ANALYSIS OF TWO GENERIC WIENER FILTERING CONCEPTS FOR BINAURAL SPEECH ENHANCEMENT IN HEARING AIDS

Klaus Reindl, Yuanhang Zheng, and Walter Kellermann

Multimedia Communications and Signal Processing
University of Erlangen-Nuremberg, Cauerstr. 7
91058 Erlangen, Germany
Email:{reindl, zheng, wk}@LNT.de

ABSTRACT

In this paper, we present a detailed analysis for two generic single-channel Wiener filtering concepts for binaural hearing aids, namely, a dual-channel filter approach (individual filters for individual channels) and a single filter approach (one filter applied to both channels). After a general description of the concept, this scheme is thoroughly discussed for a two-source scenario and especially the influence of a common noise estimate on the two different concepts is analyzed in detail. Moreover, by evaluating the interaural transfer function (ITF) comprising the interaural level differences (ILDs) as well as interaural phase/time differences (IPDs/ITDs), the influence of both Wiener filter approaches on these binaural cues is analyzed. The theoretical findings are confirmed by experimental results.

1. INTRODUCTION

Following the advent of wireless technologies for connecting both ears, many different binaural processing strategies are currently under investigation. Several binaural multi-channel Wiener filtering approaches preserving binaural cues for the speech and noise components were presented, e.g., in [1, 2]. However, these techniques require estimates for the noise components in each individual microphone. Since, in practice, it is almost impossible to obtain reliable noise estimates for nonstationary noise and interference, we investigated the combination of a common noise estimate with a single-channel Wiener-type filter in [3] to obtain binaural output signals. The common noise estimate is obtained by forcing a spatial null into the direction of the desired speaker. As this noise estimate is the complementary signal to the desired speech signal, this estimate captures the entire spatio-temporal information of all interfering point sources as well as background noise (as long as spatial filtering allows). So far, there was no detailed analysis how this common noise estimate interacts with single-channel speech enhancement algorithms and, correspondingly, how the combination affects binaural cues. Thus, in this contribution, the combination of a common noise estimate with two fundamentally different Wiener filtering concepts is analyzed in detail. Moreover, the influence of the single-channel speech enhancement techniques on binaural cues is evaluated. The most important binaural cues for speech understanding in adverse environments and localizing sound sources correctly are ILDs, ITDs, and the coherence between the signals arriving at the right and left ear. In order to preserve ITDs, only linear-phase filter approaches are considered so that the signal delay is equal in both channels. Correspondingly, only the impairment of ILDs by both Wiener filtering concepts must be analyzed.

The paper is organized as follows. The general binaural configuration is described in Sect. 2 and the Wiener filtering concepts are introduced in Sect. 3. A detailed analysis for a two-source scenario is presented in Sect. 4. Experimental results will be presented in Sect. 5 before providing concluding remarks in Sect. 6.

2. SIGNAL MODEL

Fig. 1 depicts the signal model studied in this paper. Lower-case boldface characters represent (column) vectors capturing signals or

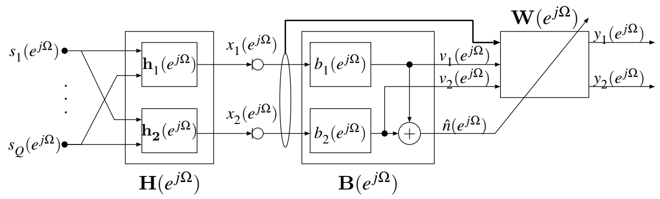

Figure 1: General binaural speech enhancement

the spectral weights of Multiple-Input-Single-Output (MISO) systems. Matrices denoting Multiple-Input-Multiple-Output (MIMO) systems are represented by upper-case boldface characters. The superscripts $(\cdot)^*$, $\{\cdot\}^T$, and $\{\cdot\}^H$ denote complex conjugation, vector or matrix transposition, and conjugate transposition, respectively.

We consider here hearing aid devices (in-the-ear (ITE) or behind-the-ear (BTE)) with a single microphone at each ear. Capturing reverberation and scattering at the user's head, Q point source signals s_q $(q = 1, \ldots, Q)$ are filtered by a MISO mixing system (one $Q \times 1$ MISO system for each ear). This can be expressed in the discrete-time Fourier transform (DTFT)-domain as

$$x_i(e^{j\Omega}) = \sum_{q=1}^{Q} h_{qi}(e^{j\Omega}) s_q(e^{j\Omega}) \qquad i \in \{1,2\}, \tag{1}$$

where $x_i(e^{j\Omega})$ is the frequency domain representation of the received source signal mixture at the microphone of the right or left hearing aid, respectively. $\Omega = \frac{2\pi f}{f_s}$ represents the normalized frequency and f_s denotes the sampling frequency. $h_{qi}(e^{j\Omega})$ denotes the transfer function between the q-th source and the i-th microphone. We assume that $s_1(e^{j\Omega})$ represents the desired source signal ('target') while the remaining $Q-1$ sources are considered as interfering point sources. For brevity, the frequency-dependency $(e^{j\Omega})$ will be omitted in the rest of the paper. Using vector/matrix notation the microphone signals are given by

$$\mathbf{x} = \mathbf{H}^T \mathbf{s}, \tag{2}$$

where $\mathbf{x} = [x_1,\, x_2]^T$ is a column vector capturing both microphone signals. $\mathbf{s} = [s_1,\, \ldots, s_Q]^T$ captures all point source signals and the acoustic mixing system is represented by the $Q \times 2$ matrix $\mathbf{H}$:

$$\mathbf{H} = [\ \mathbf{h}_1 \quad \mathbf{h}_2\ ] = \begin{bmatrix} h_{11} & h_{12} \\ \vdots & \vdots \\ h_{Q1} & h_{Q2} \end{bmatrix}. \tag{3}$$

The blocking matrix $\mathbf{B}$ that performs time-frequency as well as spatial filtering is used to generate a common noise estimate which is used to control a speech enhancement algorithm. Using the spectral weights $b_i,\ i \in \{1,2\}$, the blocking matrix is given by

$$\mathbf{b} = [\ b_1 \quad b_2\ ]^T, \qquad (4)$$
$$\mathbf{B} = [\ \mathbf{b} \quad \mathrm{diag}\{\mathbf{b}\}\ ]. \qquad (5)$$

In (5), diag$\{\mathbf{b}\}$ is necessary to obtain the signals $v_i,\ i \in \{1,2\}$. The output signals of the blocking matrix are then given by

$$[\ \hat{n} \quad v_1 \quad v_2\ ]^T = \mathbf{B}^T\mathbf{x}. \qquad (6)$$

The binaural output signals are represented by a 2×1 column vector $\mathbf{y} = [y_1,\ y_2]^T$ and can be written as

$$\mathbf{y} = \mathbf{W}\mathbf{x}. \qquad (7)$$

The matrix $\mathbf{W}$ captures the spectral weights $w_i,\ i \in \{1,2\}$ used to enhance the microphone signals $x_i,\ i \in \{1,2\}$:

$$\mathbf{W} = \mathrm{diag}\{\mathbf{w}\}, \qquad (8)$$
$$\mathbf{w} = [\ w_1 \quad w_2\ ]^T. \qquad (9)$$

The diag$\{\cdot\}$ operator applied to a vector builds a diagonal matrix with the vector entries placed on the main diagonal. The Wiener filtering concepts used to enhance the microphone signals are discussed in the following section. Note that, accounting for the non-stationarity of the involved signals, all the frequency-domain representations are based on a short-time Fourier transform (STFT) with a finite block length of 32ms.

3. WIENER FILTERING CONCEPTS

The common noise estimate $\hat{n}$ can be written as

$$\hat{n} = \mathbf{b}^T\mathbf{x} = \mathbf{b}^T\mathbf{H}^T\mathbf{s}. \qquad (10)$$

The most intuitive approach to obtain a speech enhancement filter would be a dual-channel filter, where individual filters are applied to the individual channels. If, as in the given scenario, only a common noise estimate $\hat{n}$ is available, it seems reasonable to investigate a single filter that is applied to both channels as well. Consequently, the following two concepts are analyzed:

- Dual-channel linear-phase filter (individual filters for individual channels),
- Single linear-phase filter (one filter applied to both channels).

For a dual-channel Wiener filter, the spectral weights are given by

$$w_i = \max\left[1 - \mu\frac{\hat{P}_{\hat{n}\hat{n}}}{\hat{P}_{v_iv_i}}, w_{\min}\right] \qquad i \in \{1,2\}, \qquad (11)$$

where $\hat{P}_{\hat{n}\hat{n}}$ and $\hat{P}_{v_iv_i},\ i \in \{1,2\}$ represent the Power Spectral Density (PSD) estimates of $\hat{n}$ and $v_i,\ i \in \{1,2\}$, respectively. v_i is given by [4]

$$v_i = b_i x_i \qquad i \in \{1,2\}. \qquad (12)$$

Using (2), (6), and (10), the PSD estimates can be obtained by

$$\hat{P}_{\hat{n}\hat{n}} = \mathbf{b}^H\mathbf{H}^H\hat{\mathbf{P}}_{\mathbf{ss}}\mathbf{H}\mathbf{b}, \qquad (13)$$
$$\hat{P}_{v_iv_i} = |b_i|^2\mathbf{h}_i^H\hat{\mathbf{P}}_{\mathbf{ss}}\mathbf{h}_i \qquad i \in \{1,2\}, \qquad (14)$$

where $\hat{\mathbf{P}}_{\mathbf{ss}}$ is the estimate for the (presumably diagonal) PSD matrix of the source signals, and $\mathbf{h}_i$ represents the i-th column of $\mathbf{H}$. $w_{\min}$ $(0 < w_{\min} < 1)$ denotes the minimum value of the spectral weights (spectral floor). μ is a real number and is used to achieve a trade-off between noise reduction and speech distortion [5]. For the frequency response values of a single linear-phase filter we choose:

$$w = w_1 = w_2 = \max\left[1 - \mu\frac{\hat{P}_{\hat{n}\hat{n}}}{\hat{P}_{v_1v_1} + \hat{P}_{v_2v_2}}, w_{\min}\right]. \qquad (15)$$

Since the obtained noise estimate consists of noise components from both channels, $\hat{n}$ will usually differ from the noise components in the individual channels. Thus, the spectral weights for both concepts are always given by

$$w_i = w_{\mathrm{o}_i} + \delta_{w_i} \qquad i \in \{1,2\}, \qquad (16)$$

where w_{o_i} denotes the channel-specific optimum Wiener-filter weight (which cannot be realized as the necessary individual PSDs cannot be observed) and δ_{w_i} represents the deviation. The optimum weights w_{o_i} can be written as

$$w_{\mathrm{o}_i} = 1 - \frac{\mathbf{h}_i^H\mathbf{U}^T\hat{\mathbf{P}}_{\mathbf{ss}}\mathbf{U}\mathbf{h}_i}{\mathbf{h}_i^H\hat{\mathbf{P}}_{\mathbf{ss}}\mathbf{h}_i} \qquad i \in \{1,2\}. \qquad (17)$$

With s_1 denoting the desired source signal, the $Q \times Q$ matrix $\mathbf{U}$ serves to select the interference components in $\hat{\mathbf{P}}_{\mathbf{ss}}$ and is given by

$$\mathbf{U} = \left[\begin{array}{c|c} \multicolumn{2}{c}{\mathbf{0}_{1\times Q}} \\ \hline \mathbf{0}_{(Q-1)\times 1} & \mathbf{I}_{(Q-1)\times(Q-1)} \end{array}\right], \qquad (18)$$

where $\mathbf{I}_{(Q-1)\times(Q-1)}$ represents a $(Q-1)\times(Q-1)$ identity matrix and $\mathbf{0}_{1\times Q}$ and $\mathbf{0}_{(Q-1)\times 1}$ denote zero vectors of size $1\times Q$ and $(Q-1)\times 1$, respectively.

4. DISCUSSION

For a two-source scenario $(s_1 = s,\ s_2 = n)$ as depicted in Fig. 2, the Wiener filtering concepts are discussed in more detail. s and n

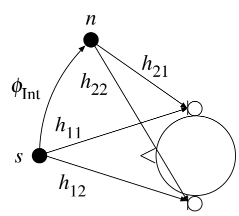

Figure 2: Investigated scenario with the desired speaker s located at 0° (broadside direction) and the interferer n located at ϕ_{Int}.

denote the desired source and the interfering source, respectively. From Fig. 1 and Fig. 2, the common noise estimate $\hat{n}$ is obtained by

$$\hat{n} = (h_{11}b_1 + h_{12}b_2)s + (h_{21}b_1 + h_{22}b_2)n := \tilde{n} + s_{\mathrm{r}}, \qquad (19)$$

where $\tilde{n}$ represents an estimate of the interfering signal n and s_{r} describes the residual desired speech components. For negligible speech components contained in the common noise estimate (i.e., a nearly perfect blocking matrix), (19) is then given by

$$\hat{n} \approx (h_{21}b_1 + h_{22}b_2)n. \qquad (20)$$

In this case, the PSD estimate (13) simplifies to

$$\hat{P}_{\hat{n}\hat{n}} = (|h_{21}b_1|^2 + |h_{22}b_2|^2 + 2\Re\{h_{21}b_1h_{22}^*b_2^*\})\hat{P}_{nn}. \qquad (21)$$

4.1 Dual-channel Wiener filter

First of all, we will discuss the influence of (21) on the dual-channel linear-phase filter approach. The spectral weights are given by

$$w_i = \max\left[1-\mu\frac{\hat{P}_{\hat{n}\hat{n}}}{|h_{1i}b_i|^2\hat{P}_{ss}+|h_{2i}b_i|^2\hat{P}_{nn}}, w_{\min}\right], \qquad (22)$$

where $\hat{P}_{\hat{n}\hat{n}}$ is given by (21). As soon as the interferer is located close to the desired speaker, both, the desired speaker as well as the interfering source are strongly attenuated by the blocking matrix and $\hat{P}_{\hat{n}\hat{n}}$ will converge to zero. Correspondingly, the spectral weights given in (22) converge to one. Because of the front-back ambiguity of a two-microphone setup, the same effect occurs for interfering sources from the rear. Next, we assume a pronounced head shadowing effect, i.e., the interferer is either located at $\phi_{\text{Int}} \approx 90^\circ$ or at $\phi_{\text{Int}} \approx 270^\circ$. It cannot be assumed that the transfer function describing the path from the interfering source to the contralateral microphone is close to zero for the entire frequency range. Consequently, the cross-correlation component $2\Re\{h_{21}b_1h_{22}^*b_2^*\}$ (21) must not be neglected and optimum spectral weights cannot be obtained for the ipsilateral channel. In fact, the term $1-\mu\frac{\hat{P}_{\hat{n}\hat{n}}}{|h_{1i}b_i|^2\hat{P}_{ss}+|h_{2i}b_i|^2\hat{P}_{nn}}$ may become negative as the common noise PSD estimate ($\hat{P}_{\hat{n}\hat{n}}$) strongly deviates from the individual noise PSD estimates given by $|h_{2i}b_i|^2\hat{P}_{nn}$. Consequently, the frequency response values (22) will be set to the spectral floor and therefore these frequencies will be more or less strongly attenuated depending on $w_{\min}$. This effect becomes pronounced for the contralateral channel as the interfering components contained in this channel are already suppressed to a certain extent because of head shadowing. Correspondingly, the true noise component in the contralateral channel and the common noise estimate are highly different. Since many frequencies will be strongly attenuated, higher noise reduction as well as a more severe speech distortion will result. This effect can be slightly reduced by choosing a smaller value for the parameter μ, but it cannot be completely prevented.

4.2 Single Wiener filter

Next, the single linear-phase filter is discussed. Assuming again negligible speech components contained in the common noise estimate so that (20) holds, the spectral weights for this concept read

$$w_1 = w_2 = \max\left[1-\mu\frac{\hat{P}_{\hat{n}\hat{n}}}{\sum_{i=1}^{2}|h_{1i}b_i|^2\hat{P}_{ss}+|h_{2i}b_i|^2\hat{P}_{nn}}, w_{\min}\right], \qquad (23)$$

where $\hat{P}_{\hat{n}\hat{n}}$ is given by (21). Only for uncorrelated signals ($\Re\{h_{21}b_1h_{22}^*b_2^*\} = 0$ in (21)), the weights represent an optimum 'average' filter. Again, if the interferer is located close to the desired speaker, the same behavior is obtained as with the dual-channel filter concept, i.e., the frequency response values converge to one. As the cross-correlation $\Re\{h_{21}b_1h_{22}^*b_2^*\}$ may become high especially for low frequencies, and (23) is also limited to $w_{\min}$, some frequencies may also be strongly attenuated. But this behavior will not be as dominant as for the dual-channel Wiener filter since for this filtering concept both microphone signals are exploited to compute the spectral weights. Thus, a lower noise reduction will be obtained, but more importantly, speech distortion will also be lower than for a dual-channel filter approach.

4.3 Influence on binaural cues

In the following, the influence of both filtering concepts on the binaural cues (interaural phase/time difference (IPD/ITD), interaural level difference (ILD)) is analyzed. For this, the interaural transfer function (ITF) is evaluated. The ITF for the single interfering source signal n at the input and at the output of the binaural noise reduction scheme depicted in Fig. 1 are given by

$$\text{ITF}_{\text{in}}^n = \frac{h_{21}n}{h_{22}n} = \frac{h_{21}}{h_{22}}, \qquad (24)$$

$$\text{ITF}_{\text{out}}^n = \frac{h_{21}w_1n}{h_{22}w_2n} = \frac{h_{21}w_1}{h_{22}w_2} = \text{ITF}_{\text{in}}^n \cdot \frac{w_1}{w_2}. \qquad (25)$$

The ITFs for the desired source signal s are given in the same way. The IPDs at the input and at the output are then given as the argument of the ITFs:

$$\text{IPD}_{\text{in}}^n = \arg\{\text{ITF}_{\text{in}}^n\}, \qquad (26)$$

$$\text{IPD}_{\text{out}}^n = \arg\{\text{ITF}_{\text{out}}^n\} = \arg\{\text{ITF}_{\text{in}}^n\} + \arg\left\{\frac{w_1}{w_2}\right\}. \qquad (27)$$

Correspondingly, the ILDs are given by the absolute value of the ITFs:

$$\text{ILD}_{\text{in}}^n = |\text{ITF}_{\text{in}}^n|, \qquad (28)$$

$$\text{ILD}_{\text{out}}^n = |\text{ITF}_{\text{out}}^n| = |\text{ITF}_{\text{in}}^n| \cdot \left|\frac{w_1}{w_2}\right|. \qquad (29)$$

From (27), it can be verified that if the speech enhancement filters applied in the individual channels have the same phase response or correspondingly the same group delay, the original IPDs/ITDs are preserved by the binaural noise reduction scheme. For the discussed filtering concepts, the preservation of IPDs/ILDs is ensured as the applied filter concepts are linear-phase systems with equal group delay (the filter lengths of the speech enhancement filters applied in the individual channels are always the same).

From (29), it can be verified that if a dual-channel filter approach (concept 1 (11)) is applied for binaural speech enhancement, it cannot be ensured that the ILD at the output of the scheme is still the same as the original ILD. Correspondingly, the dual-channel filter may drastically influence the ILDs so that a correct localization of all sources cannot be ensured. However, using a single filter (concept 2 (15)) for binaural speech enhancement, the original ILDs can be preserved and a correct localization of all sources can be guaranteed.

5. SIMULATION RESULTS

In the following, experimental results are shown and discussed in order to illustrate the previous analysis. The simulations were performed in a low-reverberation chamber ($T_{60} \approx 50$ms) and a living-room-like environment ($T_{60} \approx 250$ms). The source signal components were generated by convolving speech signals (both, target and interference components are speech signals) of duration 10s with recorded head-related impulse responses (HRIRs). The HRIRs were measured by using a pair of BTE hearing aid housings with two microphones and a single receiver (loudspeaker) inside each device (no processor). The cases were mounted on a real person and connected, via a pre-amplifier box, to a PC equipped with a multi-channel RME Multiface sound card. For these simulations only the HRIRs for the frontal microphones are used. As we assume that the location of the desired source is only approximately known, for estimating the target angular position and the interfering signal component, a BSS-based blocking matrix is applied (for more details see [6]). For all simulations the sampling frequency was set to $f_s = 16$kHz. The filter length for the blocking matrix was set to 1024, and 512 coefficients were used for the speech enhancement filters (concept 1 (11), concept 2 (15)). The trade-off parameter μ (see (11) and (15)) is set to 0.85 after informal listening tests.

5.1 Performance of Wiener filtering concepts

First of all, the Wiener filtering concepts given by (11) and (15), respectively, are compared with respect to noise reduction (NR), speech distortion (SD), and $\mathrm{SIR_{gain}}$. For $w = w_i$, noise reduction NR and speech distortion SD are given by (assuming all $q \geq 2$ are interfering noise sources)

$$\mathrm{NR}_i = \frac{\int_{\Omega=0}^{2\pi} \mathbf{h}_i^H \mathbf{U}^T \hat{\mathbf{P}}_{\mathbf{ss}} \mathbf{U} \mathbf{h}_i d\Omega}{\int_{\Omega=0}^{2\pi} |w_i|^2 \mathbf{h}_i^H \mathbf{U}^T \hat{\mathbf{P}}_{\mathbf{ss}} \mathbf{U} \mathbf{h}_i d\Omega} \quad i \in \{1,2\}, \tag{30}$$

$$\mathrm{SD}_i = \frac{\int_{\Omega=0}^{2\pi} |1 - w_i|^2 |h_{1i}|^2 \hat{P}_{ss} d\Omega}{\int_{\Omega=0}^{2\pi} |h_{1i}|^2 \hat{P}_{ss} d\Omega} \quad i \in \{1,2\}. \tag{31}$$

The $\mathrm{SIR_{gain}}$ is given by

$$\mathrm{SIR_{gain_i}} = \frac{\mathrm{SIR_{out_i}}}{\mathrm{SIR_{in_i}}} \quad i \in \{1,2\}, \tag{32}$$

$$\mathrm{SIR_{in_i}} = \frac{\int_{\Omega=0}^{2\pi} |h_{1i}|^2 \hat{P}_{ss} d\Omega}{\int_{\Omega=0}^{2\pi} \mathbf{h}_i^H \mathbf{U}^T \hat{\mathbf{P}}_{\mathbf{ss}} \mathbf{U} \mathbf{h}_i d\Omega} \quad i \in \{1,2\}, \tag{33}$$

$$\mathrm{SIR_{out_i}} = \frac{\int_{\Omega=0}^{2\pi} |w_i|^2 |h_{1i}|^2 \hat{P}_{ss} d\Omega}{\int_{\Omega=0}^{2\pi} |w_i|^2 \mathbf{h}_i^H \mathbf{U}^T \hat{\mathbf{P}}_{\mathbf{ss}} \mathbf{U} \mathbf{h}_i d\Omega} \quad i \in \{1,2\}. \tag{34}$$

For the following evaluation, the target speaker was located at $0°$ (broadside direction) and the interferer position varied from $25°$ to $335°$ in steps of $5°$ (clockwise). The range $-20° \leq \phi \leq 20°$ is defined to be the target angular range, whereas the actual target angular position in this range is estimated by the applied BSS-based blocking matrix [6]. Long-time estimates of (30) - (32) are obtained by averaging results obtained in individual data blocks. Results are only depicted for the right channel since both channels produce almost identical results. Both concepts are compared with a (nonrealizable) channel-specific optimum Wiener filter (see (17)). To compute the optimum Wiener filter we assume to have ideal individual noise estimates.

Fig. 3 depicts the results for (30) - (32) obtained for the dual-channel linear-phase filter approach (concept 1 (11)) in dB. The abbreviations LRC and LR stand for low-reverberation chamber and living-room environment, respectively. The performance of the dual-channel Wiener filter drastically decreases for interferer positions ϕ_{Int} around $150° - 210°$ because of the front-back ambiguity of a two-microphone setup. The noise reduction performance for the ipsilateral channel ($\phi_{\mathrm{Int}} \approx 90°$) is very close to the noise reduction performance of the optimum Wiener filter (see Fig. 3a). However, considering speech distortion (Fig. 3b), it can be verified that it is much higher than for the optimum Wiener filter. Given the good speech suppression performance of the applied blocking matrix (see [6] for more information) it is obvious that not only the residual speech components contained in the common noise estimate lead to this high speech distortion. This must result from strong deviations of the common noise estimate from the channel-specific noise components. Correspondingly, even for the ipsilateral channel where the interfering signal is dominant ($\phi_{\mathrm{Int}} \approx 90°$) the common noise estimate greatly differs from the channel-specific noise components and optimum weights cannot be obtained. This is also confirmed by the lower SIR gain depicted in Fig. 3c. For the contralateral channel ($\phi_{\mathrm{Int}} \approx 270°$), the noise reduction as well as the speech distortion shown in Fig. 3a and Fig. 3b, respectively, are much higher than for the optimum Wiener filter. Because of the strong influence of the noise components contained in the ipsilateral channel, the estimated common noise signal strongly deviates from the channel-specific noise component in the contralateral channel. Correspondingly, in the contralateral channel many frequencies are strongly attenuated resulting in a high suppression of speech and noise components which was addressed in Sect. 4.1. The SIR gain (Fig. 3c) is in this case close to the optimum Wiener filter but this does not give any hint about the quality of the speech as well as the noise components.

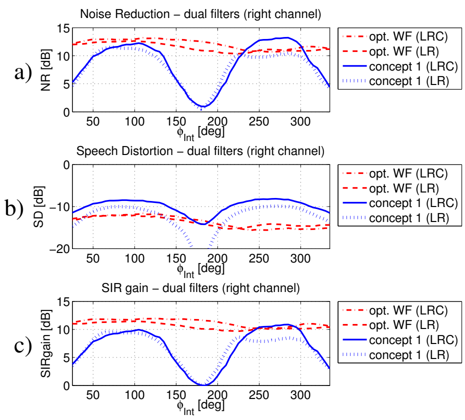

Figure 3: Performance of the dual-channel Wiener filter

Fig. 4 depicts the results for (30) - (32) for the single Wiener filter (concept 2 (15)) in dB. This approach has the same behavior for interferer positions ϕ_{Int} around $150° - 210°$ because of the front-back ambiguity of the two-microphone setup. In contrast to concept 1 (11), this approach results in a lower noise reduction performance (compare Fig. 4a and Fig. 3a) as well as a lower SIR gain (compare Fig. 4c and Fig. 3c). These quantities decrease about 2-3dB. However, more importantly, the speech distortion reduces about 3-4dB in contrast to concept 1 (compare Fig. 4b and Fig. 3b) which is very important especially for hearing aid applications. The reason for the lower performance of concept 2 compared to concept 1 in terms of noise reduction and SIR gain results from the fact that in concept 2 it is taken into account that the common noise estimate is made up of the noise components contained in both channels (the denominator in (15) is given as a sum of the auto PSDs of both microphone signals). The performance of the single filter is still lower than that of the optimum individual filter as the cross correlation of the noise components in both channels (see (21)) still influences the spectral weights. As this cross correlation cannot be easily obtained, optimum weights cannot be calculated.

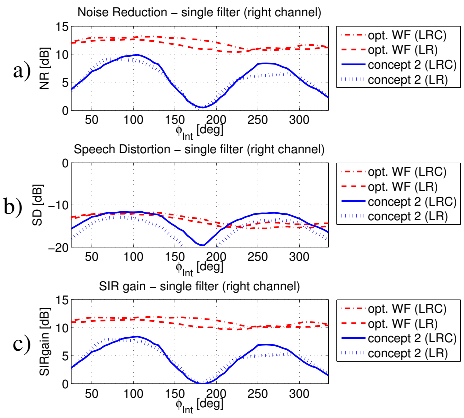

Figure 4: Performance of the single Wiener filter

In Fig. 5, the spectral weights for both concepts as well as for the optimum separate Wiener filter are depicted (for both channels). The desired speaker was located at 0° (broadside direction) and the interferer position was $\phi_{\text{Int}} \approx 90°$. These weights are obtained as

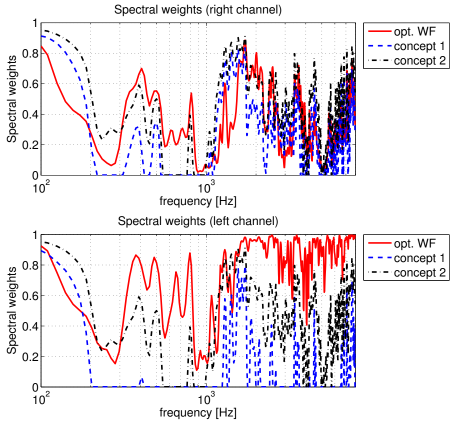

Figure 5: Spectral weights for the right and left channel of the different Wiener filter concepts and the optimum Wiener filter. The desired speaker is located at 0° and the interferer is located at 90°.

follows: The spectral weights for the individual channels were calculated according to (11), (15), and (17) for concept 1, concept 2, and the optimum Wiener filter, respectively. The depicted filter weights are long-time estimates, i.e., the filter weights obtained in individual segments are averaged. The results depicted in Fig. 5 further explain the behavior of both Wiener filtering concepts. For concept 1 (dashed line), it can be clearly verified that especially for the contralateral channel (left channel) many frequencies are strongly attenuated and thus resulting in a high noise reduction as well as a high speech distortion compared to the optimum Wiener filter (solid line). This is caused by the strong deviation of the common noise estimate of the individual noise components contained in the sensor signals. If the strong attenuation in the range of 200Hz - 1kHz is undesired, the gain factor μ could be reduced ($\mu < 1$). This effect is alleviated for concept 2 (dash-dot line) and correspondingly results in a lower noise reduction, but more importantly, also in a lower speech distortion compared to concept 1.

5.2 Influence on ILDs

As pointed out in Sect. 4.3, both Wiener filtering concepts do not have any influence on IPDs/ILDs which humans exploit for the localization of sources at low frequencies. However, it was shown that concept 1 may influence ILDs which are necessary for a correct localization of sources especially at higher frequencies whereas concept 2 does not influence the ILDs at all. In the following, this claim shall be experimentally verified. To this end, the ILDs given by (28) and (29) are calculated for both Wiener filtering concepts. The ILDs at the input and at the output of the proposed scheme are depicted in dB in Fig. 6 for both concepts. These results are valid for the low-reverberation chamber ($T_{60} \approx 50$ms). The ILDs are obtained as follows: The position of the source varied from $\phi_{\text{Int}} = 0°$ to $\phi_{\text{Int}} = 360°$ in steps of 5°. For each position, a (long-time) estimate of the PSD of the interfering speech signal is obtained by using the Welch method. For this, a frame length of 512 samples (32ms) is chosen and the overlap between consecutive frames is 50%. The ILDs are then obtained by

$$\text{ILD}_{\text{in/out}} = \begin{cases} 10\log_{10}\left(\dfrac{\hat{p}^{\text{in/out}}_{n_1 n_1}}{\hat{p}^{\text{in/out}}_{n_2 n_2}}\right) \text{ dB} & \text{for } 0° \leq \phi_{\text{Int}} \leq 180° \\ 10\log_{10}\left(\dfrac{\hat{p}^{\text{in/out}}_{n_2 n_2}}{\hat{p}^{\text{in/out}}_{n_1 n_1}}\right) \text{ dB} & \text{for } 180° < \phi_{\text{Int}} < 360° \end{cases} \tag{35}$$

for the input and output, respectively. These results reveal that the dual-channel filter strongly affects the ILDs for most of the interferer positions (dashed line Fig. 6) whereas the ILDs are preserved for the single filter (dash-dotted line Fig. 6). For source positions $30° - 150°$ as well as $210° - 330°$ the original ILDs are drastically reduced by the dual-channel filter approach (dashed line Fig. 6). The strong influence of the dual-channel filter on the ILDs may affect the localization of sources. However, the effect of influencing ILDs by this Wiener filter concept has to be assessed by listening tests with hearing impaired people.

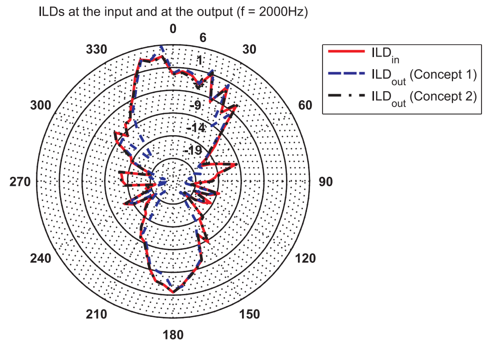

Figure 6: ILDs at the input and at the output of the proposed scheme for both Wiener filtering concepts ($f = 2$kHz)

6. CONCLUSION

An analysis of two generic Wiener filtering approaches in binaural hearing aids was presented. This analysis included the investigation of the combination of a common noise estimate with two fundamentally different Wiener-filter approaches as well as their impact on ILDs. ITDs are preserved as the speech enhancement filters are linear-phase filter approaches. We demonstrated that a dual-channel filter approach will lead to higher noise reduction and to a higher SIR gain compared to a single filter approach. In contrast to the dual-channel Wiener filter, the single Wiener filter approach achieves a lower speech distortion that is important for speech intelligibility. Besides, a dual-channel filter approach will impair the ILDs that in turn may affect a correct localization of sources. Further work will focus on the improvement of the performance of the Wiener-type filters.

REFERENCES

[1] T. J. Klasen, S. Doclo, T. Bogaert, M. Moonen, and J. Wouters, “Binaural multi-channel Wiener filtering for hearing aids: Preserving interaural time and level differences,” in *IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Toulouse, France, May 2006, pp. 145–148.

[2] T. J. Klasen, T. Bogaert, M. Moonen, and J. Wouters, “Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues,” *IEEE Trans. Signal Processing*, vol. 55, no. 4, pp. 1579–1585, 2007.

[3] K. Reindl, Y. Zheng, and W. Kellermann, “Speech enhancement for binaural hearing aids based on blind source separation,” in *4th Int. Symp. on Communications, Control, and Signal Proc. (ISCCSP)*, Limassol, Cyprus, March 2010.

[4] H. Puder, “Method and acoustic signal processing system for binaural noise reduction,” German Patent, File number 09004196.3, Application date 24.03.2009.

[5] J. Chen, J. Benesty, Y. Huang, and S. Doclo, “New Insights into the Noise Reduction Wiener Filter,” *IEEE Trans. Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1218-1234, July 2006.

[6] Y. Zheng, K. Reindl, and W. Kellermann, “BSS for improved interference estimation for blind speech signal extraction with two microphones,” in *IEEE Int. Workshop on Comp. Advances in Multi-Sensor Adapt. Proc. (CAMSAP)*, Aruba, Dutch Antilles, Dec. 2009.