BINAURAL CUE PRESERVATION FOR HEARING AIDS USING AN INTERAURAL TRANSFER FUNCTION MULTICHANNEL WIENER FILTER

Tim Van den Bogaert [1,2], *Jan Wouters*[1]

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

Simon Doclo[2], *Marc Moonen*[2]

[2]ESAT–SCD
Dept. of Electrotechn. Eng., K.U.Leuven
Kasteelpark Arenberg 10, 3001 Leuven, Belgium

ABSTRACT

This paper describes the binaural cue preservation of a noise reduction algorithm for bilateral hearing aids, namely the multichannel Wiener filter with interaural transfer function extension (MWF-ITF). An extra term is added to the cost function to preserve the binaural cues of both the speech and noise component of a signal at the cost of some noise reduction. This paper combines the theoretical analysis with objective binaural performance measures and a perceptual evaluation.

Index Terms— hearing aids, binaural hearing, noise reduction, adaptive filter, localization

1. INTRODUCTION

Noise reduction algorithms in hearing aids are crucial for hearing impaired persons to improve speech intelligibility in background noise. Multi-microphone systems are able to exploit spatial in addition to spectral information and are hence preferred to single-microphone systems [1, 2]. However, hearing aid users often localize sounds better when switching off the noise reduction in their hearing aids [3]. This is not suprising, since noise reduction algorithms currently used in hearing aids are not designed to preserve binaural cues. This puts the hearing aid user at a disadvantage as well as at risk. Sound localization is important in speech segregation in noisy environments (a.k.a. the cocktail party effect), and in certain situations, such as traffic, incorrect localization of sounds could even endanger the user. Typical adaptive multi-microphone noise reduction techniques have a single output which contains the best estimate of the desired signal. Changing from a monaural to a binaural hearing aid design, i.e. generating an output signal for both ears, may destroy the binaural cues present between the signals arriving at both eardrums. It was proven that the multi-channel Wiener filter (MWF) perfectly preserves the binaural cues of the speech component of the total signal but changes the binaural cues of the noise component to these of the speech component [4]. When extending the multichannel Wiener filter with terms related to the interaural transfer function (ITF) it is possible to preserve both the cues of the speech and a single noise source at the cost of some noise reduction. Section 2 summarizes the MWF-ITF technique and makes a simplification to the algorithm for complexity reasons. Section 3 describes the objective error measures which estimate the performance in terms of noise reduction and binaural cue preservation. A perceptual evaluation, described in section 4, was necessary to validate the algorithm and the proposed binaural objective measures. This evaluation allows to correlate the theoretical error measures with the perceptual performance. It also leads to an estimation of the optimal settings of the different parameters in the algorithm to preserve both the speech and noise cues in an optimal way. In addition, it is shown that the loss in noise reduction performance when including the extra ITF-term is stronger than the binaural unmasking effect which takes place when the binaural cues of both the speech and noise source are preserved.

2. BINAURAL MULTICHANNEL WIENER FILTER WITH ITF EXTENSION

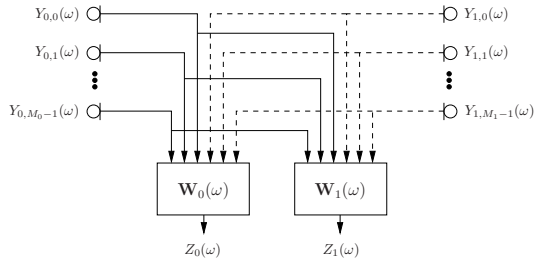


Fig. 1. Layout of a binaural noise reduction system

Consider the binaural hearing aid configuration in Figure 1, where the left and right hearing aid have a microphone array consisting of respectively M_0 and M_1 microphones. The mth microphone signal $Y_{0,m}(\omega)$ can be written in the frequency domain as

$$Y_{0,m}(\omega) = X_{0,m}(\omega) + V_{0,m}(\omega) \quad m = 0 \dots M_0 - 1, \quad (1)$$

where $X_{0,m}(\omega)$ and $V_{0,m}(\omega)$ represent the speech and the noise component at the mth microphone input of the left hearing aid. Assuming a link between both hearing aids, all microphone inputs can be used to generate an output for the left and the right ear. We define the M-dimensional signal vector $\mathbf{Y}(\omega)$, with $M = M_0 + M_1$, as

$$\mathbf{Y}(\omega) = \left[Y_{0,0}(\omega) \dots Y_{0,M_0-1}(\omega) \;\; Y_{1,0}(\omega) \dots Y_{1,M_1-1}(\omega) \right]^T.$$

The signal vector can be written as $\mathbf{Y}(\omega) = \mathbf{X}(\omega) + \mathbf{V}(\omega)$, where $\mathbf{X}(\omega)$ and $\mathbf{V}(\omega)$ are defined similarly as $\mathbf{Y}(\omega)$. The output signals for the left and the right ear are equal to

$$Z_0(\omega) = \mathbf{W}_0^H(\omega)\mathbf{Y}(\omega), \quad Z_1(\omega) = \mathbf{W}_1^H(\omega)\mathbf{Y}(\omega) \;, \quad (2)$$

with $\mathbf{W}_0(\omega)$ and $\mathbf{W}_1(\omega)$ M-dimensional complex vectors. We define the $2M$-dimensional stacked weight vector $\mathbf{W}(\omega)$ as

$$\mathbf{W}(\omega) = \begin{bmatrix} \mathbf{W}_0(\omega) \\ \mathbf{W}_1(\omega) \end{bmatrix} \;. \quad (3)$$

The output signal for the left ear can be written as

$$Z_0(\omega) = Z_{x0}(\omega) + Z_{v0}(\omega) = \mathbf{W}_0^H(\omega)\mathbf{X}(\omega) + \mathbf{W}_0^H(\omega)\mathbf{V}(\omega) \;,$$

where $Z_{x0}(\omega)$ and $Z_{v0}(\omega)$ represent the speech and the noise component. Similarly, the output signal for the right ear $Z_1(\omega) = Z_{x1}(\omega) + Z_{v1}(\omega)$. For conciseness, we will omit the frequency-domain variable ω in the remainder of the paper.

2.1. Binaural multi-channel Wiener filter (MWF)

The binaural MWF produces an MMSE (minimum-mean-square-error) estimate of the speech component in both hearing aids. The MSE cost function for the filter $\mathbf{W}_0$ estimating the speech component X_{0,r_0} in the r_0th microphone of the left hearing aid and the filter $\mathbf{W}_1$ estimating the speech component X_{1,r_1} in the r_1th microphone of the right hearing aid is equal to[1]

$$J_{MSE}(\mathbf{W}) = \mathcal{E}\left\{ \left\| \begin{bmatrix} X_{0,r_0} - \mathbf{W}_0^H \mathbf{Y} \\ X_{1,r_1} - \mathbf{W}_1^H \mathbf{Y} \end{bmatrix} \right\|^2 \right\} . \quad (4)$$

In order to provide a trade-off between speech distortion and noise reduction, the speech distortion weighted multi-channel Wiener filter (SDW-MWF) minimizes the weighted sum of the residual noise energy and the speech distortion energy. The binaural SDW-MWF cost function is equal to

$$\boxed{J_{SDW}(\mathbf{W}) = \mathcal{E}\left\{ \left\| \begin{bmatrix} X_{0,r_0} - \mathbf{W}_0^H \mathbf{X} \\ X_{1,r_1} - \mathbf{W}_1^H \mathbf{X} \end{bmatrix} \right\|^2 + \mu \left\| \begin{bmatrix} \mathbf{W}_0^H \mathbf{V} \\ \mathbf{W}_1^H \mathbf{V} \end{bmatrix} \right\|^2 \right\}} \quad (5)$$

where μ provides a trade-off between noise reduction and speech distortion. The filter minimizing $J_{SDW}(\mathbf{W})$ can be calculated using the estimated speech and noise correlation matrices, i.e. $\mathbf{R}_x = \mathcal{E}\{\mathbf{X}\mathbf{X}^H\}$ and $\mathbf{R}_v = \mathcal{E}\{\mathbf{V}\mathbf{V}^H\}$.

2.2. Extension with the interaural transfer function

However it was proven that the binaural MWF perfectly preserves the binaural cues of the speech component but distorts the cues of the noise component. In order to control the binaural cues of both the speech and the noise component, the cost function in (5) has been extended with terms related to the interaural transfer function (ITF) in [4]. The extra ITF term for preserving the binaural cues of the noise component is defined as the difference between the ITF at the output of the algorithm and the desired ITF, i.e.

$$J_{ITF}^v(\mathbf{W}) = \mathcal{E}\left\{ \left| \frac{\mathbf{W}_0^H \mathbf{V}}{\mathbf{W}_1^H \mathbf{V}} - ITF_{des}^v \right|^2 \right\} , \quad (6)$$

where the desired ITF can be calculated in a least square sense using the cross correlation matrices.

$$ITF_{des}^x = \frac{\mathcal{E}\{X_{0,r_0} X_{1,r_1}^*\}}{\mathcal{E}\{X_{1,r_1} X_{1,r_1}^*\}}, \quad ITF_{des}^v = \frac{\mathcal{E}\{V_{0,r_0} V_{1,r_1}^*\}}{\mathcal{E}\{V_{1,r_1} V_{1,r_1}^*\}} .$$

In case of a single localized source, (6) equals

$$J_{ITF}^v(\mathbf{W}) = \frac{\mathcal{E}\{|\mathbf{W}_0^H \mathbf{V} - ITF_{des}^v \mathbf{W}_1^H \mathbf{V}|^2\}}{\mathcal{E}\{|\mathbf{W}_1^H \mathbf{V}|^2\}} \quad (7)$$

The ITF cost function $J_{ITF}^x(\mathbf{W})$ for the speech component is defined similarly as $J_{ITF}^v(\mathbf{W})$, 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 binaural cue preservation is defined as

$$\boxed{J_{tot}(\mathbf{W}) = J_{SDW}(\mathbf{W}) + \alpha J_{ITF}^x(\mathbf{W}) + \beta J_{ITF}^v(\mathbf{W})} \quad (8)$$

where the parameters α and β enable to put more emphasis on binaural cue preservation for the speech and the noise component. Since no closed-form expression is available for the filter minimizing $J_{tot}(\mathbf{W})$, iterative optimization techniques should be used. To reduce the complexity of the hearing aid algorithm a less computationally expensive cost function is derived from (7) [5], i.e.

$$\boxed{J_{ITF}^v(\mathbf{W}) = \mathcal{E}\{|\mathbf{W}_0^H \mathbf{V} - ITF_{des}^v \mathbf{W}_1^H \mathbf{V}|^2\}} \quad (9)$$

3. OBJECTIVE EVALUATION

In this section objective measures are used to compute the performance (noise reduction and cue preservation) of the algorithm. Three different performance measures were used: SNR improvement, the error on the ITD cues and the error on the ILD cues.

The improvement in signal to noise ratio (SNR) is defined as the difference between the input and the output intelligibility weighted SNR, i.e. for the left hearing aid

$$\Delta SNR_0 = \sum I(\omega_i) \Delta SNR_0(\omega_i) \quad (10)$$

with $I(\omega_i)$ the importance of the i-th frequency band for speech intelligibility. The interaural time difference (ITD) is the difference in arrival time between left and right ear and is one of the two main binaural cues present in the ITF. The ITD error of the speech or noise component is calculated using the phase of the cross-correlation per freq. band. For example the ITD error of the noise component is computed as

$$\Delta ITD_v = \sum_i A(\omega_i) \frac{|\angle \mathcal{E}\{Z_{v0} Z_{v1}^*\} - \angle \mathcal{E}\{V_{0,r_0} V_{1,r_1}^*\}|}{\pi} \quad (11)$$

with $A(\omega_i)$ a weighting which only includes the frequency bands below 1500Hz. The interaural level difference (ILD) is the difference in power between the sounds arriving at the left and right ear due to the headshadow effect. This is the second main binaural cue present in an ITF. The ILD error generated by the noise reduction algorithm on the speech or noise component is defined as the difference between input and output ILD of both components. The ILD error on the noise component is defined as

$$\Delta ILD_v = \sum_i A(\omega_i) |10 \log_{10} P_{vi}^{out} - 10 \log_{10} P_{vi}^{in}| . \quad (12)$$

with $A(\omega_i)$ a frequency dependent weighing function. This function is set to 1 in this paper (all freq. bands equally important).

The left column of Figure 2 shows the speech intelligibility weighted SNR improvement when a speaker is talking from 0° (in front of the listener) with a noise source at 60° (S_0N_{60}). The unprocessed left or right front microphone was used as a reference. The microphone signals of the hearing aids were generated by convolving the speech and noise components with the appropriate transfer functions previously measured using two dual microphone hearing aids on an artificial head. The room had a reverberation time of T60=0.13s. This technique will be used throughout the rest of the paper. An FFT-size of N=256 was used in the algorithm together with a sampling frequency of fs=16kHz. Dutch VU sentences were used as the target signal and a stationary speech weighted noise was used as jammer signal. As expected the best noise reduction performance is found at low values

[1] Typically, the first microphone is used, i.e. $r_0 = r_1 = 0$.

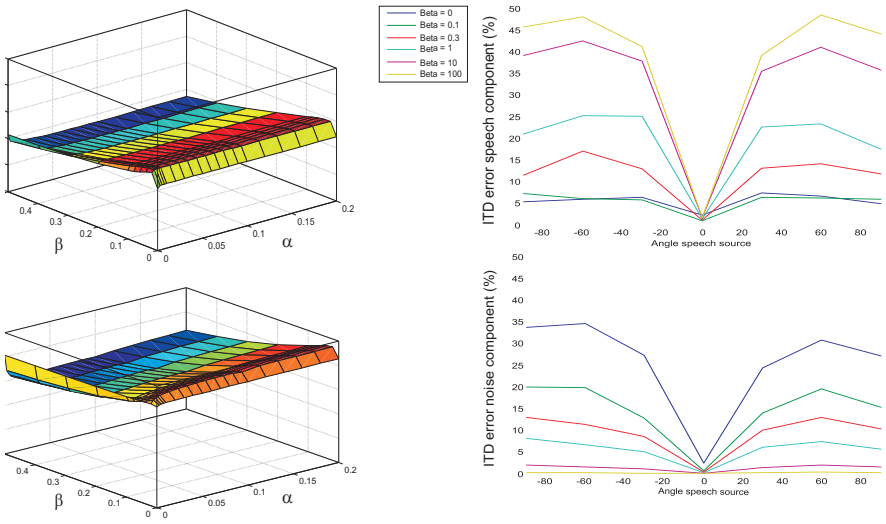


Fig. 2. Left: SNR improvement at the output of both hearing aids. Right: Angular dependent ITD error for different values of β when the angle of the speech source is varied.

of α and β. The performance gradually drops for increasing values of α and β.

The right column of Figure 2 shows the ITD error for different values of β when the position of the speech source is varied from -90° to $+90^\circ$ in steps of 30°. The position of the noise source is fixed at 0°. An input SNR of 0dB was assumed. The ITD error of the speech component is low when β is low. However, this introduces large ITD errors on the noise component. When increasing the value of β the ITD error on the noise component drops significantly, however if this value is too high ITD errors are introduced on the speech component. The same tendency was observed for the ILD errors but these data are omitted in this paper.

4. PERCEPTUAL EVALUATION

The binaural MWF-ITF technique was validated perceptually with 5 normal hearing subjects. This was necessary to validate the algorithm and to estimate the optimal parameter settings perceptually.

4.1. Noise reduction performance

The microphone signals were generated as described in the previous section (S_0N_{60}). The algorithm used these signals to calculate and store the correlation matrices and filter coefficients off-line. These filters were then used in an adaptive speech reception threshold (SRT) procedure which estimates the SNR level at which the listener understands 50 percent of the speech correctly. The target signal were dutch VU sentences and a stationary speech weighted noise was used as jammer signal. The signals were presented under headphones. The SRT improvement compared to a system without filtering is shown in Figure 3 for different settings of the parameter β. A maximum noise reduction of approx. 13dB is reached with $\beta = 0$ corresponding to the standard binaural MWF-design. A continuous decrease in noise reduction performance is observed when more emphasis is placed on the preservation of the binaural cues of the noise source. It also shows that the gain in speech perception due to restoring the spatial separation of the speech and noise source ('binaural unmasking') is not large enough to compensate for the loss in noise reduction performance of the algorithm. Still the gain remains quite large even when the emphasis on noise ITF preservation becomes very strong (11dB at $\beta = 10$).

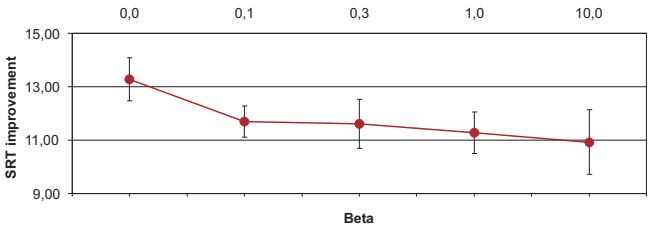


Fig. 3. SRT gain of the MWF-ITF algorithm using $\beta = 0, 0.1, 0.3, 1$ and 10 and $\alpha = 0$ in the condition S_0N_{60}.

4.2. Localization performance

Microphone signals of both hearing aids where generated for different conditions ($S_\Theta N_\Gamma$). These were used to calculate the accompanying filter coefficients using different parameter settings. Next, telephone signals arriving from the noise and speech angle were filtered with the pre-calculated coefficients, thereby simulating a filtered version of a telephone signal arriving from the noise angle and a filtered version of a telephone signal arriving from the speech angle. These signals were presented under headphones. In the tested condition the filters were calculated using a target signal located from -90° to +90° in steps of 30° and a noise source located at 0° ($S_\Theta N_0$). Subjects were instructed to localize the sound source in the frontal hemisphere between -90° and +90° in steps of 15°. Stimuli were repeated three times, randomized and a level roving of 5dB was applied during the test procedure.

Figure 5 shows the accumulation of responses for the five test subjects in two extreme cases of the condition $S_\Theta N_0$, namely $\beta = 0$ (standard binaural MWF) and $\beta = 10$, both using α=0. It illustrates clearly that the standard MWF technique ($\beta = 0$) moves the cues of the noise source to the cues of the speech source. When β is high ($\beta = 10$), the speech source is moved to the location of the noise source. Now the question arises whether an optimal parameter setting can be found which preserves both the speech and the noise cues.

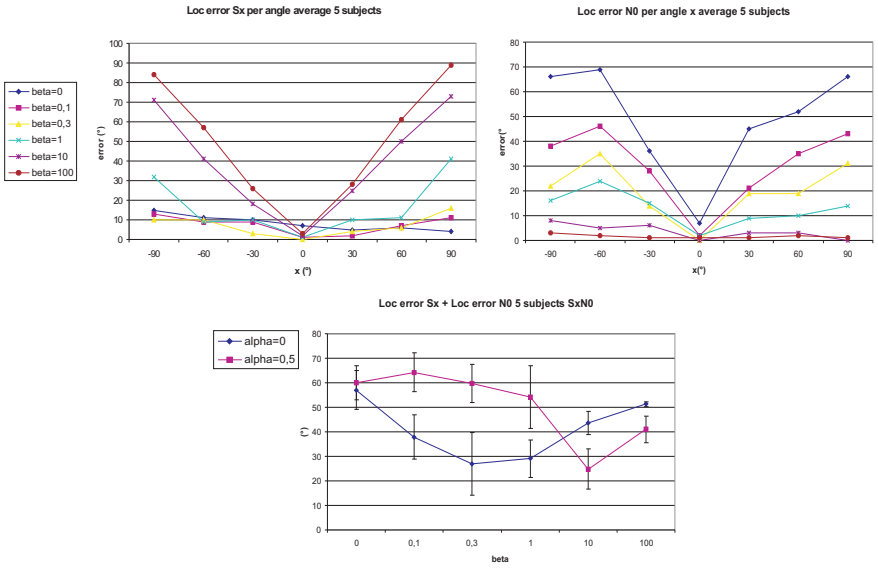


Fig. 4. Upper: Perceptual localization error on the speech and noise component for different values of β and different speech angles $S_\Theta N_0$. Lower: Sum of the mean speech and noise localization error for the condition $S_x N_0$ for the tested parameters, an optimum is reached at $\alpha = 0, \beta = 0.3$.

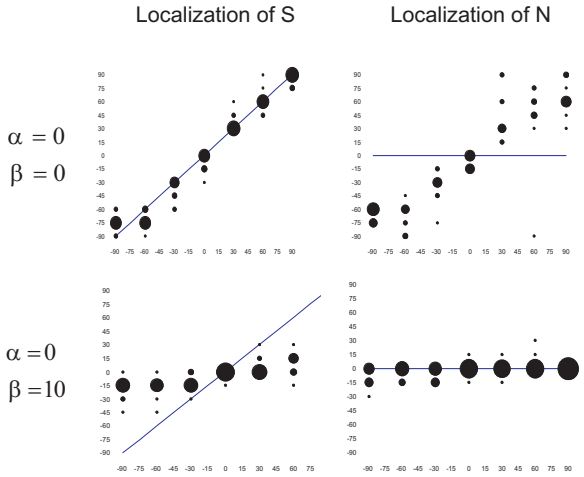


Fig. 5. Accumulation of the responses for 5 subjects at two extreme β-values of the condition $S_\Theta N_0$. The thin blue line is the correct location of the sound component.

Figure 4 shows the average localization error over five subjects for both the speech and the noise component. The x-axis representing the location of the presented speech source. Again, low values of β introduce a large localization error on the noise component whereas a high value of β introduces a large localization error on the speech component. This figure shows a large correlation with the objective measures shown in Figure 2. If we assume that each angle is equally important we can calculate the mean error on the speech and noise component over all angles. When adding the mean speech and noise error a single number is obtained per parameter setting indicating the quality of the parameter setting for the condition $S_\Theta N_0$. This is shown in Figure 4 for different parameter settings. For $\alpha = 0$ an optimal setting for localizing sound sources would be $\beta = 0.5$. For $\alpha = 0.5$ this optimal setting is $\beta = 10$.

5. CONCLUSION

Binaural hearing aid algorithms offer more flexibility than monaural hearing aid algorithms due to the amount of input microphones positioned around the head. However typical noise reduction systems producing a single output signal are not suited for binaural processing. A technique (MWF-ITF) was described that allows to find a compromise between noise reduction and the preservation of binaural cues. Using subjective evaluations an optimal setting was found for the algorithm. However it seems that the loss in noise reduction performance due to this ITF term in the cost fucntion of the multichannel Wiener filter is larger than the gain due to the binaural unmasking effect. It was also shown that the proposed objective, relative straightforward, binaural error measures correlated well with the perceptual evaluation described in this paper.

6. REFERENCES

[1] D.P. Welker, J.E. Greenberg, J.G. Desloge, and Zurek P.M., "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, November 1997.

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

[3] T. Van den Bogaert, "Localization with bilateral hearing aids: without is better than with," *J. Acoust. Soc. Am.*, vol. 119, no. 1, pp. 515–526, january 2006.

[4] 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 fucntions," in *Proc. IWAENC*, Paris, France, September 2006.

[5] T.J. Klasen, T. Van den Bogaert, M. Moonen, and J. Wouters, "Preservation of interaural time delay for binaural hearing aids through multi-channel wiener filtering based noise reduction," in *Proc. ICASSP*, Philadelphia PA, USA, March 2005, pp. 29–32.