

# Combined Noise Reduction and Coherence Reshaping for Binaural Hearing Aids

Daniel Marquardt<sup>1</sup>, Volker Hohmann<sup>2</sup>, Simon Doclo<sup>1</sup>

<sup>1</sup> *University of Oldenburg, Department of Medical Physics and Acoustics, Signal Processing Group*

<sup>2</sup> *University of Oldenburg, Department of Medical Physics and Acoustics, Medical Physics*

## 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 the 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 information can be exploited [1]. In addition to reducing noise and limiting speech distortion, another important objective of binaural noise reduction algorithms is the preservation of the listener's impression of the acoustical 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 binaural cues of the speech and the noise component.

To achieve binaural cue preservation, two main concepts for binaural noise reduction have been developed. In the first concept, the multi-channel signals are used to calculate a real-valued gain, where the same gain is applied to the reference microphone in the left, respectively right hearing aid [2]. This processing strategy allows perfect preservation of the binaural cues of both the speech and the noise component, but typically suffers from limited noise reduction performance and possible single-channel noise reduction artifacts. The second concept is to apply a complex-valued filter to all available microphone signals on the left and the right hearing aid, combining spatial and spectral filtering. Using this processing strategy, a large noise reduction performance can be achieved, but the binaural cues of the residual noise component are not guaranteed to be preserved. In [1] the binaural Speech Distortion Weighted Multi-channel Wiener Filter (MWF) has been presented. It has been theoretically proven in [3] that in case of a single speech source this technique preserves the binaural cues of the speech component but typically distorts the binaural cues of the noise component. Hence, algorithms have been proposed that aim to preserve the binaural cues of directional noise sources by adding a cue preservation term related to the Interaural Transfer Function (ITF), the Interaural Level Difference (ILD) or the Interaural Time Difference (ITD) to the basic noise reduction cost function [3, 4, 5].

In contrast to directional noise sources, the spatial characteristics of e.g. spatially isotropic noise however can not be properly described by the ITF, but rather by the Interaural Coherence (IC). In this paper we propose an extension of the MWF with a term related to the

IC preservation of the noise component. Experimental results for two scenarios show that the proposed algorithm yields a good preservation of the Interaural Coherence without significantly degrading the output SNR compared to the binaural MWF.

## Configuration and Notation

Consider 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  $Y_{0,m}(\omega)$  can be written as

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

with  $X_{0,m}(\omega)$  and  $V_{0,m}(\omega)$  representing the speech and the noise component. The  $m$ -th microphone signal in the right hearing aid  $Y_{1,m}(\omega)$  is defined similarly. For conciseness we will omit the variable  $\omega$  in the remainder of the paper.

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

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

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 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

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

where  $Z_{x,0}$  represents the speech component and  $Z_{v,0}$  represents the noise component. Similarly, the output

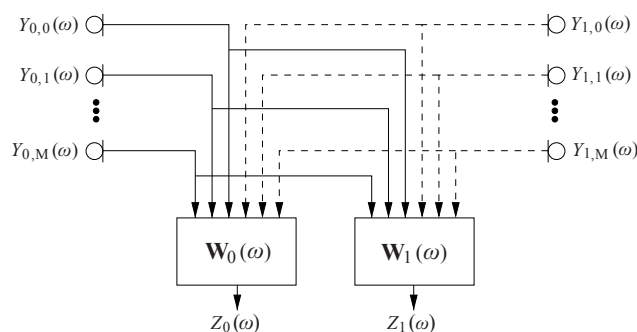


Figure 1: Binaural hearing aid configuration

signal at the right hearing aid  $Z_1$  can be defined. The correlation matrices are defined as

$$\mathbf{R}_y = \mathcal{E} \{ \mathbf{Y} \mathbf{Y}^H \}, \mathbf{R}_v = \mathcal{E} \{ \mathbf{V} \mathbf{V}^H \}, \mathbf{R}_x = \mathcal{E} \{ \mathbf{X} \mathbf{X}^H \}. \quad (5)$$

The Interaural Coherence of the input and output noise component are defined as

$$\begin{aligned} IC_v^{in} &= \frac{\mathcal{E} \{ V_0 V_1^* \}}{\sqrt{\mathcal{E} \{ V_0 V_0^* \} \mathcal{E} \{ V_1 V_1^* \}}} \\ &= \frac{\mathbf{e}_0^T \mathbf{R}_v \mathbf{e}_1}{\sqrt{\mathbf{e}_0^T \mathbf{R}_v \mathbf{e}_0 \mathbf{e}_1^T \mathbf{R}_v \mathbf{e}_1}} \end{aligned} \quad (6)$$

$$\begin{aligned} IC_v^{out} &= \frac{\mathcal{E} \{ Z_{v,0} Z_{v,1}^* \}}{\sqrt{\mathcal{E} \{ Z_{v,0} Z_{v,0}^* \} \mathcal{E} \{ Z_{v,1} Z_{v,1}^* \}}} \\ &= \frac{\mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_1}{\sqrt{\mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_0 \mathbf{W}_1^H \mathbf{R}_v \mathbf{W}_1}}. \end{aligned} \quad (7)$$

The vectors  $\mathbf{e}_0$  and  $\mathbf{e}_1$  are zero column vectors with  $\mathbf{e}_0(1) = 1$  and  $\mathbf{e}_1(M+1) = 1$ . The noise components in the reference microphones are  $V_0 = \mathbf{e}_0^T \mathbf{V}$  and  $V_1 = \mathbf{e}_1^T \mathbf{V}$ . The real-valued Magnitude Squared Coherence (MSC) is defined as  $MSC = |IC|^2$ .

## Binaural multi-channel Wiener filter (MWF)

The binaural MWF produces a minimum mean-square error (MMSE) estimate of the speech component in the reference microphone signal for both hearing aids. The MWF cost function estimating the speech components  $X_{0,0}$  and  $X_{1,0}$  in the left and the right hearing aid can be written as

$$J_{MWF}(\mathbf{W}) = \mathcal{E} \left\{ \left\| \begin{bmatrix} X_{0,0} - \mathbf{W}_0^H \mathbf{X} \\ X_{1,0} - \mathbf{W}_1^H \mathbf{X} \end{bmatrix} \right\|^2 + \mu \left\| \begin{bmatrix} \mathbf{W}_0^H \mathbf{V} \\ \mathbf{W}_1^H \mathbf{V} \end{bmatrix} \right\|^2 \right\}, \quad (8)$$

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

$$\mathbf{W}_{MWF} = \mathbf{R}^{-1} \mathbf{r}_x, \quad (9)$$

with

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_x + \mu \mathbf{R}_v & \mathbf{0}_{2M} \\ \mathbf{0}_{2M} & \mathbf{R}_x + \mu \mathbf{R}_v \end{bmatrix}, \quad \mathbf{r}_x = \begin{bmatrix} \mathbf{R}_x \mathbf{e}_0 \\ \mathbf{R}_x \mathbf{e}_1 \end{bmatrix}. \quad (10)$$

Based on the theoretical analysis in [3] we can show that in case of a single speech source the output IC of the speech and the noise component is equal to the input IC of the speech component, i.e.

$$IC_x^{out} = IC_v^{out} = IC_x^{in} = e^{j\angle \frac{A_0}{A_1}}. \quad (11)$$

where  $A_0$  and  $A_1$  are the acoustic transfer functions from the speech source to the reference microphone in the left, respectively right hearing aid and  $\angle$  denotes the phase. Equation (11) also implies that in case of a spatially isotropic noise field the residual noise component would be perceived as a point source coming from the speech direction, which is obviously undesired.

## MWF with Interaural Coherence preservation (MWF-IC)

To allow for the preservation of the Interaural Coherence of spatially isotropic noise fields, we define the following coherence preservation term

$$J_{IC}(\mathbf{W}) = \left| \frac{\mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_1}{\sqrt{\mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_0 \mathbf{W}_1^H \mathbf{R}_v \mathbf{W}_1}} - IC_v^{des} \right|^2, \quad (12)$$

where  $IC_v^{des}$  represents the desired output IC of the noise component, which can be equal to the estimated  $IC_v^{in}$  as in (6) or can be defined using e.g. HRTF measurements or head models [6]. We add this additional term to the MWF cost function in (8) i.e.,

$$J_{MWF-IC}(\mathbf{W}) = J_{MWF}(\mathbf{W}) + \lambda J_{IC}(\mathbf{W}), \quad (13)$$

where the parameter  $\lambda$  allows for a trade-off between noise reduction and coherence preservation. The resulting filter will be denoted as the MWF-IC. Since no closed-form expression is available for the filter minimizing  $J_{MWF-IC}$ , we will resort to iterative optimization techniques. We have used a *trust-region method* where analytical expressions for the gradient and the Hessian of the cost function  $J_{MWF-IC}(\mathbf{W})$  have been provided in order to improve the numerical robustness and the convergence speed. The analytical expression of the gradient and the Hessian are omitted due to space constraints.

## Experimental Results

In this section we perform simulations for an office room and a cafeteria scenario to investigate the performance of the MWF and the MWF-IC with respect to the intelligibility weighted output SNR improvement, the ILD and ITD preservation of the speech component and the IC preservation of the noise component.

### Setup

Binaural Behind-The-Ear Impulse Responses (BTE-IR) measured in an office room and a cafeteria from [7] have been used to generate the speech signals. Each hearing aid was equipped with 2 microphones, therefore in total 4 microphone signals are available. The speech source was located in front of the hearing aid user. For the office room scenario the noise components were generated using the method described in [8], where the power spectral density (PSD) of the noise components was equal to the PSD of speech-shaped noise and the coherence matrix of the binaural setup in a cylindrical isotropic noise field was estimated using the anechoic BTE-IR from [7]. For the cafeteria scenario realistic ambient noise that has been recorded in the cafeteria was added to the speech component. The signals were processed at  $f_s = 16$  kHz using a weighted overlap-add (WOLA) framework with a block size of  $N = 256$  samples and an overlap of 75% between successive blocks. The speech + noise signal had a length of 4s and was preceded by a noise-only signal of 4s length. The noise-only part was not taken into

account during evaluation. The average intelligibility weighted input SNR in the reference microphones was equal to 0 dB. The correlation matrices of the signal components are estimated as

$$\mathbf{R}_y(k) = \frac{1}{L_y} \sum_{i=0}^{L_y-1} \mathbf{Y}(k, i) \mathbf{Y}^H(k, i) \quad (14)$$

$$\mathbf{R}_v(k) = \frac{1}{L_v} \sum_{i=0}^{L_v-1} \mathbf{V}(k, i) \mathbf{V}^H(k, i) \quad (15)$$

with  $k = 0, 1, \dots, N - 1$  denoting the frequency index,  $i$  denoting the block index,  $L_y$  denoting the available signal vectors when speech is present and  $L_v$  denoting the available signal vectors when speech is absent, using a perfect Voice Activity Detector (VAD). The correlation matrix of the speech component is estimated as

$$\mathbf{R}_x(k) = \mathbf{R}_y(k) - \mathbf{R}_v(k). \quad (16)$$

The desired IC in the MWF-IC was calculated from the estimate of the noise correlation matrix  $\mathbf{R}_v$  as in (6). The parameter  $\mu$  in 9 was set to 5. The frequency-dependent trade-off parameter  $\lambda$  in the MWF-IC was determined such that the MSC error of the noise component in each frequency bin was kept below 0.1 [MWF-IC (0.1)], respectively below 0.01 [MWF-IC (0.01)].

## Performance measures

For comparing the performance of the algorithms we have used 3 objective performance measures to evaluate the binaural cue preservation of the speech and the noise component and the SNR improvement. The *intelligibility weighted SNR* [9] is defined as

$$iSNR = \sum_k I(k) 10 \log_{10} \left( \frac{P_x(k)}{P_v(k)} \right), \quad (17)$$

where  $P_x(k)$  and  $P_v(k)$  are the PSDs of the speech component, respectively noise component of the input signal (input *iSNR*), respectively the output signal (output *iSNR*).  $I(k)$  is a weighting function that takes the importance of different frequency bands for speech intelligibility into account.

To avoid a separate analysis for the real and imaginary part of the complex-valued IC we evaluated the performance using the real-valued MSC. The MSC of the input noise component was calculated during the 4 s speech + noise period, i.e.

$$MSC_v^{in}(k) = \left| \frac{\sum_{i=0}^{L_y-1} V_0(k, i) V_1^*(k, i)}{\sqrt{\sum_{i=0}^{L_y-1} |V_0(k, i)|^2 \sum_{i=0}^{L_y-1} |V_1(k, i)|^2}} \right|^2 \quad (18)$$

The output MSC of the noise component was calculated by replacing  $V$  with  $Z_v$  in (18).

For the directional speech component the MSC error is however not an appropriate objective measure. The MSC contains information about the amount of correlation of a signal in the microphones but does not contain information about the perceived direction of a directional source.

Hence, for the evaluation of the binaural cue preservation of the speech component we use an objective measure which is based on a model of binaural auditory processing [10] and has already been described in more detail and applied for binaural cue preservation evaluation in [11].

## Performance Results

In this section we describe the performance of the MWF and the MWF-IC in terms of the previously mentioned objective performance measures for both scenarios.

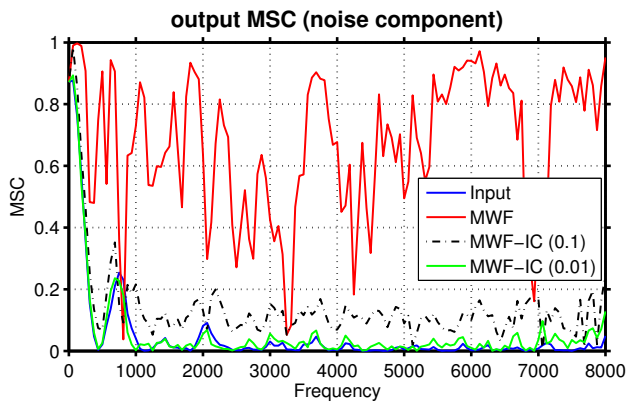
The output MSC of the noise component for the office room is depicted in Fig. 2 and the output MSC of the noise component for the cafeteria is depicted in Fig. 3. From the theoretical analysis of the output IC of the MWF, the MSC of the output noise component is expected to be 1 for all frequencies, which is not exactly the case due to estimation errors in the speech correlation matrix  $\mathbf{R}_x$ . More precisely, due to the short block length in the WOLA framework, the rank-1 assumption of the speech correlation matrix  $\mathbf{R}_x$  is violated and the theoretical results deviate from the simulation results. Nevertheless, the distortion of the MSC of the output noise component introduced by the binaural MWF is clearly visible. Using the MWF-IC the MSC of the input noise component can be preserved and the MSC preservation can be improved by increasing the trade-off parameter  $\lambda$ .

The ILD and ITD error of the speech component is depicted in Fig. 4. In the office room scenario the ILD and ITD error is slightly increased using the MWF-IC (0.01) compared to the MWF. The MWF-IC (0.1) does not increase the ILD and ITD error of the speech component compared to the MWF. Hence, a preservation of the noise component IC is possible using the MWF-IC without substantially distorting the speech component cues.

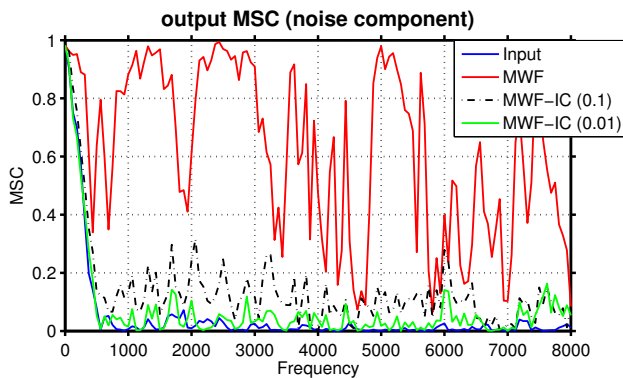
The *iSNR* improvement of the left and the right hearing aid for all algorithms is depicted in Fig. 5. As expected, the MWF yields a better performance than the MWF-IC. The *iSNR* improvement is slightly decreased for the MWF-IC, since a better preservation of the IC leads to less noise reduction due to the higher impact of the coherence preservation term in (13). Increasing the trade-off parameter  $\lambda$  in the MWF-IC leads to a better preservation of the MSC of the noise component but also degrades the output *iSNR*. Hence, the MWF-IC allows for a controllable trade-off between better IC preservation and output SNR.

## Conclusion

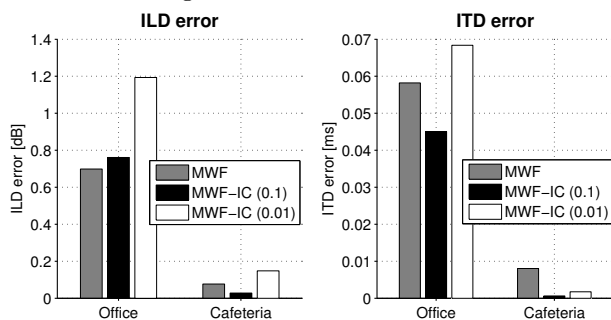
In this paper we have shown that for spatially isotropic noise fields the MWF-IC yields a better preservation of the IC of the residual noise component compared to the MWF without distorting the speech component cues. The influence of the trade-off between better IC preservation and output SNR on spatial awareness and speech intelligibility needs to be examined in future work.



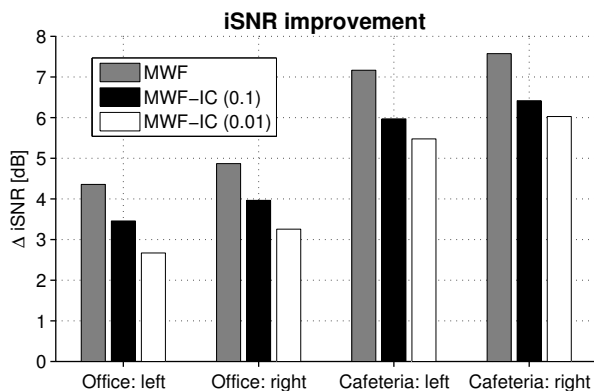
**Figure 2:** Output MSC of the noise component for the MWF and MWF-IC algorithms in the office room scenario



**Figure 3:** Output MSC of the noise component for the MWF and MWF-IC algorithms in the cafeteria scenario



**Figure 4:** Speech component ILD and ITD error for the MWF and MWF-IC algorithms



**Figure 5:** iSNR improvement for the MWF and MWF-IC algorithms

## References

[1] S. Doclo, S. Gannot, M. Moonen, and A. Spriet, “Acoustic beamforming for hearing aid applica-

tions,” in *Handbook on Array Processing and Sensor Networks*, pp. 269–302. Wiley, 2010.

- [2] T. Lotter and P. Vary, “Dual-channel speech enhancement by superdirective beamforming,” *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 1–14, 2006.
- [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, R. Dong, T. J. Klasen, J. Wouters, S. Haykin, and M. Moonen, “Extension of the multi-channel Wiener filter with localisation cues for noise reduction in binaural hearing aids,” in *Proc. IWAENC*, Eindhoven, The Netherlands, Sep. 2005.
- [5] D. Marquardt, V. Hohmann, and S. Doclo, “Binaural cue preservation for hearing aids using multi-channel Wiener filter with instantaneous ITF preservation,” in *Proc. ICASSP*, Kyoto, Japan, Mar. 2012, pp. 21–24.
- [6] M. Jeub, M. Dörbecker, and P. Vary, “A semi-analytical model for the binaural coherence of noise fields,” *Signal Processing Letters, IEEE*, vol. 18, no. 3, pp. 197–200, Mar. 2011.
- [7] H. Kayser, S. Ewert, J. Anemüller, 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.
- [8] E. A. P. Habets, I. Cohen, and S. Gannot, “Generating nonstationary multisensor signals under a spatial coherence constraint,” *Journal of the Acoustical Society of America*, vol. 124, no. 5, pp. 2911–2917, Nov. 2008.
- [9] 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.
- [10] M. Dietz, S. D. Ewert, and V. Hohmann, “Auditory model based direction estimation of concurrent speakers from binaural signals,” *Speech Communication*, vol. 53, pp. 592–605, 2011.
- [11] D. Marquardt, L. Wang, V. Hohmann, and S. Doclo, “Improved noise estimation for the binaural MWF with instantaneous ITF preservation,” in *Proc. European Signal Processing Conference (EUSIPCO)*, Bucharest, Romania, Aug. 2012, pp. 300–304.