# EXTENSION OF THE MULTI-CHANNEL WIENER FILTER WITH LOCALISATION CUES FOR NOISE REDUCTION IN BINAURAL HEARING AIDS

*Simon Doclo*[1,2], *Rong Dong*[2], *Thomas J. Klasen*[1,3], *Jan Wouters*[3], *Simon Haykin*[2] *and Marc Moonen*[1]

[1]simon.doclo@esat.kuleuven.ac.be

[1]KU Leuven, Dept. of Elec. Engineering, Kasteelpark Arenberg 10, 3001 Leuven, Belgium

[2]McMaster University, ASL, 1280 Main Street West, Hamilton ON L8S-4K1, Canada

[3]KU Leuven, Lab. Exp. ORL, Kapucijnenvoer 33, 3000 Leuven, Belgium

## ABSTRACT

This paper presents an extension of the multi-channel Wiener filter (MWF) for noise reduction in binaural hearing aids, taking into account binaural localisation cues. By adding a term related to the interaural time difference (ITD) and the interaural level difference (ILD) of the noise component to the cost function of the MWF, the ITD and ILD cues of both the speech and the noise component can be preserved, in addition to significantly improving the signal-to-noise ratio of the microphone signals.

## 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. Commonly used multi-microphone noise reduction techniques for - monaural and binaural - hearing aids are based on fixed beamforming [1], adaptive beamforming [2, 3], or multi-channel Wiener filtering [4, 5, 6, 7].

In a binaural hearing aid system, output signals for both ears are generated, either by using both hearing aids independently or by sharing information between the hearing aids. In addition to reducing noise and limiting speech distortion, another important objective of a binaural algorithm is to preserve the listener's impression of the auditory environment in order to exploit the binaural hearing advantage. This can be achieved by preserving the binaural cues, i.e. the interaural time and level difference (ITD, ILD), of the speech and the noise components.

In [1], a fixed beamforming technique has been proposed where the filter weights are optimised in order to maximise the directivity index while restricting the ITD error below some threshold. Binaural adaptive beamforming techniques, based on the Generalised Sidelobe Canceller (GSC), have been proposed in [2, 3]. In [2], the low frequencies of the left and the right signal are passed through unaltered in order to preserve the ITD cues, whereas the high frequencies are adaptively processed using the GSC and added to the low frequencies. A major drawback is that not only the speech but also the noise in the low-frequency portion is passed through, significantly reducing the noise reduction performance. In [3], the preservation of the ITD and ILD cues is restricted to an angular region around the front, while at other angles the background noise is reduced.

Simon Doclo is a postdoctoral researcher supported by the Fund for Scientific Research - Flanders. This work was carried out at the ESAT-SCD laboratory, KU Leuven, and the Adaptive Systems Laboratory, McMaster University, in the frame of the F.W.O. Project G.0233.01, the I.W.T. Projects 020540 and 040803, the Concerted Research Action GOA-AMBIORICS, and the Interuniversity Attraction Pole IUAP P5-22.

In [6], a binaural multi-channel Wiener filter, providing an enhanced output signal at both ears, has been discussed. In addition to significantly suppressing the background noise, it has been shown that this algorithm preserves the ITD cues of the speech component. On the contrary, the binaural cues of the noise component may be distorted. An extension of the MWF that partially preserves these binaural noise cues has been proposed in [7], resulting however in a considerable reduction of the noise reduction performance. Recently, another extension of the MWF has been proposed, where a term related to the noise ITD cue is added to the cost function of the MWF [8]. This paper discusses the addition of a second term related to the noise ILD cue. Experimental results show that the binaural cues of both the speech and the noise component can be preserved without compromising the noise reduction performance.

## 2. CONFIGURATION AND NOTATION

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

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

where $X_{0,m}(\omega)$ represents the speech component and $V_{0,m}(\omega)$ represents the noise component. Similarly, the $m$th microphone signal in the right hearing aid is $Y_{1,m}(\omega) = X_{1,m}(\omega)+V_{1,m}(\omega)$. Assuming that some sort of communication (e.g. wireless link) exists between both hearing aids, we are able to use all microphone inputs from both the left and the right hearing aid 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

$$\mathbf{Y}(\omega) = \begin{bmatrix} Y_{0,0}(\omega) \dots Y_{0,M_0-1}(\omega) & Y_{1,0}(\omega) \dots Y_{1,M_1-1}(\omega) \end{bmatrix}^T.$$

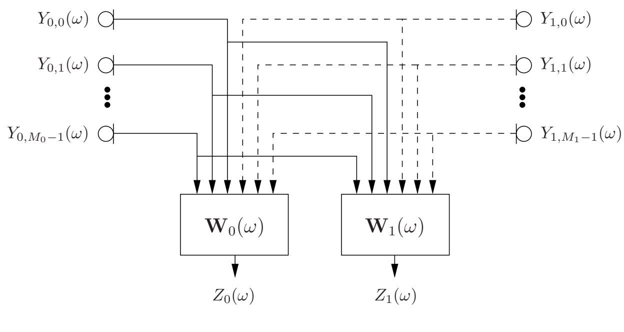


Figure 1: Binaural hearing aid configuration

The signal vector can be decomposed 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 hearing aid $Z_0(\omega)$ and $Z_1(\omega)$ are equal to

$$Z_0(\omega) = \mathbf{W}_0^H(\omega)\mathbf{Y}(\omega) = \mathbf{W}_0^H(\omega)\mathbf{X}(\omega) + \mathbf{W}_0^H(\omega)\mathbf{V}(\omega) \ ,$$
$$Z_1(\omega) = \mathbf{W}_1^H(\omega)\mathbf{Y}(\omega) = \mathbf{W}_1^H(\omega)\mathbf{X}(\omega) + \mathbf{W}_1^H(\omega)\mathbf{V}(\omega) \ ,$$

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

$$\mathbf{W}(\omega) = \begin{bmatrix} \mathbf{W}_0^T(\omega) & \mathbf{W}_1^T(\omega) \end{bmatrix}^T \ . \tag{2}$$

For conciseness, we will omit the frequency-domain variable $\omega$ in the remainder of the paper.

## 3. BINAURAL MULTI-CHANNEL WIENER FILTER

The multi-channel Wiener filter (MWF) produces a minimum mean-square error (MMSE) estimate of the speech component in one of the microphone signals, hence simultaneously reducing residual noise and limiting speech distortion [4, 5]. Hence, in a binaural hearing aid system an estimate of a speech component at both the left and the right hearing aid can be generated. The MSE cost function for the filter $\mathbf{W}_0$ estimating the speech component $X_{0,r_0}$ in the $r_0$th microphone signal of the left hearing aid is equal to

$$J_{MSE,0}(\mathbf{W}_0) = \mathcal{E}\{|X_{0,r_0} - \mathbf{W}_0^H\mathbf{Y}|^2\} \ .$$

The MSE cost function $J_{MSE,1}(\mathbf{W}_1)$ for the right hearing aid is defined similarly. The total MSE cost function is equal to

$$\boxed{J_{MSE}(\mathbf{W}) = J_{MSE,0}(\mathbf{W}_0) + J_{MSE,1}(\mathbf{W}_1)} \tag{3}$$

These cost functions can be written as

$$J_{MSE,0}(\mathbf{W}_0) = P_0 + \mathbf{W}_0^H(\mathbf{R}_x+\mathbf{R}_v)\mathbf{W}_0 - \mathbf{W}_0^H\mathbf{r}_{x0} - \mathbf{r}_{x0}^H\mathbf{W}_0,$$
$$J_{MSE,1}(\mathbf{W}_1) = P_1 + \mathbf{W}_1^H(\mathbf{R}_x+\mathbf{R}_v)\mathbf{W}_1 - \mathbf{W}_1^H\mathbf{r}_{x1} - \mathbf{r}_{x1}^H\mathbf{W}_1,$$

with

$$\mathbf{R}_x = \mathcal{E}\{\mathbf{X}\mathbf{X}^H\} \quad \mathbf{r}_{x0} = \mathcal{E}\{\mathbf{X}X_{0,r_0}^*\} \quad P_0 = \mathcal{E}\{|X_{0,r_0}|^2\}$$
$$\mathbf{R}_v = \mathcal{E}\{\mathbf{V}\mathbf{V}^H\} \quad \mathbf{r}_{x1} = \mathcal{E}\{\mathbf{X}X_{1,r_1}^*\} \quad P_1 = \mathcal{E}\{|X_{1,r_1}|^2\} \ ,$$

assuming independence between the speech and the noise component. In practice, we assume that the noise correlation matrix $\mathbf{R}_v$ can be estimated during noise-only periods, and the speech correlation matrix can be computed as

$$\mathbf{R}_x = \mathbf{R}_y - \mathbf{R}_v \ , \tag{4}$$

where the matrix $\mathbf{R}_y$ is estimated during speech and noise-periods. Using (2), the total SDW cost function in (3) can be written as

$$\boxed{J_{SDW}(\mathbf{W}) = P + \mathbf{W}^H\mathbf{R}\mathbf{W} - \mathbf{W}^H\mathbf{r} - \mathbf{r}^H\mathbf{W}} \tag{5}$$

with $P = P_0 + P_1$ and

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_x + \mathbf{R}_v & \mathbf{0}_M \\ \mathbf{0}_M & \mathbf{R}_x + \mathbf{R}_v \end{bmatrix} \quad \mathbf{r} = \begin{bmatrix} \mathbf{r}_{x0} \\ \mathbf{r}_{x1} \end{bmatrix}. \tag{6}$$

By setting the gradient of $J_{SDW}(\mathbf{W})$ equal to $\mathbf{0}$, the optimal filter minimising $J_{SDW}(\mathbf{W})$ is equal to

$$\mathbf{W}_{SDW} = \mathbf{R}^{-1}\mathbf{r} \ . \tag{7}$$

## 4. PRESERVATION OF BINAURAL CUES

Since the MWF produces an MMSE estimate of the speech component in the reference microphone signals at both hearing aids, the binaural cues, i.e. ITD and ILD, of the speech component are generally well preserved [6]. On the contrary, the binaural cues of the noise component may be distorted. In addition to reducing the noise level, it is however also important to (partially) preserve these binaural noise cues in order to exploit the binaural hearing advantage of normal hearing and hearing impaired persons or in order to further process the binaural output signals with a speech enhancement procedure that is based on a difference between speech and noise cues [9].

### 4.1. Partial estimation of the noise component

An extension of the MWF that partially preserves the binaural noise cues has been proposed in [7]. The objective is to produce an MMSE estimate of a desired signal that is equal to the sum of the speech component and a scaled version of the noise component, i.e. the cost function for the left hearing aid becomes

$$\bar{J}_{MSE,0}(\mathbf{W}_0) = \mathcal{E}\{|(X_{0,r_0} + \lambda_0 V_{0,r_0}) - \mathbf{W}_0^H\mathbf{Y}|^2\} \ , \tag{8}$$

with $0 \le \lambda_0 \le 1$. When $\lambda_0 = 0$, this cost function reduces to $J_{MSE,0}(\mathbf{W}_0)$. When $\lambda_0 = 1$, the optimal filter is equal to a vector consisting of zeros, with the $r_0$th element equal to 1, resulting in no noise reduction, but complete preservation of the binaural noise cues. It can be easily shown that all expressions derived in Section 3 remain valid when replacing $\mathbf{r}$ in (7) with

$$\mathbf{r} = \begin{bmatrix} \mathbf{r}_{x0} + \lambda_0\mathbf{r}_{v0} \\ \mathbf{r}_{x1} + \lambda_1\mathbf{r}_{v1} \end{bmatrix} \ , \tag{9}$$

with $\mathbf{r}_{v0}$ defined similarly as $\mathbf{r}_{x0}$. As will be shown in the simulations in Section 5, the ITD and the ILD cues of both the speech and the noise component can be preserved using this technique. However, this can not be achieved without considerably reducing the noise reduction performance.

### 4.2. Extension of SDW-MWF with binaural cues

In this paper, we present a different way to preserve the binaural noise cues by adding a term to the MSE cost function that is related to the ITD cue and the ILD cue of the noise component. The total cost function can then be expressed as

$$\boxed{\begin{aligned} J_{tot}(\mathbf{W}) \ = \ & J_{MSE}(\mathbf{W}) + \beta \underbrace{|ITD_{out}(\mathbf{W}) - ITD_{in}|^2}_{J_{ITD}(\mathbf{W})} \\ & + \gamma \underbrace{|ILD_{out}(\mathbf{W}) - ILD_{in}|^2}_{J_{ILD}(\mathbf{W})} \end{aligned}} \tag{10}$$

where $\beta$ and $\gamma$ are weight factors[1]. The main challenge is to come up with a perceptually relevant mathematical expression for these binaural cues.

a) We will express the ITD in the frequency-domain using the phase of the cross-correlation between two signals.The input noise cross-correlation is equal to

$$s = \mathcal{E}\{V_{0,r_0}V_{1,r_1}^*\} = \mathbf{R}_v(r_0, r_1) \ . \tag{11}$$

[1]These factors could be frequency-dependent, since it is well known that e.g for sound localisation the ITD cue is more important at low frequencies and the ILD cue is more important at high frequencies [10].

Similarly, the output noise cross-correlation is equal to

$$\mathcal{E}\{Z_{v0}Z_{v1}^*\} = \mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_1 \ . \tag{12}$$

In [8] the cost function $J_{ITD}(\mathbf{W})$ has been defined using the cosine of the phase difference $\phi(\mathbf{W})$ between the input and the output noise cross-correlation, i.e.

$$\boxed{\begin{aligned} J_{ITD}(\mathbf{W}) &= 1 - \cos\big(\phi(\mathbf{W})\big) \\ &= 1 - \frac{s_R\,(\mathbf{W}_0^H\mathbf{R}_v\mathbf{W}_1)_R + s_I\,(\mathbf{W}_0^H\mathbf{R}_v\mathbf{W}_1)_I}{\sqrt{s_R^2+s_I^2}\sqrt{(\mathbf{W}_0^H\mathbf{R}_v\mathbf{W}_1)_R^2+(\mathbf{W}_0^H\mathbf{R}_v\mathbf{W}_1)_I^2}} \end{aligned}} \tag{13}$$

where $\cdot_R$ and $\cdot_I$ denote the real and the imaginary part.
b) We will express the ILD in the frequency-domain using the power ratio of two signals. The input power ratio of the noise components in the reference microphone signals is equal to

$$\frac{\mathcal{E}\{|V_{0,r_0}|^2\}}{\mathcal{E}\{|V_{1,r_1}|^2\}} = \frac{\mathbf{R}_v(r_0,r_0)}{\mathbf{R}_v(r_1,r_1)} = P. \tag{14}$$

Similarly, the output power ratio of the noise components in the output signals is equal to

$$\frac{\mathcal{E}\{|Z_{v0}|^2\}}{\mathcal{E}\{|Z_{v1}|^2\}} = \frac{\mathbf{W}_0^H\mathbf{R}_v\mathbf{W}_0}{\mathbf{W}_1^H\mathbf{R}_v\mathbf{W}_1} \ . \tag{15}$$

We now define the cost function $J_{ILD}(\mathbf{W})$ as the squared difference between the input and the output noise power ratios, i.e.

$$\boxed{J_{ILD}(\mathbf{W}) = \left[\frac{\mathbf{W}_0^H\mathbf{R}_v\mathbf{W}_0}{\mathbf{W}_1^H\mathbf{R}_v\mathbf{W}_1} - P\right]^2} \tag{16}$$

c) Since no closed-form expression is available for the filter minimising the total cost function $J_{tot}(\mathbf{W})$, we will use iterative optimisation techniques. Many of these techniques (e.g. quasi-Newton method) are able to exploit the analytical expressions for the gradient and the Hessian, which can be derived using (5), (13) and (16). As will be shown in Section 5, the ITD and ILD cues of both the speech and the noise component can be preserved without comprimising the noise reduction performance.

## 5. EXPERIMENTAL RESULTS

### 5.1. Set-up and performance measures

The recordings used in the simulations were made in a room with dimensions 11'×11'×8'6", having a relatively low reverberation time ($T_{60} \approx 150$ ms). Two Knowles FG microphones were placed horizontally inside both ears of a KEMAR mannequin ($M_0 = M_1 = 2$), with a microphone spacing of 1 cm. The desired speech source is positioned in front of the head (0°) and consists of English sentences. The noise scenario consists of a multi-talker babble source positioned at 45°. All recordings were performed at a sampling frequency of 16 kHz. For evaluation purposes, the speech and the noise signal were recorded separately. The unbiased broadband SNR of the reference microphone signals at the left and the right hearing aid ($r_0 = r_1 = 0$) is 0 dB and $-3.2$ dB.
The FFT-size used for frequency-domain processing is $N = 256$. The noise correlation matrices $\mathbf{R}_v^n, n = 0\ldots N-1$, are estimated during noise-only periods, the matrices $\mathbf{R}_y^n$ are estimated during speech and noise-periods, and the speech correlation matrices are computed as $\mathbf{R}_x^n = \mathbf{R}_y^n - \mathbf{R}_v^n$.

As performance measures we use the *SNR improvement* between the input and the output signal at the left and the right hearing aid, and the *ITD and ILD cost function* for the noise and the speech component. The SNR improvement for the left hearing aid is defined as the mean of the SNR improvement in dB over all frequencies, i.e.

$$\Delta\mathrm{SNR}_0 = \frac{10}{N}\sum_{n=0}^{N-1} \log_{10}\frac{\mathbf{W}_0^{n,H}\mathbf{R}_x^n\mathbf{W}_0^n}{\mathbf{W}_0^{n,H}\mathbf{R}_v^n\mathbf{W}_0^n} - \log_{10}\frac{\mathbf{R}_x^n(r_0,r_0)}{\mathbf{R}_v^n(r_0,r_0)} \ .$$

The SNR improvement for the right hearing aid is defined similarly. The ITD cost function for the noise component is defined as the mean of the cost function $J_{ITD}(\mathbf{W}^n)$ in (13) over all frequencies. The ILD cost function for the noise component is defined as the mean of the cost function $J_{ILD}(\mathbf{W}^n)$ in (16) over all frequencies. The ITD and ILD cost functions for the speech component are defined similarly as for the noise component, by replacing $\mathbf{R}_v$ with $\mathbf{R}_x$ in (11), (13), (14) and (16).

### 5.2. SNR improvement and preservation of binaural cues

In the *first experiment*, we used the technique described in Section 4.1. Figure 2 shows the SNR improvement, the ITD and the ILD cost function for different values of $\lambda$ ($\lambda_0 = \lambda_1 = \lambda$). For the standard MWF, i.e. $\lambda = 0$, the ITD cost function for the speech component is quite low, but the ITD cost function for the noise component is relatively high, implying that the ITD cue for the speech component is preserved and the ITD cue for the noise component is distorted. For the standard MWF, the ILD cost function for both components is relatively low. As $\lambda$ increases, the ITD and the ILD cost functions for both the speech and the noise component decrease, but the SNR improvement is also significantly degraded (for $\lambda = 1$, $\Delta\mathrm{SNR} = 0$ and $J_{ITD} = J_{ILD} = -\infty$).
In the *second experiment*, we used the technique described in Section 4.2. Figure 3 shows the SNR improvement, the ITD and the ILD cost function for different values of $\beta$ and $\gamma$. As $\beta$ increases, the ITD cost function for the noise component decreases (almost independently of $\gamma$) and the ITD cost function for the speech component slightly increases. As $\gamma$ increases, the ILD cost function for the noise component decreases (almost independently of $\beta$) and the ILD cost function for the speech component slightly increases. The effect on the SNR improvement is relatively small ($< 1.3$ dB). As $\beta$ increases, the SNR improvement at both ears decreases. As $\gamma$ increases, the SNR improvement at the right ear decreases, but the SNR improvement at the left ear increases. This can be explained by a decreased noise level at the left ear and an increased noise level at the right ear in order to better preserve the noise ILD cue. We can conclude that the binaural cues of both the speech and the noise component can be preserved without significantly reducing the noise reduction performance.

## 6. CONCLUSION

In this paper we presented an extension of the MWF for binaural hearing aids, which is able to achieve a significant noise reduction while not distorting the binaural cues (ITD and ILD) of both the speech and the noise component.

## 7. REFERENCES

[1] J.G. Desloge, W.M. Rabinowitz, and P.M. Zurek, "Microphone-array hearing aids with binaural output–Part

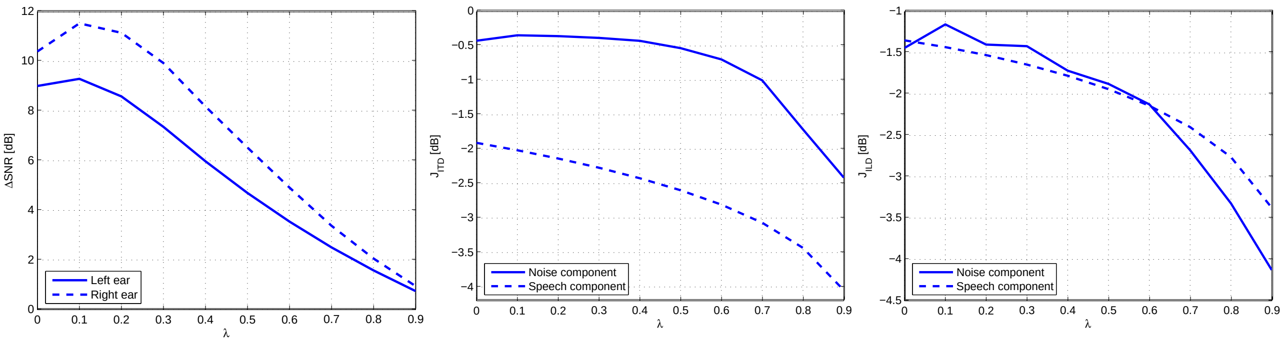


Figure 2: *SNR improvement, ITD and ILD cost function using partial estimation of the noise component* ($M = 4$, $\beta = \gamma = 0$)

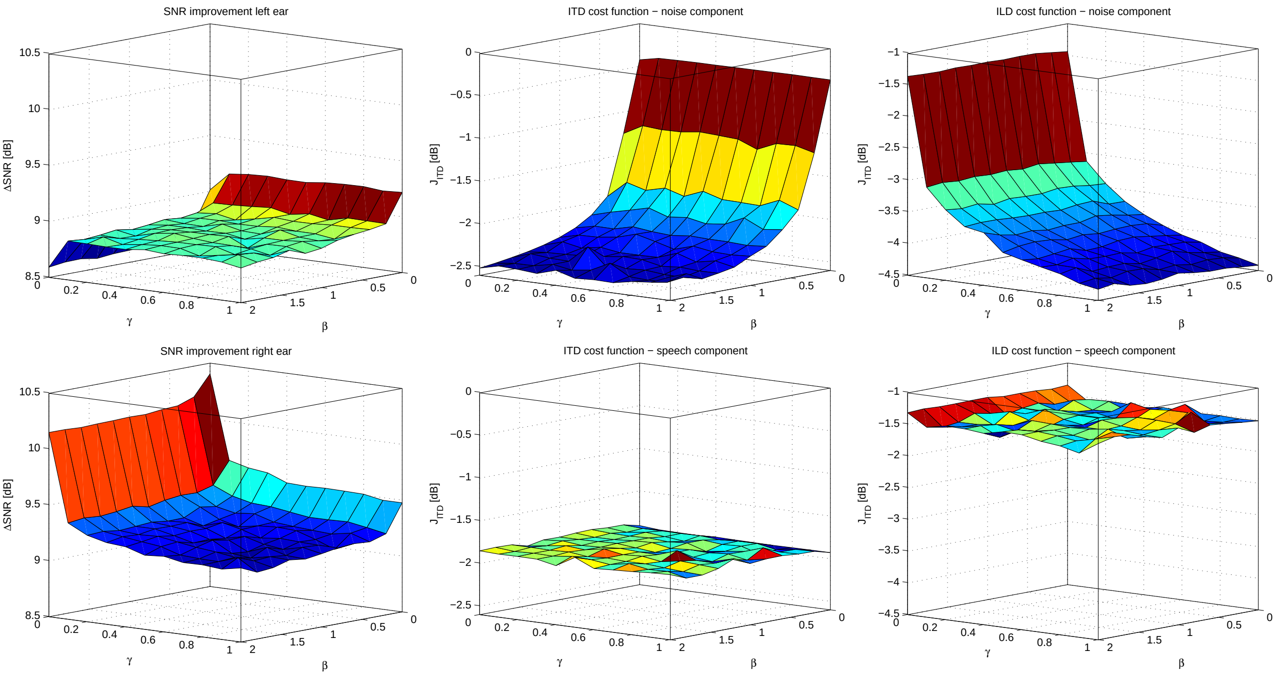


Figure 3: *SNR improvement, ITD and ILD cost function using extension of MWF with binaural cues* ($M = 4$, $\lambda_0 = \lambda_1 = 0$)


I: Fixed-processing systems," *IEEE Trans. Speech and Audio Processing*, vol. 5, no. 6, pp. 529–542, Nov. 1997.

[2] D.P. Welker, J.E. Greenberg, J.G Desloge, and P.M. Zurek, "Microphone-array hearing aids with binaural output–Part II: A two-microphone adaptive system," *IEEE Trans. Speech and Audio Processing*, vol. 5, no. 6, pp. 543–551, Nov. 1997.

[3] R. Nishimura, Y. Suzuki, and F. Asano, "A new adaptive binaural microphone array system using a weighted least squares algorithm," in *Proc. ICASSP*, Orlando, USA, May 2002, pp. 1925–1928.

[4] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Trans. Signal Processing*, vol. 50, no. 9, pp. 2230–2244, Sept. 2002.

[5] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, *Speech Distortion Weighted Multichannel Wiener Filtering Techniques for Noise Reduction*, Springer-Verlag, 2005, ch. 2 in "Speech Enhancement", pp. 199–228.

[6] T. 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*, Mar. 2005, pp. 29–32.

[7] T. Klasen, T. Van den Bogaert, M. Moonen, and J. Wouters, "Binaural noise reduction for hearing aids: Preserving interaural time delay cues," in *Proc. SPS-DARTS 2005*, Antwerp, Belgium, Apr. 2005, pp. 23–26.

[8] S. Doclo, R. Dong, T. Klasen, J. Wouters, S. Haykin and M. Moonen, "Extension of the multi-channel Wiener filter with ITD cues for noise reduction in binaural hearing aids," *Submitted to WASPAA*, New Paltz, USA, Oct. 2005.

[9] T. Wittkop and V. Hohmann, "Strategy-selective noise reduction for binaural digital hearing aids," *Speech Communication*, vol. 39, no. 1-2, pp. 111–138, Jan. 2003.

[10] F. Wightman and D. Kistler, "The dominant role of low-frequency interaural time differences in sound localization," *Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1648–1661, Mar. 1992.