ROBUST DUAL-CHANNEL NOISE POWER SPECTRAL DENSITY ESTIMATION

Marco Jeub, Christoph Nelke, Hauke Krüger, Christophe Beaugeant[⋆]*, and Peter Vary*

Institute of Communication Systems and Data Processing (ind)
RWTH Aachen University, Germany
[⋆] Intel Mobile Communications, Sophia-Antipolis, France
{jeub,nelke,krueger,vary}@ind.rwth-aachen.de christophe.beaugeant@intel.com

ABSTRACT

In this paper, we propose a novel noise power spectral density (PSD) estimator which is beneficial for speech enhancement systems with two microphones in diffuse noise environments. The algorithm has a low computational complexity and requires low memory usage. The main advantage is that arbitrary models of the noise field coherence can be employed and a scalable extension of existing single-channel speech enhancement systems to dual channels is also possible. Experiments demonstrate with simulated and measured data that the proposed algorithm outperforms related algorithms in diffuse noise conditions.

1. INTRODUCTION

Algorithms for the reduction of background noise are nowadays essential components in many speech communication systems. Most mobile phones and hearing aids have integrated single- or multi-channel algorithms to enhance the speech quality in adverse environments. Among such algorithms, a predominant principle is the spectral subtraction technique which generally requires an estimate of the power spectral density (PSD) of the unwanted background noise. During the last decades, different single-channel noise PSD estimators have been proposed, cf. [1, 2] and the references therein. Disadvantages of existing algorithms are a high computational complexity, memory consumption, and the difficulties in estimating non-stationary noise. A low complexity single-channel algorithm that is also capable of tracking non-stationary noise has recently been proposed in [3]. Multi-channel noise PSD estimators for systems with two or more microphones have not been studied very intensively. In [4], a dual-channel spectral subtraction algorithm is proposed which uses the left and right signals of a binaural hearing aid system to estimate the noise PSD. However, this algorithm assumes uncorrelated noise between the different microphones which results in an underestimation of the noise PSD in realistic conditions [5]. Recently, a further binaural estimator based on [4] was proposed in [6] which is explicitly designed for binaural hearing aids.

In this contribution, we derive a generalized dual-channel noise PSD estimator which uses knowledge about the noise field coherence. It turns out that the approach of [4] can be seen as a special case for an uncorrelated background noise assumption. The novel algorithm has a low computational complexity and can be combined with different speech enhancement systems.

In contrast to related dual-channel noise reduction systems such as [7], the new approach allows for a scalable extension of an existing single-channel noise suppression system by exploiting a secondary microphone channel for a more robust noise estimation.

2. DUAL-CHANNEL SIGNAL MODEL

The two microphone signals $x_1(k)$ and $x_2(k)$ are the inputs of the dual-channel speech enhancement system and are related to clean speech $s(k)$ and additive background noise signals $n_m(k)$ by the signal model shown in Fig.1, with $m = 1, 2$ and discrete time index k. The noisy signals are termed $x_m(k)$. The acoustic transfer functions between source and the microphones are denoted by $H_m(e^{j\Omega})$. The required noise PSD

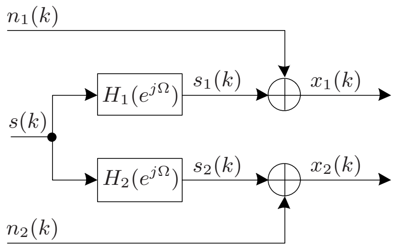

Figure 1: Dual-channel signal model.

estimate $\hat{\Phi}_{nn}(\lambda, \mu)$ for a frequency domain speech enhancement system is calculated by $x_1(k)$ and $x_2(k)$. Discrete frequency bin and frame index are denoted by μ and λ. Within the processing, the input signals $x_m(k)$ are first segmented into overlapping frames of length L. After windowing (e.g., applying a Hann window), these frames are transformed via FFT of length M ($M \geq L$) into the short-term spectral domain. The corresponding spectra are denoted by $X_m(\lambda, \mu)$.

3. COHERENCE-BASED NOISE PSD ESTIMATION

For the derivation of the dual-channel noise PSD estimator, the signal model shown in Fig.1 is considered. For the sake of brevity, we omit the frame and frequency indices (λ and μ) in the following equations.

The derivation is related to the discussions in [8], where an estimator for the *speech PSD* based on the noise field coherence was derived and incorporated in a Wiener filter rule for the reduction of diffuse background noise. The main advantage of our approach is a *noise PSD* estimate for versatile application in any spectral noise suppression rule. Throughout this paper, the complex coherence between the two microphone signals plays an important role. It is defined in the frequency domain by

$$\Gamma_{x_1x_2} = \frac{\Phi_{x_1x_2}}{\sqrt{\Phi_{x_1x_1}\,\Phi_{x_2x_2}}}, \tag{1}$$

with the auto-power spectral densities $\Phi_{x_1x_1}$ and $\Phi_{x_2x_2}$ as well as the cross-power spectral density $\Phi_{x_1x_2}$. Assuming

 - ISSN 2076-1465

that the noise sources ($n_m(k)$) are uncorrelated with the speech signals, the auto- and cross power spectra at the input of the speech enhancement system ($x_m(k)$) read

$$\Phi_{x_1x_1} = \Phi_{ss} + \Phi_{n_1n_1} \tag{2}$$
$$\Phi_{x_2x_2} = \Phi_{ss} + \Phi_{n_2n_2} \tag{3}$$
$$\Phi_{x_1x_2} = \Phi_{ss} + \Phi_{n_1n_2}, \tag{4}$$

with $\Phi_{ss} = \Phi_{s_1s_1} = \Phi_{s_2s_2}$. If the signal model of Fig.1 applies strictly, the coherence of the speech signals is $\Gamma_{s_1s_2} = 1$. In practice, the speech coherence is close to 1 if the source-microphone distance is smaller than the critical distance. The critical distance is defined as the distance from the source at which the sound energy due to the direct-path component is equal to the sound energy due to reverberation.

Furthermore, we assume that the noise field can be characterized as diffuse and hence, the coherence of the unwanted background noise $n_m(k)$ is close to zero, except for low frequencies. This was confirmed by measurements, as in [9, 8]. Additionally, we assume a homogeneous diffuse noise field, i.e.,

$$\Phi_{n_1n_1} = \Phi_{n_2n_2} = \Phi_{nn}. \tag{5}$$

Applying and reordering (1) for the noise signals $n_m(k)$ leads with (5) to

$$\Phi_{n_1n_2} = \Gamma_{n_1n_2} \sqrt{\Phi_{n_1n_1}\,\Phi_{n_2n_2}} = \Gamma_{n_1n_2} \cdot \Phi_{nn}, \tag{6}$$

with $\Gamma_{n_1n_2}$ being an arbitrary model for the noise field coherence such as (17). By using (5) and (2,3), the auto-power spectra can thus be formulated as

$$\Phi_{x_1x_1} = \Phi_{ss} + \Phi_{nn} \tag{7}$$
$$\Phi_{x_2x_2} = \Phi_{ss} + \Phi_{nn}. \tag{8}$$

It can be shown easily that the cross-power spectrum, which is obtained using (4) and (6), can be expressed as

$$\Phi_{x_1x_2} = \Phi_{ss} + \Gamma_{n_1n_2}\Phi_{nn}. \tag{9}$$

With the geometric mean of the two auto-power spectra

$$\sqrt{\Phi_{x_1x_1}\,\Phi_{x_2x_2}} = \Phi_{ss} + \Phi_{nn} \tag{10}$$

and the reordering of (9) to

$$\Phi_{ss} = \Phi_{x_1x_2} - \Gamma_{n_1n_2}\,\Phi_{nn}, \tag{11}$$

the following equation can be formulated:

$$\sqrt{\Phi_{x_1x_1}\,\Phi_{x_2x_2}} = \Phi_{x_1x_2} + \Phi_{nn}\left(1 - \Gamma_{n_1n_2}\right). \tag{12}$$

Finally, the desired real-valued noise PSD estimate reads

$$\Phi_{nn} = \frac{\sqrt{\Phi_{x_1x_1}\,\Phi_{x_2x_2}} - \mathrm{Re}\{\Phi_{x_1x_2}\}}{1 - \mathrm{Re}\{\Gamma_{n_1n_2}\}}, \tag{13}$$

where $1 - \mathrm{Re}\{\Gamma_{n_1n_2}\} > 0$ has to be ensured for the denominator, e.g., an upper threshold of the coherence $\Gamma_{\max} = 0.99$. The function $\mathrm{Re}\{\cdot\}$ returns the real part of its argument. It can be seen that for the special case of an uncorrelated noise field, i.e., $\Gamma_{n_1n_2} = 0$, the estimator reduces to the approach of [4]. In the practical implementation, the auto- and cross power spectra terms in (13) are replaced by their discrete short-time estimates using a recursive periodogram approach according to

$$\begin{aligned}\hat{\Phi}_{x_1x_1|x_2x_2}(\lambda,\mu) = \alpha\,\hat{\Phi}_{x_1x_1|x_2x_2}(\lambda-1,\mu)\\ + (1-\alpha)|X_{1|2}(\lambda,\mu)|^2,\end{aligned} \tag{14}$$

$$\begin{aligned}\hat{\Phi}_{x_1x_2}(\lambda,\mu) = \alpha\,\hat{\Phi}_{12}(\lambda-1,\mu)\\ + (1-\alpha)X_1(\lambda,\mu)\cdot X_2^*(\lambda,\mu),\end{aligned} \tag{15}$$

with smoothing factor $0 \le \alpha \le 1$ and periodograms $|X_m(\lambda,\mu)|^2$. The estimated noise PSD is denoted by $\hat{\Phi}_{nn}$.

Table 1: Main simulation parameters.

Parameter	Settings
Sampling frequency	$f_s = 16\,\mathrm{kHz}$
Frame length	$L = 320$ (20 ms)
FFT length	$M = 512$ (including zero-padding)
Frame overlap	50% (Hann window)
Smoothing factors	$\alpha = 0.9, \alpha_{\mathrm{DD}} = 0.98, \alpha_{\mathrm{nn}} = 0.9$

4. EXPERIMENTS AND RESULTS

In the experiments, the proposed coherence-based noise PSD estimator (*Proposed*) is compared to the reference noise PSD estimator (*Reference*) [4] as well as to the two single-channel noise PSD estimators: Minimum Statistics (*MS*) [1] and MMSE-based noise tracker (*MMSE*) [3]. Besides, a comparison to the dual-channel coherence-based Wiener filter rule (*Wiener*) [8] is given.

The experiment section is subdivided into three subsections. First the performance of the four noise PSD estimators in terms of the estimation error is given. Second, we use the estimators in combination with a noise suppression system which is evaluated with simulated data (*A*) as well as measured data (*B*). Besides, the influence of a possible attenuation of the desired speech signal among the two microphone channels is discussed followed by a short discussion on complexity issues in the third subsection. Since the proposed generalized concept of the noise PSD estimator is capable to employ arbitrary coherence models, we investigate the algorithm under two special cases assuming

- an uncorrelated noise field (as in [4]):

$$\Gamma_{n_1n_2} = 0 \tag{16}$$

- an ideal homogeneous spherically isotropic noise field:

$$\Gamma_{n_1n_2} = \mathrm{sinc}\left(\frac{2\pi f d_{\mathrm{mic}}}{c}\right), \tag{17}$$

  with distance d_{mic} between two omnidirectional microphones at frequency f and sound velocity c.

The generated dual-channel signals (*A*) are computed using the approach of [10], where predefined spatial coherence constraints and hence, different microphone distances, can be employed. Speech samples from the TSP speech database [11] are summed with babble noise from the ETSI background noise database [12]. Such simulation ensures reproducible results and the employment of objective evaluation measures, especially as the coherence of realistic noise fields such as a cocktail-party or office environment can be modeled by a diffuse noise field. Besides, we use binaural babble noise recordings (*B*) from the ETSI background noise database [12] which were measured with two microphones of a dummy head at $d_{\mathrm{mic}} = 0.15\,\mathrm{m}$. Even if the evaluation is performed on babble noise only, it has to be mentioned that the experiments showed the same tendency for other noise types. Further simulation parameters are listed in Tab.1. All upcoming plots use the legend shown in Fig. 2 which is omitted within the plots for visual clarity.

4.1 Estimation Accuracy

In a first experiment, the estimation error is evaluated. The signals are generated with a diffuse noise field constraint (*A*). The symmetric segmental logarithmic estimation error between the ideal noise PSD $\Phi_{nn}(\lambda,\mu)$ and the estimated noise PSD $\hat{\Phi}_{nn}(\lambda,\mu)$ is calculated according to [2] by

$$\mathrm{logErr} = \frac{1}{\mathrm{KM}} \sum_{\lambda=1}^{\mathrm{K}} \sum_{\mu=1}^{\mathrm{M}} \left| 10\log_{10}\left[\frac{\Phi_{\mathrm{nn}}(\lambda,\mu)}{\hat{\Phi}_{\mathrm{nn}}(\lambda,\mu)}\right] \right|, \tag{18}$$

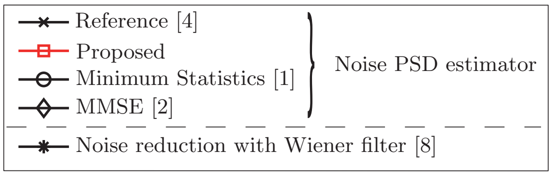

Figure 2: Legend for the evaluation plots.

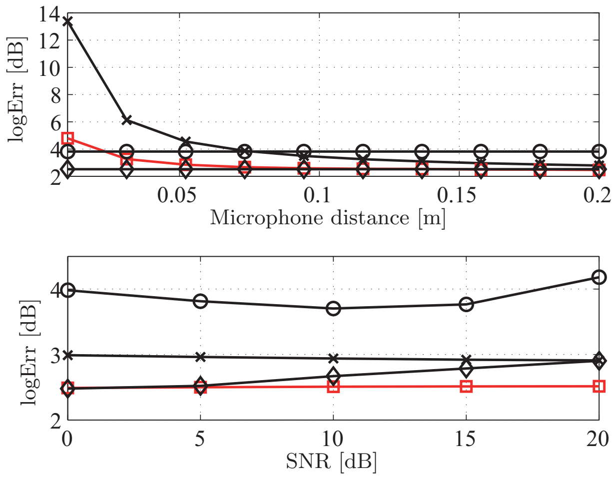

Figure 3: Noise PSD estimation error: (top) fixed SNR of 5 dB (babble noise) over microphone distance, (bottom) fixed inter-microphone distance of 0.10 m over SNR.

with total number of frames K. The ideal noise PSD is obtained using the noise periodograms smoothed over time, i.e.,

$$\Phi_{nn}(\lambda,\mu) = \alpha_{\mathrm{nn}}\,\Phi_{nn}(\lambda-1,\mu) + (1-\alpha_{\mathrm{nn}})|N_1(\lambda,\mu)|^2. \quad (19)$$

The bottom plot in Fig.3 shows the logErr at a fixed inter-microphone distance of 0.10 m over a varying signal-to-noise ratio (SNR). The results in terms of a varying inter-microphone distance at a fixed SNR of 5 dB are illustrated in the top plot. Depicted is the approach as in [4] with the uncorrelated noise field assumption and the proposed approach (proposed) using the diffuse coherence model (17). As a reference, the results for the single-channel MS and MMSE estimators are also plotted. It can clearly be seen that the proposed estimator outperforms the dual-channel estimator [4] in terms of a lower estimation error. Since the algorithm [4] assumes uncorrelated noise, the estimation error is reduced for larger microphone distances where the correlation of the noise field becomes lower for lower frequencies. However, even the proposed algorithm requires a microphone distance larger than a specific minimum which is determined by the real coherence characteristics.

The proposed algorithm outperforms Minimum Statistics independent of the inter-microphone distance and SNR. Besides, the curves show similar results as the MMSE approach for low SNR conditions and a better performance for high SNR conditions, with the expense of an additional microphone, but with the additional benefit of a lower computational complexity. We can also conclude that the estimator does not work for purely coherent noise sources and shows a larger estimation variance for lower frequencies. The scenario of mixed coherent and diffuse noise fields can be tackled, e.g., by a combination with MS or MMSE noise PSD estimators controlled by a classification method as proposed in [13].

4.2 Noise Suppression Performance

In order to demonstrate the noise suppression performance, the dual-channel speech enhancement system depicted in Fig.4 is used in the following. It consists of a single-channel noise reduction algorithm for processing the primary noisy channel $x_1(k)$ only. The required noise PSD estimate $\hat{\Phi}_{nn}(\lambda,\mu)$ is calculated from the primary channel $x_1(k)$ and a secondary noisy channel $x_2(k)$. For the reference single-channel estimators (MS) and (MMSE) the processing is performed on $x_1(k)$ only.

For the spectral gain calculation $G(\lambda,\mu)$ we use the single-channel magnitude DFT estimation procedure under the generalized gamma-model for the DFT-magnitudes proposed in [14]. The enhanced spectrum $\hat{S}(\lambda,\mu)$ is given by the multiplication of the coefficients $X_1(\lambda,\mu)$ with the spectral weighting gains. The enhanced time domain signal $\hat{s}(k)$ is obtained by using the IFFT and overlap-add. The a priori SNR estimation is performed by means of the decision-directed approach [15] with smoothing factor α_{DD} (see Tab.1).

It has to be mentioned that dual-channel spectral weighting rules as well as algorithms with a dual-channel output, e.g., for binaural hearing aids can also be employed.

For the objective evaluation, the noise attenuation (NA) minus speech attenuation (SA) measure is used (NA-SA). Besides, we employ the improvement in Perceptual Evaluation of Speech Quality (PESQ) score [16] for selected simulations. Even though this measure was initially developed for the evaluation of speech codecs it is also widely used for the assessment of speech enhancement algorithms. All results were confirmed by informal listening tests.

4.2.1 Generated Input Signals (A)

The speech enhancement results for generated input signals are shown in Figs.5 and 6. The same tendency of the logErr performance in terms of inter-microphone distance as in Fig.3 can be observed. The proposed algorithm outperforms all other noise estimators in terms of the NA-SA for distances above 0.05 m and above 0.12 m for the Δ PESQ measure. For the special case of a small inter-microphone distance the high estimation error of the proposed algorithm reflects in a low noise suppression performance. The fixed inter-microphone distance of $d_{\mathrm{mic}} = 0.15$ m for the SNR varying plots is used in order to ensure comparability with the recorded data (see Section 4.2.2).

4.2.2 Recorded Input Signals (B)

In the following experiment, recorded dual-channel (binaural) babble noise from the ETSI database is used. The noise is added to the anechoic speech signal according to the signal model shown in Fig.1 at different SNR conditions. Since the dummy head, which was used for the recording, has an influence on the noise field coherence, we use a binaural coherence model [17] for the proposed algorithm.

The results for a fixed inter-microphone distance (determined by the ETSI recordings) and varying SNR are depicted in Fig.7. From these curves we can conclude that the performance of the proposed estimator is similar to the generated data (A) and effects due to recording such as acoustic measurement noise, self-noise of the microphones and microphone mismatch do no have a significant influence on the performance, it contrast to the other algorithms. It can also be seen that the assumption of a homogeneous noise field with the same noise PSD among the microphones is valid. For the recorded background noise, the proposed algorithm outperforms all other approaches.

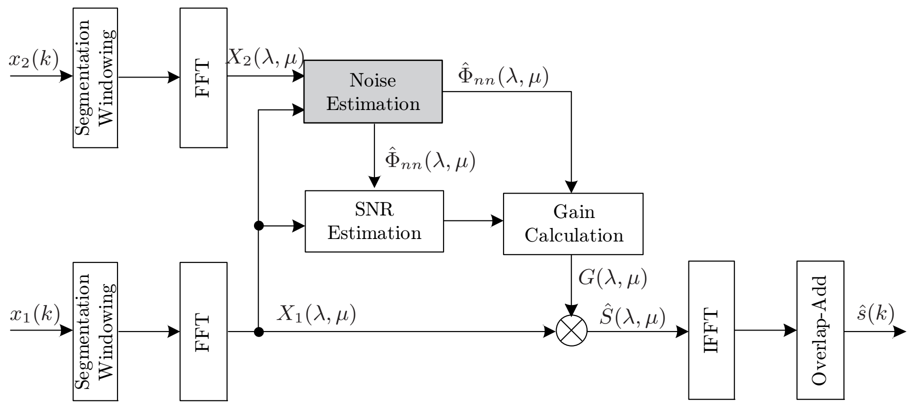

Figure 4: Block diagram of the used dual-channel noise reduction system where different estimators for the "noise estimation" block are employed.

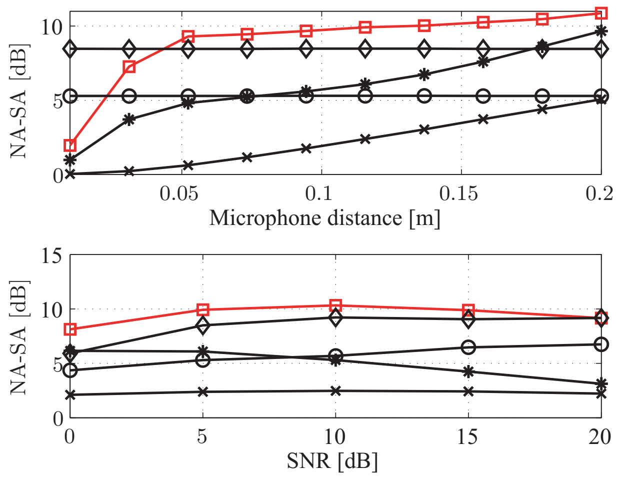

Figure 5: Speech enhancement performance in terms of noise minus speech attenuation (NA-SA): (top) plotted over the inter-microphone distance at 5 db SNR babble noise: (bottom) plotted over SNR at $d_{\text{mic}} = 0.15\,\text{m}$.

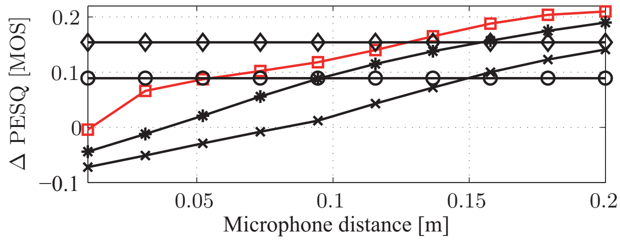

Figure 6: Speech enhancement performance in terms of improvements on PESQ score plotted over the microphone distance at 5 dB SNR babble noise.

4.3 Influence of Speech Attenuation among the Microphones

In order to investigate the influence of the idealized assumption of similar acoustic transfer functions $H_m(e^{j\Omega})$ (see Fig.1), simulations with a variable attenuation between the desired signals $s_1(k)$ and $s_2(k)$ is investigated.

Such investigation is important since for binaural hearing aids, where the desired speaker might not be located in front of the hearing impaired user, or for dual-microphone mobile phones, where one microphone can be placed on top of the device and a high attenuation of the speech by the user can be expected. Figure 8 shows the influence of a speech attenuation between $s_1(k)$ and $s_2(k)$ from 0 to 10 dB. The noise is assumed to have the same energy at the two microphones. Hence, for a given fixed SNR of 10 dB at the primary microphone, the attenuation causes an SNR of 10 to 0 dB at the secondary microphone. The NA-SA results indicate that the proposed noise PSD estimator is robust against a possible attenuation up to values of $5-6$ dB. Compared to the Wiener approach, which results in a significant performance drop even at a low attenuation factor of 2 dB, the proposed estimator is quite robust. Similar insights are obtained from PESQ improvements. The MS and MMSE curves are constant since the attenuation factor affects only the secondary microphone which is not exploited in the latter two algorithms.

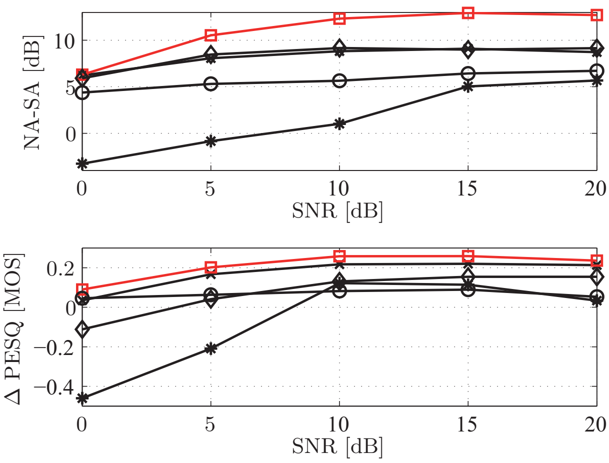

Figure 7: Speech enhancement performance plotted over the SNR for recorded babble background noise from the ETSI database: (top) noise minus speech attenuation (NA-SA), (bottom) improvement in PESQ (Δ PESQ).

4.4 Computational Complexity

The computational effort is compared in terms of the normalized processing time for all four noise PSD estimators in Matlab. The values are normalized such that the proposed method has a value of one. From Table 2 it can be seen that the proposed algorithm does not show a significant increase

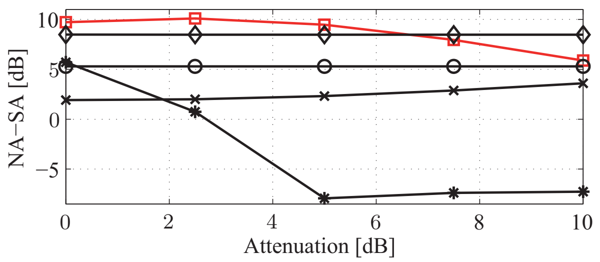

Figure 8: Influence of speech attenuation between the two microphones at $d_{\mathrm{mic}} = 0.15\,\mathrm{m}$ and an SNR of 5 dB at the primary microphone.

Table 2: Normalized processing time.

Algorithm:	Reference	Proposed	MS	MMSE
Time:	0.92	1	1.87	1.16

in complexity compared to the reference approach [4]. Both coherence-based algorithms can efficiently be implemented using first order IIR filters for the auto- and cross-PSD estimation. However, for real-time applications, the additional square-root in (13) has to be taken into account. Regarding the two single-channel approaches, the MS shows the highest processing time and memory consumption due to the large tracking window of 1.5 s. The MMSE requires an additional calculation of the decision-directed approach and the computation of an incomplete gamma function, which however could be avoided by incorporating suitable look-up tables. A long-term memory for minimum tracking in the MMSE approach causes a high computational load as well as a large amount of memory. However, a realization based on subwindows can be employed to significantly reduce both.

5. CONCLUSIONS

In this contribution, we have derived a generalized expression for the coherence-based noise PSD estimator first proposed in [4]. We have shown that by exploiting a priori knowledge of the noise field coherence, a dual-channel speech enhancement system using the novel noise PSD estimator can reduce unwanted background noise in diffuse noise-field conditions. In comparison to the dual-channel noise PSD estimator [4] and the single-channel Minimum Statistics estimator, the novel approach shows a lower estimation error, independent of inter-microphone distance and SNR. For inter-microphone distances larger than $0.10 - 0.15\,\mathrm{m}$, it outperforms the single-channel MMSE noise tracking algorithm. The same tendency was observed in terms of the noise reduction performance. Hence, the novel algorithm is favorable if the design allows for a specific minimum inter-microphone distance.

The main benefit of the proposed algorithm is the low complexity and memory consumption compared to related approaches. Besides, the algorithm can be employed to various speech enhancement applications. A further improvement can be a combination with MS or MMSE noise PSD estimators in mixed coherence and diffuse noise fields.

REFERENCES

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

[2] R. C. Hendriks, J. Jensen, and R. Heusdens, "Noise tracking using DFT domain subspace decompositions," *IEEE Trans.on Audio, Speech, and Language Process.*, vol. 16, no. 3, pp. 541–553, 2008.

[3] R. C. Hendriks, R. Heusdens, and J. Jensen, "MMSE based noise PSD tracking with low complexity," in *Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Dallas, TX, USA, 2010, pp. 4266–4269.

[4] M. Dörbecker and S. Ernst, "Combination of two channel spectral subtraction and adaptive wiener post-filtering for noise reduction and dereverberation," in *Proc. European Signal Processing Conference (EUSIPCO)*, Trieste, Italy, 1996.

[5] V. Hamacher, "Comparison of advanced monaural and binaural noise reduction algorithms for hearing aids," in *Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Orlando, Florida, USA, 2002, vol. 4.

[6] 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.on Audio, Speech, and Lang. Process.*, vol. 17, no. 4, pp. 521–533, 2009.

[7] A. Guerin, R. Le Bouquin, and G. Faucon, "A two-sensor noise reduction system: Applications for hands-free car kit," *EURASIP Journal on Applied Signal Processing*, , no. 11, pp. 1125–1134, 2003.

[8] I.A. McCowan and H. Bourlard, "Microphone array post-filter based on noise field coherence," *IEEE Trans.on Speech and Audio Process.*, vol. 11, no. 6, pp. 709–716, 2003.

[9] T. Lotter, C. Benin, and P. Vary, "Multichannel direction-independent speech enhancement using spectral amplitude estimation," *EURASIP Journal on Applied Signal Processing*, vol. 11, pp. 1147–1156, 2003.

[10] E.A.P. Habets, I. Cohen, and S. Gannot, "Generating non-stationary multisensor signals under a spatial coherence constraint," *Journal of the Acoustical Society of America*, vol. 124, no. 5, pp. 2911–2917, 2008.

[11] P. Kabal, "TSP speech database," Tech. Rep., Department of Electrical & Computer Engineering, McGill University, Montreal, Quebec, Canada, 2002.

[12] ETSI 202 396-1, *Speech and multimedia Transmission Quality (STQ); Part 1: Background noise simulation technique and background noise database*, 03 2009, V1.2.3.

[13] M. Jeub, C. M. Nelke, C. Beaugeant, and P. Vary, "Blind estimation of the coherent-to-diffuse energy ratio from noisy speech signals," in *Proc. European Signal Processing Conference (EUSIPCO)*, Barcelona, Spain, 2011.

[14] J. S. Erkelens, R. C. Hendriks, and R. Heusdens, "On the estimation of complex speech DFT coefficients without assuming independent real and imaginary parts," *IEEE Signal Processing Letters*, vol. 15, pp. 213–216, 2008.

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

[16] ITU-T Rec. P.862, *Perceptual evaluation of speech quality (PESQ)*, ITU, Geneva, 2001.

[17] M. Jeub, M. Dörbecker, and P. Vary, "A semi-analytical model for the binaural coherence of noise fields," *IEEE Signal Processing Letters*, vol. 18, no. 3, March 2011.