Binaural noise reduction through Wiener filtering for hearing aids preserving interaural transfer functions

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

***Abstract*— This paper presents an extension of the binaural multi-channel Wiener filtering based noise reduction algorithm discussed in [2]. The goal of this extension 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 underlying 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. Additionally, the desired ITFs can be replaced by known ITFs for a specific direction of arrival. Preserving these desired ITFs allows one to change the direction of arrival of the speech and noise sources.**

EDICS Category: AUD-LMAP, AUD-AUDI, and SPE-ENHA

I. Introduction

HEARING impaired persons localize sounds better without their bilateral hearing aids than with them [3]. In addition, noise reduction algorithms currently used in hearing aids are not designed to preserve localization cues [4]. 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 [5].

Studies have shown that preserving the spatial separation between the speech and noise sources contributes to an improvement in speech intelligibility by a process referred to as spatial release from masking [6], [7]. Therefore the benefit of a noise reduction algorithm that preserves localization cues is twofold. First, noise reduction leads to an improvement in intelligibility. Additionally, preserving localization cues preserves the spatial separation of the target speech and noise sources, resulting again in an improvement in intelligibility. Furthermore, an algorithm that intentionally *distorts* the localization cues of the speech and noise sources to improve the spatial separation of speech and noise could lead to improvements in intelligibility.

Part of this manuscript has been presented at ICASSP 2006 in Toulouse, France [1]. This research work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven, in the frame of the Belgian Program on Inter-university Attraction Poles, initiated by the Belgian Federal Science Policy Office IUAP P5/22 ('Dynamical Systems and Control: Computation, Identification and Modeling'), 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'). The scientific responsibility is assumed by its authors.

A hearing impaired person wearing a monaural hearing aid on each ear is said to be using bilateral hearing aids. Each monaural hearing aid uses its own microphone inputs to generate an output for its respective ear. No information is shared between the hearing aids. Contrastingly, binaural hearing aids use the microphone inputs from both the left and right hearing aids, possibly through a wireless link, to generate an output for the left and right ear.

Interaural time delay (ITD) and interaural level difference (ILD) are the basis for horizontal sound localization [8]. 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. 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 binaural noise reduction algorithm that allows one to control the ITD and ILD cues at the output.

In [9], the Wiener filter cost function used in a noise reduction procedure 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 compromising the noise reduction performance. However, iterative optimization techniques are required 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 deals with the problem of binaural cue preservation by preserving the ITF. In this paper an extension of the binaural Wiener filter [2] 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 [10]–[12]. The remaining two terms aim at preserving the ITFs of the speech and noise component. Contrary to the Wiener filter extensions proposed in [2], this algorithm co-designs the right and left filter. The purpose of this paper is to explore the performance of this algorithm in a well-defined scenario.

This paper is comprised of seven sections. In section II, the

system model is introduced. Additionally, the notation used in this paper is presented. The ITF is defined in section III. In section IV, the original speech distortion weighted binaural Wiener filtering cost function is reviewed. Next, the cost function is extended by adding two terms to control the ITFs of the speech and noise component at the output. Performance measures and the experimental setup are presented in section V. Finally, the simulation results and future research questions are discussed.

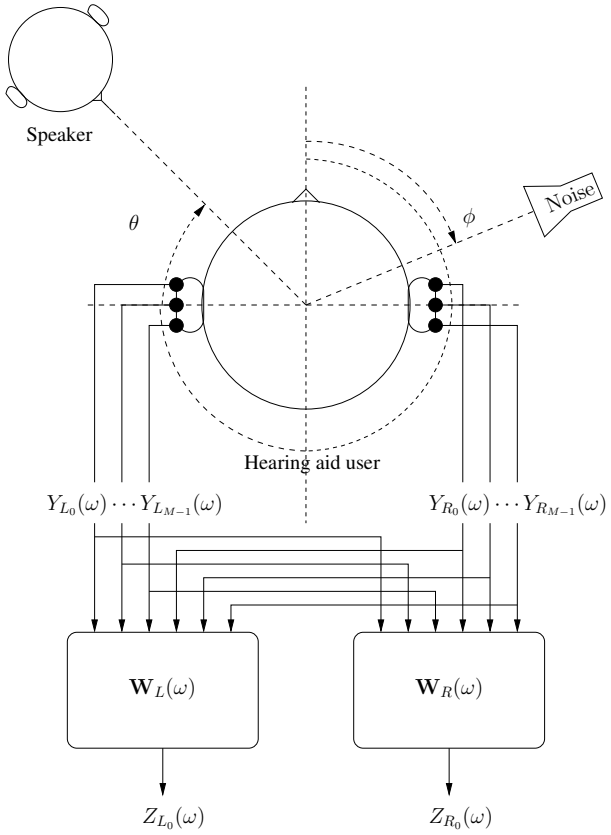


Fig. 1. Binaural hearing aid user in a single noise source listening scenario

II. System Model

Figure 1 shows a binaural hearing aid user in a single noise source 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 as

$$Y_{L_m}(\omega) = X_{L_m}(\omega) + V_{L_m}(\omega) \quad (1)$$
$$Y_{R_m}(\omega) = X_{R_m}(\omega) + V_{R_m}(\omega), \quad (2)$$

where $X_{L_m}(\omega)$ 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.

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.

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

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 *2M*-dimensional filters for the left and right hearing aid.

$$\mathbf{W}_L(\omega) = \left[W_{L_0}(\omega)\dots W_{L_{2M-1}}(\omega)\right]^T \quad (4)$$
$$\mathbf{W}_R(\omega) = \left[W_{R_0}(\omega)\dots W_{R_{2M-1}}(\omega)\right]^T \quad (5)$$

Using (4) and (5), we write the *4M*-dimensional stacked filter,

$$\mathbf{W}(\omega) = \begin{bmatrix} \mathbf{W}_L(\omega) \\ \mathbf{W}_R(\omega) \end{bmatrix}. \quad (6)$$

The outputs of the left and right filters are written below.

$$Z_{L_0}(\omega) = \mathbf{W}_L^H(\omega)\mathbf{Y}(\omega) \quad Z_{R_0}(\omega) = \mathbf{W}_R^H(\omega)\mathbf{Y}(\omega) \quad (7)$$

In this paper the outputs of the left and right filters are the estimates of the speech components in the first microphone pair. Nevertheless, the algorithm could be designed to estimate any microphone pair. For clarity the frequency domain variable, ω, will be omitted, whenever possible, throughout the remainder of this paper.

III. Interaural Transfer Function

In this section we define the desired ITFs of the speech and noise components. The cost function in section IV will incorporate these desired ITFs. In order to preserve the ITFs of the speech and noise components, we simply have to set the desired ITFs equal to the original ITFs of the input signals. Correspondingly, the localization cues, ITD and ILD cues, of the speech and noise components can be preserved. Alternatively, the perceived location of the speech and noise component can be manipulated by choosing the appropriate desired ITFs.

The ITF is defined as the ratio of the signal in the left ear to the signal in the right ear. The input speech and noise ITFs are written below.

$$ITF_{X_{in}} = \frac{X_{L_0}}{X_{R_0}} \qquad ITF_{V_{in}} = \frac{V_{L_0}}{V_{R_0}}. \quad (8)$$

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

$$ITF_{X_{out}}(\mathbf{W}) = \frac{\mathbf{W}_L^H\mathbf{X}}{\mathbf{W}_R^H\mathbf{X}} \quad ITF_{V_{out}}(\mathbf{W}) = \frac{\mathbf{W}_L^H\mathbf{V}}{\mathbf{W}_R^H\mathbf{V}}. \quad (9)$$

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 (8) are constant[1]

[1]In the case of a single source, this ITF is equal to the ratio of the acoustic transfer functions between the source and the reference microphone signals at the left and right hearing aids. 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.

and can be estimated, in a least-squares sense, using the microphone signals.

$$ITF_{X_{des}} = \frac{\mathcal{E}\left\{X_{L_0}X_{R_0}^*\right\}}{\mathcal{E}\left\{X_{R_0}X_{R_0}^*\right\}} \quad ITF_{V_{des}} = \frac{\mathcal{E}\left\{V_{L_0}V_{R_0}^*\right\}}{\mathcal{E}\left\{V_{R_0}V_{R_0}^*\right\}} \quad (10)$$

Alternatively, the direction of arrival of the speech and noise components can be controlled by choosing any set of head-related transfer functions (HRTFs). For simplicity, the desired ITFs of the speech and noise components are written in function of the desired angles of the speech and noise components, θ_X and θ_V, and frequency, ω.

$$ITF_{X_{des}} = \frac{HRTF_{X_L}(\omega,\theta_X)}{HRTF_{X_R}(\omega,\theta_X)} \quad (11)$$

$$ITF_{V_{des}} = \frac{HRTF_{V_L}(\omega,\theta_V)}{HRTF_{V_R}(\omega,\theta_V)} \quad (12)$$

$HRTF_{X_L}(\omega,\theta_X)$ and $HRTF_{X_R}(\omega,\theta_X)$ are the HRTFs for the speech component of the left and right ear. Similarly, $HRTF_{V_L}(\omega,\theta_V)$ and $HRTF_{V_R}(\omega,\theta_V)$ are the HRTFs for the noise component of the left and right ear.

In this paper we will address both situations. First we will look at the performance of the algorithms when trying to preserve the original ITFs. Later the possibility of manipulating the ITFs of the speech and noise components will be explored.

IV. Binaural multi-channel Wiener filtering

In this section we derive the binaural multi-channel Wiener filter that performs noise reduction, while at the same time attempting preserving the ITFs of the speech and noise component. We begin by looking at the binaural expansion of the speech distortion weighted cost function discussed in [10]–[12]. Using the reasoning from Section III the cost function is manipulated to incorporate two terms used to preserve the ITFs of the speech and noise components. The final cost function contains the original speech distortion weighted terms plus two additional terms for the ITFs of the speech and noise components.

A. Original cost function

The multi-channel Wiener filter generates a minimum mean-square error estimate of the speech component in the first microphone pair[2] [2], [10], [11], [11], [12]. The original binaural cost function is written as,

$$J(\mathbf{W}) = \mathcal{E}\left\{\left\|\begin{bmatrix} X_{L_0} - \mathbf{W}_L^H\mathbf{Y} \\ X_{R_0} - \mathbf{W}_R^H\mathbf{Y}\end{bmatrix}\right\|^2\right\}. \quad (13)$$

In [10]–[12] the original cost function is split into two terms. The first term quantifies speech distortion and the second residual noise. Next a weight, μ, is added to introduce a trade-off between speech distortion and noise reduction. Analogously, this reasoning can be applied to the binaural cost function in (13). The binaural speech distortion weighted cost function is expressed below.

$$J(\mathbf{W}) = \mathcal{E}\left\{\underbrace{\left\|\begin{bmatrix} X_{L_0} - \mathbf{W}_L^H\mathbf{X} \\ X_{R_0} - \mathbf{W}_R^H\mathbf{X}\end{bmatrix}\right\|^2}_{Speech\ Distortion} + \mu\underbrace{\left\|\begin{bmatrix} \mathbf{W}_L^H\mathbf{V} \\ \mathbf{W}_R^H\mathbf{V}\end{bmatrix}\right\|^2}_{Residual\ Noise}\right\} \quad (14)$$

B. Cost function incorporating ITFs

In order to incorporate the ITFs of the speech and noise components, the output speech and residual noise vectors, $\begin{bmatrix} \mathbf{W}_L^H\mathbf{X} \\ \mathbf{W}_R^H\mathbf{X}\end{bmatrix}$ and $\begin{bmatrix} \mathbf{W}_L^H\mathbf{V} \\ \mathbf{W}_R^H\mathbf{V}\end{bmatrix}$, are broken into components that are parallel and perpendicular to the desired ITF vector. Since only the direction of the desired ITF vector is important, we can write the desired noise ITF vector as,

$$\begin{bmatrix} V_{L_0} \\ V_{R_0}\end{bmatrix} \parallel to \begin{bmatrix} HRTF_{V_L}(\omega,\theta_V) \\ HRTF_{V_R}(\omega,\theta_V)\end{bmatrix} \parallel to \begin{bmatrix} ITF_{V_{des}} \\ 1\end{bmatrix}. \quad (15)$$

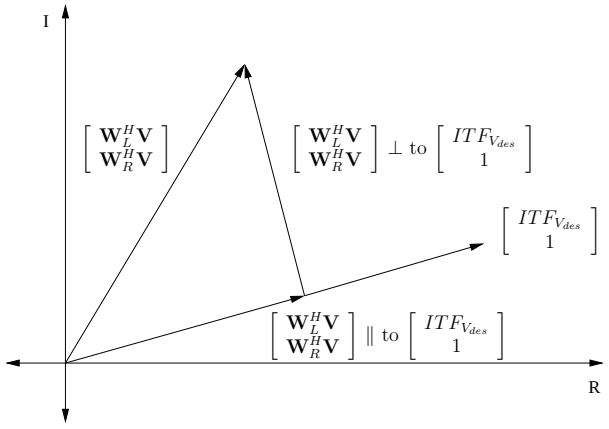


Fig. 2. Decomposition of residual noise vector

The decomposition of the residual noise vector is depicted in Figure 2. A similar decomposition can be obtained for the output speech 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 output speech and residual noise vectors need to be parallel to the desired speech and noise ITF vectors. This can be done by putting a positive weight on the perpendicular terms. Therefore our cost function is now

$$J(\mathbf{W}) = \mathcal{E}\Bigg\{\underbrace{\left\|\begin{bmatrix} X_{L_0} - \mathbf{W}_L^H\mathbf{X} \\ X_{R_0} - \mathbf{W}_R^H\mathbf{X}\end{bmatrix}\right\|^2 + \mu\left\|\begin{bmatrix} \mathbf{W}_L^H\mathbf{V} \\ \mathbf{W}_R^H\mathbf{V}\end{bmatrix}\right\|^2}_{Original\ SDW\ Cost\ Function} + \underbrace{\alpha\left\|\begin{bmatrix} \mathbf{W}_L^H\mathbf{X} \\ \mathbf{W}_R^H\mathbf{X}\end{bmatrix}_\perp\right\|^2 + \beta\left\|\begin{bmatrix} \mathbf{W}_L^H\mathbf{V} \\ \mathbf{W}_R^H\mathbf{V}\end{bmatrix}_\perp\right\|^2}_{Additional\ ITF\ Terms}\Bigg\}. \quad (17)$$

Using the definition of the cross product, (17) can be written as (16). Next, we take the derivative of (16), set the derivative to zero, and solve for $\mathbf{W}$. The solution is expressed in matrix

[2]The mth microphone pair can also be used.

$$J(\mathbf{W}) = \mathcal{E}\left\{ \left\| \begin{bmatrix} X_{L_0} - \mathbf{W}_L^H \mathbf{X} \\ X_{R_0} - \mathbf{W}_R^H \mathbf{X} \end{bmatrix} \right\|^2 + \mu \left\| \begin{bmatrix} \mathbf{W}_L^H \mathbf{V} \\ \mathbf{W}_R^H \mathbf{V} \end{bmatrix} \right\|^2 + \alpha \frac{\left| \mathbf{W}_L^H \mathbf{X} - ITF_{X_{des}} \mathbf{W}_R^H \mathbf{X} \right|^2}{\left\| \begin{bmatrix} ITF_{X_{des}} \\ 1 \end{bmatrix} \right\|^2} + \beta \frac{\left| \mathbf{W}_L^H \mathbf{V} - ITF_{V_{des}} \mathbf{W}_R^H \mathbf{V} \right|^2}{\left\| \begin{bmatrix} ITF_{V_{des}} \\ 1 \end{bmatrix} \right\|^2} \right\}. \tag{16}$$

form below.

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

where,

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

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

$$\mathbf{R}_{\mathbf{R}_{XC}} = \frac{\begin{bmatrix} \mathbf{R}_X & -ITF_{X_{des}}^* \mathbf{R}_X \\ -ITF_{X_{des}} \mathbf{R}_X & |ITF_{X_{des}}|^2 \mathbf{R}_X \end{bmatrix}}{\left\| \begin{matrix} ITF_{X_{des}} \\ 1 \end{matrix} \right\|^2}$$

$$\mathbf{R}_{\mathbf{R}_{VC}} = \frac{\begin{bmatrix} \mathbf{R}_V & -ITF_{V_{des}}^* \mathbf{R}_V \\ -ITF_{V_{des}} \mathbf{R}_V & |ITF_{V_{des}}|^2 \mathbf{R}_V \end{bmatrix}}{\left\| \begin{matrix} ITF_{V_{des}} \\ 1 \end{matrix} \right\|^2}$$

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. Additionally, if α and β are chosen to be zero, then the left and right filter design becomes independent. And the filters are those from the original binaural speech distortion weighted cost function in (14).

V. Simulations

A. Experimental setup

Two sets of simulations were run. The first set of simulations attempted to show the algorithm's ability to preserve the original ITFs of the speech and noise components. The second set of simulations showed how altering the algorithm's desired ITFs can shift the perceived location of the noise source.

The recordings used in the 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 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 [13].

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 offline, and access to a perfect voice activity detection (VAD) algorithm was assumed. An FFT length of 256 was used.

For the first set of simulations the speech source was located in front of the artificial head, $-15°$, and the noise source was located at $60°$. The input speech intelligibility weighted signal-to-noise ratio was 2.9dB in the front left microphone and -6.9dB in the front right microphone. The parameter controlling the ITF of the speech component, α, was varied from 0 to 100 and the parameter controlling the ITF of the noise component, β, was varied from 0 to 100. The parameter governing noise reduction, μ, was held constant at 1.

The same setup was used for the second set of simulations. However, this time the desired noise ITF was not the least squares estimate of the actual noise ITF, but the ITF for a source located at $270°$. This ITF was calculated using the HRTFs for a source located at $270°$. Again, α, was varied from 0 to 100 and β was varied from 0 to 100. The noise reduction parameter, μ, was held constant at 1.

B. Performance measures

The purpose of the simulations is to show the effect of the parameters on ITD error, ILD error, output SNR, and ITF error. The ITD metric is written as

$$\text{ITD Error} = \frac{1}{N} \sum_{i=1}^{N} F(\omega_i) \Delta \text{ITD}_x(\omega_i),$$

where $\Delta \text{ITD}_x(\omega_i)$ is computed as the cosine of the phase difference between the frequency-domain cross-correlation between the left and the right hearing aid, and $F(\omega_i)$ a weighting function equal to 1 for frequencies below $f_c = 1500$,Hz and equal to 0 for frequencies above $f_c = 1500\,\text{Hz}$ (since ITD errors at high frequencies do not lead to incorrect localization.

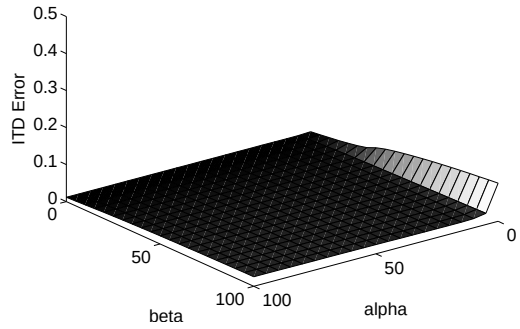
The second measure, expressed below, assessed the preservation of the ILD cues. The average over frequency bins of the absolute difference of the ILD of the input signals and ILD of the output signals is used.

$$\text{ILD Error} = \frac{1}{N} \sum_{i=1}^{N} F(\omega_i) \Delta \text{ILD}_x(\omega_i)$$

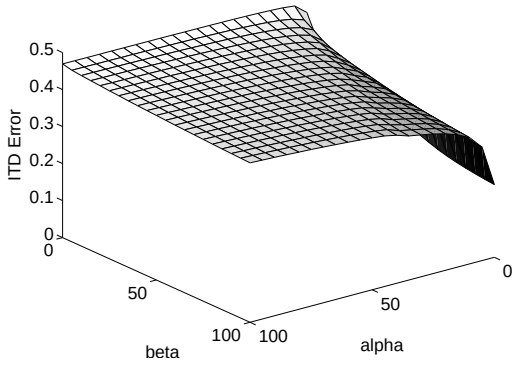
$$= \frac{1}{N} \sum_{i=1}^{N} F(\omega_i) \left| 10 \log_{10} \frac{P_{L_{in}}(\omega_i)}{P_{R_{in}}(\omega_i)} - 10 \log_{10} \frac{P_{L_{out}}(\omega_i)}{P_{R_{out}}(\omega_i)} \right|$$

P stands for power and ILD error is averaged over the N frequency bins. In this case the frequency weighting function $F(\omega_i)$ is always equal to 1.

In order to quantify the noise reduction performance, the speech intelligibility weighted signal-to-noise-ratio, defined in

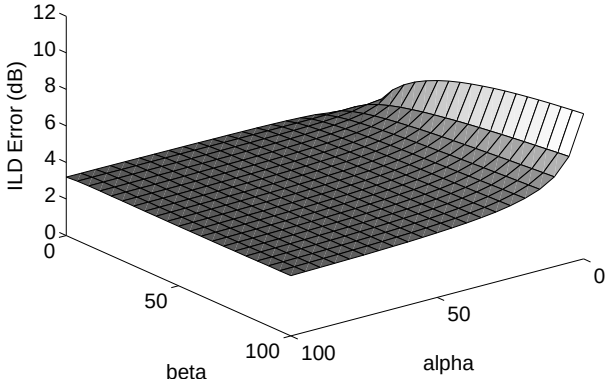


(a) Speech component

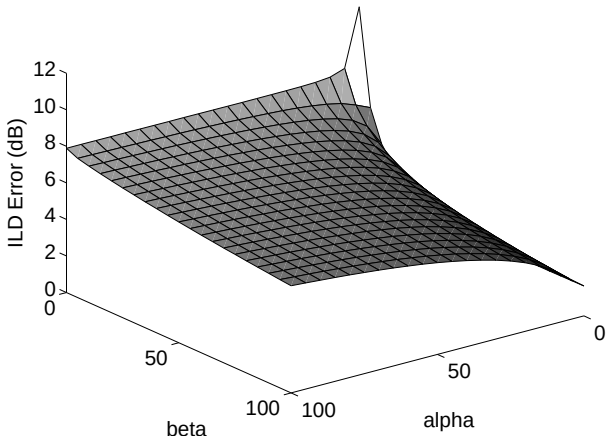


(b) Noise component

Fig. 3. Absolute ITD Error



(a) Speech component



(b) Noise component

Fig. 4. Mean squared error ILD

[14], is used.

$$\mathrm{SNR}_{\mathrm{INT}} = \sum_{j=1}^{J} w_j SNR_j$$

The weight, w_j, emphasizes the importance of the jth $\frac{1}{3}$-octave frequency band's overall contribution to intelligibility, and SNR_j is the signal-to-noise-ratio of the jth $\frac{1}{3}$-octave frequency band. The band definitions and the individual weights of the J frequency bands are given in [15].

Finally, the ITF error corresponds to the ITF terms of the speech and noise component in the cost function, $\mathbf{W}^H\mathbf{R}_{\mathbf{R}_{XC}}\mathbf{W}$ and $\mathbf{W}^H\mathbf{R}_{\mathbf{R}_{VC}}\mathbf{W}$.

C. Results and discussion[3]

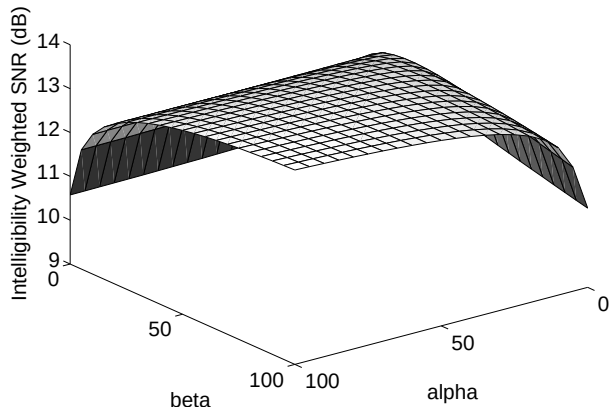
The first set of simulations attempted to show the algorithm's ability to preserve the original ITFs of the speech and noise components. Figure 3 shows the ITD error for the speech and noise component. The ILD error is depicted in Figure 4. The output speech intelligibility weighted SNR can be seen in Figure 5. Finally, Figure 6 illustrates the ITF error for the speech and noise component.

One should begin by looking at the ITF error of the speech and noise component. Clearly, it can be seen from Figure 6(a) that the ITF error of the speech component decreases as α increases. However, increasing the parameter β, attempting to preserve the ITF of the noise component, causes the ITF error of the speech component to increase. Similarly, the ITF error of the noise component decreases as β increases. Again, the other parameter, α, has a negative influence on the ITF error. Clearly, there is a trade off between preserving the ITF of the speech and noise component. This trade off is also visible in other performance metrics.

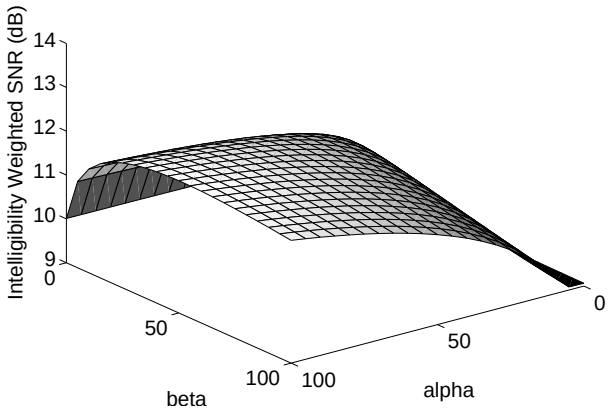
Now we turn our attention to the ITD cues of the speech components in Figure 3(a). First, it is important to notice that the speech ITD error is very small, 0.01, for the original binaural multi-channel Wiener filtering scheme, $\alpha = 0$ and $\beta = 0$, proposed in [2]. Moreover, the speech ITD error very small for almost all combinations of α and β. They are only distorted when β is large and $alpha$ is small.

In Figure 3(b), clearly the ITD cues of the noise component are distorted for the original binaural multi-channel Wiener filtering approach. Here the ITD error is 0.4. Increasing the

[3]Sounds samples can be found on
www.esat.kuleuven.be/∼tklasen/soundsamples/

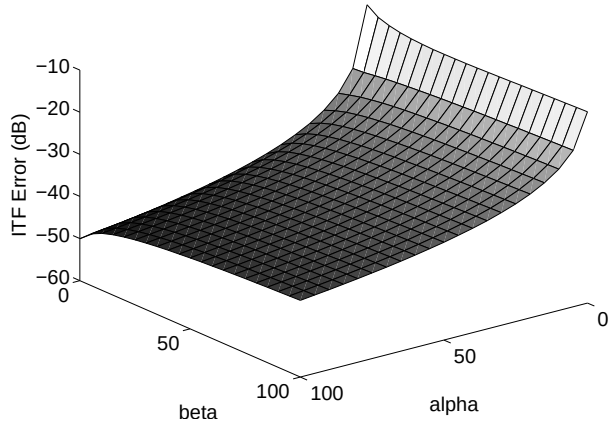


(a) Left ear

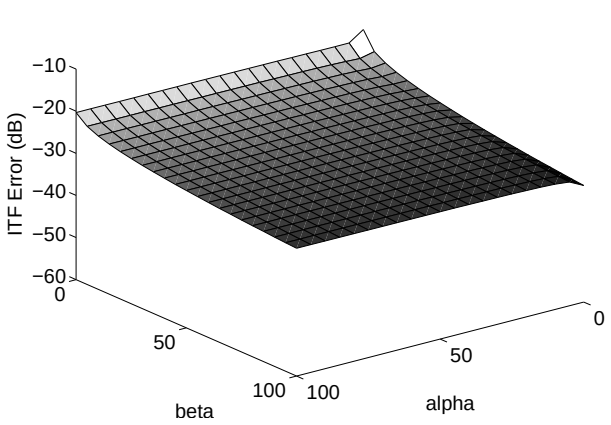


(b) Right ear

Fig. 5. Output Speech Intelligibility Weighted SNR



(a) Speech component



(b) Noise component

Fig. 6. ITF Error

parameter β leads to a reduction in noise ITD error. However, the noise ITD error is dependent on α. A small increase in α can cause the noise ITD cues to be distorted.

The ILD error of the speech component, depicted in Figure 4(a). Although the ILD error of the speech component for the original binaural multi-channel Wiener filtering algorithm is not visible in Figure 4(a), it is small, 0.7dB. A small amount of distortion is visible when β is increased.

On the other hand, the ILD cues of the noise component are clearly distorted, the ILD error is 11.4dB , when α and β are both zero. As β is increased, the ILD error of the noise component decreases. The parameter α has some influence on the ILD error of the noise component for $\beta > 0$.

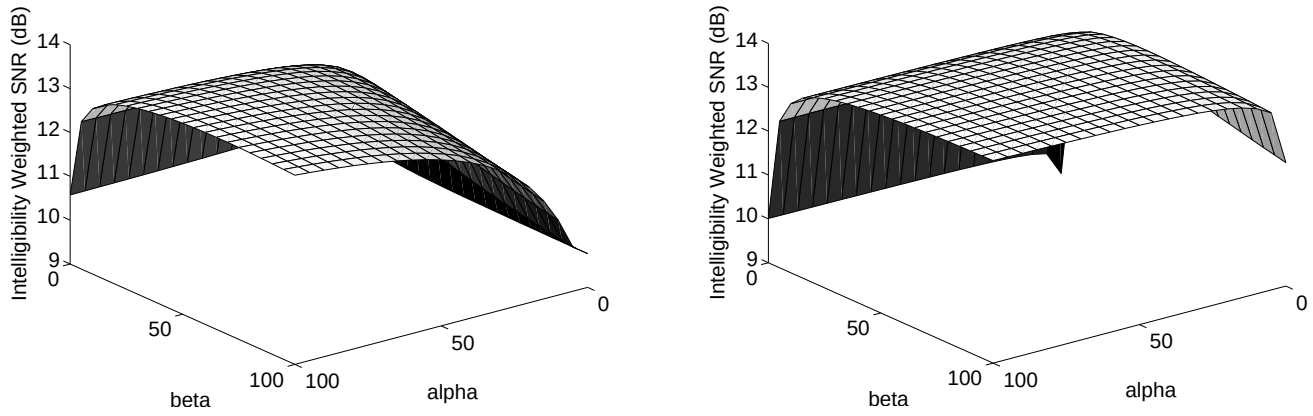
Finally, the output speech intelligibility weighted SNR for the left and right ear is shown in Figure 5. Clearly, regardless of the values of α and β this algorithm improves the SNR. Varying α and β causes some fluctuation in output SNR.

The second set of simulations is designed to show how altering the algorithm's desired ITFs can shift the perceived location of a source. In this case we focus on shifting the noise source from its original location at 60° to a new location at 270°. The main performance measure we will use is the value of the ITF terms from the cost function. The ITF error is plotted in Figure 8, and the output speech intelligibility weighted SNR is depicted in Figure 7.

Figure 8 shows that the noise component can be shifted from a location of 60° to a perceived location of 270°, while preserving the ITF of the speech source. Again, as β is increased, the ITF error decreases. Additionally, by looking at Figure 7 it is clear that even while altering the perceived location of the noise source, an improvement in speech intelligibility weighted SNR is possible.

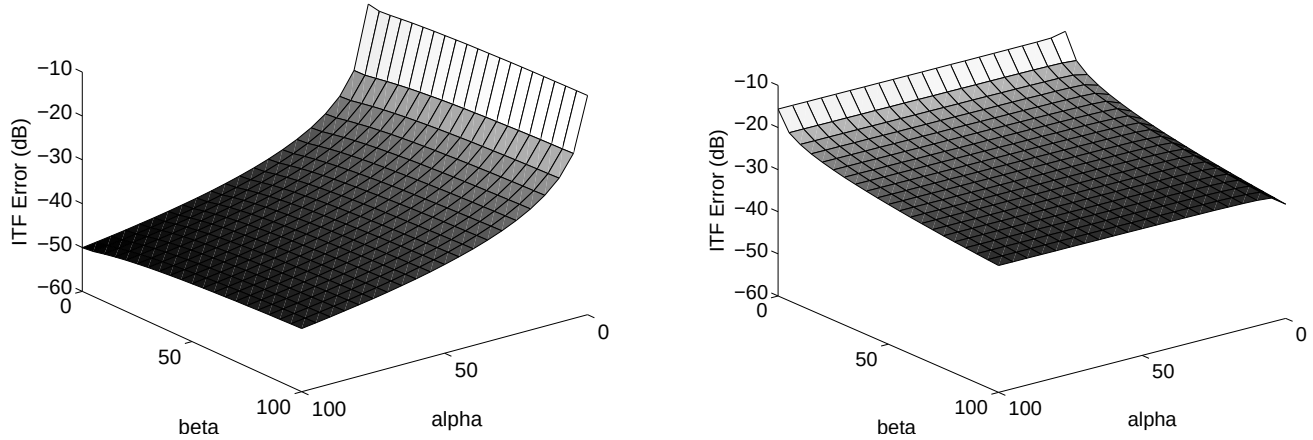
VI. Future research

We have shown that for the correct choice of parameters it is possible to preserve the current acoustical situation. It is even possible to alter the current acoustical situation to a more favorable one by moving noise sources. However, how does one choose these parameters in real life? Future research will focus on developing a method for choosing the parameters automatically in function of the current acoustical situation.



(a) Left ear (b) Right ear

Fig. 7. Output Speech Intelligibility Weighted SNR with desired noise ITF located at 270°



(a) Speech component (b) Noise component

Fig. 8. ITF Error with desired noise ITF located at 270°

Moreover, we have only focused on parameters that are uniform over frequency. It is possible, and probably advantageous, to choose α, β, and μ to be frequency dependent. One possibility would be to choose these parameter in function of the speech and noise power in each frequency bin. It does not make sense to try to preserve the ITF of a component in a frequency bin where that component is not present. Conversely, it would be beneficial to make sure the ITF of the component is preserved when a frequency bin contains a large amount of that component. This could lead to better preservation of the localization cues and help reduce the interdependencies among the parameters.

Throughout this paper, we have been dealing with a single noise source scenario. However, in the real world the noise scenarios will be much more complex, and will certainly consist of multiple noise sources. Future research will look into preserving the ITFs of multiple noise sources.

Finally, the performance of this algorithm relies on VAD. A discussion of the influence of VAD errors on monaural multi-channel Wiener filtering techniques can be found in [16]. Future work will look into the influence of VAD errors on the binaural extension. Moreover, we only explored the effect of the noise reduction techniques on ITD and ILD cues using error metrics. Further research will be conducted to explore this technique perceptually, specifically its effect on speech reception thresholds and localization performance. It is possible that these results will also lead to some insight into the parameters.

VII. Conclusion

This paper presented a binaural Wiener filter based noise reduction procedure 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 these 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. Additionally, it has been shown that the algorithm can even shift the noise source to a new location, by using a different desired ITF

for the noise source, while maintaining good noise reduction performance.

REFERENCES

[1] 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. of the 2006 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toulouse, France*, vol. V, May 2006, pp. 145–148.

[2] T. Klasen, T. Van den Bogaert, M. Moonen, and J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," accepted to *IEEE Trans. on Signal Processing*, available at ftp://ftp.esat.kuleuven.ac.be/pub/SISTA/∼tklasen/reports/05-10.pdf.

[3] T. Van den Bogaert, T. Klasen, L. Van Deun, J. Wouters, and M. Moonen, "Horizontal localization with bilateral hearing aids: without is better than with," *J. Acoust. Soc. Amer.*, vol. 119, no. 1, pp. 515–526, Jan. 2006.

[4] J. Desloge, W. Rabinowitz, and P. Zurek, "Microphone-Array Hearing Aids with Binaural Output-Part I: Fixed-Processing Systems," *IEEE Trans. Speech Audio Processing*, vol. 5, no. 6, pp. 529–542, Nov. 1997.

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

[6] M. L. Hawley, R. Y. Litovsky, and J. F. Culling, "The benefit of binaural hearing in a cocktail party: Effect of localization and type of interferer," *J. Acoust. Soc. Amer.*, vol. 115, no. 2, pp. 833–843, Feb. 2004.

[7] J. Peissig and B. Kollmeier, "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners," *J. Acoust. Soc. Amer.*, vol. 101, no. 3, pp. 1660–1670, 1997.

[8] W. Hartmann, "How We Localize Sound," *Physics Today*, pp. 24–29, Nov. 1999.

[9] S. Doclo, R. Dong, T. 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, pp. 221–224.

[10] 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.

[11] S. Doclo and M. Moonen, "GSVD-Based Optimal Filtering for Single and Multi-Microphone Speech Enhancement," *IEEE Trans. Signal Processing*, vol. 50, no. 9, pp. 2230–2244, Sep. 2002.

[12] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, *Speech Enhancement*. Springer-Verlag, 2005, ch. Speech Distortion Weighted Multichannel Wiener Filtering Techniques for Noise Reduction, pp. 199–228.

[13] M. Nilsson, S. Soli, and J. 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, 1994.

[14] J. Greenberg, P. Peterson, and P. Zurek, "Intelligibility-weighted measures of speech-to-interference ratio and speech system performance," *J. Acoust. Soc. Amer.*, vol. 94, no. 5, pp. 3009–3010, Nov. 1993.

[15] Acoustical Society of America, "American National Standard Methods for Calculation of the Speech Intelligibility Index," in *ANSI S3.5-1997*, 1997.

[16] A. Spriet, M. Moonen, and J. Wouters, "The impact of speech detection errors on the noise reduction performance of multi-channel Wiener filtering and Generalized Sidelobe Cancellation," *Signal Processing*, vol. 85, no. 6, pp. 1073–1088, Jun. 2005.