BINAURAL CUE PRESERVATION FOR HEARING AIDS USING MULTI-CHANNEL WIENER FILTER WITH INSTANTANEOUS ITF PRESERVATION

Daniel Marquardt[1], *Volker Hohmann*[2], *Simon Doclo*[1]

[1]University of Oldenburg, Institute of Physics, Signal Processing Group, Oldenburg
[2]University of Oldenburg, Institute of Physics, Medical Physics, Oldenburg

ABSTRACT

An important objective of binaural noise reduction algorithms is the preservation of the binaural cues. In this paper an extension of the Multi-channel Wiener filter with binaural cue preservation (MWF-ITF) is presented, where the average noise ITF preservation term is replaced by an instantaneous noise ITF preservation term. This framework in addition allows to impose perfect ITF preservation, leading to a hard-constraint formulation. Experimental results show that the proposed technique yields a better performance in preserving the binaural cues of both the noise component and the speech component compared to the MWF-ITF, without degrading the output SNR.

Index Terms— Hearing aids, binaural cues, noise reduction

1. INTRODUCTION

Noise reduction algorithms in hearing aids are crucial to improve speech understanding in background noise for hearing impaired persons. For binaural hearing aids algorithms that exploit multiple microphone signals from both the left and the right hearing aid are considered to be promising techniques for noise reduction, because in addition to spectral information spatial sound information can be exploited. In addition to reducing noise and limiting speech distortion, another important objective of binaural noise reduction algorithms [1, 2, 3] is the preservation of the listeners impression of the auditory scene, in order to exploit the binaural hearing advantage and to avoid confusions due to a mismatch between the acoustical and the visual information. This can be achieved by preserving the Interaural Transfer Function (ITF) of the speech and the noise component, comprising both the Interaural Time Difference (ITD) and Interaural Level Difference (ILD) binaural cues.

A binaural Speech Distortion Weighted Multi-channel Wiener Filter (MWF) has been presented in [4]. It has been theoretically proven in [3] that this technique preserves the ITF of the speech component for a single speech source. On the contrary, the ITF of the noise component is distorted, such that the ITF of the residual noise component equals the ITF of the speech component. In addition, an extension of the MWF, namely the MWF-ITF, has been presented, by imposing a soft constraint on the preservation of the ITF of the noise component. Theoretical and experimental results in [3] have shown that a better preservation of the ITF of the noise component leads to a distortion of the ITF of the speech component, depending on the input SNR and a trade-off parameter. Hence using the MWF-ITF it is not possible to preserve the speech and the noise ITF simultaneously.

This paper describes a novel variation of the MWF-ITF, by replacing the average noise ITF preservation term used in the MWF-ITF by an instantaneous (time-varying) noise ITF preservation term, moreover allowing to transform the soft-constraint problem into a hard-constraint problem. Although this instantaneous noise ITF preservation term largely relies on the noise estimation procedure, experimental results show that a better preservation of the binaural cues of the noise component is achieved, whereas the distortion on the binaural cues of the speech component are reduced compared to the traditional MWF-ITF without degrading the output SNR.

This work was partly funded by the BMBF project "Modellbasierte Hörsysteme"

2. CONFIGURATION AND NOTATION

Considering the binaural hearing aid configuration in Figure 1 consisting of the left and the right microphone array with M microphones each, the frequency-domain representation of the m-th microphone signal in the left hearing aid $Y_{0,m}(k,l)$ can be written as

$$Y_{0,m}(k,l) = X_{0,m}(k,l) + V_{0,m}(k,l), \quad m = 0 \ldots M-1, \quad (1)$$

with $X_{0,m}(k,l)$ representing the speech component, $V_{0,m}(k,l)$ representing the noise component, k denoting the frequency index and l the block index. The m-th microphone signal in the right hearing aid $Y_{1,m}(k,l)$ is defined similarly. For conciseness we will omit the frequency variable k and the block index l in the remainder of the paper, except where explicitly required.

We define the $2M$-dimensional signal vector $\mathbf{Y}$ as

$$\mathbf{Y} = \left[Y_{0,0} \ldots Y_{0,M-1} Y_{1,0} \ldots Y_{1,M-1}\right]^T. \quad (2)$$

The signal vector can be written as $\mathbf{Y} = \mathbf{X} + \mathbf{V}$, where $\mathbf{X}$ and $\mathbf{V}$ are defined similarly as $\mathbf{Y}$. Furthermore, we define the $4M$-dimensional complex stacked weight vector $\mathbf{W}$ as

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

The output signal at the left hearing aid Z_0 is equal to

$$Z_0 = \mathbf{W}_0^H \mathbf{Y} = \mathbf{W}_0^H \mathbf{X} + \mathbf{W}_0^H \mathbf{V} = Z_{x,0} + Z_{v,0}, \quad (4)$$

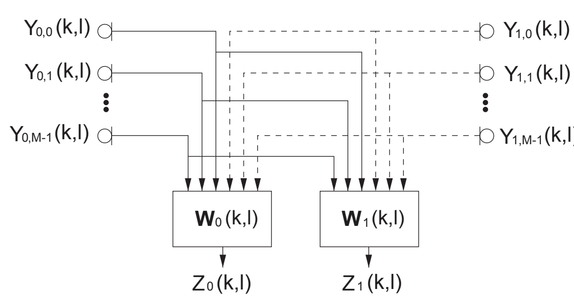

Fig. 1. Binaural hearing aid configuration

where $Z_{x,0}$ represents the speech component and $Z_{v,0}$ represents the noise component in the output signal. Similarly, the output signal at the right hearing aid can be written as

$$Z_1 = \mathbf{W}_1^H\mathbf{Y} = \mathbf{W}_1^H\mathbf{X} + \mathbf{W}_1^H\mathbf{V} = Z_{x,1} + Z_{v,1}. \qquad (5)$$

The instantaneous input and output Interaural Transfer Function (ITF) of the speech and the noise components are defined as

$$ITF_x^{in} = \frac{X_{0,0}}{X_{1,0}} \qquad ITF_x^{out} = \frac{Z_{x,0}}{Z_{x,1}} = \frac{\mathbf{W}_0^H\mathbf{X}}{\mathbf{W}_1^H\mathbf{X}} \qquad (6)$$

$$ITF_v^{in} = \frac{V_{0,0}}{V_{1,0}} \qquad ITF_v^{out} = \frac{Z_{v,0}}{Z_{v,1}} = \frac{\mathbf{W}_0^H\mathbf{V}}{\mathbf{W}_1^H\mathbf{V}}. \qquad (7)$$

Assuming that the speech and noise components are independent, the correlation matrices of the signal components are defined as

$$\mathbf{R}_y = \mathcal{E}\left\{\mathbf{Y}\mathbf{Y}^H\right\}, \mathbf{R}_v = \mathcal{E}\left\{\mathbf{V}\mathbf{V}^H\right\}, \mathbf{R}_x = \mathbf{R}_y - \mathbf{R}_v, \qquad (8)$$

which in the remainder of the paper are estimated as

$$\mathbf{R}_y(k) = \frac{1}{L_y}\sum_{i=0}^{L_y-1}\mathbf{Y}(k,i)\mathbf{Y}^{\mathrm{H}}(k,i) \quad \text{speech present,} \qquad (9)$$

$$\mathbf{R}_v(k) = \frac{1}{L_v}\sum_{i=0}^{L_v-1}\mathbf{V}(k,i)\mathbf{V}^{\mathrm{H}}(k,i) \quad \text{speech absent,} \qquad (10)$$

i.e. the mean value of the L_y available signal vectors when speech is present, respectively the L_v available signal vectors when speech is absent, depending on the decision of a perfect Voice Activity Detector (VAD).

3. BINAURAL NOISE REDUCTION ALGORITHMS

In this section we discuss the cost functions for the binaural MWF and the MWF-ITF proposed in [4] and [3].

3.1. Binaural multi-channel Wiener filter (MWF)

The binaural MWF produces a minimum mean-square error (MMSE) estimate of the speech component in one of the microphone signals for both hearing aids, hence simultaneously reducing noise and limiting speech distortion. The MWF cost function estimating the speech component $X_{0,0}$ in the left hearing aid and the speech component $X_{1,0}$ in the right hearing aid can be written as

$$J_{MWF}(\mathbf{W}) = \mathcal{E}\left\{\left\|\begin{bmatrix} X_{0,0} - \mathbf{W}_0^H\mathbf{X} \\ X_{1,0} - \mathbf{W}_1^H\mathbf{X}\end{bmatrix}\right\|^2 + \mu\left\|\begin{bmatrix}\mathbf{W}_0^H\mathbf{V} \\ \mathbf{W}_1^H\mathbf{V}\end{bmatrix}\right\|^2\right\}, \qquad (11)$$

where μ provides a trade-off between noise reduction and speech distortion and the first microphone is used as reference microphone. The filter minimizing $J_{MWF}(\mathbf{W})$ is equal to

$$\mathbf{W}_{MWF} = \mathbf{R}^{-1}\mathbf{r}_x, \qquad (12)$$

with

$$\mathbf{R} = \begin{bmatrix}\mathbf{R}_x + \mu\mathbf{R}_v & \mathbf{0}_{2M} \\ \mathbf{0}_{2M} & \mathbf{R}_x + \mu\mathbf{R}_v\end{bmatrix}, \quad \mathbf{r}_x = \begin{bmatrix}\mathbf{R}_x\mathbf{e}_{0,0} \\ \mathbf{R}_x\mathbf{e}_{1,0}\end{bmatrix}. \qquad (13)$$

The vectors $\mathbf{e}_{0,0}$ and $\mathbf{e}_{1,0}$ are zero column vectors with the first element equal to 1. It has been shown in [3] that for a single speech source the ITF of the output speech and noise component are the same and equal to ITF_x^{in}, implying that all components are perceived as coming from the speech direction, which is obviously undesired.

3.2. MWF with binaural cue preservation (MWF-ITF)

To reduce the distortion of the output noise ITF, an extension of the MWF cost function with a quadratic term related to the ITF of the noise component has been proposed and analyzed in [3]. The ITF preservation term is defined as

$$J_{ITF}^v(\mathbf{W}) = \mathcal{E}\left\{\left|\frac{\mathbf{W}_0^H\mathbf{V}}{\mathbf{W}_1^H\mathbf{V}} - ITF_v^{des}\right|^2\right\}, \qquad (14)$$

where the desired ITF can be calculated as

$$ITF_v^{des} = \frac{\mathbf{e}_{0,0}^T\mathbf{R}_v\mathbf{e}_{1,0}}{\mathbf{e}_{1,0}^T\mathbf{R}_v\mathbf{e}_{1,0}}, \qquad (15)$$

which can be interpreted as an average input ITF, which is constant in case of a single noise source. The total cost function of the MWF-ITF is defined as

$$J_{MWF-ITF}(\mathbf{W}) = J_{MWF}(\mathbf{W}) + \delta J_{ITF}^v(\mathbf{W}), \qquad (16)$$

where the parameter δ controls the emphasis on the ITF preservation term for the noise component in the total cost function. Since no closed-form expression is available for the filter minimizing $J_{MWF-ITF}(\mathbf{W})$, a simplified ITF extension has been introduced in [3] as

$$J_{ITF}^v(\mathbf{W}) = \mathcal{E}\left\{\left|\mathbf{W}_0^H\mathbf{V} - ITF_v^{des}\mathbf{W}_1^H\mathbf{V}\right|^2\right\}, \qquad (17)$$

which can be interpreted as an average ITF preservation term. The filter minimizing the simplified cost function is equal to

$$\mathbf{W}_{MWF} = \left(\mathbf{R} + \delta\mathbf{R}_{vt}\right)^{-1}\mathbf{r}_x, \qquad (18)$$

with

$$\mathbf{R}_{vt} = \begin{bmatrix}\mathbf{R}_v & -ITF_v^{des,*}\mathbf{R}_v \\ -ITF_v^{des}\mathbf{R}_v & \left|ITF_v^{des}\right|^2\mathbf{R}_v\end{bmatrix}. \qquad (19)$$

It has been proven in [3] that the ITF of the output speech component and the ITF of the output noise component are still equal. If $\delta \to 0$, the MWF-ITF is equal to the MWF, hence $ITF_v^{out} = ITF_x^{in}$. If $\delta \to \infty$, the desired ITF of the noise component is perfectly preserved but the ITF of the output speech component equals ITF_v^{des}, implying that all components are perceived as coming from the noise direction. Therefore the solution is always a trade-off between preserving the binaural cues of the speech component and preserving the binaural cues of the noise component depending on the parameter δ and the input SNR.

4. INSTANTANEOUS ITF PRESERVATION

In this section a modification of the MWF-ITF is presented by defining an instantaneous ITF preservation term, i.e.

$$J_{ITFin}^v(\mathbf{W}) = \left|\mathbf{W}_0^H\mathbf{V} - ITF_v^{des}\mathbf{W}_1^H\mathbf{V}\right|^2, \qquad (20)$$

instead of the average ITF preservation term in (17). By adding this term to the MWF cost function, the optimal filter MWF-ITFin, similarly to (18), is equal to

$$\mathbf{W}_{MWF-ITFin} = \left(\mathbf{R} + \delta\mathbf{R}_{vt-in}\right)^{-1}\mathbf{r}_x, \qquad (21)$$

with

$$\mathbf{R}_{vt-in} = \begin{bmatrix}\mathbf{V}\mathbf{V}^H & -ITF_v^{des,*}\mathbf{V}\mathbf{V}^H \\ -ITF_v^{des}\mathbf{V}\mathbf{V}^H & \left|ITF_v^{des}\right|^2\mathbf{V}\mathbf{V}^H\end{bmatrix}. \qquad (22)$$

Contrary to $\mathbf{R}_{vt}$ in (19), the matrix $\mathbf{R}_{vt-in}$ in (22) is updated in each frame, since it depends on the instantaneous noise vector $\mathbf{V}$. Note however that the matrix $\mathbf{R}$ in (21) is still the estimated (mean) correlation matrix in (13).

4.1. Hard constraint

Using the instantaneous ITF preservation term defined in (20), it is now possible to impose perfect noise ITF preservation, corresponding to $\delta \to \infty$ in (21), by imposing a linear constraint on the ITF preservation, i.e.

$$\min_{\mathbf{W}} \mathcal{E}\left\{\left\|\begin{bmatrix} X_{0,0} - \mathbf{W}_0^H\mathbf{X} \\ X_{1,0} - \mathbf{W}_1^H\mathbf{X}\end{bmatrix}\right\|^2 + \mu\left\|\begin{bmatrix}\mathbf{W}_0^H\mathbf{V} \\ \mathbf{W}_1^H\mathbf{V}\end{bmatrix}\right\|^2\right\} \tag{23}$$

$$\text{subject to} \quad \frac{\mathbf{W}_0^H\mathbf{V}}{\mathbf{W}_1^H\mathbf{V}} = ITF_v^{des}. \tag{24}$$

The solution to this constrained optimization problem (MWF-ITFhc) is given by

$$\mathbf{W}_{MWF-ITFhc} = \mathbf{R}^{-1}\,\mathbf{r}_x - \frac{\mathbf{R}^{-1}\,\mathbf{C}^H\,\mathbf{C}\,\mathbf{R}^{-1}\,\mathbf{r}_x}{\mathbf{C}\,\mathbf{R}^{-1}\,\mathbf{C}^H} \tag{25}$$

with

$$\mathbf{C} = \begin{bmatrix}\mathbf{V}^H & -ITF_v^{des,*}\,\mathbf{V}^H\end{bmatrix}. \tag{26}$$

With this formulation of the optimization problem a perfect preservation of ITF_v^{des} in each frequency bin k and time segment l can be achieved. Note that the first part of (25) is fixed as in the MWF solution and the second part contains the vector $\mathbf{C}$ which is updated in each frame. Since both (21) and (25) require the noise vector $\mathbf{V}$ to be available - which is obviously not the case during speech segments - an estimate of the noise vector $\hat{\mathbf{V}}$ is required.

4.2. Noise estimation

For estimating the noise vector, we will again use the MWF, where now $V_{0,m}$ and $V_{1,m}$ are the desired signals for the left, respectively right hearing aid, such that for each microphone m in both hearing aids an estimate of the noise signal is computed, i.e.

$$\mathbf{W}_v^m = \mathbf{R}_s^{-1}\mathbf{r}_v^m, \quad m = 0\dots M-1, \tag{27}$$

with

$$\mathbf{R}_s = \begin{bmatrix}\mathbf{R}_v + \eta\mathbf{R}_x & \mathbf{0}_{2M} \\ \mathbf{0}_{2M} & \mathbf{R}_v + \eta\mathbf{R}_x\end{bmatrix}, \quad \mathbf{r}_v^m = \begin{bmatrix}\mathbf{R}_v\mathbf{e}_{0,m} \\ \mathbf{R}_v\mathbf{e}_{1,m}\end{bmatrix} \tag{28}$$

with the parameter η allowing to put less emphasis on speech suppression resulting in less distortion of the noise estimate. The noise estimates are calculated as

$$\begin{bmatrix}\hat{V}_0^m \\ \hat{V}_1^m\end{bmatrix} = \begin{bmatrix}\mathbf{W}_{v,0}^{m,H} \\ \mathbf{W}_{v,1}^{m,H}\end{bmatrix}\mathbf{Y} \tag{29}$$

Note that in this case M additional binaural MWF filter need to be computed, resulting in an increased computational complexity.

5. EXPERIMENTAL RESULTS

In this section we perform simulations to investigate the performance of the MWF, MWF-ITF, MWF-ITFin and MWF-ITFhc on ILD distortion and on the intelligibility weighted output SNR for a scenario consisting of one speech source and one noise source.

5.1. Setup

Binaural Behind-The-Ear Head-Related Impulse Responses measured in an office room ($T_{60} \approx 300\,ms$) from [5] have been used to generate the speech and the noise signals. Each hearing aid was equipped with 2 microphones, therefore in total 4 microphone signals are available. The speech source (one sentence taken from the OLSA sentence material) was located in front of the listener 0° and the interfering white noise source was positioned on a azimuthal angle of 60° (right side of the head). The signals were processed at $f_s = 16$ kHz using an overlap-add framework with a block size of 256 samples and an overlap of 50% between successive blocks. The correlation matrices $\mathbf{R}_y$ and $\mathbf{R}_v$ are estimated using (9) and (10) and $\mathbf{R}_x = \mathbf{R}_y - \mathbf{R}_v$. $\mathbf{R}_v$ was estimated using 3 seconds of a noise-only signal, preceding the noisy speech signal. The noise-only part was not taken into account when evaluating the performance of the algorithms. The parameter μ in (13) was set to 1 and the parameter η in (28) was set to 0.25. The performance was evaluated for an intelligibility weighted input SNR in the first microphone of the left hearing aid ranging from -6 dB to 6 dB.

5.2. Performance measures

For comparing the performance of the algorithms we have used 3 objective performance measures. The *intelligibility weighted output SNR* [6] of the left hearing aid is defined as

$$SNR_0 = \sum_k I(k)\frac{Pz_{x0}(k)}{Pz_{v0}(k)}, \tag{30}$$

where $Pz_{x0}(k)$ and $Pz_{v0}(k)$ are the power spectral densities of the left output speech component, respectively output noise component. $I(k)$ is a weighting function that takes the importance of different frequency bands for the speech intelligibility into account. The output SNR on the right hearing aid is defined in a similar way.
To evaluate the binaural cue preservation performance, we define the *ILD error of the noise component* ΔILD_v as

$$\Delta ILD_v = \frac{1}{KL}\sum_k\sum_l\left|10\log_{10}\left(\frac{ILD_v^{out}(k,l)}{ILD_v^{des}(k,l)}\right)\right|, \tag{31}$$

and the *ILD error of the speech component* ΔILD_x as

$$\Delta ILD_x = \frac{1}{KL}\sum_k\sum_l\left|10\log_{10}\left(\frac{ILD_x^{out}(k,l)}{ILD_x^{in}(k,l)}\right)\right|, \tag{32}$$

$$\text{with} \quad ILD(k,l) = |ITF(k,l)|^2. \tag{33}$$

Since we obtained similar results for the ITD error as for the ILD error, we will only discuss the performance in ILD preservation and output SNR.

5.3. Performance in ILD preservation and output SNR

The results in preserving the binaural cues of the speech and the noise component are depicted in Figure 2. The MWF shows the best performance for ΔILD_x but also introduces a large error ΔILD_v. In theory the MWF perfectly preserves the ILD of the speech component, which is not the case here due to estimation errors in $\mathbf{R}_x$. For a sufficient high value $\delta = 10^5$ the MWF-ITF perfectly preserves the ILD of the noise component at the cost of high distortions in the ILD of the speech component. Decreasing δ to 2 decreases the ILD error of the speech component but also increases the ILD error of the noise component. For $\delta = 2$, the MWF-ITFin outperforms the MWF-ITF while the MWF-ITFhc further decreases the error in ΔILD_v, while introducing a slightly higher distortion on the ILD of the speech component compared to the MWF-ITFin. Note that the MWF-ITFhc fails to perfectly preserve the ILD of the noise component due to the erroneous estimate of the noise vector $\hat{\mathbf{V}}$.
To further investigate the performance of the MWF-ITF compared to the MWF-ITFhc, the results for the MWF, MWF-ITF and the MWF-ITFhc are depicted in Figure 3. On the one hand we have chosen the parameter δ in the MWF-ITF such that for an input SNR of 0 dB we can achieve the same performance in ΔILD_v as the MWF-ITFhc, which is the case for $\delta = 4.5$. In this case, ΔILD_x is significantly higher (up to 4 dB) for the MWF-ITF compared to the MWF-ITFhc.

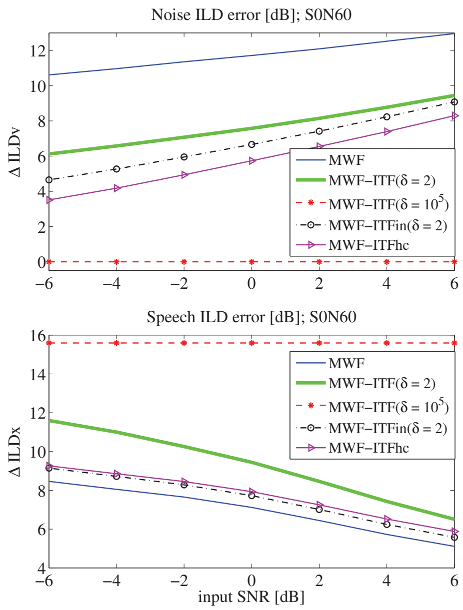

Fig. 2. ILD error for noise and speech component

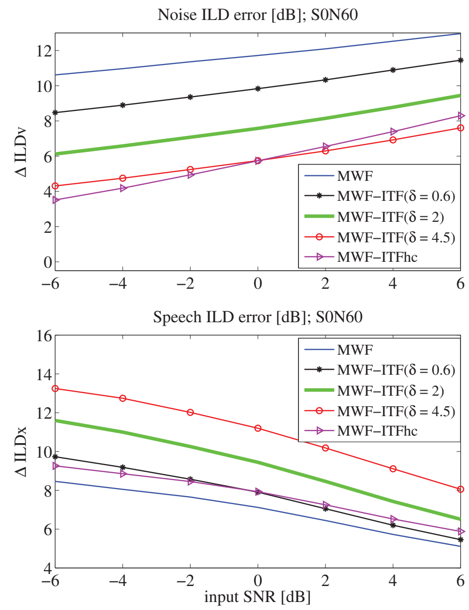

Fig. 3. ILD error for MWF, MWF-ITF and MWF-ITFhc

On the other hand we have chosen the parameter δ in the MWF-ITF such that for an input SNR of 0 dB we can achieve the same performance in ΔILD_x as the MWF-ITFhc, which is the case for $\delta = 0.6$. In this case, ΔILD_v is significantly higher (up to 4.5 dB) for the MWF-ITF compared to the MWF-ITFhc. From this experiment we can conclude that for the MWF-ITF the best achievable trade-off between ILD distortion of the noise and the speech component can be achieved by setting the parameter δ to 2. However, in this case the MWF-ITFhc outperforms the MWF-ITF in preserving the binaural cues of the speech and the noise component.

In Figure 4 the output SNR vs. the input SNR is depicted for the algorithms compared in Figure 2. For the left hearing aid the MWF-ITFin and the MWF-ITFhc show the best performance but the advantage decreases with increasing input SNR. The output SNR of the MWF-ITF with $\delta = 10^5$ is very low compared to the MWF, which can be explained by the decreasing output speech power in the left hearing aid due to the ILD distortion of the speech component. The output SNR on the right hearing aid is approximately the same for all algorithms.

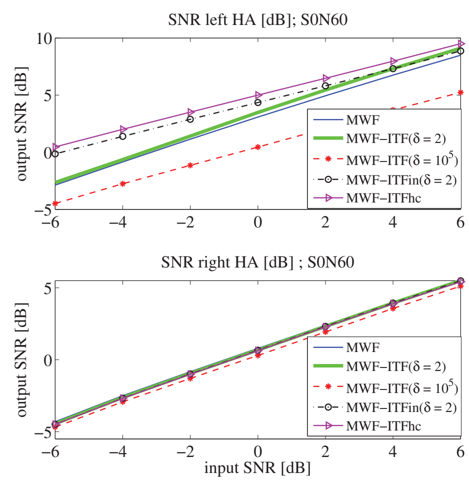

Fig. 4. output SNR vs. input SNR

6. CONCLUSION

In this paper we have shown that a better preservation of the binaural cues of both the speech and the noise component can be achieved by adding an instantaneous noise ITF preservation term to the MWF cost function, while even slightly improving the output SNR. Improving the estimation of the noise vector, which is required in the proposed algorithm, remains a topic of further research.

7. REFERENCES

[1] T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Robustness analysis of binaural hearing aid beamformer algorithms by means of objective perceptual quality measures," in *Proc. WASPAA*, New Paltz, NY, Oct. 2007, pp. 315–318.

[2] K. Reindl, Y. Zheng, and W. Kellermann, "Analysis of two generic wiener filtering concepts for binaural speech enhancement in hearing aids," in *Proc. EUSIPCO*, Aalborg, Denmark, Aug. 2010, pp. 989–993.

[3] B. Cornelis, S. Doclo, T. Van den Bogaert, J. Wouters, and M. Moonen, "Theoretical analysis of binaural multi-microphone noise reduction techniques," *IEEE Trans. on Audio, Speech and Language Proc.*, vol. 18, no. 2, pp. 342–355, Feb. 2010.

[4] S. Doclo, S. Gannot, M. Moonen, and A. Spriet, "Acoustic beamforming for hearing aid applications," in *Handbook on Array Processing and Sensor Networks*, pp. 269–302. Wiley, 2010.

[5] H. Kayser, S. Ewert, J. Annemueller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Database of multichannel In-Ear and Behind-The-Ear Head-Related and Binaural Room Impulse Responses," *Eurasip Journal on Advances in Signal Processing*, vol. 2009, pp. 10 pages, 2009.

[6] J. E. Greenberg, P. M. Peterson, and P. M. Zurek, "Intelligibility-weighted measures of speech-to-interference ratio and speech system performance," *Journal of the Acoustical Society of America*, vol. 94, no. 5, pp. 3009–3010, Nov. 1993.