuven, Belgium
{tim.vandenbogaert,jan.wouters}@uz.kuleuven.be

ABSTRACT

This paper presents an extension of the binaural multi-channel Wiener filtering algorithm discussed in [1]. The goal of this paper is to preserve both the interaural time difference (ITD) and interaural level difference (ILD) of the speech and noise components. This is done by extending the cost function to incorporate terms for the interaural transfer functions (ITF) of the speech and noise components. Using weights, the emphasis on the preservation of the ITFs can be controlled in addition to the emphasis on noise reduction. Adapting these parameters allows one to preserve the ITFs of the speech and noise component, and therefore ITD and ILD cues, while enhancing the signal-to-noise ratio.

1. INTRODUCTION

Hearing impaired persons localize sounds better without their bilateral hearing aids than with them [2]. In addition, noise reduction algorithms currently used in hearing aids are not designed to preserve localization cues. The inability to correctly localize sounds puts the hearing aid user at a disadvantage. The sooner the user can localize a speech signal, the sooner the user can begin to exploit visual cues. Generally, visual cues lead to large improvements in intelligibility for hearing impaired persons [3]. Furthermore, preserving the spatial separation between the target speech and the interfering signals leads to an improvement in speech understanding [4].

Interaural time delay (ITD) and interaural level difference (ILD) help listeners localize sounds horizontally [5]. ITD is the time delay in the arrival of the sound signal between the left and right ear, and ILD is the intensity difference between the two ears. Owing to the fact that ITD is caused by the sound waves diffracting around the head, ITD cues are more reliable in low frequencies. On the other hand, ILD is more prominent in high frequencies, since it stems from the scattering of the sound waves by the head. The goal of this paper is to design a noise reduction algorithm that does not introduce any adverse processing artefacts, such as distorting ITD and ILD cues.

In [6], the cost function has been extended, and includes terms related to ITD and ILD cues of the noise component. The ITD cost function is expressed as the phase difference between the output noise cross-correlation and the input noise cross-correlation. The ILD cost function is expressed as the difference between the output noise power ratio and the input noise power ratio. It has been shown that it is possible to preserve the binaural cues of both the speech and noise components without significantly compromizsing the noise reduction performance. However, iterative optimization techniques have to be used to compute the filter.

Clearly, the interaural transfer function (ITF), which is the ratio between the speech components (noise components) in the microphone signals at the left and right ear, captures all information between the two ears including ITD and ILD cues. Accordingly, this paper attacks the problem of binaural cue preservation by preserving the ITF. If the algorithm preserves the ITFs of the speech and noise components then the algorithm preserves the ITD and ILD cues of the speech and noise component. An extension of the binaural Wiener filter [1] is presented, where the cost function is comprised of four terms. The first two terms are present in the monaural speech distortion weighted Wiener filter proposed by [7]. The remaining two terms aim at preserving the ITFs of the speech and noise component. Contrary to the Wiener filter extensions proposed in [1], this algorithm co-designs the right and left filter.

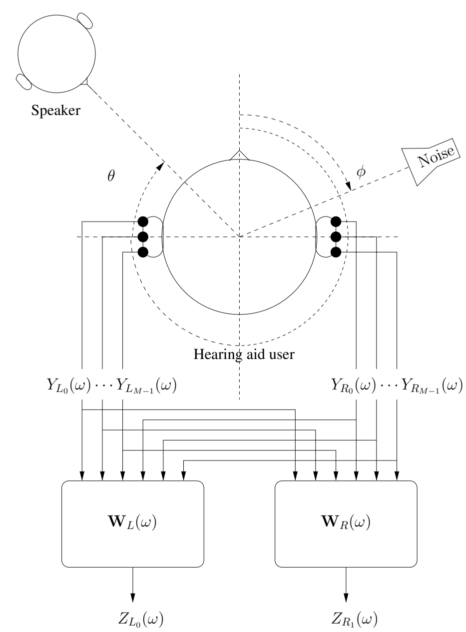


Fig. 1. Typical setup

This research work was carried out at the ESAT laboratory of the Katholieke Universitei1t Leuven, in the frame of the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office IUAP P5/22 ('Dynamical Systems and Control: Computation, Identification and Modelling'), the Concerted Research Action GOA-AMBioRICS, and the Research Project FWO nr.G.0233.01 ('Signal processing and automatic patient fitting for advanced auditory prostheses'). Simon Doclo is a Postdoctoral Fellow of the Research Foundation - Flanders (FWO - Vlaanderen).

2. SYSTEM MODEL

Figure 1 shows a binaural hearing aid user in a typical listening scenario. The speaker speaks intermittently in the continuous background noise caused by the noise source. There are M microphones on each hearing aid. We refer to the mth microphone of the left hearing aid and the mth microphone of the right hearing aid as the mth microphone pair. The received signals at the mth microphone pair are expressed in frequency domain below.

$$Y_{L_m}(\omega) = X_{L_m}(\omega) + V_{L_m}(\omega) \tag{1}$$

$$Y_{R_m}(\omega) = X_{R_m}(\omega) + V_{R_m}(\omega) \tag{2}$$

In (1) and (2), $X_{L_m}(\omega)$ and $X_{R_m}(\omega)$ represent the speech component in the mth microphone pair. Likewise, $V_{L_m}(\omega)$ and $V_{R_m}(\omega)$ represent the noise component of the mth microphone pair. Additionally, Figure 1 depicts a binaural hearing aid setup. All received microphone signals are used to design the filters, $\mathbf{W}_L(\omega)$ and $\mathbf{W}_R(\omega)$, and to generate an output for the left and right ear, $Z_{L_0}(\omega)$ and $Z_{R_0}(\omega)$.

The following definitions will be used in the derivation of the Wiener filter extension. First, we define the *2M*-dimensional signal vector.

$$\mathbf{Y}(\omega) = \left[Y_{L_0}(\omega) \ldots Y_{L_{M-1}}(\omega) Y_{R_0}(\omega) \ldots Y_{R_{M-1}}(\omega)\right]^T \tag{3}$$

In a similar fashion we write $\mathbf{X}(\omega)$ and $\mathbf{V}(\omega)$, where $\mathbf{Y}(\omega) = \mathbf{X}(\omega) + \mathbf{V}(\omega)$. Next, we define the filters for the left and right hearing aid.

$$\mathbf{W}_L(\omega) = \left[W_{L_0}(\omega) \ldots W_{L_{2M-1}}(\omega)\right]^T \tag{4}$$

Again, $\mathbf{W}_R$ is defined analogously. Using (4), we write $\mathbf{W}(\omega) = \begin{bmatrix} \mathbf{W}_L(\omega) \\ \mathbf{W}_R(\omega) \end{bmatrix}$. For clarity the frequency domain variable, ω, will be omitted throughout the remainder of this paper.

3. INTERAURAL TRANSFER FUNCTION

This paper presents a technique for controlling binaural noise cues, using the ITF. The ITFs of the input speech and noise components are written below.

$$ITF_{X_{des}} = \frac{X_{L_0}}{X_{R_0}} \qquad ITF_{V_{des}} = \frac{V_{L_0}}{V_{R_0}}. \tag{5}$$

Similarly, the ITFs of the output speech and noise components are,

$$ITF_{X_{out}}(\mathbf{W}) = \frac{\mathbf{W}_L^H \mathbf{X}}{\mathbf{W}_R^H \mathbf{X}} \qquad ITF_{V_{out}}(\mathbf{W}) = \frac{\mathbf{W}_L^H \mathbf{V}}{\mathbf{W}_R^H \mathbf{V}}. \tag{6}$$

Next, we can write the desired ITFs of the speech and noise components, in function of the desired angles of the speech and noise components, θ_X and θ_V, and frequency, ω.

$$ITF_{X_{des}} = \frac{HRTF_{X_L}(\omega, \theta_X)}{HRTF_{X_L}(\omega, \theta_X)} \tag{7}$$

$HRTF_{X_L}(\omega, \theta_X)$ and $HRTF_{X_R}(\omega, \theta_X)$ are the head-related transfer functions (HRTF) for the speech component of the left and right ear. Similarly, the ITF of desired output ITF of the noise component, $ITF_{V_{des}}$, can be defined. Any set of HRTFs can be chosen. Therefore the direction of arrival of the speech and noise components can be controlled. In order to preserve the binaural cues of the speech and noise components, the original ITFs are selected as the desired ITFs. We assume that the original ITFs (5) to be constant [1] and can be computed using the microphone signals.

$$ITF_{X_{des}} = \frac{\mathcal{E}\left\{X_{L_0} X_{R_0}^*\right\}}{\mathcal{E}\left\{X_{R_0} X_{R_0}^*\right\}} \qquad ITF_{V_{des}} = \frac{\mathcal{E}\left\{V_{L_0} V_{R_0}^*\right\}}{\mathcal{E}\left\{V_{R_0} V_{R_0}^*\right\}} \tag{8}$$

4. BINAURAL WIENER FILTERING

In this section we derive a binaural Wiener filter that suppresses the noise component, while preserving the desired ITFs of the speech and noise component. We begin by looking at a binaural expansion of the speech distortion weighted cost function discussed in [7].

$$J(\mathbf{W}) = \mathcal{E}\left\{ \underbrace{\left\| \begin{bmatrix} X_{L_0} - \mathbf{W}_L^H \mathbf{X} \\ X_{R_0} - \mathbf{W}_R^H \mathbf{X} \end{bmatrix} \right\|^2}_{Speech\ Distortion} + \mu \underbrace{\left\| \begin{bmatrix} \mathbf{W}_L^H \mathbf{V} \\ \mathbf{W}_R^H \mathbf{V} \end{bmatrix} \right\|^2}_{Residual\ Noise} \right\} \tag{9}$$

The speech distortion and residual noise vectors can be broken into components that are parallel and perpendicular to the desired ITFs. This decomposition is depicted in Figure 2 for the residual noise

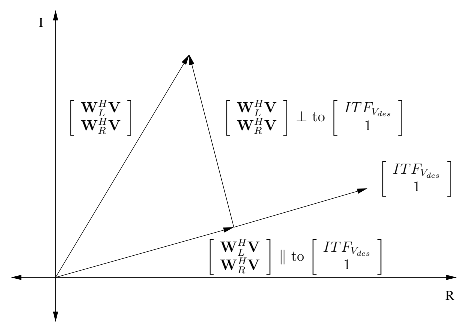


Fig. 2. Decomposition of residual noise vector

vector. Remember that this decomposition is performed for each frequency bin. In order to preserve the desired ITFs of the speech and noise components, the speech distortion and residual noise vectors need to be parallel to the desired ITF vectors. This can be done by putting a positive weight on the perpendicular terms. Therefore our cost function is now

$$J(\mathbf{W}) = \mathcal{E}\left\{ \left\| \begin{bmatrix} X_{L_0} - \mathbf{W}_L^H \mathbf{X} \\ X_{R_0} - \mathbf{W}_R^H \mathbf{X} \end{bmatrix}_{\parallel} \right\|^2 + \alpha_X^1 \left\| \begin{bmatrix} X_{L_0} - \mathbf{W}_L^H \mathbf{X} \\ X_{R_0} - \mathbf{W}_R^H \mathbf{X} \end{bmatrix}_{\perp} \right\|^2 + \mu \left(\left\| \begin{bmatrix} \mathbf{W}_L^H \mathbf{V} \\ \mathbf{W}_R^H \mathbf{V} \end{bmatrix}_{\parallel} \right\|^2 + \alpha_V^1 \left\| \begin{bmatrix} \mathbf{W}_L^H \mathbf{V} \\ \mathbf{W}_R^H \mathbf{V} \end{bmatrix}_{\perp} \right\|^2 \right) \right\}. \tag{10}$$

The residual noise terms in (10) can be rewritten as

$$\mu \left(\left\| \begin{bmatrix} \mathbf{W}_L^H \mathbf{V} \\ \mathbf{W}_R^H \mathbf{V} \end{bmatrix} \right\|^2 + (\alpha_V^1 - 1) \left\| \begin{bmatrix} \mathbf{W}_L^H \mathbf{V} \\ \mathbf{W}_R^H \mathbf{V} \end{bmatrix}_{\perp} \right\|^2 \right). \tag{11}$$

[1] In the case of a single noise source, this desired ITF is equal to the ratio of the acoustic transfer functions between the noise source and the reference microphone signals, i.e. $H_{0,r_0}/H_{0,r_1}$. In this case, it can also be easily 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.

A similar step can be taken for the speech distortion vector. Note, $\left\|\begin{bmatrix} X_{L_0}-\mathbf{W}_L^H\mathbf{X} \\ X_{R_0}-\mathbf{W}_R^H\mathbf{X}\end{bmatrix}_\perp\right\|$ and $\left\|\begin{bmatrix} \mathbf{W}_L^H\mathbf{X} \\ \mathbf{W}_R^H\mathbf{X}\end{bmatrix}_\perp\right\|$, both perpendicular to $\begin{bmatrix} X_{L_0} \\ X_{R_0}\end{bmatrix}$, are equivalent. Armed with this statement and defining new weights, α and β, the cost function, consisting of a speech distortion term, a noise reduction term and two ITF terms, is

$$J(\mathbf{W})=\mathcal{E}\Bigg\{\underbrace{\left\|\begin{bmatrix} X_{L_0}-\mathbf{W}_L^H\mathbf{X} \\ X_{R_0}-\mathbf{W}_R^H\mathbf{X}\end{bmatrix}\right\|^2+\mu\left\|\begin{bmatrix} \mathbf{W}_L^H\mathbf{V} \\ \mathbf{W}_R^H\mathbf{V}\end{bmatrix}\right\|^2}_{Original\ SDW\ Cost\ Function}+$$

$$\underbrace{\alpha\left\|\begin{bmatrix} \mathbf{W}_L^H\mathbf{X} \\ \mathbf{W}_R^H\mathbf{X}\end{bmatrix}_\perp\right\|^2+\beta\left\|\begin{bmatrix} \mathbf{W}_L^H\mathbf{V} \\ \mathbf{W}_R^H\mathbf{V}\end{bmatrix}_\perp\right\|^2}_{Additional\ ITF\ Terms}\Bigg\}. \quad (12)$$

Using the definition of the cross product, (12) can be written as

$$\mathbf{J}(\mathbf{W})=\mathcal{E}\Bigg\{\left\|\begin{bmatrix} X_{L_0}-\mathbf{W}_L^H\mathbf{X} \\ X_{R_0}-\mathbf{W}_R^H\mathbf{X}\end{bmatrix}\right\|^2+\mu\left\|\begin{bmatrix} \mathbf{W}_L^H\mathbf{V} \\ \mathbf{W}_R^H\mathbf{V}\end{bmatrix}\right\|^2+$$

$$\alpha\frac{\left|\mathbf{W}_L^H\mathbf{X}-ITF_{X_{des}}\mathbf{W}_R^H\mathbf{X}\right|^2}{\left\|\begin{bmatrix} ITF_{X_{des}} \\ 1\end{bmatrix}\right\|^2}+\beta\frac{\left|\mathbf{W}_L^H\mathbf{V}-ITF_{V_{des}}\mathbf{W}_R^H\mathbf{V}\right|^2}{\left\|\begin{bmatrix} ITF_{V_{des}} \\ 1\end{bmatrix}\right\|^2}\Bigg\}.$$

Next, we take the derivative of the above equation, set the derivative to zero, and solve for $\mathbf{W}$. The solution is expressed in matrix form below.

$$\mathbf{W}=\left(\mathcal{E}\Big\{\mathbf{R}_{\mathbf{R}_X}+\mu\mathbf{R}_{\mathbf{R}_V}+\alpha\mathbf{R}_{\mathbf{R}_{XC}}+\beta\mathbf{R}_{\mathbf{R}_{VC}}\Big\}\right)^{-1}\mathcal{E}\Big\{\mathbf{r}_X\Big\},$$

where,

$$\mathbf{r}_X=\begin{bmatrix} X_{L_0}^*\mathbf{X} \\ X_{R_0}^*\mathbf{X}\end{bmatrix} \quad \begin{matrix}\mathbf{R}_X=\mathbf{X}\mathbf{X}^H \\ \mathbf{R}_V=\mathbf{V}\mathbf{V}^H\end{matrix}$$

$$\mathbf{R}_{\mathbf{R}_X}=\begin{bmatrix} \mathbf{R}_X & \mathbf{0}_{2M} \\ \mathbf{0}_{2M} & \mathbf{R}_X\end{bmatrix} \quad \mathbf{R}_{\mathbf{R}_V}=\begin{bmatrix} \mathbf{R}_V & \mathbf{0}_{2M} \\ \mathbf{0}_{2M} & \mathbf{R}_V\end{bmatrix}$$

$$\mathbf{R}_{\mathbf{R}_{XC}}=\begin{bmatrix} \mathbf{R}_X & -ITF_{X_{des}}^*\mathbf{R}_X \\ -ITF_{X_{des}}\mathbf{R}_X & |ITF_{X_{des}}|^2\mathbf{R}_X\end{bmatrix}$$

$$\mathbf{R}_{\mathbf{R}_{VC}}=\begin{bmatrix} \mathbf{R}_V & -ITF_{V_{des}}^*\mathbf{R}_V \\ -ITF_{V_{des}}\mathbf{R}_V & |ITF_{V_{des}}|^2\mathbf{R}_V\end{bmatrix}$$

Note, because the desired ITFs are considered to be constant, the norm-squared terms are absorbed by the weights, α and β. This notation allows us to gain some crucial insight into the filter design. Clearly, if there is no correlation between the signals at the right and left ear, the filter design is decoupled. This is logical since there are no cues to preserve.

5. SIMULATIONS

5.1. Experimental setup

The recordings used in the following simulations were made in a reverberant room, $T_{60} = 0.76$sec. Two GN ReSound Canta behind the ear (BTE) hearing aids were placed on a CORTEX MK2

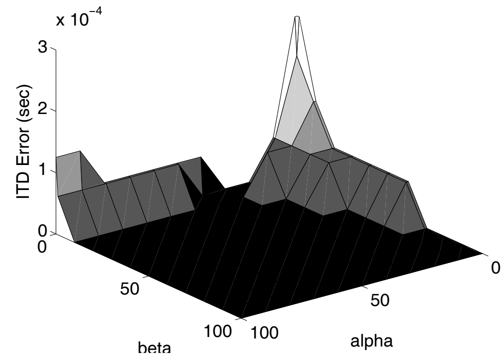


Fig. 3. Absolute ITD Error Noise component

artificial head. Each hearing aid had two omni-directional microphones. The sound level measured at the center of the dummy head was 70dB SPL. Speech and noise sources were recorded separately. All recordings were performed at a sampling frequency of 16kHz. HINT sentences and HINT noise were used for the speech and noise signals.

In the simulations both microphone signals from each hearing aid were used, $M = 2$, to estimate the speech component in the first microphone pair. The statistics were calculated off-line, and access to a perfect voice activity detection (VAD) algorithm was assumed. An FFT length of 512 was used. The parameters controlling the ITF of the speech and noise components, α and β, were varied from 0 to 100, while the parameter governing noise reduction, μ, was held constant at 1.

5.2. Performance measures

The purpose of the simulations is to show the effect of the parameters on ITD error, ILD error, and SNR improvement. The ITD metric used was the absolute difference between the ITD of the input signals and the output signals. ITD was calculated by cross correlation.

$$Absolute\ ITD\ Error\ =\ |ITD_{in}-ITD_{out}|$$

The second measure, expressed below, assessed the preservation of the ILD cues.

$$ILD\ Error=\frac{1}{N}\sum_{i=1}^{N}10\log_{10}\left(\left(\frac{P_{L_{in}}(i)}{P_{R_{in}}(i)}-\frac{P_{L_{out}}(i)}{P_{R_{out}}(i)}\right)^2\right)$$

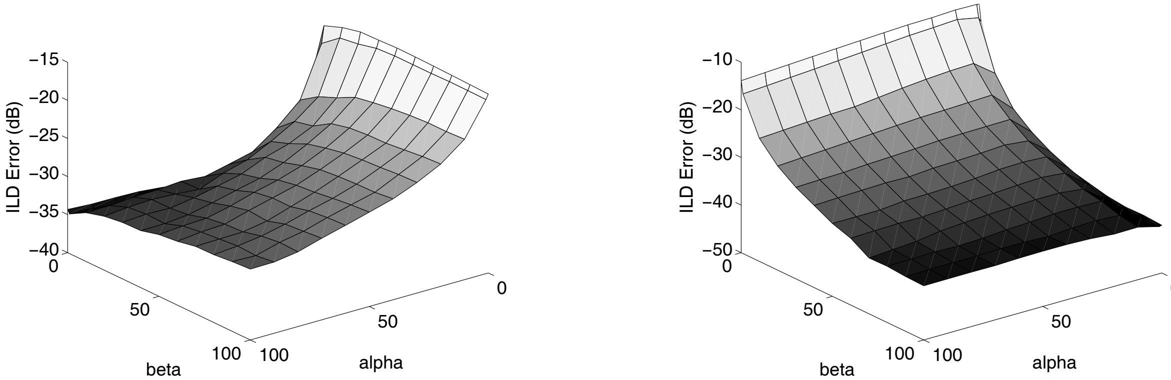
P stands for power and ILD error is averaged over the N frequency bins. In order to quantify the noise reduction performance, the speech intelligibility weighted signal-to-noise-ratio is used.

$$SNR_{INT}=\sum_{j=1}^{J}w_jSNR_j$$

The weight, w_j, emphasizes the importance of the jth $\frac{1}{3}$-octavefrequency band's overall contribution to intelligibility, and SNR_j is the signal-to-noise-ratio of the jth $\frac{1}{3}$-octave frequency band.

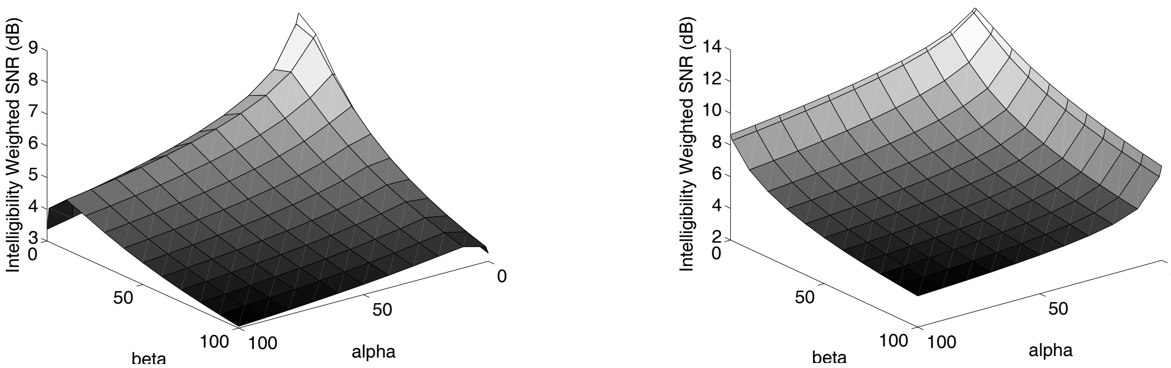
5.3. Results and discussion

First, the absolute ITD error of the speech component is not shown, since it is zero for all values of α and β. The absolute ITD error of



(a) Speech component (b) Noise component

Fig. 4. Mean squared error ILD



(a) Left ear (b) Right ear

Fig. 5. Improvement in Speech Intelligibility Weighted SNR

the noise component is depicted in Figure 3. Clearly, β can be chosen to preserve the ITD of the noise component. Figure 4 shows the average mean square ILD error of the speech and noise component. we note that with appropriate values of α and β the ILD cues of the speech and noise components are preservable.

Finally, we turn our attention to Figure 5. Expectedly, as more emphasis is placed on preserving the ITF of the speech and noise components, the improvement in speech intelligibility weighted SNR decreases. Nevertheless, respectable gains in SNR are achieved.

Unfortunately, we are left with the dilemma of choosing α and β. Naturally, this decision depends on the user and the situation. Additionally, further research could focus on moving the noise source to a desired position. This would guarantee a separation between speer and noise source, and would lead to improvements in intelligibility.

6. CONCLUSION

This paper presented a binaural Wiener filter extended by incorporating two terms in the cost function that account for the ITFs of the speech and noise components. Using weights, the emphasis on the preservation of the ITF of the speech and noise component can be controlled in addition to the emphasis on noise reduction. Adapting theses parameters allows one to preserve the ITF of the speech and noise component, and therefore ITD and ILD cues, while enhancing the signal-to-noise ratio.

7. REFERENCES

[1] T.J. Klasen, T. Van den Bogaert, M. Moonen, and J. Wouters, “Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues,” Submitted Jan., 2005.

[2] T. Van den Bogaert, T.J. Klasen, L. Van Deun, J. Wouters, and M. Moonen, “Horizontal localization with bilateral hearing aids: without is better than with,” Accepted for publication in *J. Acoust. Soc. Amer.* 2005.

[3] N.P. Erber, “Auditory-visual perception of speech,” *J. Speech Hearing Dis.*, vol. 40, pp. 481–492, 1975.

[4] Monica L. Hawley, Ruth Y. Litovsky, and John F. Culling, “The benefit of binaural hearing in a cocktail party: Effect of locaiton and type of interferer,” *J. Acoust. Soc. Amer.*, vol. 115, no. 2, pp. 833–843, Feb. 2004.

[5] W.M. Hartmann, “How We Localize Sound,” *Physics Today*, pp. 24–29, Nov. 1999.

[6] S. Doclo, R. Dong, T.J. Klasen, J. Wouters, S. Haykin, and M. Moonen, “Extension of the multi-channel Wiener filter with ITD and ILD cues for noise reduction in binaural hearing aids,” in Proc. IWAENC*, Eindhoven, The Netherlands*, Sep. 2005.

[7] A. Spriet, M. Moonen, and J. Wouters, “Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction,” *Signal Processing*, vol. 84, no. 12, pp. 2367–2387, Dec. 2004.