NOISE POWER SPECTRAL DENSITY ESTIMATION FOR BINAURAL NOISE REDUCTION EXPLOITING DIRECTION OF ARRIVAL ESTIMATES

Daniel Marquardt, Simon Doclo

University of Oldenburg, Department of Medical Physics and Acoustics and Cluster of Excellence Hearing4All, Oldenburg, Germany
{daniel.marquardt,simon.doclo}@uni-oldenburg.de

ABSTRACT

Noise reduction algorithms for head-mounted assistive listening devices are crucial to improve speech quality and intelligibility in background noise. For binaural hearing devices with one microphone per device, the noise power spectral density (PSD) is commonly estimated using various assumptions about the acoustic scenario. Since these methods lack robustness if the underlying assumptions are not satisfied, alternatively the noise PSD can be estimated at the output of a blocking matrix, however requiring an estimate of the relative transfer function (RTF) or direction of arrival (DOA) of the desired speech source. For constructing the blocking matrix, in this paper we exploit RTF estimates using the covariance whitening method and DOA estimates obtained from a binaural DOA estimator using anechoic prototype acoustic transfer functions (ATFs). Simulation results in a realistic cafeteria scenario show that exploiting DOA estimates for binaural noise PSD estimation leads to an improved noise reduction performance, especially in the presence of directional interfering speakers.

Index Terms— Binaural noise reduction, binaural hearing aids, binaural cues, noise PSD, DOA estimation

1. INTRODUCTION

For head-mounted assistive listening devices (e.g., hearing aids, cochlear implants), algorithms that use the microphone signals from both the left and the right hearing device are considered to be promising techniques for noise reduction, because the spatial information captured by all microphones can be exploited [1]. In addition to reducing noise and limiting speech distortion, another important objective of binaural noise reduction algorithms is the preservation of the listener's impression of the acoustical scene, in order to exploit the binaural hearing advantage and to avoid confusions due to a mismatch between acoustical and visual information. To achieve binaural noise reduction with binaural cue preservation two main concepts have been developed. The first concept is to apply a complex-valued filter to all available microphone signals on the left and the right hearing device using binaural extensions of spatial filtering techniques [2, 3, 4, 5, 6, 7]. In the second concept, which is considered in this paper, a common real-valued, spectro-temporal gain is applied to the reference microphone signals in the left and the right hearing device. This common gain is either calculated based on the output of a multi-microphone noise reduction algorithm [8, 9] or by explicitly exploiting the spatial information between the microphones using several assumptions about the acoustic scenario [10, 11, 12]. This processing strategy allows for perfect preservation of the instantaneous binaural cues of both the speech and the noise component, but inevitably introduces distortions of the speech and the residual noise component. However, if only one microphone per hearing device is available, these techniques are of particular interest since the noise reduction performance of distortionless algorithms is rather limited, especially for diffuse noise fields which are commonly encountered in binaural applications.

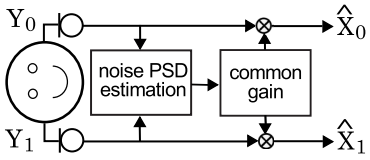

Figure 1: Binaural noise reduction scheme using a common spectro-temporal gain.

In this paper, we consider two strategies for binaural noise PSD estimation, which on the one hand is based on solving a set of equations using the assumption of a single desired speech source in a diffuse noise field similarly to [10] and on the other hand based on the output of a blocking matrix similarly to [12, 11]. For constructing the blocking matrix, we propose a binaural DOA estimator based on the generalized cross-correlation with phase transform (GCC-PHAT) features [13]. The estimated noise PSD is then used in the speech-distortion-weighted Wiener filter [14], where the trade-off parameter is selected such that speech distortion and noise reduction are jointly minimized in the L-curve sense [15]. The speech enhancement performance of the considered noise PSD estimators is evaluated in a realistic cafeteria scenario. The objective performance measures PESQ and STOI indicate that exploiting DOA estimates for binaural noise PSD estimation leads to an improved speech quality and intelligibility.

2. CONFIGURATION AND NOTATION

Consider the binaural configuration in Figure 1, consisting of one microphone per hearing device. In the short-time Fourier transform (STFT) domain, the microphone signal of the left hearing device $Y_0(l,k)$ can be written as

$$Y_0(l,k) = X_0(l,k) + N_0(l,k), \tag{1}$$

with l the frame index, k the frequency index, $X_0(l,k)$ the speech component and $N_0(l,k)$ the noise component. The microphone signal of the right hearing device $Y_1(l,k)$ is defined similarly. For the sake of readability the frequency index k will be omitted in the remainder of this paper, except where explicitly required. We define the 2-dimensional signal vector $\mathbf{Y}(l)$ as

$$\mathbf{Y}(l) = [Y_0(l), Y_1(l)]^T, \qquad \mathbf{Y}(l) = \mathbf{X}(l) + \mathbf{N}(l), \tag{2}$$

where the vectors $\mathbf{X}(l)$ and $\mathbf{N}(l)$ are defined similarly as $\mathbf{Y}(l)$. Considering an acoustical scenario with one desired speech source

This work was supported in part by the joint Lower Saxony-Israeli Project ATHENA financially supported by the State of Lower Saxony and the Cluster of Excellence 1077 "Hearing4All", funded by the German Research Foundation (DFG).

978-1-5386-1631-4/17/$31.00 ©2017 IEEE

$S_{\mathrm{x}}(l)$, the speech component $\mathbf{X}(l)$ can be written as

$$\mathbf{X}(l) = S_{\mathrm{x}}(l)\mathbf{A}(l), \tag{3}$$

with $\mathbf{A}(l)$ the ATF vector between the speech source and the microphones of the left and the right hearing device. Assuming statistical independence between the speech and the noise component, the correlation matrix of the microphone signals $\mathbf{R}_{\mathrm{y}}(l)$ can be written as

$$\mathbf{R}_{\mathrm{y}}(l) = \mathcal{E}\left\{\mathbf{Y}(l)\mathbf{Y}^H(l)\right\} = \mathbf{R}_{\mathrm{x}}(l) + \mathbf{R}_{\mathrm{n}}(l) = \begin{bmatrix} \Phi_{\mathrm{y},0}(l) & \Phi_{\mathrm{y},01}(l) \\ \Phi^*_{\mathrm{y},01}(l) & \Phi_{\mathrm{y},1}(l) \end{bmatrix}, \tag{4}$$

with $\mathbf{R}_{\mathrm{x}}(l)$ the speech correlation matrix, $\mathbf{R}_{\mathrm{n}}(l)$ the noise correlation matrix,

$$\Phi_{\mathrm{y},0}(l) = \mathcal{E}\left\{|Y_0(l)|^2\right\} = \Phi_{\mathrm{x},0}(l) + \Phi_{\mathrm{n},0}(l), \tag{5}$$

$$\Phi_{\mathrm{y},1}(l) = \mathcal{E}\left\{|Y_1(l)|^2\right\} = \Phi_{\mathrm{x},1}(l) + \Phi_{\mathrm{n},1}(l), \tag{6}$$

$$\Phi_{\mathrm{y},01}(l) = \mathcal{E}\left\{Y_0(l)Y_1(l)^*\right\} = \Phi_{\mathrm{x},01}(l) + \Phi_{\mathrm{n},01}(l), \tag{7}$$

the PSDs and Cross Spectral Density (CSD) of the signal component, $\Phi_{\mathrm{x},0}$, $\Phi_{\mathrm{x},1}$ and $\Phi_{\mathrm{x},01}$ the PSDs and CSD of the speech component, and $\Phi_{\mathrm{n},0}$, $\Phi_{\mathrm{n},1}$ and $\Phi_{\mathrm{n},01}$ the PSDs and CSD of the noise component. In the case of a diffuse noise field, the noise correlation matrix $\mathbf{R}_{\mathrm{n}}(l)$ can be written as

$$\mathbf{R}_{\mathrm{n}}(l) = \Phi_{\mathrm{n}}(l)\mathbf{\Gamma}, \qquad \Phi_{\mathrm{n}}(l) = \Phi_{\mathrm{n},0}(l) = \Phi_{\mathrm{n},1}(l), \tag{8}$$

with $\mathbf{\Gamma}$ the time-invariant spatial coherence matrix of a diffuse noise field and $\Phi_{\mathrm{n}}(l)$ the time-varying diffuse noise PSD. In the case of a single desired speech source (cf. (3)), the speech PSDs and CSD are related as

$$\Phi_{\mathrm{x},0}(l) = \Phi_{\mathrm{x},01}(l)H^*(l) = \Phi_{\mathrm{x},1}(l)|H(l)|^2, \tag{9}$$

with

$$H(l) = \frac{A_0(l)}{A_1(l)}, \tag{10}$$

the RTF between the speech component in both microphones. In the next section, we address binaural noise PSD estimation using the system of equations (5) - (7) and the output of a blocking matrix.

3. BINAURAL NOISE PSD ESTIMATION

In [10, 12, 11] two different approaches for binaural noise PSD estimation based on the assumption of a single desired speech source and a diffuse noise field have been presented. Using (8) and (9) the system of equations (5) - (7) can be written as

$$\Phi_{\mathrm{y},0}(l) = \Phi_{\mathrm{x},0}(l) + \Phi_{\mathrm{n}}(l), \tag{11}$$

$$\Phi_{\mathrm{y},1}(l) = \Phi_{\mathrm{x},0}(l)|H(l)|^{-2} + \Phi_{\mathrm{n}}(l), \tag{12}$$

$$\Phi_{\mathrm{y},01}(l) = \Phi_{\mathrm{x},0}(l)(H^*(l))^{-1} + \Phi_{\mathrm{n}}(l)\Gamma_{01}, \tag{13}$$

containing 3 unknowns, i.e. $\Phi_{\mathrm{x},0}(l), \Phi_{\mathrm{n}}(l)$ and $H(l)$. Solving for the noise PSD $\Phi_{\mathrm{n}}(l)$ leads to [10]

$$\Phi_{\mathrm{n}}(l) = \frac{b(l) - \sqrt{b^2(l) + 4a\left(|\Phi_{\mathrm{y},01}(l)|^2 - \Phi_{\mathrm{y},0}(l)\Phi_{\mathrm{y},1}(l)\right)}}{2a}, \tag{14}$$

with

$$a = 1 - |\Gamma_{01}|^2, \;\; b(l) = \Phi_{\mathrm{y},0}(l) + \Phi_{\mathrm{y},1}(l) - 2\Re\left\{\Phi_{\mathrm{y},01}(l)\Gamma^*_{01}\right\}. \tag{15}$$

This solution is equivalent to calculating the smallest eigenvalue of the prewhitened signal correlation matrix $\mathbf{\Gamma}^{-1}\mathbf{R}_{\mathrm{y}}(l)$, which has been used for late reverberant PSD estimation in [16]. It should be noted that the noise PSD in (14) only requires the assumption of a diffuse noise field and a single desired speech. However, in realistic scenarios with reverberation and arbitrary noise fields, the underlying assumptions are not necessarily fulfilled. Furthermore, the expected value of the noisy speech correlation matrix $\mathbf{R}_{\mathrm{y}}(l)$ is typically approximated using recursive averaging with a short time constant, typically capturing some correlation between the speech and the noise component. Hence, it might be beneficial to also exploit information about the desired speech source, like the RTF or the DOA, to first construct a blocking matrix and then estimate the noise PSD based on the blocking matrix output as proposed in [12, 11]. The PSD at the output of the blocking matrix $\Phi_{\mathrm{b}}(l)$ can be calculated as

$$\Phi_{\mathrm{b}}(l) = \mathcal{E}\left\{|Y_0(l) - H(l)Y_1(l)|^2\right\}. \tag{16}$$

Assuming perfect blocking of the speech component, i.e., a perfect estimate of the RTF, the blocking matrix output is equal to

$$\Phi_{\mathrm{b}}(l) = \Phi_{\mathrm{n},0}(l) + |H(l)|^2\Phi_{\mathrm{n},1}(l) - 2\Re\left\{H(l)^*\Phi_{\mathrm{n},01}(l)\right\}. \tag{17}$$

Again, assuming a diffuse noise field (cf. (8)), the noise PSD can be calculated as [12, 11]

$$\Phi_{\mathrm{n}}(l) = \frac{\Phi_{\mathrm{b}}(l)}{1 + |H(l)|^2 - 2\Re\left\{H^*(l)\Gamma_{01}\right\}}. \tag{18}$$

Comparing (14) with (18), both PSD estimators require the assumption of a single speech source and a diffuse noise field, while the blocking matrix based estimator in (18) additionally requires an estimate of the RTF of the speech component $H(l)$, which is inherently included in the estimator in (14). Furthermore, the estimator in (14) also requires the assumption that the speech and the noise component are uncorrelated, which is not required in (18), however replaced by the assumption of a perfect blocking of the speech component. Consequently, both estimators contain several potential sources of error and hence a comparison of the performance of both estimators for speech enhancement in a realistic cafeteria scenario is provided in Section 6.

In [11], the RTF, required for the estimator in (18), has been estimated by means of blind system identification approaches, however resulting in a biased estimate in the case of a noisy scenario. An unbiased estimate of the RTF can be directly calculated from the system of equations in (11) - (13), which is equivalent to using the covariance whitening method [17] for the special case of a diffuse noise field. Since estimating an unbiased RTF inherently requires an estimate of the noise PSD and contains the same possible sources of error as the noise PSD estimator in (14), alternatively, instead of directly estimating the RTF, we also estimate the DOA of the desired speech source and use anechoic prototype RTFs related to the estimated DOA to approximate $H(l)$. A comparison of the performance of both, the RTF and the DOA based noise PSD estimators, is provided in Section 6.

4. BINAURAL DOA ESTIMATION

Several procedures have been proposed for binaural DOA estimation, e.g., based on (biased) RTFs [18], beamforming [19] or pretrained models [20]. In this section, we propose a binaural DOA estimator using the generalized cross-correlation with phase transform (GCC-PHAT) features [13]. The normalized cross-correlation between the two reference microphones is equal to

$$GCC(l,k) = \frac{\Phi_{\mathrm{y},01}(l,k)}{|\Phi_{\mathrm{y},01}(l,k)|} = e^{j\angle\left(\Phi_{\mathrm{x},01}(l,k) + \Phi_{\mathrm{n},01}(l,k)\right)}, \tag{19}$$

with $\Phi_{\mathrm{y},01}$ defined in (7). In the case of a general noise field, i.e. $\Phi_{\mathrm{n},01} \neq 0$, (19) reflects a biased estimate of the phase of the RTF. In the case of a spatially white noise field, i.e. $\Phi_{\mathrm{n},01} = 0$, (19) is equal to the normalized RTF of the desired speech source, i.e.

$$GCC(l,k) = e^{j\angle H(l,k)}, \tag{20}$$

with $H(l)$ defined in (10). Since the distance between the microphones of the left and the right hearing device is typically between

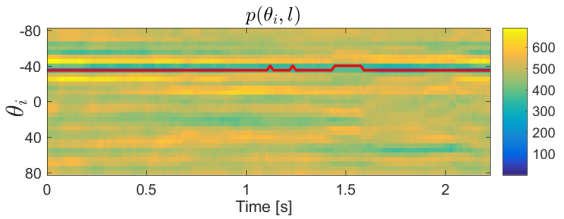

Figure 2: Across-frequency averaged squared norm $p(\theta_i, l)$ and estimated DOA $\theta_s(l)$ (solid red line) for a speech source at -35° in a cafeteria environment with recorded ambient noise at an average input iSNR of -10 dB (cf. Section 6 for detailed information).

$15 - 20$ cm, in the case of a diffuse noise field the assumption of a spatially white noise field is quite valid, especially for frequencies above $1 - 2$ kHz, motivating the use of GCC-PHAT features for binaural applications. From the GCC-PHAT features, the DOA is typically estimated by identifying the (oversampled) delay related to the location of the largest peak in the time-domain representation of the normalized cross-correlation. In the free-field case, mapping this delay to the estimated DOA can be realized using a plane-wave model [13], while for the binaural case the mapping function can be calculated using anechoic IR measurements or head models. Alternatively, instead of only relying on the largest peak of the normalized cross-correlation, we use the entire information contained in the GCC-PHAT features by calculating the squared norm between the GCC-PHAT features and normalized anechoic prototype RTFs for each time index l and frequency index k, i.e.,

$$P(\theta_i, l, k) = \left| GCC(l,k) - e^{j\angle H_d(\theta_i,k)} \right|^2, \tag{21}$$

with $H_d(\theta_i, k)$ the anechoic prototype RTF for direction θ_i. Similarly as for the mapping function, these anechoic prototype RTFs can be obtained using head models or measured IRs. Contrary to peak picking, using (21) allows for a narrowband DOA estimate as, e.g., in [18, 19], which may be beneficial in the case of multiple sources due to the sparse representation of directional sources in the STFT domain. Since in this paper we specifically aim to estimate the DOA of a single desired speech source in very noisy conditions, in order to increase the robustness of the DOA estimator we average $P(\theta_i, l, k)$ across frequency in order to obtain a broadband DOA estimate, i.e.,

$$p(\theta_i, l) = \frac{1}{K-1} \sum_{k=1}^{K} P(\theta_i, l, k), \tag{22}$$

and the DOA of the desired speech source can then be calculated as

$$\theta_s(l) = \underset{\theta_i}{\operatorname{argmin}}\, p(\theta_i, l). \tag{23}$$

Since it is quite difficult to differentiate between sources from the frontal and the rear hemisphere, we use the common assumption that the desired speech source is located in the frontal hemisphere. Hence, we limit the subset of considered DOAs in (23) to angles ranging from -80° to 80°. Figure 2 exemplary depicts the estimated DOA for a source at -35° in a cafeteria environment with recorded ambient noise at an average intelligibility-weighted input SNR (iSNR) [21] of -10 dB (cf. Section 6 for detailed information).

5. SPECTRAL FILTERING

For speech enhancement in the left and the right hearing device we use the speech-distortion-weighted Wiener filter [14], i.e.,

$$G_0(l,k) = \frac{\xi_0(l,k)}{\mu_0(l,k) + \xi_0(l,k)}, \; G_1 = \frac{\xi_1(l,k)}{\mu_1(l,k) + \xi_1(l,k)}, \tag{24}$$

where the parameters $\mu_0(l,k)$ and $\mu_1(l,k)$ provide a trade-off between noise reduction and speech distortion and $\xi_0(l,k)$ and $\xi_1(l,k)$ are the a-priori SNRs in the left and the right hearing device, respectively. The trade-off parameters $\mu_0(l,k)$ and $\mu_1(l,k)$ are chosen equal to

$$\mu_0(l,k) = \frac{1}{\xi_0(l,k)}, \qquad \mu_1(l,k) = \frac{1}{\xi_1(l,k)}, \tag{25}$$

jointly minimizing speech distortion and noise reduction in the L-curve sense [15]. The a-priori SNR is estimated using the decision directed approach [22], exploiting the noise PSD estimates.

In order to preserve the binaural cues of the speech and the residual noise component a common gain is calculated as the geometrical mean of the Wiener gains G_0 and G_1, i.e.,

$$G(l,k) = \sqrt{G_0(l,k)G_1(l,k)} = \frac{\xi_0(l,k)\xi_1(l,k)}{\sqrt{(1+\xi_0^2(l,k))(1+\xi_1^2(l,k))}}, \tag{26}$$

and the minimum gain is set to $G_{\min} = 0.1$. The speech component in the left and the right hearing device is then estimated as

$$\widehat{X}_0(l,k) = G(l,k)Y_0(l,k), \quad \widehat{X}_1(l,k) = G(l,k)Y_1(l,k). \tag{27}$$

6. SIMULATIONS

In this section, we present simulation results for a cafeteria scenario comparing the performance of the considered binaural noise PSD estimators.

The binaural input signals have been generated using measured impulse responses for a binaural hearing aid setup mounted on a dummy head in a cafeteria ($T_{60} \approx 1250$ ms) [23], where each hearing aid was equipped with 1 microphone. For different spatial scenarios, the angular positions of the speech source and the distance between the dummy head and the speaker are presented in Table 1. As desired speech signal, sentences from the British Oldenburg Sentence Test (OLSA) database were chosen, where each sentence has a length of about 2 s. For the first experiment, we added recorded ambient noise (including babble noise, clacking plates and occasionally occurring interfering speakers), recorded in the same cafeteria [23], to the speech signal. For the second experiment, additionally two constantly active interfering speakers, positioned at -90° and 150°, have been added to the mixture. For each spatial scenario (cf. Table 1), 30 OLSA sentences have been concatenated and evaluated. For each OLSA sentence, the average iSNR was set to -10 dB and 0 dB and for the second experiment the average intelligibility-weighted input Signal-to-Interference Ratio (iSIR) was set to 0 dB for each interfering source.

To calculate the anechoic prototype RTFs required for DOA estimation (cf. Section 4) and the blocking matrix in (16), we use anechoic ATFs measured on the same dummy head [23]. The ATFs where measured for angles ranging from -180° to 175° in steps of 5°. The spatial coherence between both microphones Γ_{01}, required in the noise PSD estimators (14) and (18), is calculated using spatially averaged auto and cross-correlations of the anechoic ATFs [5]. The

	Experiment 1		Experiment 2	
	SC1	SC2	SC3	SC4
Desired Source	-35°	0°	-35°	0°
Distance	117.5 cm	102 cm	117.5 cm	102 cm
Interfering Sources	-		$\{-90, 150\}^\circ$	
iSNR	$\{-10, 0\}$ dB			
iSIR	-		0 dB	

Table 1: Spatial scenarios (0° - frontal direction. -90° - left hand side. 90° - right hand side).

signals are processed using a weighted overlap-add STFT framework with a frame size of 512 samples and an overlap of 50% at a sampling frequency of 16 kHz. For noise PSD estimation, the PSDs $\Phi_{\mathrm{y},0}$, $\Phi_{\mathrm{y},1}$, Φ_{b} and the CSD $\Phi_{\mathrm{y},01}$ are estimated using recursive averaging with a time constant of 40 ms, while for RTF and DOA estimation the PSDs $\Phi_{\mathrm{y},0}$, $\Phi_{\mathrm{y},1}$ and the CSD $\Phi_{\mathrm{y},01}$ are estimated using recursive averaging with a time constant of 200 ms. The time constant for the recursive averaging parameter in the decision directed approach was set to 200 ms.
The considered noise PSD estimators are denoted as:

- **EIG**: Noise PSD estimator according to (14).
- **DOA, DOA-OPT**: Noise PSD estimator according to (18), based on the estimated DOA as described in Section 4 (**DOA**) and based on the ground truth DOA (**DOA-OPT**).
- **RTF, RTF-OPT**: Noise PSD estimator according to (18), based on the estimated RTF using the covariance whitening method [17] (**RTF**) and the ground truth RTF calculated from the cafeteria impulse responses (**RTF-OPT**).
- **OPT**: Optimal noise PSD estimated from the noise component using recursive averaging with a time constant of 40 ms.

The performance was evaluated using the objective measures PESQ [24] and STOI [25].
Experiment 1: The results for SC1 and SC2 are depicted in Figure 3. All noise PSD estimators generally show an improvement in PESQ and STOI compared to the input signal, while the DOA based noise PSD estimator (**DOA**) generally outperforms the eigenvalue based estimator (**EIG**) and the RTF based estimator (**RTF**). Especially for an input iSNR of 0 dB, **RTF** generally shows a slightly better performance compared to **EIG**. For an input iSNR of -10 dB, using the ground truth DOA (**DOA-OPT**) only slightly improves the performance compared to **DOA**, indicating a very good performance of the proposed DOA estimator even for very noisy scenarios. Contrary, using the ground truth RTF (**RTF-OPT**) significantly improves the performance compared to **RTF** and expectedly shows a better performance then using the ground truth DOA.
Experiment 2: The results for SC3 and SC4, additionally comprising directional interfering sources, are depicted in Figure 4. For both scenarios, **EIG** and **RTF** generally decrease the PESQ and the STOI score, where especially **RTF** shows a very non-robust performance to deviations from the assumed signal model. Contrary, the DOA-based noise PSD estimator (**DOA**) shows a rather robust performance and generally shows an improvement of up to 0.08 (STOI) and 0.4 (PESQ) compared to the input signal. Contrary to the first experiment, using the ground truth DOA (**DOA-OPT**) generally shows an improvement compared to **DOA**, especially for low input iSNRs, due to the additional presence of two directional interfering speakers.
For both experiments, considering the performance of **OPT**, i.e., using the recursively averaged noise PSD, there is quite some room for improvement which is subject to further research.

7. CONCLUSIONS

In this paper we proposed a binaural DOA estimator which has been used in a blocking matrix to estimate the noise PSD for binaural hearing devices. Simulation results in a realistic cafeteria scenario show that exploiting DOA estimates for binaural noise PSD estimation improves the noise reduction performance, especially in the case of additional directional interfering speakers in a diffuse noise field. Generalizing the method for binaural devices with multiple microphones in combination with distortionless beamformers such as the binaural LCMV remains subject for further research.

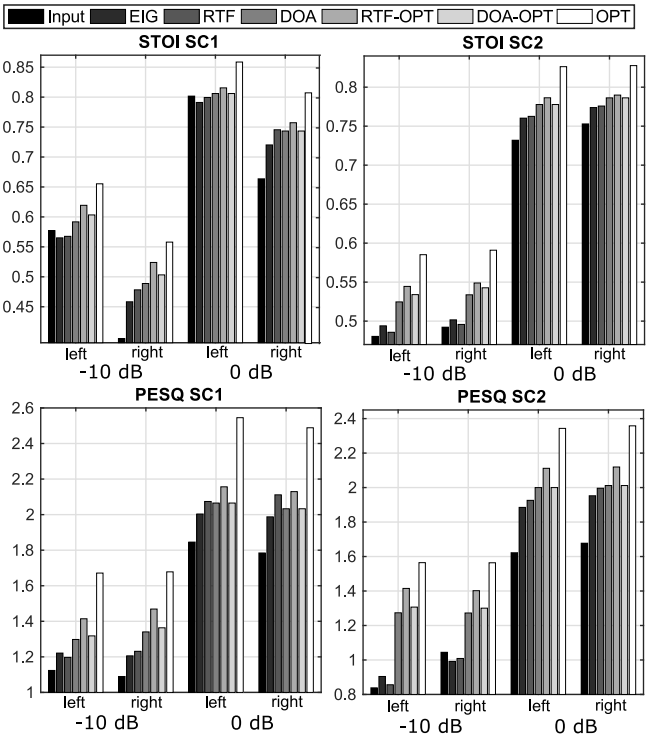

Figure 3: STOI and PESQ results for the left and the right hearing device for experiment 1 (SC1 and SC2).

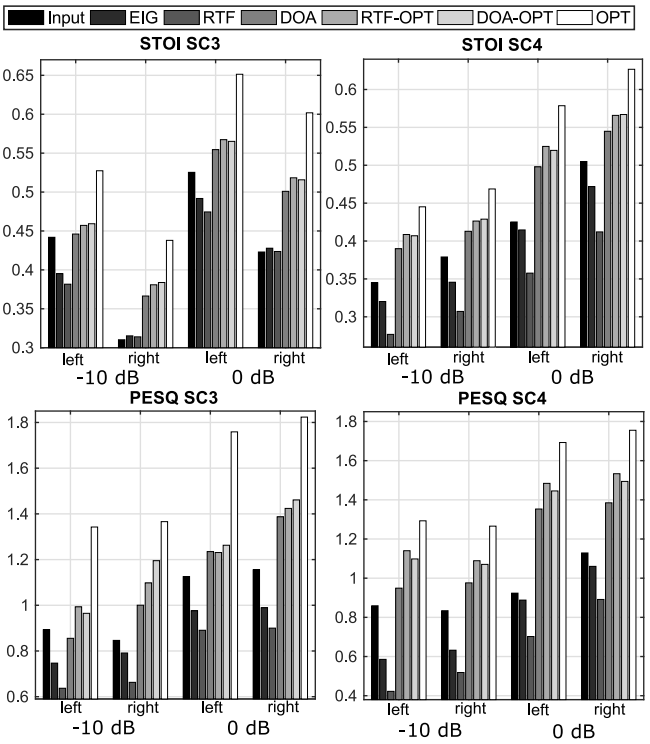

Figure 4: STOI and PESQ results for the left and the right hearing device for experiment 2 (SC3 and SC4).

8. REFERENCES

[1] S. Doclo, W. Kellermann, S. Makino, and S. Nordholm, "Multichannel Signal Enhancement Algorithms for Assisted Lis-

tening Devices: Exploiting spatial diversity using multiple microphones," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 18–30, Mar. 2015.

[2] R. Aichner, H. Buchner, M. Zourub, and W. Kellermann, "Multi-channel source separation preserving spatial information," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Honolulu HI, USA, Apr. 2007, pp. 5–8.

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

[4] D. Marquardt, V. Hohmann, and S. Doclo, "Interaural Coherence Preservation in Multi-channel Wiener Filtering Based Noise Reduction for Binaural Hearing Aids," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2162–2176, Dec. 2015.

[5] D. Marquardt, E. Hadad, S. Gannot, and S. Doclo, "Theoretical Analysis of Linearly Constrained Multi-channel Wiener Filtering Algorithms for Combined Noise Reduction and Binaural Cue Preservation in Binaural Hearing Aids," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2384–2397, Dec. 2015.

[6] E. Hadad, D. Marquardt, S. Doclo, and S. Gannot, "Theoretical Analysis of Binaural Transfer Function MVDR Beamformers with Interference Cue Preservation Constraints," *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, vol. 23, no. 12, pp. 2449–2464, Dec. 2015.

[7] E. Hadad, D. Marquardt, W. Pu, S. Gannot, S. Doclo, Z.-Q. Luo, I. Merks, and T. Zhang, "Comparison of two binaural beamforming approaches for hearing aids," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, USA, Mar. 2017, pp. 236–240.

[8] T. Lotter and P. Vary, "Dual-channel speech enhancement by superdirective beamforming," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 1–14, 2006.

[9] R. Baumgärtel, M. Krawczyk-Becker, D. Marquardt, C. Völker, H. Hu, T. Herzke, G. Coleman, K. Adiloğlu, S. M. A. Ernst, T. Gerkmann, S. Doclo, B. Kollmeier, V. Hohmann, and M. Dietz, "Comparing binaural signal processing strategies I: Instrumental evaluation." *Trends in Hearing*, vol. 19, pp. 1–16, 2015.

[10] A. Kamkar-Parsi and M. Bouchard, "Improved Noise Power Spectrum Density Estimation for Binaural Hearing Aids Operating in a Diffuse Noise Field Environment," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 4, pp. 521–533, May 2009.

[11] M. Azarpour, G. Enzner, and R. Martin, "Binaural noise PSD estimation for binaural speech enhancement," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, May 2014, pp. 7068–7072.

[12] K. Reindl, Y. Zheng, A. Schwarz, S. Meier, R. Maas, A. Sehr, and W. Kellermann, "A stereophonic acoustic signal extraction scheme for noisy and reverberant environments," *Computer Speech & Language*, vol. 27, no. 3, pp. 726 – 745, 2013.

[13] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, Aug. 1976.

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

[15] I. Kodrasi, D. Marquardt, and S. Doclo, "Curvature-based optimization of the trade-off parameter in the speech distortion weighted multichannel Wiener filter," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, Apr. 2015, pp. 315–319.

[16] I. Kodrasi and S. Doclo, "Late reverberant power spectral density estimation based on an eigenvalue decomposition," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, USA, Mar. 2017, pp. 611–615.

[17] S. Markovich-Golan and S. Gannot, "Performance analysis of the covariance subtraction method for relative transfer function estimation and comparison to the covariance whitening method," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, Apr. 2015.

[18] S. Braun, W. Zhou, and E. A. P. Habets, "Narrowband direction-of-arrival estimation for binaural hearing aids using relative transfer functions," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct. 2015, pp. 1–5.

[19] M. Zohourian, G. Enzner, and R. Martin, "On the use of beamforming approaches for binaural speaker localization," in *Proc. ITG Symposium on Speech Communication*, Oct 2016, pp. 1–5.

[20] H. Kayser and J. Annemüller, "A discriminative learning approach to probabilistic acoustic source localization," in *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*, Juan-les-Pins, France, Sep. 2014, pp. 99–103.

[21] J. E. Greenberg, P. M. Peterson, and P. M. Zurek, "Intelligibility-weighted measures of speech-to-interference ratio and speech system performance," *Journal of the Acoustical Society of America*, vol. 94, no. 5, pp. 3009–3010, Nov. 1993.

[22] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.

[23] H. Kayser, S. Ewert, J. Annemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Database of multichannel In-Ear and Behind-The-Ear Head-Related and Binaural Room Impulse Responses," *Eurasip Journal on Advances in Signal Processing*, vol. 2009, p. 10 pages, 2009.

[24] ITU-T, *Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs P.862*, International Telecommunications Union (ITU-T) Recommendation, Feb. 2001.

[25] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An Algorithm for Intelligibility Prediction of Time-Frequency Weighted Noisy Speech," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2125–2136, Sep. 2011.