

Preserving binaural hearing of hearing impaired subjects with binaural noise reduction systems for hearing aids

T. Van den Bogaert¹, S. Doclo^{2,3}, J. Wouters¹, M. Moonen²

¹ *ExpORL, K.U.Leuven, Herestraat 49 bus 721, B-3000 Leuven, Belgium, Email: tim.vandenbogaert@med.kuleuven.be*

² *ESAT-SCD, K.U.Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium;*

³ *Presently at NXP Semiconductors, Interleuvenlaan 80 (C034), 3001 Leuven, Belgium;*

Introduction

Noise reduction algorithms in hearing aids are important 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], [4]. This is not surprising, since noise reduction algorithms currently used in hearing aids, typically bilateral adaptive noise reduction systems, are designed to optimize signal to noise ratio (SNR) in a monaural way, and not to preserve binaural cues. Binaural cues are especially important for sound source localization and for speech segregation in noisy environments (a.k.a. 'the cocktail party effect').

The progress made in wireless communication between hearing aids is slowly enabling the design of a full binaural hearing aid. Changing from a bilateral to a binaural hearing aid design, i.e. generating an output signal for both ears using all microphone signals, could enhance the amount of noise reduction and could increase the capability to control the adaptive processes to preserve the binaural cues between left and right hearing aid. An important limitation of most noise reduction array systems studied thus far is that most of them are designed to have a single or monaural output. Extending these to a binaural output is not always trivial.

In this manuscript we present an overview of work done on two different binaural adaptive multi-microphone noise reduction algorithms: the multichannel Wiener filter (MWF) and the MWF with partial noise estimation (MWF-N). A classic bilateral adaptive directional microphone (ADM) was taken as a reference.

Algorithms

In general, the goal of a single channel or multichannel Wiener filter is to filter out noise corrupting a desired signal given all the available microphone signals. A Wiener filter in hearing aids is typically calculated by first estimating and then using the second-order statistical properties of the desired signal and the noise components. It generates an output signal which approaches the target signal as closely as possible in a mean-square error (MSE) sense. See [9] for an overview on Wiener filtering.

Doing noise reduction in a binaural framework leads to

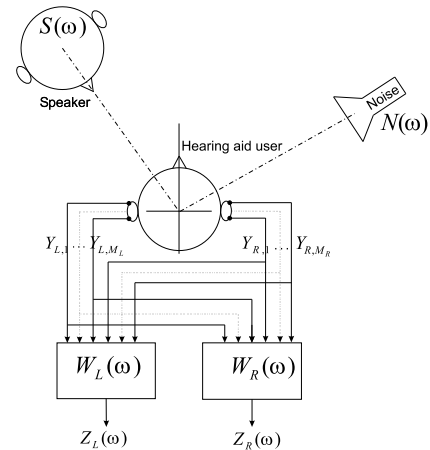


Figure 1: Noise reduction in a binaural framework. Calculate the filters \mathbf{W}_L and \mathbf{W}_R to remove the noisy part of the signal and which preserve the binaural information.

the system illustrated in Figure 1. A target signal and one or multiple noise sources are present in an acoustic environment. A mixture of both components, $\mathbf{Y}(\omega)$ is recorded by the microphones. $\mathbf{Y}(\omega)$ is a complex vector with size $1 \times M$ with M the number of microphones present in the system. We define the M -dimensional signal vector $\mathbf{Y}(\omega)$, with $M = M_{Left} + M_{Right}$, as

$$\mathbf{Y}(\omega) = [Y_{L,1}(\omega) \dots Y_{L,M_{Left}}(\omega) Y_{R,1}(\omega) \dots Y_{R,M_{Right}}(\omega)]^T.$$

$\mathbf{Y}(\omega)$ equals the sum of the recorded target component $\mathbf{X}(\omega)$ and noise component $\mathbf{V}(\omega)$ which are defined similarly as $\mathbf{Y}(\omega)$. For conciseness, the frequency domain variable ω is now omitted from the manuscript.

MWF

In the proposed MWF approach, an estimate of the unknown target signal picked up at the front microphone of the left ($X_{L,1}$) and the right ($X_{R,1}$) hearing aid is produced for the left and the right ear respectively. This is done by taking all microphone signals into account. Mathematically the problem can be formulated by minimizing the following cost function:

$$J_{MSE}(\mathbf{W}) = \mathcal{E} \left\{ \left\| \begin{bmatrix} X_{L,1} - \mathbf{W}_L^H \mathbf{Y} \\ X_{R,1} - \mathbf{W}_R^H \mathbf{Y} \end{bmatrix} \right\|^2 \right\}. \quad (1)$$

By rewriting the vector \mathbf{Y} as a sum of the target component \mathbf{X} and the noise component \mathbf{V} and by introducing a trade-off parameter μ (see [7]), this cost

function can be rewritten as

$$J_{MWF}(\mathbf{W}) = \mathcal{E} \left\{ \left\| \begin{bmatrix} X_{L,1} - \mathbf{W}_L^H \mathbf{X}_L \\ X_{R,1} - \mathbf{W}_R^H \mathbf{X}_R \end{bmatrix} \right\|^2 + \mu \left\| \begin{bmatrix} \mathbf{W}_L^H \mathbf{V}_L \\ \mathbf{W}_R^H \mathbf{V}_R \end{bmatrix} \right\|^2 \right\} \quad (2)$$

The parameter μ was introduced in monaural hearing aid research to increase the amount of noise reduction by allowing the system to introduce some speech distortion. In eq. 2, it can be seen that a large μ emphasizes the amount of noise reduction (the second term of the equation) while allowing some speech distortion (the first term of the equation). The solution to this minimization problem is known from Wiener filter theory and equals

$$\begin{bmatrix} \mathbf{W}_L \\ \mathbf{W}_R \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{x,L} + \mu \mathbf{R}_{v,L} & \mathbf{0}_M \\ \mathbf{0}_M & \mathbf{R}_{x,R} + \mu \mathbf{R}_{v,R} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{R}_{x,L} e_L \\ \mathbf{R}_{x,R} e_R \end{bmatrix} \quad (3)$$

with e_L and e_R being vectors with one element equal to 1 and the other elements equal to zero, defining the reference microphones used at both hearing aids, i.e., in case of the front omnidirectional microphone $e_L(1) = 1$ and $e_R(1) = 1$. \mathbf{R}_x and \mathbf{R}_v , which are at present still unknown, are defined as the $M \times M$ -dimensional speech and noise correlation matrices, containing the auto-correlations and cross-correlations (or the statistical information) of, respectively, the speech and noise components \mathbf{X} and \mathbf{V} over the different input channels, e.g., $\mathbf{R}_{x,L} = \mathcal{E}\{\mathbf{X}_L \mathbf{X}_L^H\}$. To find the filters \mathbf{W} using eq. (8), a voice activity detector (VAD) is used to discriminate between 'speech and noise periods' and 'noise only periods'. The noise correlation matrix \mathbf{R}_v can be calculated during the 'noise only periods'. By assuming a sufficient stationary noise signal, the speech correlation matrix \mathbf{R}_x can be estimated during speech and noise periods by subtracting \mathbf{R}_v from the correlation matrix \mathbf{R}_y for the noisy signal Y . By using these correlation matrices, the filters \mathbf{W} can be found.

Using a MWF in a binaural reference framework has the advantage that the binaural cues of the target component are always preserved by the algorithm. This was mathematically proven in [8] and can be easily understood by the fact that a binaural MWF estimates the target signal present at the left and right front microphone as well as possible for respectively the left and right ear. However, in [8] it was also proven that the binaural cues of the remaining noise component are changed into those of the original target component.

MWF-N

A binaural Wiener filter with partial noise estimation (MWF-N) is designed to remove only part $(1-\eta)$ of the noisy signal. The remaining part (η) of the noise component is preserved and presented to the hearing aid user. The reasoning for this approach is the assumption that the small but significant amount of unprocessed sound will result in a more natural perception of the surrounding sound environment and that the binaural cues of the unprocessed sound will enable the user to localize both the target component and the noise component(s). Increasing η off course degrades the

obtained improvement in SNR by the MWF algorithm. The minimization criterion of eq. 1 now changes into

$$J_{MSE\eta}(\mathbf{W}) = \mathcal{E} \left\{ \left\| \begin{bmatrix} X_{L,1} + \eta V_{L,1} - \mathbf{W}_L^H \mathbf{Y}_L \\ X_{R,1} + \eta V_{R,1} - \mathbf{W}_R^H \mathbf{Y}_R \end{bmatrix} \right\|^2 \right\} \quad (4)$$

The solution to this problem can be written similar as eq.3. However, there is also a simple relationship between the filter output of the MWF and the MWF-N:

$$Z_{MWF,\eta,L}(\eta, \mu) = \eta Y_{L,1} + (1 - \eta) Z_{MWF,L}(\mu) \quad (5)$$

$$Z_{MWF,\eta,R}(\eta, \mu) = \eta Y_{R,1} + (1 - \eta) Z_{MWF,R}(\mu) \quad (6)$$

During the evaluation of the algorithms we have chosen to put η to $\eta = 0.2$, thereby removing 80% of the noise component. Algorithms were always evaluated using a perfect VAD.

Reference

Two different reference conditions were evaluated: a condition without noise reduction and a condition in which a bilateral ADM was used. In the latter condition a single tap ADM was running independently for each ear which used both microphone inputs of the left and right dual microphone hearing aid respectively.

Localization

A CORTEX MK2 manikin was placed in an array of 13 loudspeakers. The speakers were located from -90° (at the left side) to $+90^\circ$ (at the right side) of the manikin with a spacing of 15° between them. All the speakers were located at a distance of 1m from the manikin. Then, impulse responses between the loudspeakers and the microphones of two dual microphone behind-the-ear hearing aids placed on the manikin were measured. Three different spatial scenarios were generated with these impulse responses: $S_0 N_{60}$, $S_{45} N_{-45}$, $S_{90} N_{-90}$. The spatial scenarios are denoted as $S_x N_y$ with x the location of the target component and y the location of a single noise component. A steady state noise, weighted by the average spectrum of a dutch male talker, was used as target component (S) and a multitalker babble (Auditec of St. Louis) was used as noise source (N). To evaluate localization performance in an unprocessed reference condition, the speech and noise signal were convolved with the impulse responses from angles x and y , measured with the front left and the front right microphone. These stimuli were then presented to the left and the right ear of the listener using headphones. The listeners were sitting in the loudspeaker array and had to localize the stimulus. An accumulation of these results is given in the top left figure of Figure 2. The x-axis presents where the stimulus was played (e.g. N-45 denotes that in this test a multitalker babble was played arriving from -45° of the subject) while the y-axis shows where the sound was localized. Figure 2 shows that localization is almost perfect in the unprocessed condition with a slight degradation at the sides of the subject.

When evaluating the noise reduction algorithms, the mixture of the speech and the noise component of $S_x N_y$

were first processed by the noise reduction algorithms. This way, the adaptive filters had the opportunity to adapt and convergence to that specific condition. Afterwards, the filters were fixed and the speech and noise source were filtered separately through the filters. The filtered speech and noise source were then presented to the listener who had to localize the sound. Target and noise source were presented separately to the subject since this way interaction effects such as masking, localizing two sound sources is different from localizing a single source, etc., were avoided. This enabled a clear interpretation on what the filters were doing to the target and the noise component. The accumulated responses for all three noise reduction algorithms are given in Figure 2.

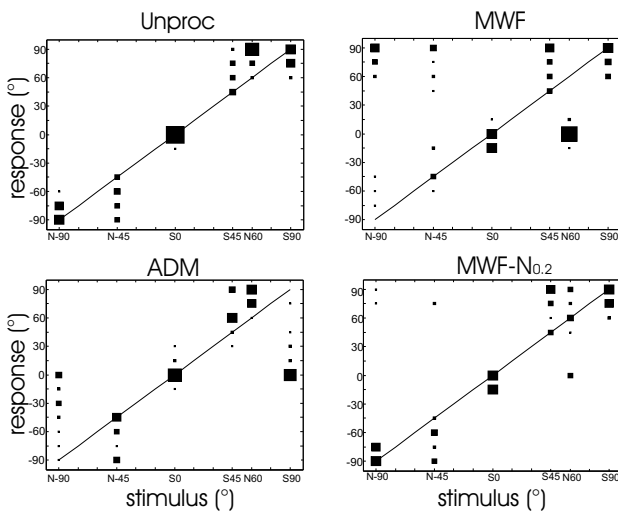


Figure 2: Accumulation of the responses of 5 normal hearing subjects on the localization task in the frontal horizontal hemisphere. Three different spatial scenarios (S_0N_{60} , $S_{45}N_{-45}$, $S_{90}N_{-90}$) were processed by three different algorithms. The x-axis shows where the sound was presented. The size of the square is proportional to the number of responses given at this location.

The top right of Figure 2 shows that all target signals (S) were localized correctly when using the MWF algorithm. The noise components (N) however were localized at the angle of where the speech source was played. E.g. both S and N are localized around $+90^\circ$ in the spatial condition $S_{90}N_{-90}$. This is in accordance with the theory of [8]. When using the MWF-N algorithm (bottom right of Figure 2), it was observed that both the speech and noise components were localized correctly. This proves that preserving 20% of the original noise component was sufficient to preserve a correct sound source localization. The bottom left figure of Figure 2 shows the results obtained when using a bilateral ADM. This figure shows that all sounds between -60° and $+60^\circ$ were localized correctly when using the ADM. This is because an ADM used in hearing aids is constrained to preserve the sound signal arriving from the front of the subject. Hence the signal and its binaural cues were not distorted. If a sound source, target or noise component, is arriving from the sides of the head, then localization deteriorated. For these angles of arrival, subjects localized the sound around 0° . However, when asked, subjects identified

the sound signals as being diffuse with no directional information present. This can be explained by the fact that the ADM tries to remove sound signals arriving from these angles. Hence, a diffuse sound component remains.

A final remark should be that in [5] it was shown that when filtering a noise component with the MWF algorithm, typically a "dual sound" is perceived. Due to the fact that a very small part (typically the frequency regions with a very low SNR) of the noise component are wrongly classified as target component (e.g. due to errors in the VAD algorithm), the binaural information of this part is preserved. On the other hand, the binaural cues of the part of the noise component which was correctly classified, are changed into those of the target component and a "dual sound" is heard. If speech and noise are presented simultaneously to the subject, masking effects occur. The combination of misclassification and masking effects can result in a correct localization of the noise signal even when using the MWF algorithm. The misclassification of the noise component of course reduces the noise reduction performance of the algorithm.

Noise reduction

Speech reception thresholds (SRTs) in multitalker babble of ten normal hearing subjects were measured using an adaptive test procedure and dutch sentence material. The obtained gain in SRT, ΔSRT_{algo} , is given in Table 1 for each algorithm. Three different spatial scenarios were generated by using the impulsresponses measured with the manikin (see previous section). Stimuli were processed off-line. First, the mixture of the speech and the noise component (S_xN_y) was processed by the different adaptive algorithms. After the filters reached convergence, the filters were fixed and stored. This was done for a range of different SNRs between $+10$ and -30 dB. Afterwards, an adaptive test procedure used the converged filters to produce the stimuli which were presented to the subjects. Three different spatial scenarios were evaluated: two single noise source scenarios, i.e. S_0N_{60} and $S_{90}N_{-90}$, and one scenario with three noise sources, $S_0N_{90/180/270}$. Similar to the localization study, three algorithms were evaluated: the ADM, the MWF and the MWF-N with 20% noise preservation, i.e. MWF- $N_{0.2}$. In contrast with the previous study, the number of microphones which was used by the MWF algorithms was varied. All MWF algorithms made use of the two microphones present on the ipsilateral hearing aid. In some conditions microphone signals from the contralateral hearing aid were added (M_C). The algorithms are denoted as MWF $_{2+M_C}$. For example, MWF $_{2+1}$ shown in Table 1 illustrates that this MWF algorithm used one additional microphone from the contralateral hearing aid.

Objective measurements, fully described in [6], were also performed to analyze the performance of the noise reduction algorithms. In these measurements, a speech intelligibility weighted improvement in SNR (ΔSNR_{SI}) was calculated for the different algorithms using different microphone combinations. SNR improvements were

calculated for both the left and the right hearing aid. The measured ΔSNR_{SI} are also added to Table 1.

Bilat/bin SRT [dB]	S_0N_{60}			S_0N_{270}			$S_0N_{90/180/270}$		
	perceptual	Left	Right	perceptual	Left	Right	perceptual	Left	Right
ADM	2.1 ± 1.9	2.7	2.8	$-4.3 \pm 1.3^*$	4.3	-3.2	1.3 ± 1.4	6.0	5.9
MWF ₂₊₂	$4.3 \pm 1.5^*$	4.9	9.6	0.7 ± 1.4	10.0	2.5	$4.6 \pm 0.8^*$	7.1	7.2
MWF ₂₊₁	$3.8 \pm 1.6^*$	4.0	6.2	0.3 ± 2.0	9.6	2.1	$4.0 \pm 1.5^*$	6.6	6.0
MWF ₂₊₀	$1.0 \pm 0.7^*$	1.9	3.3	-1.2 ± 1.6	3.8	1.0	$2.8 \pm 1.3^*$	5.1	4.9
MWF _{2+2-N_{0,2}}	$3.6 \pm 1.4^*$	3.3	5.4	$2.0 \pm 1.4^*$	4.3	1.9	$3.2 \pm 0.8^*$	4.1	4.2
MWF _{2+1-N_{0,2}}	$2.7 \pm 1.3^*$	2.6	3.0	1.5 ± 1.6	3.9	1.6	$3.4 \pm 0.8^*$	3.7	3.3
MWF _{2+0-N_{0,2}}	1.0 ± 2.1	1.1	0.9	0.0 ± 1.5	1.0	0.7	$2.3 \pm 1.4^*$	2.8	2.6
SNR-unproc[dB]	-6.2±1.8			-9.1±1.7			-7.2±1.6		

Table 1: The gain in SRT, ΔSRT_{algo} , averaged over ten normal hearing subjects. The bottom row show the SNR at which the unprocessed reference SRT was measured. A '*' depicts a significant noise reduction performance compared to the unprocessed condition. ΔSNR_{SI} , calculated for the left and right hearing aid in the objective evaluation, is also added to the table.

Several conclusions can be drawn from Table 1. First it is observed that the objective performance measure, ΔSNR_{SI} , gives a good indication of the perceptual outcome of the algorithm, ΔSRT_{algo} . However, ΔSNR_{SI} is a performance measure originating from monaural hearing aid research and therefore some of the effects of listening with two ears have to be taken into account. First, one has to take into account the best ear benefit. In an asymmetric spatial scenario, ΔSNR_{SI} will be different for both ears. ΔSRT_{algo} correlates only with the ΔSNR_{SI} measured at the ear which has the best input SNR. Typically, this is the ear with the lowest gain in SNR. E.g. in condition S_0N_{60} , the perceptual results correlate best with ΔSNR_{SI} obtained at the left ear which is the ear with the best input SNR. Second, ΔSNR_{SI} does not take into account spatial release from masking effects which explains some of the small differences between ΔSNR_{SI} and ΔSRT .

Second, it is observed that a two microphone MWF and ADM have approximately the same performance, except when the target signal is not arriving from the front direction, i.e. $S_{90}N_{90}$. Since the ADM is designed to preserve signals arriving from the front, it will start to remove both the noise and the target component in this spatial scenario.

Third, it is observed that feeding contralateral microphone signals into the noise reduction schemes significantly improves noise reduction performance. The largest gain is observed when the first contralateral microphone is added to the system. A lower gain is observed when adding the second contralateral microphone to the signal. This can be intuitively explained by the fact that the first contralateral microphone signal adds more new information to the noise reduction scheme in comparison with the second contralateral microphone signal.

Fourth, it is observed that the decrease in SRT performance, when replacing the MWF into an MWF-N algorithm is smaller than predicted by the objective SNR measurements. In condition $S_{90}N_{90}$, the MWF-N even outperforms the MWF algorithm. This may be due to the fact that the listener is able to localize both the target and the noise component correctly when using the MWF-N algorithm. This leads to an improved speech perception due to release from masking effects.

Conclusions

This study presented objective and perceptual measurements done with a binaural MWF and MWF-N noise reduction scheme. An unprocessed condition and a bilateral ADM were taken as a reference. First, it was shown that in contrast with an ADM, the MWF preserved the ability to localize the target signal independent of its angle of arrival. When using the MWF-N algorithm, both the target and the noise component were accurately localized. However, this comes at the cost of SNR improvement. Later it was shown that this decrease not necessarily implies a loss in speech perception due to spatial release from masking. Further it was shown that adding microphone signals of the contralateral hearing aid to a noise reduction scheme running on an ipsilateral hearing aid improves noise reduction performance.

References

- [1] Welker, D.P. et al. "Microphone-array hearing aids with binaural output-part II: A two-microphone adaptive system," IEEE Trans. Speech and Audio Processing **5** (1997), 543-551
- [2] Lotter, T., "Single and multimicrophone speech enhancement for hearing aids," Ph.D. thesis, RWTH Aachen (2004)
- [3] Keidser, G. et al. "The effect of multi-channel wide dynamic range compression, noise reduction, and the directional microphone on horizontal localization performance in hearing aid wearers," Int. J. Audiol. **45**, 563-579. (2006)
- [4] Van den Bogaert, T. et al. "Localization with bilateral hearing aids: without is better than with," J. Acoust. Soc. Am. **119**, 515-526. (2006)
- [5] Van den Bogaert, T. et al. "The effect of multi-microphone noise reduction systems on sound source localization in binaural hearing aids", Journal of the Acoustical Society of America, 124(1), pp. 484-497, (2008)
- [6] Van den Bogaert, T. et al. "Speech enhancement with multichannel Wiener filter techniques in binaural hearing aids", Journal of the Acoustical Society of America, 125(1), pp. 360-371. (2009)
- [7] Doclo, S. et al. "GSVD-based optimal filtering for single and multi-microphone speech enhancement," IEEE Trans. Signal Process. **50**, 2230-2244 (2002).
- [8] Doclo, S. et al. "Theoretical analysis of binaural cue preservation using multichannel Wiener filtering and interaural transfer functions," in Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC), Paris, France, pp. 1-4. (2006).
- [9] Haykin, S., Adaptive Filter Theory 4th ed., chapter 2 pp94-135. Prentice Hall, New Jersey, 2002