Reduced-bandwidth Multi-channel Wiener Filter based binaural noise reduction and localization cue preservation in binaural hearing aids

Bram Cornelis[1,2,*], *Student Member, IEEE*, Marc Moonen[1,2], *Fellow, IEEE*, and Jan Wouters[3]

[1]ESAT-SCD, Dept. Electr. Eng.,
[2]IBBT Future Health Dept.,
Katholieke Universiteit Leuven,
Kasteelpark Arenberg 10,
3001 Heverlee, Belgium
email: bram.cornelis@gmail.com,
marc.moonen@esat.kuleuven.be

[3]ExpORL, Dept. Neurosciences,
Katholieke Universiteit Leuven,
Herestraat 49/721,
3000 Leuven, Belgium
email: jan.wouters@med.kuleuven.be

Abstract

Binaural hearing aids allow for a wireless exchange of microphone signals between a left and a right device. As microphone signals from both devices then become available in a binaural noise reduction procedure, a significant noise reduction performance improvement can be achieved compared to a monaural configuration (a single device) or a bilateral configuration (in which the left and the right device work independently). In addition, the localization cues can also be better preserved in a binaural procedure, in particular the Interaural Time Differences (ITDs) and Interaural Level Differences (ILDs) by which the brain can localize sound sources in the horizontal plane. It was previously proven that a binaural noise reduction procedure based on the Speech Distortion Weighted Multi-channel Wiener Filter (SDW-MWF) indeed preserves the speech localization cues, if all microphone signals can be exchanged. However, in practice it may not be feasible to exchange all microphones signals between the

B. Cornelis is funded by a Ph.D. grant of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen).

This research work was carried out at the ESAT Laboratory of Katholieke Universiteit Leuven in the frame of the Belgian Programme on Interuniversity Attraction Poles initiated by the Belgian Federal Science Policy Office IUAP P6/04 (DYSCO, 'Dynamical systems, control and optimization', 2007-2011), Concerted Research Action GOA-MaNet, research project FWO nr. G.0600.08 ('Signal processing and network design for wireless acoustic sensor networks') and research project IBBT. The scientific responsibility is assumed by its authors.

devices, so that reduced-bandwidth SDW-MWF schemes (where only filtered combinations of microphone signals are exchanged) have to be utilized. In this paper, it is shown that a straightforward reduced-bandwidth SDW-MWF scheme still preserves the speech ITD cues, but distorts the speech ILD cues. Novel reduced-bandwidth SDW-MWF schemes, which make use of a common spectral postfilter, are therefore introduced. Experiments in a reverberant environment demonstrate that the novel schemes reduce the ILD distortion, without severely degrading the noise reduction performance.

Index Terms

Noise reduction, speech enhancement, microphone arrays, hearing aids, binaural hearing aids, localization cues, Multi-channel Wiener Filter, spectral postfilter

I. INTRODUCTION

Degraded speech understanding in noise is a frequent complaint of people suffering from sensorineural hearing loss [1]. Noise reduction in hearing aids has therefore been an active area of research for many years. Modern hearing aids are fitted with multiple microphones, so that spatial information can be utilized in addition to temporal and spectral information to reduce the noise [2].

A current trend [2] is to develop so-called binaural hearing aids, where a wireless link enables the exchange of parameters or even microphone signals between a left and a right device. If microphone signals from both sides of the head can be shared, a significant noise reduction performance improvement can be achieved compared to a monaural configuration (a single device) or a bilateral configuration (in which the left and the right device work independently). The localization cues of the target speech and residual noise can also be better preserved in a binaural procedure, in particular the Interaural Time Differences (ITDs) and Interaural Level Differences (ILDs) by which the brain can localize sound sources in the horizontal plane [3]–[5]. In addition to sound localization, the ITDs and ILDs also improve speech understanding in noise due to the so-called binaural unmasking effect, which leads to Speech Reception Threshold (SRT) improvements of up to 3 dB [6].

Many existing binaural noise reduction techniques apply identical real-valued spectral weights to one microphone signal on the left device and one microphone signal on the right device [7]–[11], so that the ITDs and ILDs are indeed preserved. Although the outputs of a beamformer can be utilized to derive the spectral weights, in essence these techniques can be viewed as spectral filtering approaches. It is therefore plausible that these techniques can offer only limited speech intelligibility improvements to hearing aid users, similarly to (single-microphone) spectral filtering approaches [12]. Another class of

binaural techniques [13], [14] combines the outputs of an (adaptive) beamformer with the unprocessed (lowpass-filtered) microphone signals. Although true beamforming is performed, there is necessarily a trade-off between noise reduction performance and localization cue preservation.

The binaural Speech Distortion Weighted Multi-channel Wiener Filter (SDW-MWF) [15] also performs true beamforming with the microphone signals and thus belongs to the second class of binaural techniques. If all microphone signals are exchanged between the devices, it can be proven that the binaural SDW-MWF preserves the speech localization cues [16], while the localization cues of the residual noise can also be preserved by including extensions to the SDW-MWF cost function [17]. Power consumption will however always be a limiting factor so that in practice, only a subset of microphone signals can be exchanged. Several reduced-bandwidth binaural SDW-MWF schemes, where only filtered combinations of microphone signals are exchanged, have therefore been proposed [18]. Although the previous work shows the potential Signal-to-Noise Ratio (SNR) improvements of the reduced-bandwidth schemes, the localization cue preservation by these approaches has not been studied.

The theoretical reduced-bandwidth framework of [18], [19] is utilized in this paper to study the speech localization cue preservation of reduced-bandwidth binaural SDW-MWF schemes. As in [19], it is shown that the reduced-bandwidth schemes do not necessarily preserve the speech localization cues. In particular, it is proven theoretically and demonstrated through experiments that the speech ITD cues are preserved by all reduced-bandwidth SDW-MWF schemes, while the speech ILD cues can be significantly distorted in certain scenarios. This even applies to a bilateral SDW-MWF, i.e. where no signals are exchanged and the devices work independently. Although a strong cue can dominate over a weaker cue in rivalry experiments, all cues should be kept consistent for optimal sound reproduction [5]: if the conflicts are too large, the resulting source image can be perceived to be diffuse or in the wrong location [4], [5]. The speech ILD distortion introduced by the reduced-bandwidth SDW-MWF schemes should therefore be reduced.

It is illustrated that by adapting the speech distortion parameter or by including a so-called partial noise estimation parameter, the ILD distortion can sometimes be reduced, but never completely eliminated. Two novel reduced-bandwidth schemes, which in principle completely eliminate the speech ILD distortion, are therefore introduced in this paper: a scheme for a low-bandwidth binaural link which does not allow for full audio-streaming, and a scheme where one microphone signal can be streamed in full-duplex. The schemes make use of a decomposed filter structure in which the SDW-MWF is structured as a Minimum Variance Distortionless Response (MVDR) beamformer followed by a spectral postfilter. It is proven theoretically and demonstrated through experiments that by using a common spectral postfilter at the left

and right devices, speech ILD distortion is effectively eliminated.

The paper is organized as follows. In Section II, the notation and general framework of the reduced-bandwidth schemes is given. A review of the binaural SDW-MWF and alternative (decomposed) expressions is given in Section III. In Section IV, the speech ILD distortion introduced by the reduced-bandwidth SDW-MWF schemes is analyzed theoretically, using the general framework of Section II. The novel reduced-bandwidth SDW-MWF schemes with common postfilter are presented in Section V. The new schemes are evaluated by experiments in a reverberant environment in Section VI. Finally, conclusions are given in Section VII.

II. Configuration and notation

A. Microphone signals and output signals

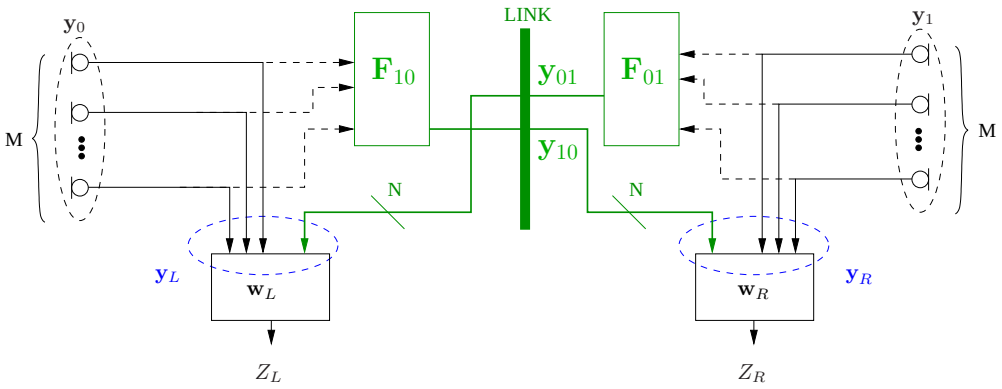


Fig. 1. General reduced-bandwidth binaural processing scheme

The general reduced-bandwidth binaural processing scheme proposed in [18], [19] is depicted in Figure 1. Both hearing aids have a microphone array consisting of M microphones[1]. The mth microphone signal in the left hearing aid $Y_{0,m}(\omega)$ can be specified in the frequency domain as

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

where $X_{0,m}(\omega)$ represents the speech component and $V_{0,m}(\omega)$ represents the noise component. Similarly, the mth microphone signal in the right hearing aid is equal to $Y_{1,m}(\omega) = X_{1,m}(\omega) + V_{1,m}(\omega)$. For conciseness, we will omit the frequency-domain variable ω from now on. We define the M-dimensional

[1]It is possible to use different array sizes as in [16], but here both arrays are assumed to have M microphones for the sake of simplicity.

stacked microphone signal vectors $\mathbf{y}_0$ and $\mathbf{y}_1$ and the $2M$-dimensional signal vector $\mathbf{y}$ as[2]

$$\mathbf{y}_0 = \begin{bmatrix} Y_{0,0} \\ \vdots \\ Y_{0,M-1} \end{bmatrix}, \quad \mathbf{y}_1 = \begin{bmatrix} Y_{1,0} \\ \vdots \\ Y_{1,M-1} \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \end{bmatrix} . \tag{2}$$

The signal vector $\mathbf{y}$ can be written as $\mathbf{y} = \mathbf{x} + \mathbf{v}$, where $\mathbf{x}$ and $\mathbf{v}$ are defined similarly as $\mathbf{y}$.

In the binaural processing scheme, signals are exchanged between the two devices over a wireless link. However, due to bandwidth limitations, not every microphone signal can be exchanged, so that each device does not have access to the full signal vector $\mathbf{y}$. In general, we assume that each device receives linear combinations of the contralateral microphone signals, i.e.

$$\mathbf{y}_{10} = \mathbf{F}_{10}^H \mathbf{y}_0 \ , \quad \mathbf{y}_{01} = \mathbf{F}_{01}^H \mathbf{y}_1 \ , \tag{3}$$

where $\mathbf{F}_{10}$ and $\mathbf{F}_{01}$ are $M \times N$ dimensional complex matrices with $1 \leq N < M$. The available signals at the left and right device[3] can then be defined as:

$$\mathbf{y}_L = \begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_{01} \end{bmatrix} = \left[\begin{array}{c:c} \mathbf{I}_M & \mathbf{0}_{M\times M} \\ \hdashline \mathbf{0}_{N\times M} & \mathbf{F}_{01}^H \end{array}\right] \mathbf{y} = \mathbf{Q}_L^H \mathbf{y} \ , \tag{4}$$

$$\mathbf{y}_R = \begin{bmatrix} \mathbf{y}_{10} \\ \mathbf{y}_1 \end{bmatrix} = \left[\begin{array}{c:c} \mathbf{F}_{10}^H & \mathbf{0}_{N\times M} \\ \hdashline \mathbf{0}_{M\times M} & \mathbf{I}_M \end{array}\right] \mathbf{y} = \mathbf{Q}_R^H \mathbf{y} \ , \tag{5}$$

where $\mathbf{Q}_L$ and $\mathbf{Q}_R$ are $2M \times (M + N)$ matrices which compress the signal vector $\mathbf{y}$ into the lower-dimensional $\mathbf{y}_L$ and $\mathbf{y}_R$. The speech components $\mathbf{x}_L$ and $\mathbf{x}_R$ and the noise components $\mathbf{v}_L$ and $\mathbf{v}_R$ are similarly defined. For the special case of a bilateral setup (i.e. no signals are exchanged), the same framework can be used, where then $N = 0$ and so $\mathbf{Q}_L$ and $\mathbf{Q}_R$ are then $2M \times M$ matrices, defined as:

$$\mathbf{Q}_L^H = \left[\begin{array}{c:c} \mathbf{I}_M & \mathbf{0}_{M\times M} \end{array}\right] , \tag{6}$$

$$\mathbf{Q}_R^H = \left[\begin{array}{c:c} \mathbf{0}_{M\times M} & \mathbf{I}_M \end{array}\right] . \tag{7}$$

The correlation matrix $\mathbf{R}_y$, the speech correlation matrix $\mathbf{R}_x$ and the noise correlation matrix $\mathbf{R}_v$ are defined as

$$\mathbf{R}_y = \mathcal{E}\{\mathbf{y}\mathbf{y}^H\}, \quad \mathbf{R}_x = \mathcal{E}\{\mathbf{x}\mathbf{x}^H\}, \quad \mathbf{R}_v = \mathcal{E}\{\mathbf{v}\mathbf{v}^H\} \ , \tag{8}$$

[2]Although the signal vectors contain complex-valued frequency-domain variables, they are denoted with lower-case letters throughout the paper to distinguish them from matrices.

[3]In contrast to [17], [18], a distinction is made here between the microphone signal vectors $\mathbf{y}_0$ and $\mathbf{y}_1$, and the signal vectors $\mathbf{y}_L$ and $\mathbf{y}_R$ which represent the available signals at the left and right device.

where $\mathcal{E}$ denotes the expected value operator. Assuming that the speech and the noise components are uncorrelated, $\mathbf{R}_y = \mathbf{R}_x + \mathbf{R}_v$. By using definitions (4) and (5), the left and right correlation matrices (i.e. the correlation matrices estimated at the left and right device) can be defined as

$$\mathbf{R}_{yL} = \mathbf{Q}_L^H \mathbf{R}_y \mathbf{Q}_L, \quad \mathbf{R}_{yR} = \mathbf{Q}_R^H \mathbf{R}_y \mathbf{Q}_R \, . \tag{9}$$

The speech correlation matrices $\mathbf{R}_{xL}$ and $\mathbf{R}_{xR}$ and the noise correlation matrices $\mathbf{R}_{vL}$ and $\mathbf{R}_{vR}$ can be similarly defined.

One signal of the left device and one signal of the right device are the so-called reference signals for the noise reduction algorithms. The reference signals at the left and the right device are denoted as $Y_{L,\mathrm{ref}}$ and $Y_{R,\mathrm{ref}}$, which are then equal to

$$Y_{L,\mathrm{ref}} = \mathbf{e}_L^H \mathbf{y}_L, \quad Y_{R,\mathrm{ref}} = \mathbf{e}_R^H \mathbf{y}_R \, , \tag{10}$$

where $\mathbf{e}_L$ and $\mathbf{e}_R$ are $(M+N)$-dimensional vectors with only one element equal to 1 and the other elements equal to 0. Typically, the front microphone signals are used as reference signals, which corresponds to $\mathbf{e}_L(0) = \mathbf{e}_R(N) = 1$ (assuming zero-based indexing). The reference signals can also be written as $Y_{L,\mathrm{ref}} = X_{L,\mathrm{ref}} + V_{L,\mathrm{ref}}$ and $Y_{R,\mathrm{ref}} = X_{R,\mathrm{ref}} + V_{R,\mathrm{ref}}$.

The output signals Z_L and Z_R at the left and the right device are obtained by filtering and summing the left and right signal vectors, i.e.

$$Z_L = \mathbf{w}_L^H \mathbf{y}_L, \quad Z_R = \mathbf{w}_R^H \mathbf{y}_R \, , \tag{11}$$

where $\mathbf{w}_L$ and $\mathbf{w}_R$ are $(M+N)$-dimensional complex weight vectors. The output signal at the left hearing aid can also be written as $Z_L = Z_{xL} + Z_{vL} = \mathbf{w}_L^H \mathbf{x}_L + \mathbf{w}_L^H \mathbf{v}_L$, where Z_{xL} represents the speech component and Z_{vL} represents the noise component of the output signal. Similarly, the output signal at the right hearing aid can be written as $Z_R = Z_{xR} + Z_{vR} = \mathbf{w}_R^H \mathbf{x}_R + \mathbf{w}_R^H \mathbf{v}_R$.

B. Special case: single speech source

In the case of a single speech source, the speech signal vector can be modelled as

$$\mathbf{x} = \mathbf{a} S \, , \tag{12}$$

where the $2M$-dimensional steering vector $\mathbf{a}$ contains the acoustic transfer functions from the speech source to the microphones (including room acoustics, microphone characteristics and head shadow effect) and S denotes the speech signal. The vectors $\mathbf{a}_0$, $\mathbf{a}_1$ and $\mathbf{a}$ can be defined in a similar manner as $\mathbf{y}_0$, $\mathbf{y}_1$ and $\mathbf{y}$ in (2). As in (4) and (5), the vectors $\mathbf{a}_L$ and $\mathbf{a}_R$ can be defined as $\mathbf{a}_L = \mathbf{Q}_L^H \mathbf{a}$ and $\mathbf{a}_R = \mathbf{Q}_R^H \mathbf{a}$.

With assumption (12), the speech correlation matrix is a rank-1 matrix, i.e.

$$\mathbf{R}_x = P_s \mathbf{a}\mathbf{a}^H \ , \tag{13}$$

with $P_s = \mathcal{E}\{|S|^2\}$ the power of the speech signal. The speech correlation matrices $\mathbf{R}_{xL}$ and $\mathbf{R}_{xR}$ (i. e. the correlation matrices at the left and right devices) can be similarly written as $\mathbf{R}_{xL} = P_s \mathbf{a}_L \mathbf{a}_L^H$ and $\mathbf{R}_{xR} = P_s \mathbf{a}_R \mathbf{a}_R^H$.

The reference microphone signals at the left and the right hearing aid can be written as

$$Y_{L,\mathrm{ref}} = A_{L,\mathrm{ref}} S + V_{L,\mathrm{ref}}, \quad Y_{R,\mathrm{ref}} = A_{R,\mathrm{ref}} S + V_{R,\mathrm{ref}} \ , \tag{14}$$

with $A_{L,\mathrm{ref}}$ the reference element of $\mathbf{a}_L$ (i.e. $A_{L,\mathrm{ref}} = \mathbf{e}_L^H \mathbf{a}_L$) and $A_{R,\mathrm{ref}}$ the reference element of $\mathbf{a}_R$ (i.e. $A_{R,\mathrm{ref}} = \mathbf{e}_R^H \mathbf{a}_R$).

C. Theoretical performance measures

For the theoretical analysis, the theoretical performance measures proposed in [17] are extended for the general reduced-bandwidth binaural processing scheme of Section II-A. The *local input SNR* (i.e. per frequency bin) is defined as the power ratio of the speech and noise component in the reference signals, i.e.

$$\mathrm{SNR}_L^{in} = \frac{\mathcal{E}\{|X_{L,\mathrm{ref}}|^2\}}{\mathcal{E}\{|V_{L,\mathrm{ref}}|^2\}} = \frac{\mathbf{e}_L^H \mathbf{R}_{xL} \mathbf{e}_L}{\mathbf{e}_L^H \mathbf{R}_{vL} \mathbf{e}_L} \ , \tag{15}$$

and similarly for SNR_R^{in}. The *local output SNR* is defined as the power ratio of the speech and noise component in the output signals, i.e.

$$\mathrm{SNR}_L^{out} = \frac{\mathcal{E}\{|Z_{xL}|^2\}}{\mathcal{E}\{|Z_{vL}|^2\}} = \frac{\mathbf{w}_L^H \mathbf{R}_{xL} \mathbf{w}_L}{\mathbf{w}_L^H \mathbf{R}_{vL} \mathbf{w}_L} \ , \tag{16}$$

and similarly for SNR_R^{out}.

The *input and output Interaural Transfer Functions (ITFs)* of the speech component are defined as the ratio of the speech component at the left and right device, i.e.[4]

$$\mathrm{ITF}_x^{in} = \frac{X_{L,\mathrm{ref}}}{X_{R,\mathrm{ref}}} = \frac{\mathbf{e}_L^H \mathbf{x}_L}{\mathbf{e}_R^H \mathbf{x}_R} = \frac{\mathbf{e}_L^H \mathbf{R}_{xLR} \mathbf{e}_R}{\mathbf{e}_R^H \mathbf{R}_{xR} \mathbf{e}_R} = \frac{\mathbf{e}_L^H \mathbf{R}_{xL} \mathbf{e}_L}{\mathbf{e}_R^H \mathbf{R}_{xRL} \mathbf{e}_L} \ , \tag{17}$$

$$\mathrm{ITF}_x^{out} = \frac{Z_{xL}}{Z_{xR}} = \frac{\mathbf{w}_L^H \mathbf{x}_L}{\mathbf{w}_R^H \mathbf{x}_R} = \frac{\mathbf{w}_L^H \mathbf{R}_{xLR} \mathbf{w}_R}{\mathbf{w}_R^H \mathbf{R}_{xR} \mathbf{w}_R} = \frac{\mathbf{w}_L^H \mathbf{R}_{xL} \mathbf{w}_L}{\mathbf{w}_R^H \mathbf{R}_{xRL} \mathbf{w}_L} \ , \tag{18}$$

where $\mathbf{R}_{xLR} = \mathbf{Q}_L^H \mathbf{R}_x \mathbf{Q}_R$ and $\mathbf{R}_{xRL} = \mathbf{R}_{xLR}^H = \mathbf{Q}_R^H \mathbf{R}_x \mathbf{Q}_L$. These input and output ITFs are complex valued scalars, of which the amplitude and phase can be defined as the (square root of the) *Interaural Level Differences (ILDs)* and *Interaural Time Differences (ITDs)*, cfr. [17].

[4]In the following formulas it is implicitly assumed that there is a single speech source in free field as in Section II-B.

III. Review of binaural Multi-channel Wiener Filter

The Speech Distortion Weighted Multi-channel Wiener Filter (SDW-MWF) minimizes a weighted sum of the mean-squared residual noise and speech distortion [15], [20]. The optimal SDW-MWF[5] for the left device is equal to

$$\mathbf{w}_{\mathrm{MWF},L} = (\mathbf{R}_{xL} + \mu \mathbf{R}_{vL})^{-1} \mathbf{R}_{xL} \mathbf{e}_L \ , \tag{19}$$

where the speech distortion parameter μ provides a trade-off between noise reduction and speech distortion. The optimal SDW-MWF for the right device is similarly defined.

For a single speech source, it can be shown [21]–[25] that equivalent rank-one SDW-MWF expressions can be derived, which still only make use of the speech and noise correlation matrices as in (19). In this paper, we make use of the Spatial Prediction SDW-MWF (SP-MWF), which was discussed in [23], [24] and originally proposed as the Distortionless Multi-channel Wiener Filter in [25]. The SP-MWF for the left device is equal to

$$\mathbf{w}_{\mathrm{SP-MWF},L} = \mathbf{R}_{vL}^{-1} \mathbf{R}_{xL} \mathbf{e}_L \ . \ \frac{\mathbf{e}_L^H \mathbf{R}_{xL} \mathbf{e}_L}{\mu \ \mathbf{e}_L^H \mathbf{R}_{xL} \mathbf{e}_L + \mathrm{Tr}\{\mathbf{R}_{vL}^{-1} \mathbf{R}_{xL} \mathbf{e}_L \mathbf{e}_L^H \mathbf{R}_{xL}\}} \ , \tag{20}$$

where $\mathrm{Tr}\{.\}$ is the Trace operator. The SP-MWF for the right device is similarly defined.

In [15], [17], the binaural SDW-MWF cost function is extended by means of a so-called partial noise estimation parameter η in order to preserve the residual noise localization cues. The optimal so-called MWF-η filter for the left device is equal to [17]:

$$\mathbf{w}_{\mathrm{MWF}\eta,L} = (1-\eta) \ \mathbf{w}_{\mathrm{MWF},L} + \eta \ \mathbf{e}_L \ , \tag{21}$$

the MWF-η filter for the right device is similarly obtained. The SP-MWF (20) can also be extended in this manner.

As shown in [16], [24], in the case of a single speech source (13), the filters (19) and (20) are equivalent and equal to

$$\mathbf{w}_{\mathrm{MWF},L} = \mathbf{R}_{vL}^{-1} \mathbf{a}_L \ \frac{P_s}{\mu + \rho_L} A_{L,\mathrm{ref}}^* \ , \tag{22}$$

where ρ_L is equal to the local output SNR (16), i.e.

$$\rho_L = P_s \mathbf{a}_L^H \mathbf{R}_{vL}^{-1} \mathbf{a}_L = \mathrm{SNR}_L^{out} \ . \tag{23}$$

$\mathbf{w}_{\mathrm{MWF},R}$ and ρ_R are similarly defined. As this paper focuses on the speech localization cue preservation for a single target speech source, a single analysis using (22) can be made for all filter expressions.

[5]For conciseness, *SDW-MWF* is abbreviated to *MWF* in the formulas in this paper.

As shown in [26], [27], the filter (22) can be decomposed into a spatial filter part, which is equivalent to the Minimum Variance Distortionless Response (MVDR) beamformer [25], [28] and the Transfer Function Ratio Generalized Sidelobe Canceller (TF-GSC) [29], followed by a (single-channel) spectral postfilter part, i.e.

$$\mathbf{w}_{\mathrm{MWF},L} = \underbrace{\mathbf{R}_{vL}^{-1}\mathbf{a}_L \, \frac{A_{L,\mathrm{ref}}^*}{\mathbf{a}_L^H\mathbf{R}_{vL}^{-1}\mathbf{a}_L}}_{\mathrm{MVDR}} \cdot \underbrace{\frac{\rho_L}{\mu+\rho_L}}_{\mathrm{postfilter}} , \tag{24}$$

and similarly for $\mathbf{w}_{\mathrm{MWF},R}$. The speech distortion parameter μ only appears in the spectral postfilter and fulfills the same role as in single-channel constrained Wiener filters [30], [31] or spectral oversubtraction [30], [32]. The decomposed structure of (24) is conceptually interesting, but requires explicit estimation (or prior knowledge) of the speech power P_s and of the steering vector $\mathbf{a}_L$. The SP-MWF (20) is also explicitly decomposed into a spatial filter and spectral postfilter (see also Section V-A), but only requires a relatively easy estimation of the speech and noise correlation matrices. The decomposed structure of the SP-MWF allows control over the spectral postfilter. This property will be exploited in Section V to derive novel reduced-bandwidth SDW-MWF schemes, which better preserve the speech ILD cues.

IV. Speech localization cue preservation by reduced bandwidth SDW-MWF and MWF-η

A. Output speech localization cues

By using the general framework of Section II-A, and by assuming a single speech source as in Section II-B, the output speech ITF, ILD and ITD can be calculated for any $\mathbf{Q}_L$ and $\mathbf{Q}_R$. By plugging the optimal filter (22) into the ITF definitions (17)-(18), and using definition (23), we obtain:

$$\mathrm{ITF}_x^{out} = \frac{\frac{P_s}{\mu+\rho_L}\,\mathbf{a}_L^H\mathbf{R}_{vL}^{-1}\mathbf{a}_L\;A_{L,\mathrm{ref}}}{\frac{P_s}{\mu+\rho_R}\,\mathbf{a}_R^H\mathbf{R}_{vR}^{-1}\mathbf{a}_R\;A_{R,\mathrm{ref}}} = \frac{1+\frac{\mu}{\rho_R}}{1+\frac{\mu}{\rho_L}}\;\mathrm{ITF}_x^{in} , \tag{25}$$

$$\mathrm{ILD}_x^{out} = \frac{(1+\frac{\mu}{\rho_R})^2}{(1+\frac{\mu}{\rho_L})^2}\;\mathrm{ILD}_x^{in} , \tag{26}$$

$$\mathrm{ITD}_x^{out} = \angle\left(\frac{1+\frac{\mu}{\rho_R}}{1+\frac{\mu}{\rho_L}}\;\mathrm{ITF}_x^{in}\right) = \mathrm{ITD}_x^{in} . \tag{27}$$

It is seen from (27) that the speech ITD cues are never distorted by the reduced-bandwidth SDW-MWF, regardless of which signals are exchanged over the wireless link, or even for the bilateral case where no signals are exchanged (6), (7). However, (26) shows that the output speech ILD is scaled compared to the input speech ILD, so that the speech ILD cues are distorted by the reduced-bandwidth SDW-MWF.

For the special case $\rho_L = \rho_R$, which means that the obtained left and right local output SNRs are equal, (26) shows that the speech ILD cues are also undistorted. This is the case if all microphone signals can be exchanged over the wireless link, i.e. $\mathbf{Q}_L = \mathbf{Q}_R = \mathbf{I}_{2M}$. This result was also obtained in [16], [17].

However, if not all microphone signals are exchanged ($N < M$), then in general $\rho_L \neq \rho_R$, leading to ILD distortions. By increasing the number of exchanged signals, the local output SNRs ρ_L and ρ_R generally increase while the mismatch $|\rho_L - \rho_R|$ decreases [18]. As a result, the ILD distortion will also decrease (cfr. Appendix A). It can therefore be concluded that, in addition to leading to better noise reduction performance, a configuration using a binaural link will generally introduce lower speech ILD distortion compared to a bilateral configuration.

B. Influence of μ

As can be seen from (26), $\text{ILD}_x^{out} = \text{ILD}_x^{in}$ for $\mu = 0$. From (24) it is seen that this parameter choice corresponds to an MVDR where the speech distortion is enforced to 0, so that the speech localization cues will indeed not be distorted. For $\mu > 0$, the reduced-bandwidth SDW-MWF performs additional spectral postfiltering, so that the residual noise is further attenuated. Although the local output SNRs ρ_L and ρ_R are independent of μ, the global output SNRs can then be increased as in single-channel spectral subtraction algorithms [30], thereby increasing listening comfort. It can be shown (cfr. Appendix A) that the speech ILD error strictly increases as μ increases. Improving the noise reduction performance therefore comes at a cost: as more emphasis is put on noise reduction, the speech ILD cues will be more distorted.

In previous perceptual evaluations of the (full-bandwidth) binaural SDW-MWF [33], [34], μ was always fixed to a constant value (i.e. $\mu = 5$), in each time-frequency point. As explained, this parameter choice leads to speech ILD distortion in a reduced-bandwidth scheme. The ILD distortion can however be reduced by decreasing μ in time-frequency points with high speech power, while larger values of μ are kept in time-frequency points with low speech power (so that the noise is still sufficiently attenuated). By making μ dependent on the estimated segmental SNR (as in [32]), so that the spectral postfilter effectively becomes similar to classical single-channel spectral subtraction, the speech ILD distortion could thus be reduced.

C. Influence of η

By plugging (21)-(22) into the ITF definitions (17)-(18), we again find that the speech ITD is undistorted by the MWF-η, while the output speech ILD is equal to

$$\mathrm{ILD}_x^{out} = \frac{(1+\frac{\mu}{\rho_R})^2\ (1+\frac{\eta\mu}{\rho_L})^2}{(1+\frac{\mu}{\rho_L})^2\ (1+\frac{\eta\mu}{\rho_R})^2}\ \mathrm{ILD}_x^{in}\ . \tag{28}$$

For $\eta = 1$, the MWF-η output signals are equal to the reference microphone signals so that $\mathrm{ILD}_x^{out} = \mathrm{ILD}_x^{in}$, whereas for $\eta = 0$, (26) is obviously obtained. In Appendix A, it is shown that for intermediate values of η, the ILD error strictly decreases if η increases.

In addition to better preserving the residual noise localization cues, the MWF-η thus also reduces the speech ILD error, at the cost of a lower noise reduction performance. In [33], [34], a (small) value ($\eta = 0.2$) was proposed as a good parameter setting: the residual noise ILD and ITD errors are small, while the noise reduction performance is still adequate. For this setting, the speech ILD error can however not be eliminated, as this would only be achieved for the limit case $\eta = 1$, where no noise reduction is performed. Furthermore, if there are no clear noise ITD/ILD cues (for example in a diffuse noise field), it would also not be appropriate to use the MWF-η solution. The MWF-η by itself is therefore not a sufficient solution for eliminating the speech ILD distortion.

V. Reduced-bandwidth SDW-MWF with Common Postfilter

A. Introduction

The previous analysis reveals that the speech ILD distortion is introduced by the spectral postfilter of the SDW-MWF, which depends on the speech distortion parameter μ (24). The ILD distortion can be reduced by adapting μ based on the segmental SNR and/or by using a partial noise estimation extension (MWF-η), but with these approaches it is not possible to completely eliminate the distortion. Moreover, the noise reduction performance may be affected by these approaches, as μ may be erroneously adapted when the segmental SNR is poorly estimated, or because some of the unprocessed (low SNR) signal is mixed with the MWF output signal by the MWF-η.

An alternative solution, where the speech ILD distortion can be completely eliminated without significantly compromising the noise reduction performance, is to use the same spectral postfilter at the left and right device, which is then applied to the outputs of the MVDR-filtered left and right output signals. This approach is somewhat similar to [7]–[11], where a common spectral postfilter is applied to the raw microphone signals of the left and right device. The difference is that in our proposed approach, an MVDR beamforming, which utilizes all the ipsilateral microphone signals and possibly also contralateral

signal(s), is performed at both devices separately in a first stage. This MVDR stage is then followed by the common spectral postfiltering stage. In [7]–[11], only a spectral filtering of one microphone signal in the left device and one microphone signal in the right device is performed, whereas no beamforming is performed in the direct signal path.

Two Common Postfilter (CP) schemes, i.e. schemes where the same spectral postfilter part (24) is used in the left and right device, will be proposed in the following subsections. As previously mentioned, (24) does not allow for an easy implementation, while the standard SDW-MWF (19) does not allow for manipulating the postfilter part. The SP-MWF (20) provides a good alternative as it only relies on the speech and noise correlation matrices as in (19), but is also explicitly structured as a spatial filter followed by a spectral postfilter as in (24).

The SP-MWF (20) can be structured as an MVDR (by setting $\mu = 0$) followed by a postfilter depending on μ as in (24), which may then be replaced by a CP in the left and right device, leading to

$$\mathbf{w}_{\mathrm{SP-MWF-CP},L} = \mathbf{R}_{vL}^{-1}\mathbf{R}_{xL}\mathbf{e}_L \cdot \frac{\mathbf{e}_L^H\mathbf{R}_{xL}\mathbf{e}_L}{\mathrm{Tr}\{\mathbf{R}_{vL}^{-1}\mathbf{R}_{xL}\mathbf{e}_L\mathbf{e}_L^H\mathbf{R}_{xL}\}} \cdot \mathrm{CP}\ (\mu)\,, \tag{29}$$

$$\mathbf{w}_{\mathrm{SP-MWF-CP},R} = \underbrace{\mathbf{R}_{vR}^{-1}\mathbf{R}_{xR}\mathbf{e}_R \cdot \frac{\mathbf{e}_R^H\mathbf{R}_{xR}\mathbf{e}_R}{\mathrm{Tr}\{\mathbf{R}_{vR}^{-1}\mathbf{R}_{xR}\mathbf{e}_R\mathbf{e}_R^H\mathbf{R}_{xR}\}}}_{\mathrm{MVDR}} \cdot \mathrm{CP}\ (\mu)\,, \tag{30}$$

We propose two schemes for constructing the CP, namely a scheme for a low-bandwidth binaural link which does not allow to exchange microphone signals between the devices (bilateral MWF with CP), and a scheme where one microphone signal can be exchanged between the devices (binaural MWF-front with CP).

B. Bilateral MWF with Common Postfilter

If the binaural link does not allow to exchange microphone signals between the devices [cfr. (6) and (7)], it can still be utilized to derive a CP for the two devices. We consider the scheme of Figure 2, where only the ipsilateral microphone signals are available for bilateral MVDR filtering (MWF with $\mu = 0$), but where an estimate of the local output SNRs is exchanged between the devices. The CP is then derived in each frequency bin as:

$$\mathrm{CP}\ (\mu) = \frac{\max\{\widehat{\mathrm{SNR}}_L^{out},\ \widehat{\mathrm{SNR}}_R^{out}\}}{\mu + \max\{\widehat{\mathrm{SNR}}_L^{out},\ \widehat{\mathrm{SNR}}_R^{out}\}}\,, \tag{31}$$

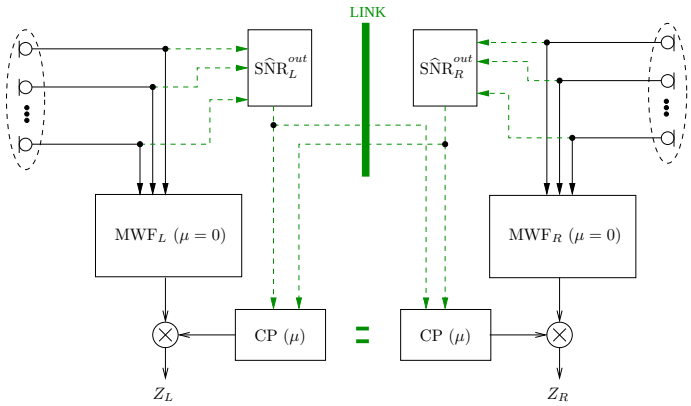


Fig. 2. Bilateral MWF with Common Postfilter (CP)

where the estimation of $\widehat{\mathrm{SNR}}_L^{out}$ and $\widehat{\mathrm{SNR}}_R^{out}$ is a by-product of the MVDR in (29) and (30), i.e.

$$\widehat{\mathrm{SNR}}_L^{out} = \frac{\mathrm{Tr}\{\mathbf{R}_{vL}^{-1}\mathbf{R}_{xL}\mathbf{e}_L\mathbf{e}_L^H\mathbf{R}_{xL}\}}{\mathbf{e}_L^H\mathbf{R}_{xL}\mathbf{e}_L}, \tag{32}$$

$$\widehat{\mathrm{SNR}}_R^{out} = \frac{\mathrm{Tr}\{\mathbf{R}_{vR}^{-1}\mathbf{R}_{xR}\mathbf{e}_R\mathbf{e}_R^H\mathbf{R}_{xR}\}}{\mathbf{e}_R^H\mathbf{R}_{xR}\mathbf{e}_R}. \tag{33}$$

By selecting the maximum output SNR estimate in (31), the performance at the best-ear side (which is crucial for intelligibility) is not compromised, whereas the noise will be less attenuated at the worst-ear side (compared to the case where each side has its own postfilter).

The required bandwidth for the bilateral MWF with CP can be kept low in a number of ways:

- The CP is only enforced at a subset of frequencies, for example only at frequencies above 1000 Hz (as ILD cues are then the dominant cues [3]).
- The output SNRs are exchanged and updated at a rate lower than the frame rate, i.e. the same SNR value is used in a number of subsequent frames.
- The frequency bins are grouped in bands, where one average output SNR estimate is used for an entire band, thus reducing the number of exchanged values.
- If only simplex instead of duplex transmission is feasible, a CP can still be obtained if the receiving device only uses the contralateral SNR estimates instead of its own SNR estimates to compute its postfilter. The approach can be improved by monitoring which is the best-ear side (for example based on simple global SNR estimates), so that the device at the best-ear side can be chosen as transmitter.

C. Binaural MWF-front with Common Postfilter

In Figure 3, a binaural reduced-bandwidth MWF algorithm with CP is depicted. In this scheme, the front microphone signal of each device is transmitted to the other device (in full-duplex). This was referred to as MWF-front in [18], and corresponds to

$$\mathbf{F}_{10}^H = \mathbf{F}_{01}^H = \Big[1 \underbrace{0 \dots 0}_{M-1}\Big] \tag{34}$$

in the general framework of (4) and (5). Each device can thus construct an $(M+1)$-microphone MVDR using its own M microphone signals and the front microphone signal of the other device.

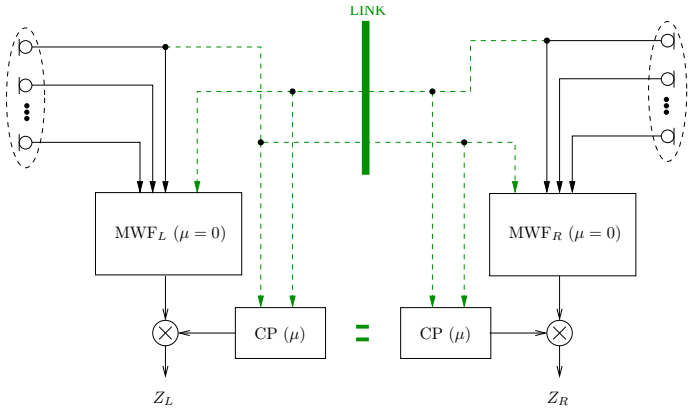


Fig. 3. Binaural MWF-front with Common Postfilter (CP)

As both devices have access to both front microphone signals, each device can additionally compute a two-microphone MVDR with these two front microphone signals (whereby each device selects the same best-ear side microphone signal as reference). The output SNR of this two-microphone MVDR can then be computed in both devices as:

$$\widehat{\mathrm{SNR}}_{front}^{out} = \frac{\mathrm{Tr}\{\mathbf{R}_{v,front}^{-1}\mathbf{R}_{x,front}\mathbf{e}_{front}\mathbf{e}_{front}^H\mathbf{R}_{x,front}\}}{\mathbf{e}_{front}^H\mathbf{R}_{x,front}\mathbf{e}_{front}}\,, \tag{35}$$

where $\mathbf{e}_{front}$ selects the best-ear side front microphone signal as reference, and where $\mathbf{R}_{v,front}$ and $\mathbf{R}_{x,front}$ are defined as:

$$\mathbf{R}_{v,front} = \mathcal{E}\{\begin{bmatrix} V_{0,0} \\ V_{1,0} \end{bmatrix} \cdot \begin{bmatrix} V_{0,0}^H & V_{1,0}^H \end{bmatrix}\}, \quad \mathbf{R}_{x,front} = \mathcal{E}\{\begin{bmatrix} X_{0,0} \\ X_{1,0} \end{bmatrix} \cdot \begin{bmatrix} X_{0,0}^H & X_{1,0}^H \end{bmatrix}\}\,. \tag{36}$$

The CP is then calculated in each device as

$$\mathrm{CP}\ (\mu) = \frac{\widehat{\mathrm{SNR}}_{front}^{out}}{\mu + \widehat{\mathrm{SNR}}_{front}^{out}}\,. \tag{37}$$

By making use of the QRD-RLS based SDW-MWF implementation proposed in [35], it is not even required to calculate a separate two-microphone SNR estimate as in (35) to obtain the CP. Namely, in Appendix B, it is shown that the CP can be extracted from the $(M+1)$-microphone MVDR solutions with only a few simple operations.

It is noted that the required bandwidth can be further reduced if only the low frequency part (e.g. up to 4000 Hz) of the front microphone signals is exchanged as in [36]. Although the bilateral array achieves sufficient noise reduction at the frequencies above 4000 Hz [36], it can still introduce a significant speech ILD distortion, which is moreover the dominant localization cue at these frequencies [3]. At the frequencies above 4000 Hz, the low-bandwidth bilateral MWF with CP of Section V-B should therefore be used.

VI. Experimental results

A. ILD estimation

In [17], [19], the speech input and output ILDs were measured using theoretical performance measures based on the speech and noise correlation matrices. These measures quantify the long-term average ILD per frequency bin. As the effect of different spectral postfilters and their introduced frame-by-frame ILD fluctuations are now also under investigation, a short-term per-frame ILD estimate is needed. Unfortunately, such a per-frame ILD estimate is unreliable, especially in the presence of reverberation [37]. As suggested in [37], [38], a solution for reducing the estimation variance is to only select the frames with an interaural coherence above a certain threshold, which is justified by the precedence effect [3]. The determination of the optimal threshold is non-trivial as it depends on both the frequency and the direction-of-arrival of the source signal [38]. It was however illustrated in [38] that a good estimate for the threshold can be obtained by fixing it to the 90th percentile of the interaural coherence values, and this procedure will therefore be utilized in this paper.

In addition to the per-frame ILD calculation and selection, a model of the middle and inner ear is also incorporated so that the physical ILD levels are converted to physiological levels. As in [38], the input and output signals are resampled to 16 kHz, divided into frames of 20 ms and decomposed into 24 critical bands using a Gammatone cochlear filterbank [39] with center frequencies according to the Glasberg and Moore model [40]. As in [37], each critical band signal is subsequently processed using the envelope compression model of [41], in order to perform a neural transduction. The different models were implemented using the downloadable binaural toolboxes from Slaney [42] and Akeroyd [43]. The

ILD (in dB) per frame is then equal to [37], [38]:

$$\text{ILD } [i,k] = 10 \ . \ \log_{10}\left(\frac{E_L[i,k]}{E_R[i,k]}\right) \ , \tag{38}$$

where i is the critical band index, k is the frame index, and where $E_L[i,k]$ and $E_R[i,k]$ are the smoothed energies of the left and right compressed critical band signals [37]. By this definition, a source arriving from the left side of the head obtains a positive ILD (in dB).

B. Setup and stimuli

We consider a binaural hearing aid configuration with two behind-the-ear devices connected by a wireless link. Each device has two omnidirectional microphones ($M = 2$), with an intermicrophone distance of approximately 1 cm.

Head-related transfer functions (HRTFs) were measured in a reverberant room ($\text{RT}_{60} = 0.61s$ [33], [34]) on a CORTEX MK2 manikin, so that the head-shadow effect is taken into account. Four spatial scenarios are considered, i.e. S0N60, S0N120, S0 N[60-120-180-210] and S90N270, where the target speech (S) and interfering noise (N) source(s) are positioned at the specified azimuthal angles (with 0° in front of the head, 90° to the right of the head). To generate the microphone signals, the noise and speech signals were convolved with the HRTFs corresponding to their angles of arrival, before being added together.

In all experiments, five consecutive sentences (with five-second periods of silence between sentences) of the English Hearing-In-Noise Test (HINT) [44] were used as speech stimulus. Multitalker babble (Auditec [45]) is used as noise stimulus. In the scenario with multiple noise sources, different time-shifted versions of this signal were generated to obtain uncorrelated noise sources. The signals were scaled so that the input SNR is 0 dB in the reference microphone signal of the best ear (left ear in S0N60, S0N120 and S0 N[60-120-180-210], right ear in S90 N270).

To assess the performance (both SNR improvement and speech ILD distortion), only the data frames corresponding to the last three sentences are selected, to allow the filters to converge.

C. Considered algorithms

Nine different algorithms are considered and a description of each can be found in Table I. The algorithms denoted with *BIL* (=bilateral) assume that exchanging microphone signals is not possible, so that each device can only use its own two microphone signals for noise reduction. The algorithms denoted with *FRONT* assume that the front microphone signal can be exchanged, so that three microphone

Notation	Description
BIL$-\mu$=5	Bilateral MWF ($M = 2$), fixed μ, eq. (6)-(7), (20)
BIL$-\mu$=5$-$CP	Bil. MWF with common postfilter, fixed μ, Section V-B
BIL$-\mu$=adapt	Bil. MWF, adaptive μ, eq. (6)-(7), (20)
BIL$-\mu$=$5-\eta$	Bil. MWF, fixed μ, partial noise estimation, eq. (6)-(7), (20), (21)
FRONT$-\mu$=5	Binaural MWF-front (exchange front mic. signals), fixed μ, eq. (34), (20)
FRONT$-\mu$=5$-$CP	Bin. MWF-front with common postfilter, fixed μ, Section V-C and Appendix B
FRONT$-\mu$=adapt	Bin. MWF-front, adaptive μ, eq. (34), (20)
FRONT$-\mu$=$5-\eta$	Bin. MWF-front, fixed μ, partial noise estimation, eq. (34), (20), (21)
ALL$-\mu$=5	Bin. MWF (all mic. signals available, $N = M$), [17]

TABLE I

CONSIDERED ALGORITHMS

signals are available for noise reduction in each device. The algorithm denoted with *ALL* is the reference condition where all microphone signals are available in each device, as in [17]. The speech distortion parameter is always fixed to $\mu = 5$, except for two algorithmic conditions where μ is adapted according to the segmental SNR as in [32]. To improve the performance for colored noise, the segmental SNR is estimated in different frequency-bands as in [46], so that a different μ can be used in different bands. The implementation of this multi-band approach was based on the code provided with [30]. The frequency spectrum was hereby divided in eight equal bands (bandwidth of 1250 Hz) as in [46]. Furthermore, in two other algorithmic conditions, a partial noise estimation is performed as in (21), where the partial noise parameter is set to $\eta = 0.2$ as in [33], [34]. Finally, in two algorithmic conditions the proposed Common Postfilter (CP) schemes of Section V-B (for bilateral MWF) and Section V-C (for MWF-front) are tested. For the bilateral CP, we keep in mind that the required bandwidth can be reduced as explained in Section V-B, although these alternatives are not tested in these experiments. To derive the MWF-front CP, the efficient scheme described in Appendix B is used.

In all experiments, the signals are sampled at $f_s = 20480$ Hz and the filter length (=DFT size) is set to $L = 128$. The algorithms are implemented in a Weighted Overlap-Add (WOLA) filterbank [47], whereby the microphone signals are segmented into frames of L samples with 50% overlap, and windowed by a Hann window.

D. Batch versus adaptive implementation

In [17], [19], a batch implementation of the SDW-MWF was evaluated. In the batch implementation, the correlation matrices are estimated in an off-line procedure using the complete microphone signals,

which results in an optimal average filter. It is however possible to recursively update the correlation matrices and filters in an online procedure, which allows tracking a moderately changing spatial scenario [20]. In this adaptive SDW-MWF implementation, the correlation matrices are estimated at the left device as (similarly for the right device):

- In speech+noise frames:
$$\hat{\mathbf{R}}_{yL}[k+1] = \lambda_y \hat{\mathbf{R}}_{yL}[k] + (1-\lambda_y)\mathbf{y}_L[k+1]\mathbf{y}_L^H[k+1] \,, \tag{39}$$
$$\hat{\mathbf{R}}_{vL}[k+1] = \hat{\mathbf{R}}_{vL}[k] \,,$$

- In noise-only frames:
$$\hat{\mathbf{R}}_{yL}[k+1] = \hat{\mathbf{R}}_{yL}[k] \,,$$
$$\hat{\mathbf{R}}_{vL}[k+1] = \lambda_v \hat{\mathbf{R}}_{vL}[k] + (1-\lambda_v)\mathbf{v}_L[k+1]\mathbf{v}_L^H[k+1] \,, \tag{40}$$

where the exponential forgetting factors are set to $\lambda_y = \lambda_v = 0.999$. The speech correlation matrix is then found as $\hat{\mathbf{R}}_{xL} = \hat{\mathbf{R}}_{yL} - \hat{\mathbf{R}}_{vL}$. As in [17], [22], it is assumed[6] that a voice activity detection (VAD) algorithm is available, which correctly classifies frames as speech+noise or noise-only frames. Although estimation errors because of VAD errors are therefore eliminated, it is still possible that the speech correlation matrix is badly estimated at certain frequencies, e.g. because of speech absence [22]. As a result, the values in the denominator of the SP-MWF (20) might decay to zero, so that they have to be kept above certain thresholds [22]. When using the MDVR spatial filter parts of (29), (30), it is also possible to use precalibrated steering vectors $\mathbf{a}_L$ and $\mathbf{a}_R$ instead of $\mathbf{R}_{xL}\mathbf{e}_L$ and $\mathbf{R}_{xR}\mathbf{e}_R$, at the time-frequency points where the left or right speech correlation matrices are badly estimated (a bad estimation is detected if any of the diagonal entries of $\mathbf{R}_{xL}$ or $\mathbf{R}_{xR}$ are negative). The MVDR does not depend on the speech power P_s as can be seen from (24), which simplifies the precalibration. To construct the steering vector, the HRTF corresponding to 0° is used. As in practice the actual reverberation time of the environment will be unknown, an HRTF measured in an anechoic room was used for the precalibration[7].

E. Speech ILD per critical band and normalized cross-correlation function

In Figure 4 (a), the obtained mean ILDs per critical band are shown for the (batch) algorithms BIL$-\mu$=5 (MWF-bil), Front$-\mu$=5 (MWF-front) and ALL$-\mu$=5 (MWF-all), for the S0N60 spatial scenario. The

[6]The aim is to study the speech ILD cue distortion introduced by the different reduced-bandwidth SDW-MWF schemes. Any additional effects, which could be introduced by the VAD, are therefore set aside. An analysis of the SDW-MWF performance under the presence of VAD errors can be found in [48], [49].

[7]Simulations indicate that there is no significant performance difference if a steering vector, calibrated in the actual reverberant environment, is used instead.

ILD was measured on a frame-by-frame basis as outlined in Section VI-A, whereby the average standard deviations in the ILD estimation are indicated in the figure caption. The ILD measured on the unprocessed front microphone signals is also shown as a reference. It can be seen that the bilateral MWF obtains a high positive ILD for most frequencies, which means that the apparent sound location is shifted towards the left side of the head. For the MWF-front, an ILD distortion is still observed, although the effect is smaller than for the bilateral case. It is still significant however, as by [4], a difference as small as 0.5 dB can be noticeable. The MWF-all does not introduce a noticeable ILD distortion, which is in agreement with [17].

In Figure 4 (b) the normalized cross-correlation function, calculated between the left and right output signals, is shown. As in [33], the signals were filtered by a low-pass filter with a cut-off frequency of 1000 Hz, as only the lower frequencies are important for the ITD [3]. The delay for which the maximum of the cross-correlation function occurs gives an estimate of the ITD. It is thus seen that the different reduced-bandwidth algorithms obtain a correct ITD, which is in agreement with the theoretical analysis.

In Figure 5, the ILD per band and normalized cross-correlation functions for the S90N270 scenario are shown. Again, the reduced-bandwidth algorithms introduce ILD distortions, while the ITD is preserved. It is however also observed that even the MWF-all does not obtain the same ILD values as the unprocessed microphone signals, at the higher frequencies. This effect can be explained by the fact that, in addition to noise reduction, the MWF also achieves a certain amount of dereverberation. This is illustrated by the fact that the normalized cross-correlation function of the MWF-all obtains a higher maximum value (more coherence), and by the fact that the standard deviations in the ILD estimation are smaller, compared to the unprocessed microphone signals. As reverberation decreases the magnitude of the ILDs [50], the MWF-all obtains ILDs with a higher magnitude due to its inherent dereverberation effect. The experiment was also repeated for a low reverberation time ($\mathrm{RT}_{60} = 0.21s$, [33], [34]). It was then indeed observed that the dereverberation effect had less impact so that the ILD values coincided better. The bilateral MWF however still obtained ILD values with unnaturally large magnitudes.

F. Average speech ILD error and global SNR improvement

In this Section, the performances of all the considered algorithms of Table I are compared, both for a batch as for an adaptive implementation. In Figures 6-9, the average mean ILD errors (in absolute value) and standard deviations (indicated by errorbars) are shown for the different speech-in-noise scenarios, together with the left and right global SNR improvements (compared to the unprocessed front microphone signals). The global SNR is hereby calculated on the speech and noise components of the time-domain

input and output signals. As the ILD cues are dominant at higher frequencies [3], only the critical bands with center frequencies over 1000 Hz are included in the calculation of the average ILD errors.

In Figure 6, the performances of the batch and adaptive implementations of the nine considered algorithms are shown for the S0N60 scenario. As in [34], adding contralateral microphone signal(s) to the noise reduction procedure (FRONT and ALL) leads to significant SNR improvements over the bilateral case, both for the batch as for the adaptive implementation. For the batch implementation (Figure 6 (a),(c)), it can be seen that by adapting μ or using a partial noise estimate (MWF-η), the average ILD error is decreased compared to the BIL-μ=5 or FRONT-μ=5 algorithms, but not completely eliminated. The CP algorithms are best at preserving the ILD cues, while they are not outperformed by BIL-μ=5 or FRONT-μ=5 in terms of SNR improvement. If the batch results are compared with the adaptive results (Figure 6 (b),(d)), it can be observed that the average ILD error is higher for the adaptive implementations, and that there are more fluctuations in the results as the standard deviations are larger. The CP algorithms still preserve the ILD cues adequately, but show a degraded SNR improvement compared to the BIL-μ=5 or FRONT-μ=5 algorithms (although in the case of the MWF-front, the difference is within 1 dB).

In Figure 7, the performances for a scenario with four babble noise interferers (at 60, 120, 180 and 210 degrees) is shown. Overall, the same observations as for the S0N60 scenario can be made. As four babble noises are active at the same time, the overall noise level is more stationary than for a single noise interferer. As a result, there is not much performance difference between the batch and adaptive implementations, in contrast to the S0N60 scenario.

In Figure 8, the performances for the S0N120 scenario are shown. As the speech and noise source are now farther apart, a large SNR improvement is found, even for the bilateral MWF. As in [19], none of the batch algorithms introduce a significant speech ILD error. It was therefore concluded in [19] that speech ILD errors only occur for scenarios, where the obtained output SNRs at the left and right devices are very different. These are in fact the scenarios where a bilateral (close-spaced) microphone array is insufficient to reduce the noise, namely scenarios where (some of) the speech and noise sources are too close together, or alternatively, scenarios where speech and noise are positioned in two different hemispheres (such as S90N270) [34]. For the adaptive implementations (Figure 8 (d)), it is seen that ILD errors can still occur, especially for the BIL$-\mu=\mathrm{adapt}$ and FRONT$-\mu=\mathrm{adapt}$ algorithms. Again the speech ILD error is lowest for the CP algorithms.

Finally, the performances of the S90N270 scenario are shown in Figure 9. Although the precalibrated steering vector (which is used if estimation errors occur) assumes a frontal speech source, the performance does not completely degrade for the adaptive implementations with CP. Again, especially the FRONT$-\mu=$

$5-$CP algorithm achieves a good SNR improvement, while at the same time only a low speech ILD error is introduced.

VII. Conclusion

This paper featured an analysis of the speech localization cue preservation by reduced-bandwidth SDW-MWF schemes using a general theoretical framework. It was proven that reduced-bandwidth SDW-MWF schemes always preserve the speech ITD cues (regardless of which signals are exchanged over the link), but distort the speech ILD cues in scenarios where the obtained left and right output SNRs are different. This result also applies for a bilateral configuration (in which no microphone signals are exchanged between the devices). It was proven that the introduced ILD error strictly increases if the speech distortion parameter μ increases. If a partial noise estimation parameter η is included, it was similarly proven that the speech ILD error strictly decreases if η increases. The speech ILD distortion can thus be reduced by decreasing μ or by including partial noise estimation. However, these approaches can never completely eliminate the speech ILD distortion, while they necessarily degrade the noise reduction performance in order to reduce the ILD error. Two novel reduced-bandwidth SDW-MWF schemes with a so-called common postfilter, which in principle completely eliminates the speech ILD distortion, were therefore proposed. The novel schemes make use of a decomposed filter structure where the SDW-MWF is structured as a spatial filter followed by a common spectral postfilter (i.e. a postfilter which is equal for the left and right device). The two schemes have different bandwidth requirements, but both re-use the calculations of the spatial filter part in an efficient manner in order to derive the common postfilter. Experiments in a reverberant environment indicate that significant speech intelligibility improvements are obtained with the common postfilter schemes, while the speech ILD cues are preserved adequately, even in a realistic adaptive SDW-MWF implementation.

Appendix

A. Relative speech ILD error of SDW-MWF and MWF-η

In this section, expressions for the relative speech ILD error introduced by the binaural SDW-MWF and MWF-η is given. The influence of parameters η and μ on the relative ILD error is then studied.

The output speech ILD of the MWF-η (28) can be rewritten as:

$$\mathrm{ILD}_x^{out} = \left(1 + \frac{(\rho_L - \rho_R)\ \mu\ (1-\eta)\Big(\mu(1+\eta)(\rho_L+\rho_R) + 2\rho_L\rho_R + 2\mu^2\eta\Big)}{(\mu+\rho_L)^2\ (\mu\eta+\rho_R)^2}\right)\mathrm{ILD}_x^{in}\ . \qquad (41)$$

The relative ILD error is then equal to:

$$\Delta\text{ILD}_x = \frac{|\text{ILD}_x^{out} - \text{ILD}_x^{in}|}{\text{ILD}_x^{in}} , \tag{42}$$

$$= \frac{|\rho_L - \rho_R|\ \mu\ (1-\eta)\Big(\mu(1+\eta)(\rho_L+\rho_R) + 2\rho_L\rho_R + 2\mu^2\eta\Big)}{(\mu+\rho_L)^2\ (\mu\eta+\rho_R)^2} , \tag{43}$$

where $(1-\eta) \geq 0$ for $0 \leq \eta \leq 1$. As the partial derivative of (43) with respect to η is negative for $\rho_L \neq \rho_R$, $\rho_L > 0$, $\rho_R > 0$ and $\mu > 0$, i.e.

$$\frac{\delta}{\delta\eta}\left(\Delta\text{ILD}_x\right) = \frac{-2\ |\rho_L - \rho_R|\ \mu\ (\mu+\rho_R)^2(\mu\eta+\rho_L)}{(\mu+\rho_L)^2(\mu\eta+\rho_R)^3} < 0 , \tag{44}$$

the ILD error strictly decreases as η increases. Thus, by sacrificing noise reduction performance, the speech ILD error can be reduced by the MWF-η (in addition to reducing the noise ITD and ILD errors [17]).

The relative ILD error of the SDW-MWF is found by setting $\eta = 0$ in (43):

$$\Delta\text{ILD}_{x,\ \eta=0} = \frac{|\rho_L - \rho_R|\ \mu\ (\mu(\rho_L+\rho_R) + 2\rho_L\rho_R)}{(\mu+\rho_L)^2\ \rho_R^2} . \tag{45}$$

By calculating the derivative from (45) with respect to μ,

$$\frac{\delta}{\delta\mu}\left(\Delta\text{ILD}_{x,\ \eta=0}\right) = \frac{2\ |\rho_L - \rho_R|\ (\mu+\rho_R)\ \rho_L^2}{(\mu+\rho_L)^3\ \rho_R^2} > 0 , \tag{46}$$

it is seen that the speech ILD error strictly increases as μ increases. Increasing the noise reduction performance (increasing μ) therefore comes at the cost of introducing more speech ILD distortion.

B. Common Postfilter for MWF-front

In this section it is shown that the CP can be efficiently calculated for the binaural MWF-front scheme, by making use of the QRD-RLS scheme proposed in [35]. In the QRD-RLS scheme, the upper triangular Cholesky factor $\mathbf{R}_{v\Delta}$ is stored and updated instead of the noise correlation matrix (with $\mathbf{R}_v = \mathbf{R}_{v\Delta}^H\mathbf{R}_{v\Delta}$). The column vectors $\mathbf{b}_L$ and $\mathbf{b}_R$ are also defined as:

$$\mathbf{b}_L = \mathbf{R}_{v\Delta L}^{-H}\mathbf{R}_{yL}\mathbf{e}_L, \quad \mathbf{b}_R = \mathbf{R}_{v\Delta R}^{-H}\mathbf{R}_{yR}\mathbf{e}_R . \tag{47}$$

It is shown in [35] that $\mathbf{b}_L$ and $\mathbf{b}_R$ can be updated together with $\mathbf{R}_{v\Delta L}$ and $\mathbf{R}_{v\Delta R}$ by applying sequences of unitary Givens rotations. Using these definitions, the SP-MWF (20) is found to be equivalent to [35]:

$$\mathbf{w}_{SP-MWF,L} = (\mathbf{m}_{vyL} - \mathbf{e}_L)\ .\ \frac{<\mathbf{R}_{v\Delta L}\mathbf{e}_L\ ,\ \mathbf{b}_L - \mathbf{R}_{v\Delta L}\mathbf{e}_L>}{<\mathbf{b}_L + (\mu-1)\mathbf{R}_{v\Delta L}\mathbf{e}_L\ ,\ \mathbf{b}_L - \mathbf{R}_{v\Delta L}\mathbf{e}_L>} , \tag{48}$$

$$\mathbf{w}_{SP-MWF,R} = (\mathbf{m}_{vyR} - \mathbf{e}_R)\ .\ \frac{<\mathbf{R}_{v\Delta R}\mathbf{e}_R\ ,\ \mathbf{b}_R - \mathbf{R}_{v\Delta R}\mathbf{e}_R>}{<\mathbf{b}_R + (\mu-1)\mathbf{R}_{v\Delta R}\mathbf{e}_R\ ,\ \mathbf{b}_R - \mathbf{R}_{v\Delta R}\mathbf{e}_R>} , \tag{49}$$

where the dotproduct is defined as $< \mathbf{v}_1, \mathbf{v}_2 >= \mathbf{v}_2^H \mathbf{v}_1$. These expressions can again be structured as an MVDR followed by a spectral postfilter as in (24), i.e.

$$\mathbf{w}_{SP-MWF,L} = \underbrace{(\mathbf{m}_{vyL} - \mathbf{e}_L) \; . \; \frac{< \mathbf{R}_{v\Delta L}\mathbf{e}_L \; , \; \mathbf{b}_L - \mathbf{R}_{v\Delta L}\mathbf{e}_L >}{||\mathbf{b}_L - \mathbf{R}_{v\Delta L}\mathbf{e}_L||^2}}_{\text{MVDR}} \times \underbrace{\frac{||\mathbf{b}_L - \mathbf{R}_{v\Delta L}\mathbf{e}_L||^2}{< \mathbf{b}_L + (\mu - 1)\mathbf{R}_{v\Delta L}\mathbf{e}_L \; , \; \mathbf{b}_L - \mathbf{R}_{v\Delta L}\mathbf{e}_L >}}_{\text{postfilter}} , \tag{50}$$

and similarly for $\mathbf{w}_{SP-MWF,R}$. In expressions (48)-(50), $\mathbf{m}_{vyL}$ and $\mathbf{m}_{vyR}$ are found by solving the following systems of equations, which can be done by a simple backsubstitution:

$$\mathbf{R}_{v\Delta L} \; \mathbf{m}_{vyL} = \mathbf{b}_L, \quad \mathbf{R}_{v\Delta R} \; \mathbf{m}_{vyR} = \mathbf{b}_R \; . \tag{51}$$

The signal vectors $\mathbf{y}_L$ and $\mathbf{y}_R$ in (2), (4), (5) are now redefined as

$$\mathbf{y}_L = \left[\begin{array}{c} Y_{0,0} \\ Y_{1,0} \\ \hdashline Y_{0,1} \\ \vdots \\ Y_{0,M-1} \end{array}\right], \quad \mathbf{y}_R = \left[\begin{array}{c} Y_{0,0} \\ Y_{1,0} \\ \hdashline Y_{1,1} \\ \vdots \\ Y_{1,M-1} \end{array}\right] , \tag{52}$$

so that the front microphone signals are now the first two entries of each vector (in the same order). $\mathbf{e}_L$ and $\mathbf{e}_R$ are also modified accordingly. As a result of this reordering, the upper-left 2-by-2 submatrices in $\mathbf{R}_{v\Delta L}$ and $\mathbf{R}_{v\Delta R}$ are equal, and by definition (47), it can be seen that the first two entries of vectors $\mathbf{b}_L$ and $\mathbf{b}_R$ only depend on the front microphone signals (as the upper-left 2-by-2 submatrices of $\mathbf{R}_{v\Delta L}^{-H}$ and $\mathbf{R}_{v\Delta R}^{-H}$ are also equal and only depend on the front microphone signals). As a result, these quantities can be calculated in both devices and moreover, they are identical to the values which would be obtained by calculating a separate (QRD-RLS) SP-MWF on only the front microphone signals. To derive the CP, it is thus only necessary to store and update (the first two entries of) $\mathbf{b}_L$ and $\mathbf{b}_R$ in both devices, and then select the same vector at both sides (for example the vector of the best ear). For example, assuming the left side is the best ear, the CP is found as:

$$\text{CP } (\mu) = \frac{||\mathbf{b}_L^{''} - \mathbf{R}_{v\Delta L}^{''}\mathbf{e}_L||^2}{< \mathbf{b}_L^{''} + (\mu - 1)\mathbf{R}_{v\Delta L}^{''}\mathbf{e}_L \; , \; \mathbf{b}_L^{''} - \mathbf{R}_{v\Delta L}^{''}\mathbf{e}_L >} \quad (\text{left device}) , \tag{53}$$

$$= \frac{||\mathbf{b}_L^{''} - \mathbf{R}_{v\Delta R}^{''}\mathbf{e}_L||^2}{< \mathbf{b}_L^{''} + (\mu - 1)\mathbf{R}_{v\Delta R}^{''}\mathbf{e}_L \; , \; \mathbf{b}_L^{''} - \mathbf{R}_{v\Delta R}^{''}\mathbf{e}_L >} \quad (\text{right device}) , \tag{54}$$

where

$$\mathbf{R}_{v\Delta L}^{''} = \left[\begin{array}{c:c} \mathbf{I}_2 & \mathbf{0}_{2\times(M-1)} \end{array}\right] \mathbf{R}_{v\Delta L}, \quad \mathbf{b}_L^{''} = \left[\begin{array}{c:c} \mathbf{I}_2 & \mathbf{0}_{2\times(M-1)} \end{array}\right] \mathbf{b}_L \tag{55}$$

are the first two rows of $\mathbf{R}_{v\Delta L}$ and $\mathbf{b}_L$. The CP can thus be extracted from the $(M+1)$-microphone MVDR solution by calculating two simple dotproducts.

REFERENCES

[1] H. Dillon, *Hearing Aids*. Boomerang Press, Australia, 2001.

[2] V. Hamacher, J. Chalupper, J. Eggers, E. Fischer, U. Kornagel, H. Puder, and U. Rass, "Signal processing in high-end hearing aids: state of the art, challenges, and future trends," *EURASIP J. Appl. Signal Process.*, pp. 2915–2929, 2005.

[3] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localisation* (Revised Edition). MIT Press, 1996.

[4] W. Hartmann, "How we localize sound," *Physics Today*, vol. 52, no. 11, pp. 24–29, Nov. 1999.

[5] V. Algazi and R. Duda, "Headphone-based spatial sound," *IEEE Signal Process. Mag.*, vol. 28, no. 1, pp. 33–42, 2011.

[6] P. Zurek, *Binaural advantages and directional effects in speech intelligibility*. 2nd Ed., Boston, MA, Allyn and Bacon, 1992, ch. 15 in "Acoustical factors affecting hearing aid performance", pp. 255–276.

[7] B. Kollmeier, J. Peissig, and V. Hohmann, "Real-time multiband dynamic compression and noise reduction for binaural hearing aids," *J. Rehabil. Res. Develop.*, vol. 30, no. 1, pp. 82–94, 1993.

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

[9] T. Lotter and P. Vary, "Dual-channel speech enhancement by superdirective beamforming," *EURASIP J. Appl. Signal Process.*, pp. 1–14, 2006.

[10] K. Reindl, Y. Zheng, and W. Kellermann, "Analysis of two generic Wiener filtering concepts for binaural speech enhancement in hearing aids," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Aalborg, Denmark, Aug. 2010, pp. 989–993.

[11] A. H. Kamkar-Parsi and M. Bouchard, "Instantaneous binaural target PSD estimation for hearing aid noise reduction in complex acoustic environments," *IEEE Trans. Instr. Meas.*, vol. 60, no. 4, pp. 1141–1154, 2011.

[12] T. D. Trine and D. Van Tasell, "Digital hearing aid design: Fact vs. fantasy," *Hear. J.*, vol. 55, no. 2, pp. 36–38, 40–42, 2002.

[13] J. Desloge, W. Rabinowitz, and P. Zurek, "Microphone-array hearing aids with binaural output–Part I: Fixed-processing systems," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 6, pp. 529–542, Nov. 1997.

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

[15] T. J. Klasen, T. Van den Bogaert, M. Moonen, and J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1579–1585, Apr. 2007.

[16] 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 functions," in *Proc. Int. Workshop Acoust. Echo Noise Control (IWAENC)*, Paris, France, Sep. 2006.

[17] B. Cornelis, S. Doclo, T. Van den Bogaert, M. Moonen, and J. Wouters, "Theoretical analysis of binaural multimicrophone noise reduction techniques," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 2, pp. 342–355, Feb. 2010.

[18] S. Doclo, T. Van den Bogaert, J. Wouters, and M. Moonen, "Reduced-bandwidth and distributed MWF-based noise reduction algorithms for binaural hearing aids," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 1, pp. 38–51, Jan. 2009.

[19] B. Cornelis, M. Moonen, and J. Wouters, "Binaural cue preservation in binaural hearing aids with reduced-bandwidth multichannel Wiener filter based noise reduction," in *Proc. Int. Workshop Acoust. Echo Noise Control (IWAENC)*, Tel-Aviv, Israel, Sep. 2010.

[20] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Frequency-Domain Criterion for Speech Distortion Weighted Multichannel Wiener Filter for Robust Noise Reduction," *Speech Commun.*, vol. 49, no. 7–8, pp. 636–656, July-August 2007.

[21] E. Habets, J. Benesty, I. Cohen, S. Gannot, and J. Dmochowski, "New insights into the MVDR beamformer in room acoustics," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 1, pp. 158–170, Jan. 2010.

[22] M. Souden, J. Benesty, and S. Affes, "On optimal frequency-domain multichannel linear filtering for noise reduction," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 2, pp. 260–276, Feb. 2010.

[23] B. Cornelis, M. Moonen, and J. Wouters, "Comparison of frequency domain noise reduction strategies based on Multichannel Wiener Filtering and Spatial Prediction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Tai Pei, Taiwan, April 2009, pp. 129–132.

[24] ——, "Performance analysis of multichannel Wiener filter based noise reduction in hearing aids under second order statistics estimation errors," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 5, pp. 1368–1381, Jul. 2011.

[25] J. Benesty, J. Chen, and Y. Huang, *Noncausal (frequency-domain) optimal filters*. Springer-Verlag, 2008, ch. 6 in "Microphone Array Signal Processing", pp. 115–137.

[26] K. U. Simmer, J. Bitzer, and C. Marro, *Post-Filtering Techniques*. Springer-Verlag, May 2001, ch. 3 in "Microphone Arrays: Signal Processing Techniques and Applications" (Brandstein, M. S. and Ward, D. B., Eds.), pp. 39–60.

[27] A. Spriet, S. Doclo, M. Moonen, and J. Wouters, "A unification of adaptive multi-microphone noise reduction systems," in *Proc. Int. Workshop Acoust. Echo Noise Control (IWAENC)*, Paris, France, Sep. 2006.

[28] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.

[29] S. Gannot, D. Burshtein, and E. Weinstein, "Signal Enhancement Using Beamforming and Non-Stationarity with Applications to Speech," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.

[30] P. C. Loizou, *Speech enhancement: Theory and Practice*. CRC press, New York, USA, 2007.

[31] Y. Ephraim and H. L. Van Trees, "A Signal Subspace Approach for Speech Enhancement," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 4, pp. 251–266, Jul. 1995.

[32] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 4, April 1979, pp. 208–211.

[33] T. Van den Bogaert, S. Doclo, M. Moonen, and J. Wouters, "The effect of multi-microphone noise reduction systems on sound source localization in binaural hearing aids," *J. Acoust. Soc. Amer.*, vol. 124, no. 1, pp. 484–497, 2008.

[34] ——, "Speech enhancement with multichannel Wiener filter techniques in multi-microphone binaural hearing aids," *J. Acoust. Soc. Amer.*, vol. 125, no. 1, pp. 360–371, 2009.

[35] B. Cornelis, M. Moonen, and J. Wouters, "A QRD-RLS based frequency-domain Multichannel Wiener Filter algorithm for noise reduction in hearing aids," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Aalborg, Denmark, Aug. 2010, pp. 1953–1957.

[36] S. Srinivasan, "Low-bandwidth binaural beamforming," *Electronics Letters*, vol. 44, no. 22, pp. 1292–1293, 2008.

[37] C. Faller and J. Merimaa, "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence," *J. Acoust. Soc. Amer.*, vol. 116, no. 5, pp. 3075–3089, 2004.

[38] M. Jeub, M. Schafer, T. Esch, and P. Vary, "Model-based dereverberation preserving binaural cues," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 7, pp. 1732–1745, Sep. 2010.

[39] R. Patterson, M. Allerhand, and C. Gigure, "Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform," *J. Acoust. Soc. Amer.*, vol. 98, no. 4, pp. 1890–1894, 1995.

[40] B. Glasberg and B. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.*, vol. 47, pp. 103 – 138, 1990.

[41] L. R. Bernstein, S. van de Par, and C. Trahiotis, "The normalized interaural correlation: Accounting for NoSπ thresholds obtained with Gaussian and "low-noise" masking noise," *J. Acoust. Soc. Amer.*, vol. 106, no. 2, pp. 870–876, 1999.

[42] M. Slaney, "Auditory toolbox version 2," *Technical Report no.1998-010, Interval Research Corp.*, Palo Alto, CA, "*https://engineering.purdue.edu/~malcolm/interval/1998-010/*", 1998.

[43] M. Akeroyd, "A binaural cross-correlogram toolbox for MATLAB," "*www.ihr.mrc.ac.uk/downloads/products/matlab/binauraltoolbox*", 2001.

[44] M. Nilsson, S. D. Soli, and A. Sullivan, "Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise," *J. Acoust. Soc. Amer.*, vol. 95, no. 2, pp. 1085–1099, Feb. 1994.

[45] Auditec, "Auditory tests (revised)," *Compact Disc, Auditec, St. Louis*, 1997.

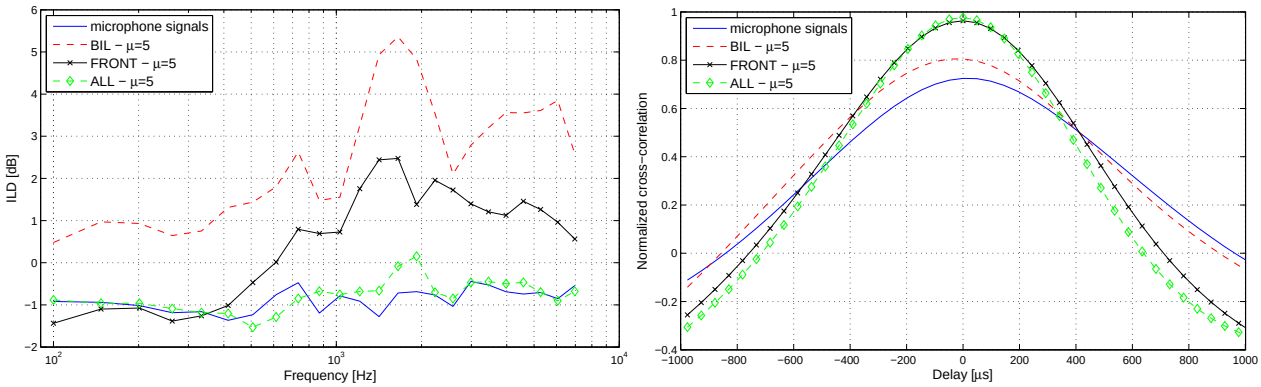
[46] S. Kamath and P. C. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Orlando, Florida, May 2002.

[47] R. Crochiere, "A weighted overlap-add method of short-time Fourier analysis/synthesis," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 28, no. 1, pp. 99 – 102, Feb. 1980.

[48] A. Spriet, M. Moonen, and J. Wouters, "The impact of speech detection errors on the noise reduction performance of multi-channel Wiener filtering and Generalized Sidelobe Cancellation," *Signal Process.*, vol. 85, no. 6, pp. 1073–1088, Jun. 2005.

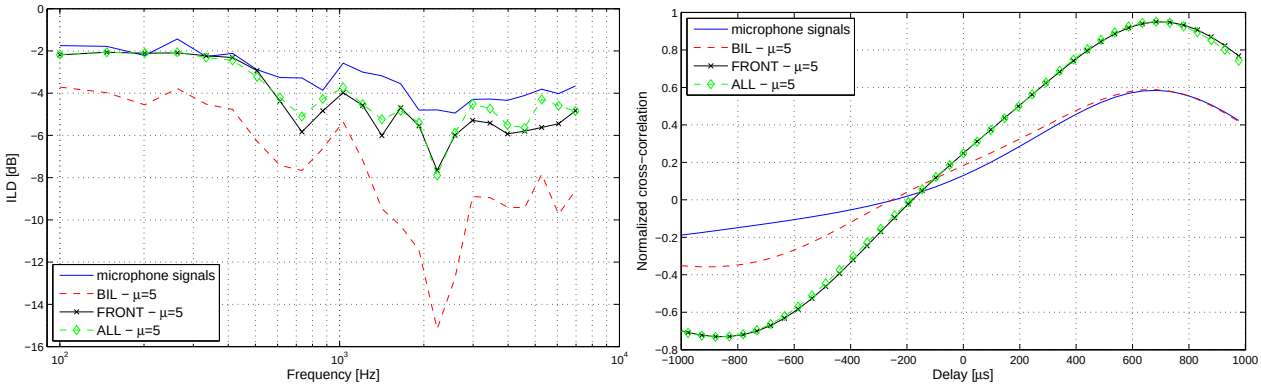
[49] B. Cornelis, M. Moonen, and J. Wouters, "A VAD-robust Multichannel Wiener Filter algorithm for noise reduction in hearing aids," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 281–284.

[50] B. G. Shinn-Cunningham, N. Kopco, and T. J. Martin, "Localizing nearby sound sources in a classroom: Binaural room impulse responses," *J. Acoust. Soc. Amer.*, vol. 117, no. 5, pp. 3100–3115, 2005.



(a) **(b)**

Fig. 4. **S0 N60**; ILD per critical band (a) and normalized cross-correlation (b) of the speech component. The average ILD standard deviations are 0.71 dB (microphone signals), 0.73 dB (BIL), 0.41 dB (FRONT) and 0.28 dB (ALL).



(a) **(b)**

Fig. 5. **S90 N270**; ILD per critical band (a) and normalized cross-correlation (b) of the speech component. The average ILD standard deviations are 1.22 dB (microphone signals), 1.27 dB (BIL), 0.43 dB (FRONT) and 0.42 dB (ALL).

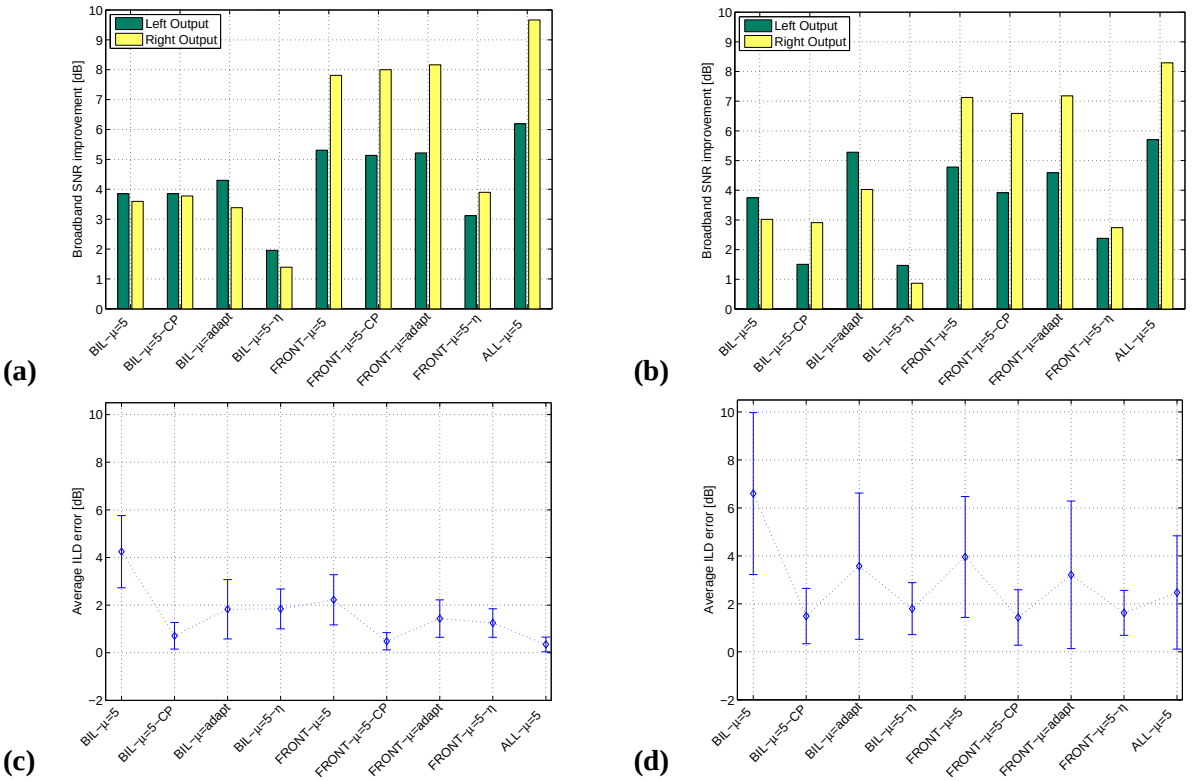


Fig. 6. **S0 N60**; Global SNR improvement for batch (a) and adaptive (b) MWF; average speech ILD error for batch (c) and adaptive (d) MWF

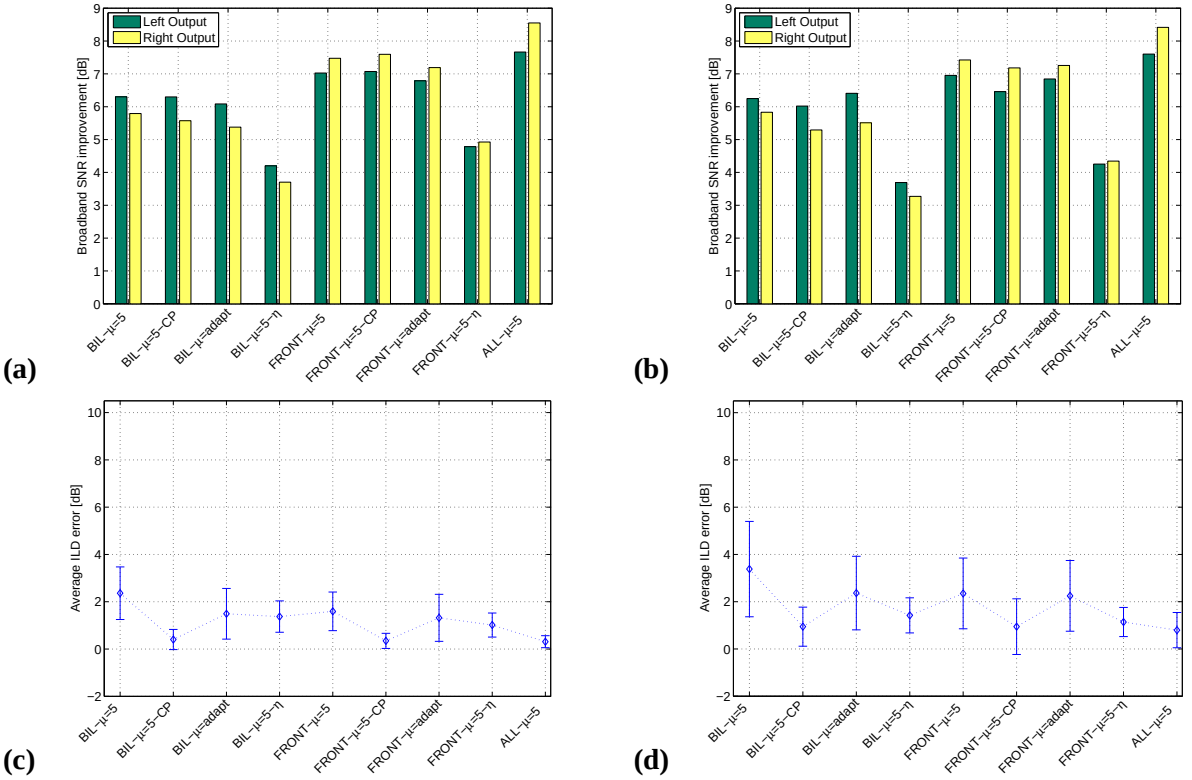


Fig. 7. **S0 N [60-120-180-210]**; Global SNR improvement for batch (a) and adaptive (b) MWF; average speech ILD error for batch (c) and adaptive (d) MWF

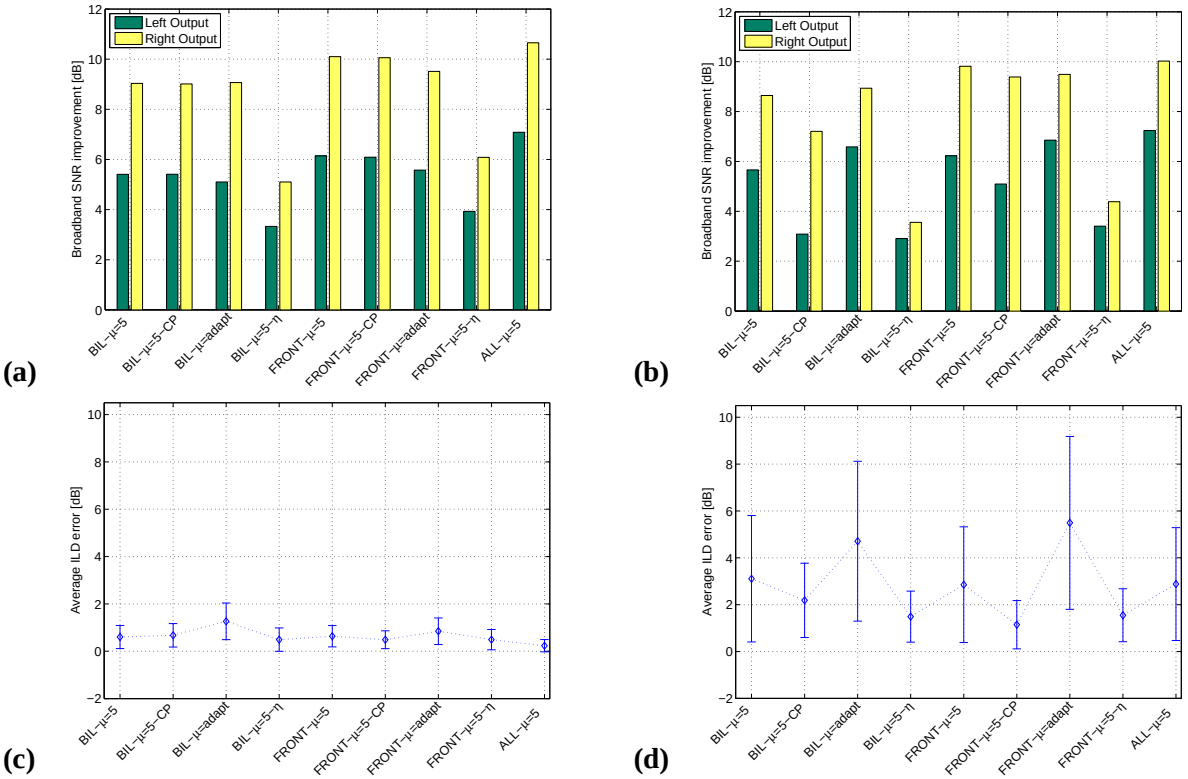


Fig. 8. **S0 N120**; Global SNR improvement for batch (a) and adaptive (b) MWF; average speech ILD error for batch (c) and adaptive (d) MWF

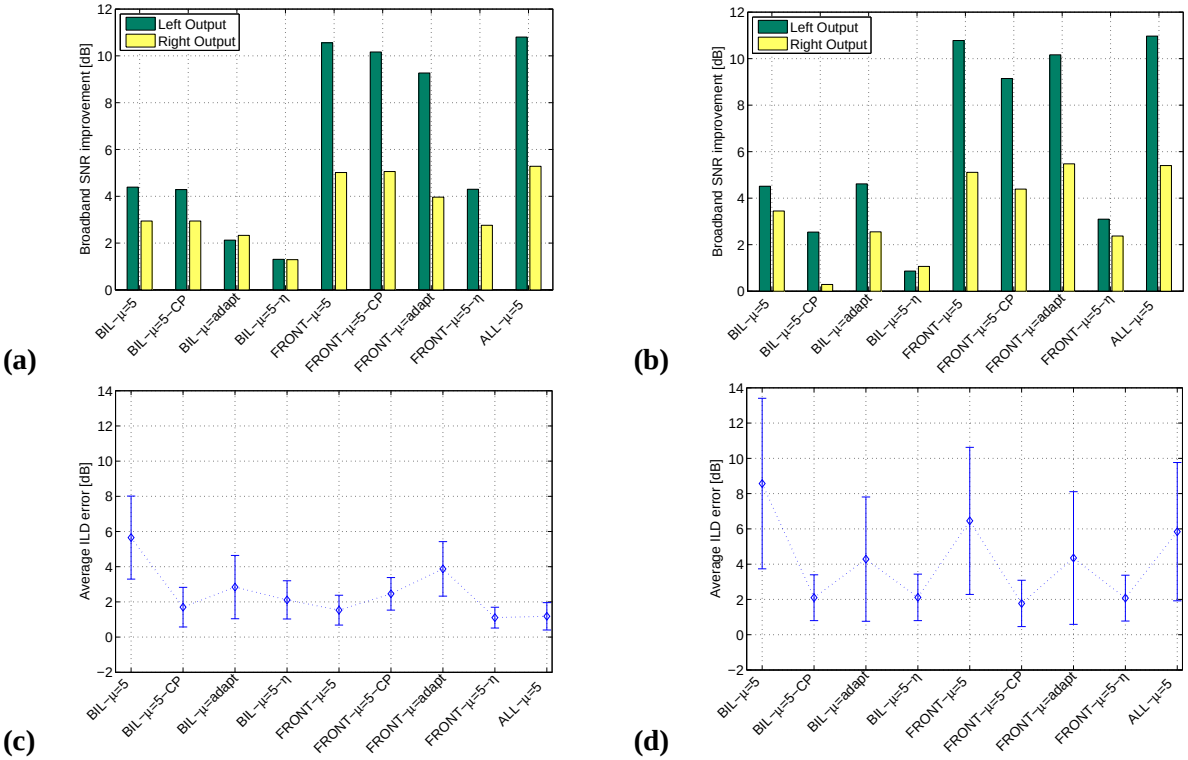


Fig. 9. **S90 N270**; Global SNR improvement for batch (a) and adaptive (b) MWF; average speech ILD error for batch (c) and adaptive (d) MWF