BINAURAL NOISE PSD ESTIMATION FOR BINAURAL SPEECH ENHANCEMENT

Masoumeh Azarpour, Gerald Enzner, and Rainer Martin

Institute of Communication Acoustics
Ruhr-Universität Bochum, Germany
{masoumeh.azarpour, gerald.enzner, rainer.martin}@rub.de

ABSTRACT

In this paper we propose a novel binaural algorithm to estimate the power spectral density (PSD) of the background noise at the left and right ear separately. Inspired by equalization-cancelation considered in binaural hearing, the target speech is canceled at both left and right ears by means of the FLMS (fast least-mean square) algorithm. Assuming the ideal equalization, the output error of the blocking filter is a biased estimation of noise PSD. The estimated noise PSD is further corrected by exploiting the estimated left and right interaural transfer functions and the coherence model of the noise field. In addition to noise power estimation assessment, the estimated noise PSD is integrated in a binaural speech enhancement framework in order to evaluate the overall noise reduction performance.

Index Terms— Binaural speech processing, Noise PSD estimation, Blocking Filter

1. INTRODUCTION AND RELATION TO PRIOR WORK

The performance of voice communication devices remarkably degrades in the presence of ambient noise, which is often found in daily communication scenarios. Therefore, noise reduction algorithms are an important component of modern communication system, for instance hearing aids. In many noise reduction algorithms primary knowledge of noise statistics is a prerequisite for proper target speech enhancement [1, 2]. Thus, different noise power estimators have been proposed, which can be generally categorized in: 1) single-channel and 2) dual-channel methods. Even though several well established single-channel estimators have been introduced thus far (e.g., [3–6]), the performance of these estimators which use only one input signal, is fairly limited as they do not exploit the spatial information, e.g., coherence properties. Furthermore, they cannot perform well in some cases where the background noise statistics vary rapidly in time.

In [7], a dual-channel noise power estimator based on cross-power spectral density of the left and right noisy signals has been proposed. It has been assumed that the left and right noise signals are uncorrelated. Later the authors in [8] derived a generalized dual-channel noise PSD estimator based on [7] taking the coherence model (spherically isotropic) of the noise signal into account. Recently, a further dual channel estimator based on the work in [8] has been proposed as a combination of two methods operating in different frequency bands [9]. In low frequencies a single-channel noise PSD estimator based on speech presence probability (SPP) [10] was utilized. The noise power estimator in high frequencies is a dual-channel estimator that exploits the knowledge about the coherence of noise and speech signals. The coherence of noise and speech were estimated and updated recursively in speech presence and noise dominant frames respectively. The combination of the single-channel estimator in low frequencies and the dual-channel one in high frequencies was found to be an efficient approach to take advantage of both classes. Furthermore, it has been proposed to use algorithms with an intermediate blocking filter, e.g., [11–14], in order to estimate the power spectrum of the interference signal. In [13, 15], the blind source separation (BSS) has been utilized to suppress the target speech and consequently estimate the noise power. The compensation gain has been derived based on the BSS demixing matrix to reduce the bias of estimation.

In this paper we propose a binaural noise PSD estimator for binaural hearing devices based on target signal blocking. While there is a fairly good number of noise power estimators available in the literature that can be utilized in binaural speech enhancement [16, 17], a binaural noise PSD estimator that can provide the binaural estimates of noise power at the left and the right ear separately has not been proposed. In almost all of the algorithms the noise power at the left and right ear have been assumed to be equal. The proposed algorithm utilizes the binaural interaural transfer functions, which are estimated by means of the FLMS algorithm. Depending on the signal-to-noise ratio, the binaural interaural transfer function blocking (ITFB) only partly acts as the desired target blocking filter and hence the output error is biased with respect to the noise estimation goal. However, in a generalized sense, we can still deduce two equations for two unknowns of the binaural noise PSDs. Further considering the prior assumption about the coherence properties of noise signals, different approaches are utilized to correct the estimated binaural noise PSD. This method does not require any voice activity detection or speech presence estimation. Moreover, in this paper different noise estimation errors are considered: 1) noise signal distortion due to blocking, 2) noise underestimation owing to uncorrelated noise assumption and 3) noise overestimation due to the speech leakage as a result of imperfect blocking. The first two error sources will be discussed and compensated in this work.

The remainder of this paper is organized as follows. In Sec. 2 the binaural signal model is described. The motivation of deriving binaural noise PSD estimates will be elaborated in Sec. 3. Then, Sec. 4 reviews the proposed binaural noise estimator. The experimental results and conclusions are presented in Secs. 5 and 6, respectively.

2. BINAURAL SIGNAL MODEL

The left and right binaural microphone signals $y_i(k) = s(k) * h_i(k) + n_i(k) = x_i(k) + n_i(k)$, $i \in \{r, l\}$, at sampling time index k are expressed as a convolution of target speech $s(k)$ with binaural room impulse responses $h_i(k)$ immersed in ambient background noises $n_i(k)$. In the algorithms proposed in this paper, the vectors $\mathbf{y}_i(k) = \begin{bmatrix} y_i(k) & y_i(k-1) & \ldots & y_i(k-L+1) \end{bmatrix}^T$ of L successive samples are used, where superscript $(.)^T$ denotes vector transposition. The signals $\mathbf{x}_i(k)$ and $\mathbf{n}_i(k)$ are defined in the same way as $\mathbf{y}_i(k)$, thus $\mathbf{y}_i(k) = \mathbf{x}_i(k) + \mathbf{n}_i(k)$. It is assumed

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

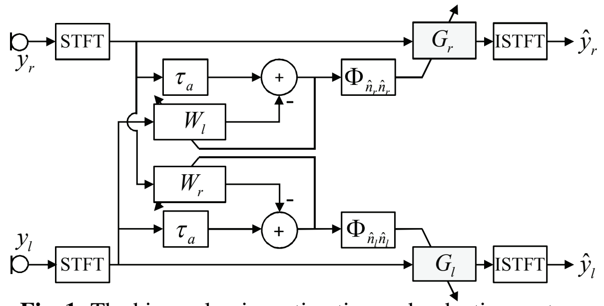


Fig. 1. The binaural noise estimation and reduction system.

that signals $\mathbf{n}_i(k)$ and $\mathbf{x}_i(k)$ are zero mean and uncorrelated. The corresponding frequency-domain signal model is expressed as

$$Y_i(\mu, \kappa) = X_i(\mu, \kappa) + N_i(\mu, \kappa), \qquad i \in \{l, r\}, \tag{1}$$

where $\kappa \in \mathbb{Z}$ and $\mu = 1, ..., M$ are the frame index and the frequency bin, respectively. The time-varying power spectral densities (PSDs) of noise and noisy signals are defined as $\Phi_{n_in_i}(\mu, \kappa) = E\{|N_i(\mu, \kappa)|^2\}$ and $\Phi_{y_iy_i}(\mu, \kappa) = E\{|Y_i(\mu, \kappa)|^2\}$ respectively, where $E\{.\}$ denotes the statistical expectation operator.

3. MOTIVATION OF THE BINAURAL NOISE PSD

In a real acoustic scenario the left and right noise PSDs are not exactly the same especially when point sources are presented. However, they are theoretically equal in the left and right ear in ideal homogeneous spherically isotropic noise fields. In order to clarify the motivation of separate estimation of left and right PSDs, the logarithmic error between the power spectral density of left and right noise signals is calculated in two frequency bands as

$$LogErr = \frac{1}{MK} \sum_{\mu=\mu_0}^{M_0} \sum_{\kappa=1}^{K} \left| 10 \log_{10} \frac{\hat{\Phi}_{N_l(\mu,\kappa)}}{\hat{\Phi}_{N_r(\mu,\kappa)}} \right|. \tag{2}$$

In the lower band we have $\mu_0 = 0$ and $M_0 = M\frac{f_c}{f_s}$ and in the upper band $\mu_0 = M\frac{f_c}{f_s} + 1$ and $M_0 = M$ ($f_c = 1$ kHz and $M = 512$). The obtained logarithmic error for different noise types, recursive and block-wise estimation procedures are illustrated in Fig. 2. The investigated noise signals are taken from the ETSI database [18]. Moreover, the computer-generated WGN (white Gaussian noise) and babble noise were produced using the algorithm proposed in [19] considering the ideal spherically isotropic noise field model. As can be seen, the left and right recursively estimated noise PSD ($\alpha = 0.9$) are not equal even in the computer-generated case. The block-wise PSD estimation is performed using the Welch method [20] and the logarithmic error ratio is averaged over frequency bins. The equal noise PSDs can only be found in stationary computer-generated WGN noise if a large amount of data is available for block-wise PSD computation, which is not granted in real scenarios.

4. BINAURAL NOISE ESTIMATION AND REDUCTION

Figure 1 presents the block diagram of the proposed binaural noise reduction algorithm. It contains two main parts: 1) the blocking via estimated interaural transfer function W_l and W_r and hence 2) the binaural noise power estimation and speech enhancement.

As can be seen, the adaptive identification FLMS algorithm for W_l and W_r operates with respect to a causality delay τ_a. Due to the limitation of space, we will not describe it in details here. Basic information can be found in [21–23]. In this work, the filter order is $L = 256$, the causality delay is $\tau_a = 60$, the forgetting factor for signal power smoothing is $\gamma = 0.9$ and the step size is $\mu = 0.1$. The single-channel Wiener filter is used to obtain the enhanced left and right signals with a $G_{min} = -20$ dB spectral floor via

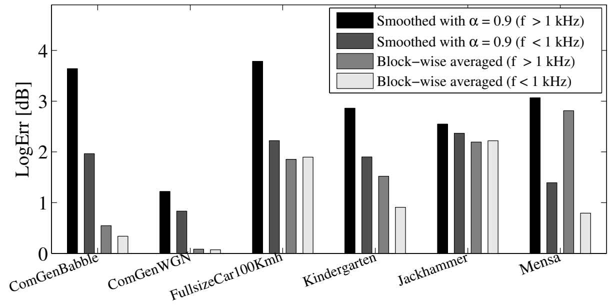


Fig. 2. Comparison of LogErr between left and right target noise PSDs in high and low frequencies for different noise types

$$G_i = \max(1 - \frac{\hat{\Phi}_{n_in_i}}{\Phi_{y_iy_i}}, G_{min}), \quad i \in \{l, r\}. \tag{3}$$

4.1. Interaural Adaptive Blocking Filter

The interaural transfer function estimation error of the FLMS algorithm is written as

$$e_l(k) = y_l(k) - \hat{\mathbf{w}}_r^T(k)\mathbf{y}_r(k) \tag{4a}$$

$$e_r(k) = y_r(k) - \hat{\mathbf{w}}_l^T(k)\mathbf{y}_l(k), \tag{4b}$$

where $\hat{\mathbf{w}}_i$ is the time domain representation of estimated filter $\hat{W}_i$. The FLMS was implemented in the frequency domain and the estimated filter $\hat{W}_i$ is adaptively controlled to minimize the mean-squared-error signals [21, 7.23]. Assuming perfect interaural transfer function estimation in noise-free conditions and therefore ideal equalization, the target microphone signal will be canceled such that:

$$\begin{aligned} e_l(k) &= x_l(k-\tau_a) + n_l(k-\tau_a) - \hat{\mathbf{w}}_r^T(k)\mathbf{x}_r(k) - \hat{\mathbf{w}}_r^T(k)\mathbf{n}_r(k) \\ &\approx n_l(k-\tau_a) - \hat{\mathbf{w}}_r^T(k)\mathbf{n}_r(k) \end{aligned} \tag{5a}$$

$$\begin{aligned} e_r(k) &= x_r(k-\tau_a) + n_r(k-\tau_a) - \hat{\mathbf{w}}_l^T(k)\mathbf{x}_l(k) - \hat{\mathbf{w}}_l^T(k)\mathbf{n}_l(k) \\ &\approx n_r(k-\tau_a) - \hat{\mathbf{w}}_l^T(k)\mathbf{n}_l(k), \end{aligned} \tag{5b}$$

By computing the PSD of (5), a system of equations including the left and right noise PSD can be obtained. As was previously mentioned, the output of blocking filtering can be considered as a biased estimation of the noise PSD, i.e., $\Phi_{\hat{n}_i\hat{n}_i} = \Phi_{e_ie_i}$:

$$\Phi_{e_le_l} = \Phi_{n_ln_l} + \left|\hat{W}_r\right|^2 \Phi_{n_rn_r} - 2\text{Re}\{e^{j\frac{2\pi}{M}\mu\tau_a}\hat{W}_r\Phi_{n_ln_r}\} \tag{6a}$$

$$\Phi_{e_re_r} = \Phi_{n_rn_r} + \left|\hat{W}_l\right|^2 \Phi_{n_ln_l} - 2\text{Re}\{e^{j\frac{2\pi}{M}\mu\tau_a}\hat{W}_l\Phi_{n_ln_r}\} \tag{6b}$$

where $\Phi_{e_ie_i}$ is estimated using the first-order recursive equation:

$$\hat{\Phi}_{e_ie_i}(\mu, \kappa) = \alpha\hat{\Phi}_{e_ie_i}(\mu, \kappa - 1) + (1 - \alpha)\left|E_i(\mu, \kappa)\right|^2. \tag{7}$$

and $E_i(\mu, \kappa)$ is the STFT (Short-time Fourier transform) of the output error of the adaptive filter as in (5). The power spectral density of the left and the right noise signals, $\hat{\Phi}_{n_ln_l}$ and $\hat{\Phi}_{n_rn_r}$, are derived by solving the simultaneous equations in (6) and, consequently, the noise signal distortion due to blocking will be corrected. In this process, two different coherence models are investigated: 1) an uncorrelated noise field, and 2) ideal homogeneous spherically isotropic (diffuse) noise. For sake of clarity, the frame index κ and frequency bin μ will be omitted hereafter.

4.2. Uncorrelated Noise Assumption

Assuming uncorrelated noises in the left and right microphone signals, which is a reasonable assumption for a diffuse noise field above a cut-off frequency f_c, $\Phi_{n_l n_r}$ is equal to zero. Therefore, (6) will be a system of linear equations. By solving the equations, the PSD of the left and the right noise signal can be derived as

$$\hat{\Phi}_{n_l n_l} = \frac{\hat{\Phi}_{e_l e_l} - \left|\hat{W}_r\right|^2 \hat{\Phi}_{e_r e_r}}{1 - \left|\hat{W}_l\right|^2 \left|\hat{W}_r\right|^2} \tag{8}$$

$$\hat{\Phi}_{n_r n_r} = \frac{\hat{\Phi}_{e_r e_r} - \left|\hat{W}_l\right|^2 \hat{\Phi}_{e_l e_l}}{1 - \left|\hat{W}_l\right|^2 \left|\hat{W}_r\right|^2}.$$

The important point which should be mentioned here is the ill-conditioned case that might occur. In very high SNR conditions, where the FLMS performs ideally well, the estimated left and right interaural transfer functions are inverses of each other and thus $\left|\hat{W}_l \hat{W}_r\right| \approx 1$. It has been observed that in very high SNR the denominator in (8) drops to about -25 dB. However, since the algorithm is proposed for noise reduction where the highest SNRs are not the main concern, the above mentioned problem will not affect the performance of the algorithm in terms of noise PSD estimation error. Nevertheless, in order to avoid getting negative value and a division by small values, the absolute values of numerator and denominator in (8) are limited to 0 and 0.01 respectively.

Unfortunately, the uncorrelated noise assumption results in underestimation of noise PSDs especially in low frequencies where the noise components are always correlated. Therefore, the estimated PSD should be compensated at low frequencies. This will be addressed in the following.

4.3. Low-frequency Compensation (Diffuse Model)

In order to overcome the underestimation of noise power at low frequencies, the uncorrelated noise signal assumption must be modified. Since several practical noise situations can be reasonably modeled as a diffuse noise field, we employ the diffuse noise assumption for the considered noise signals. In this case the coherence of the noise signals is given by [24]

$$\Gamma_{n_l n_r}(\mu) = \mathrm{sinc}\left(\frac{2\mu f_s d}{Mc}\right), \tag{9}$$

where $c = 340$ m/s is the sound velocity and $d = 0.17$m is the distance between microphones.

Moreover, it has been observed that the assumptions of equal noise PSDs at two microphones are more plausible at low frequencies than at high frequencies, see for instance Fig. 2. Therefore, assuming equal noise PSD $\Phi_{n_l n_l} = \Phi_{n_r n_r}$ at two microphones in equation (6) helps to simplify the linear equation. Consequently, $\Phi_{n_l n_r}$ can be expressed based on the left and right noise PSD and the diffuse field coherence function, i.e., $\Phi_{n_l n_r} = \Gamma_{n_l n_r}\Phi_{n_l n_l} = \Gamma_{n_l n_r}\Phi_{n_r n_r}$. Therefore, the PSD can be obtained as

$$\hat{\Phi}_{n_l n_l} = \frac{\hat{\Phi}_{e_l e_l}}{1 + \left|\hat{W}_l\right|^2 - 2\mathrm{Re}\{e^{j\frac{2\pi}{M}\mu\tau_a}\hat{W}_l \Gamma_{n_l n_r}\}} \tag{10}$$

$$\hat{\Phi}_{n_r n_r} = \frac{\hat{\Phi}_{e_r e_r}}{1 + \left|\hat{W}_r\right|^2 - 2\mathrm{Re}\{e^{j\frac{2\pi}{M}\mu\tau_a}\hat{W}_r \Gamma_{n_l n_r}\}}.$$

5. EXPERIMENTAL RESULTS

In order to comprehensively evaluate the proposed algorithm, we compared the proposed binaural algorithm with two dual-channel estimators and a single-channel one: the cross-power-spectral-density method (**CPSD**) [7], the improved CPSD method (**ImCPSD**) [16] and the single-channel SPP-based method (**SC-SPP**) [10]. Our method that directly relies on the blocking-error of the interaural-transfer-function (i.e., without bias-correction) is termed **ITFBo**. The bias-corrected blocking method employing the diffuse noise model assumption (10) is given the acronym **ITFBd**. The combination of ITFBd in low frequencies and the low-coherence-based noise PSD according to (8) in high frequencies is referred to as **ITFBc**. In addition, the unprocessed signal and the enhanced signal using a reference noise PSD are termed as **Unprocessed** and **Ref**.

5.1. Experimental Setup

The experiments have been conducted using the measured binaural meeting room impulse response ($T_{60} = 210$ ms) considering the head shadowing effect, taken from the Aachen room impulse response database [25]. As we do not consider explicitly the dereverberation task in this work, the BRIRs (binaural room impulse responses) have been cut off based on 70% of total energy in order to eliminate long reverberation tails. The binaural microphone outputs are generated by convolving the source signal with the BRIRs. The target source signal was a concatenation of five sentences taken from the TIMIT database [26], leading to a total of 60 s of data. In terms of the noise signal, the binaural babble noise generated by the algorithm proposed in [19] and the kindergarten and mensa noise from ETSI database [18] were used.

5.2. Algorithms Parameters

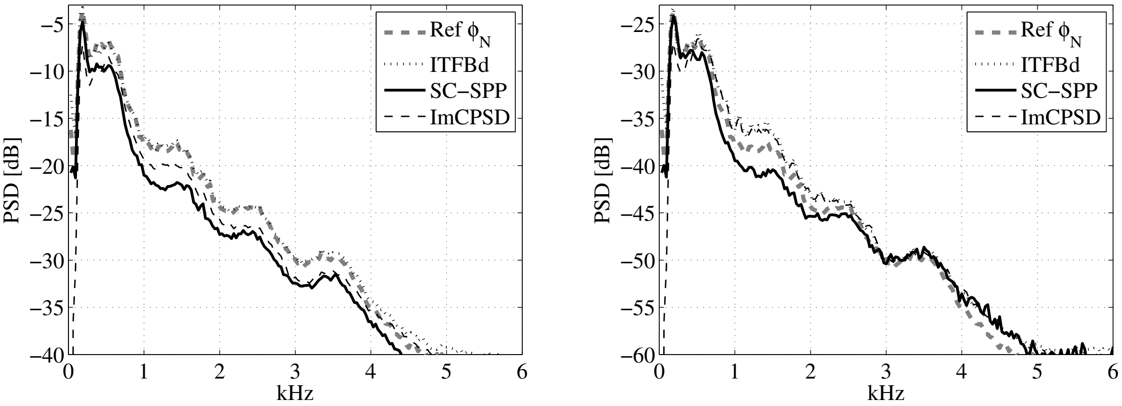
Throughout the experiments the considered signals, sampled at $f_s = 16$ kHz, are segmented into overlapping frames of length L'. The windowed frames (using a square-root Hanning window) are then transformed into the frequency domain via a discrete short-time Fourier transform (STFT) of length M with a 50% overlap and $L' = M = 2L = 512$. In order to make the comparison consistent, the smoothing factor for estimating (cross-) power spectral densities is set to the same fixed value ($\alpha = 0.8$) in all algorithms. The same error signal, i.e., (4a) was utilized to implement ImCPSD [16, $53-54$]. It has been realized that the phase compensation, as was introduced in (6), should be also taken into account in ImCPSD. Therefore, the ImCPSD was implemented considering the $e^{j\frac{2\pi}{M}\mu\tau_a}$ term in corresponding equations. Moreover, the same coherence model as described in (9) were used in the implementation of ImCPSD.

5.3. Evaluation

The performance of the algorithm has been evaluated in terms of speech enhancement and noise PSD estimation separately. The objective evaluation results have been averaged over all noise types. We compared the estimated noise PSD with a reference PSD in terms of LogErr defined as follows

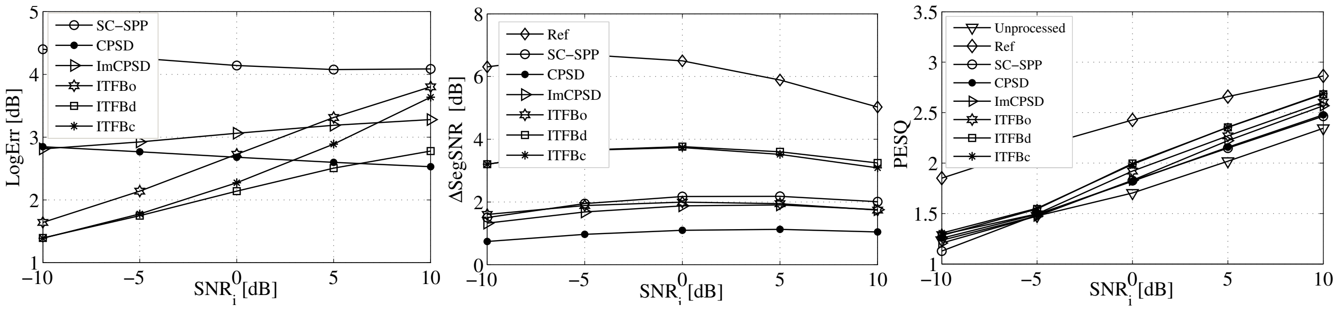
$$LogErr_{total} = \frac{1}{2MK}\sum_{i=l,r}\sum_{\mu=1}^{M}\sum_{\kappa=1}^{K}\left|10\log_{10}\frac{\Phi_{n_i n_i}(\mu,\kappa)}{\hat{\Phi}_{n_i n_i}(\mu,\kappa)}\right|. \tag{11}$$

The reference noise PSD is estimated similar to (7) and as mentioned before, the smoothing factor α was set to 0.8. The evaluation is performed when the transient phase of the adaptive FLMS algorithm



(a) (b)

Fig. 3. Comparison of reference noise PSD and estimated PSD of different algorithms a) SNR = -10 dB and b) SNR = 10 dB



(a) (b) (c)

Fig. 4. Comparison in terms of (a) LogErr , (b) ΔSegSNR, and (c) PESQ for different algorithms averaged over several noise types

has finished. The segmental SNR [2] and the perceptual evaluation of speech quality (PESQ) [27] were used to assess the overall speech enhancement performance of the algorithm. The evaluation measures are averaged over both ears.

The spectral noise estimation quality comparing the ImCPSD, SC-SPP and ITFBd algorithm at $+10$ dB and -10 dB SNR is given in Fig. 3. It can be seen from Fig. 3(a) that the ITFBd can provide a very good estimate of the reference noise signal at -10 dB. While in this case the ITFBd outperforms the other algorithms, it suffers from an slightly overestimation in the $+10$ dB SNR case, i.e., Fig. 3(b).

Figure 4(a) shows the total estimation error for each noise PSD estimator at different input SNR. The proposed ITFB algorithms perform best up to 5 dB SNR. In particular the estimation error is considerably decreased by applying the correction according to (8) and (10), i.e., ITFBd and ITFBc in comparison to ITFBo baseline. The ITFB performances are decreased at $+10$ dB due to increasing speech leakage through the blocking stage. In contrast to the proposed algorithm, the estimation error of the other algorithms especially CS-SPP is high in most of the cases and flat over input SNRs.

The result of segmental SNR improvement for different input SNR is illustrated in Fig. 4(b). The proposed algorithm outperforms all other investigated algorithms. Since the uncorrelated noise assumption is not valid for the investigated noise signals, the basic CPSD method fails to obtain a good segmental SNR improvement.

Figure 4(c) shows the PESQ score obtained by the algorithms. It can be clearly seen that all investigated algorithms improve the speech quality in terms of predicted PESQ score. The proposed ITFBd and ITFBc are slightly superior to the other algorithms. In all investigated algorithms, PESQ improves as the input SNR is increased. Comparing Fig. 4(a), 4(b) and 4(c) shows that the investigated algorithms were ranked differently in different measures. It can be seen that the lowest noise power estimation error does not necessarily guarantee significantly better quality of the enhanced speech in terms of PESQ. While PESQ was not originally introduced for evaluating noise reduction algorithms, our informal listening confirms the PESQ result.

6. CONCLUSION

We proposed a binaural noise PSD estimation algorithm for binaural signal processing. The proposed algorithm is based on blocking filtering that uses the estimated interaural transfer functions. The interaural transfer functions were estimated adaptively by means of the FLMS algorithm. The direct output of the blocking stage was considered as a biased estimate of the noise signal. The biased noise PSD was further corrected assuming two different coherence models. The inherent properties of the FLMS algorithm makes the blocking filter capable of suppressing the target speech signal at high SNRs and restoring the individual noise signal of left and right ear properly at low SNRs. The quality of the noise estimator then turns into consistent speech improvement results, but it has been also found that the noise estimation performance and the speech enhancement quality are not one-to-one related. Furthermore, we still find a significant gap between the best performing algorithm and the reference condition, which indicates a need for further improvements of noise PSD estimation.

7. REFERENCES

[1] P. Vary and R. Martin, *Digital Speech Transmission. Enhancement, Coding and Error Concealment*, John Wiley-Verlag, 2006.

[2] P. C. Loizou, *Speech Enhancement: Theory and Practice*, Boca Raton, 1st ed. edition, 2007.

[3] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech and Audio Processing*, vol. 9, no. 5, pp. 504–512, 2001.

[4] T. Gerkmann and R.C. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," *IEEE Trans. Audio, Speech and Language Processing*, vol. 20, no. 4, pp. 1383–1393, 2012.

[5] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *IEEE Trans. Speech and Audio Processing*, vol. 11, pp. 466–475, 2003.

[6] R. Yu, "A low-complexity noise estimation algorithm based on smoothing of noise power estimation and estimation bias correction," in *Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2009, pp. 4421–4424.

[7] M. Dörbecker and S. Ernst, "Combination of two-channel spectral subtraction and adaptive Wiener post-filtering for noise reduction and dereverberation," in *Proc. of European Signal Processing Conf. (EUSIPCO)*, 1996, p. 995998.

[8] M. Jeub, C. Nelke, H. Krüger, C. Beaugeant, and P. Vary, "Robust dual-channel noise power spectral density estimation," in *European Signal Processing Conference (EUSIPCO)*, 2011, pp. 2304– 2308.

[9] C. M. Nelke, C. Beaugeant, and P. Vary, "Dual microphone noise PSD estimation for mobile phones in hands-free position exploiting the coherence and speech presence probability," in *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013.

[10] T. Gerkmann and R. C. Hendriks, "Noise power estimation based on the probability of speech presence," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2011, pp. 145–148.

[11] J. Benesty, S. Makino, and J. Chen, *Speech Enhancement*, Springer, Berlin, 2005.

[12] A. Schwarz, K. Reindl, and W. Kellermann, "On blocking matrix-based dereverberation for automatic speech recognition," in *Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, 2012, pp. 1–4.

[13] 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.

[14] J. Li, S. Sakamoto, S. Hongo, M. Akagi, and Y. Suzuki, "Two-stage binaural speech enhancement with Wiener filter for high-quality speech communication," *Speech Communication*, vol. 53, no. 5, pp. 677 – 689, 2011.

[15] L. Wang, T. Gerkmann, and S. Doclo, "Noise PSD estimation using blind source separation in a diffuse noise field," in *Proc. Int. Workshop on Acoustic Signal Enhancement (IWAENC)*, 2012, pp. 1–4.

[16] A. H. Kamkar-Parsi and M. Bouchard, "Improved noise power spectrum density estimation for binaural hearing aids operating in a diffuse noise field environment," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 17, no. 4, pp. 521–533, 2009.

[17] Y. Ji, Y. C. Park, D. W. Kim, and J. Sohn, "Robust noise PSD estimation for binaural hearing aids in time-varying diffuse noise field," in *Proc. Int. Conf on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013, pp. 7264 – 7268.

[18] Speech ETSI EG 202 396-1 V1.2.3, multimedia Transmission Quality (STQ); Speech quality performance in the presence of background noise; Part 1: Background noise simulation technique, and background noise database, 2009.

[19] E. A. P. Habets, I. Cohen, and S. Gannot, "Generating nonstationary multisensor signals under a spatial coherence constraint," *J. Acoust. Soc. Am.,*, vol. 124, no. 5, pp. 2911–2917, 2008.

[20] P.D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio Electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.

[21] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Upper Saddle River, New Jersey, 2001.

[22] M. Azarpour, G. Enzner, and R. Martin, "Distortionless-response vs. matched-filter-array processing for adaptive binaural noise reduction," in *Int. Workshop on Acoustic Signal Enhancement (IWAENC)*, 2012, pp. 1–4.

[23] M. Azarpour, G. Enzner, and R. Martin, "Adaptive binaural noise reduction based on matched-filter equalization and post-filtering," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013, pp. 7259 – 7263.

[24] H. Kuttruff, *Room Acoustics*, Spon Press, 5th ed edition, 2009.

[25] M. Jeub, M. Schäfer, and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," in *Proc. of Int. Conf. on Digital Signal Processing (DSP)*, 2009, pp. 1–4.

[26] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgren, "DARPA TIMIT acoustic phonetic continuous speech corpus CDROM," 1993.

[27] A.W. Rix, J.G. Beerends, M.P. Hollier, and A.P. Hekstra, "Perceptual evaluation of speech quality (PESQ) - a new method for speech quality assessment of telephone networks and codecs," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2001, vol. 2, pp. 749–752.