DUAL MICROPHONE NOISE PSD ESTIMATION FOR MOBILE PHONES IN HANDS-FREE POSITION EXPLOITING THE COHERENCE AND SPEECH PRESENCE PROBABILITY

Christoph Matthias Nelke, Christophe Beaugeant⋆*, and Peter Vary*

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

ABSTRACT

This contribution addresses the enhancement of noisy speech signals picked up by a dual microphone mobile phone in *hands-free* position. A novel technique to estimate the noise power spectral density is presented which combines two methods: a single microphone algorithm based on the speech presence probability and a dual microphone technique exploiting the coherence properties of the target signal and the background noise. Due to the novel approach, the weakness of both methods can be overcome. Since the proposed method requires knowledge of the current coherence properties, a technique is presented which estimates the coherence of the speech and noise signals, which is usually not known in practice.

***Index Terms*—** speech enhancement, dual microphone noise PSD estimation, noise reduction, coherence estimation

1. INTRODUCTION

Making a phone call in a noisy environment often leads to significant degradations of the quality and intelligibility of the speech signal. This is even more severe during a phone call in *hands-free* position where the signal-to-background-noise ratio (SNR) is often much lower. Hence, it is necessary to reduce the distortions in the transmitted signals by means of noise reduction techniques. One crucial part of noise reduction systems is the estimation of the background noise power spectral density (PSD), in particular in the presence of speech. Given one microphone only, several well established methods exist (e.g., [1], [2], [3]). However, the performance of single microphone algorithms is somehow limited especially in the case of fast changing background noise. Hence, in the latest generation of mobile phones there are frequently two microphones for the suppression of background noise available. While the first microphone is placed at the bottom of the phone, the second microphone is usually placed at the top or back of the device. Signals captured by the second microphone can be used for a better differentiation between the desired speech signal and the background noise. The proposed system makes use of the different coherence properties of the desired speech and the background noise. In most conditions the noise can be seen as a diffuse sound field, while the speech is rather coherent in both microphones. A dual microphone system for noise reduction was proposed in [4] which exploits the coherence properties of the captured signals and can be applied for the *hands-free* conditions. This system has two drawbacks: For small microphone distances (e.g., 10 cm) the reduction of low frequency noise is quite inefficient because the speech and noise signals show similar coherence values. Furthermore, in practice deviations from the assumed coherence properties significantly degrades the performance. In order to overcome these drawbacks a novel approach is presented which is a very efficient combination of a single and dual microphone noise estimation method. More detailed information about the signal model is given in Sec. 2. In Sec. 3 the proposed noise reduction system is described and evaluation results are shown in Sec. 4. This contribution closes with the relation to prior work (Sec. 5) and a conclusion (Sec. 6).

2. ACOUSTICAL ENVIRONMENT

2.1. Dual Microphone Signal Model

The underlying signal model for using a mobile phone in a *hands-free* position is shown in Fig. 1. The two microphones are situated at the bottom and top of the device marked by the red circles. For this alignment a microphone distance of at least 10 cm can be assumed.

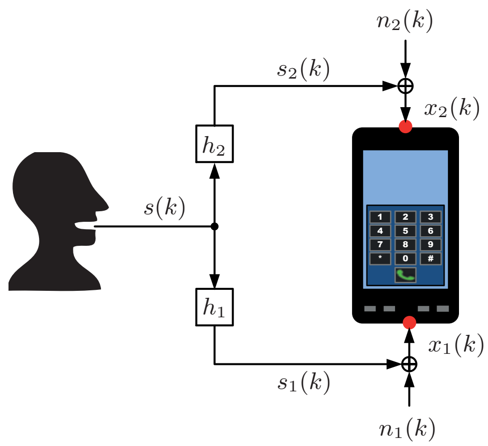

Fig. 1. Dual microphone set-up

The two microphone signals $x_{1|2}(k)$ are given by a superposition of the speech signals $s_{1|2}(k)$ and the noise signals $n_{1|2}(k)$, where the indices $1|2$ stand for the first and second microphone. While the noise signals are assumed to be homogeneous and diffuse the two speech signals are versions of the clean speech signal $s(k)$ filtered with the impulse responses $h_{1|2}$. Since the mobile phone is in *hands-free* position, there are only small power level differences for the speech as well as the noise signals which have no significant influence on the coherence properties of the speech and noise signals.

IEEE ICASSP, May 2013, Vancouver, Canada, DOI: 10.1109/ICASSP.2013.6639076

2.2. Coherence Analysis

The coherence properties of the speech and noise signals are exploited for the proposed noise reduction scheme. The coherence function of two signals $y_1(k)$ and $y_2(k)$ is defined as

$$\Gamma_y = \frac{\Phi_{y_1 y_2}}{\sqrt{\Phi_{y_1 y_1}\Phi_{y_2 y_2}}} \tag{1}$$

where $\Phi_{y_1 y_2}$ and $\Phi_{y_1 y_1}$, $\Phi_{y_2 y_2}$ are the cross- and auto PSDs of $y_1(k)$ and $y_2(k)$. For an **ideal diffuse noise field** this function can be derived [5] as

$$\Gamma_{n,dif} = \text{sinc}\left(\frac{2\pi f d_{mic}}{c}\right) \tag{2}$$

with the distance d_{mic} between two omnidirectional microphones at frequency f and the sound velocity c. The speech signal is often assumed to be **coherent** ($\Gamma_{s,coh} = 1$). However, these conditions are not fulfilled in many real environments. One constraint is that the microphones do not show an omnidirectional characteristic due to the mounting into the mobile phone. This effect as well as reflections and reverberation have an impact on the coherence properties of both the speech and noise signals [6], [7] yielding in deviations of measured coherence functions from the theoretical curves as shown in Fig. 2. The blue and red lines depict typical magnitude squared coherence (MSC) of speech and noise signals, respectively recorded with a dual microphone mock-up phone in a *hands-free* position. More detailed information on the set-up is given in Sec. 4.1. The theoretical curves for Eq. 2 and $\Gamma_s = 1$ for noise and speech respectively are represented by the black dashed lines. An analysis of the coherence of the superposition of speech and noise is given in [8].

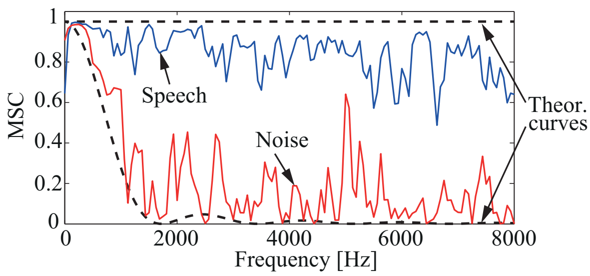

Fig. 2. Measured and theoretical coherence functions

It can be seen that except for the low frequency range it is possible to distinguish between speech and noise based on the measured coherence. Besides, there are partially large deviations between the measured curves and the theoretical values for both the speech and noise coherence.

3. SYSTEM OVERVIEW

The proposed speech enhancement system is realized in an overlap-add structure and is shown in Fig. 3. First the two noisy input signals $x_{1|2}(k)$ are segmented, windowed and transformed in the frequency domain using the Fast Fourier Transform (FFT) to yield $X_{1|2}(\lambda, \mu)$, with λ as the frame index and μ as the discrete frequency bin. The noise estimation is realized in two steps: The first stage is the single microphone **SPP based noise estimation** method [3] (see Sec. 3.1). This estimate of the noise PSD $\widehat{\Phi}_{N,SPP}(\lambda, \mu)$ is applied for the low frequency range for which the coherence of speech and noise are both high and thus can not used for an efficient noise PSD estimation (see Fig. 2). The speech presence probability (SPP) ρ of the first stage and the input signals $X_{1|2}(\lambda, \mu)$ are needed for the second stage, the **coherence based noise estimation**, which then produces the final noise PSD estimate $\widehat{\Phi}_N(\lambda, \mu)$ (see Sec. 3.2). For the spectral gain calculation $G(\lambda, \mu)$ we use the single-microphone magnitude DFT estimation procedure under the generalized gamma-model for the DFT-magnitudes proposed in [9]. The enhanced spectrum $\widehat{S}(\lambda, \mu)$ is given by the multiplication of the coefficients $X_1(\lambda, \mu)$ with the spectral weighting gains. The enhanced time domain signal $\widehat{s}(k)$ is obtained by using the IFFT and overlap-add.

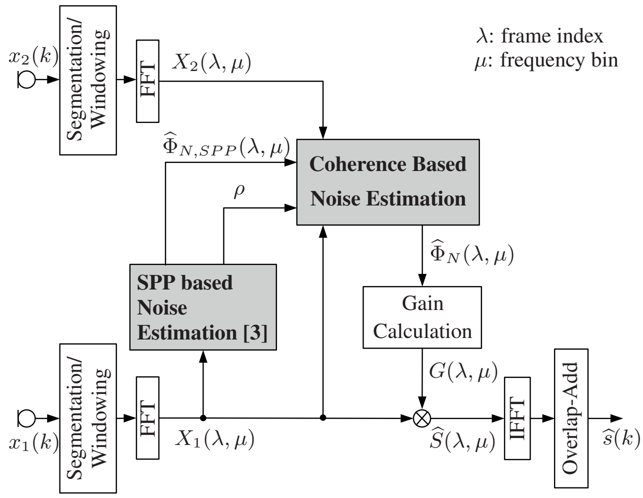

Fig. 3. Speech enhancement system

The novel concept presented in this paper is the two-stage noise estimation in the two highlighted blocks in Fig. 3.

3.1. SPP Based Noise Estimation

The method proposed in [3] estimates the noise PSD of a single microphone speech signal based on a soft decision voice activity detection (VAD). A criterion to distinguish between speech pauses and speech activity is given by the SPP ρ which is defined as [3]

$$\rho = \left(1 + (1+\xi_{opt})\exp\left(-\frac{|X_1(\lambda,\mu)|^2}{\widehat{\Phi}_{N,SPP}(\lambda-1,\mu)}\frac{\xi_{opt}}{\xi_{opt}+1}\right)\right)^{-1} \tag{3}$$

and results into values between 0 and 1, where 1 indicates speech presence. $\widehat{\Phi}_{N,SPP}(\lambda-1,\mu)$ is the noise estimate of the previous frame and ξ_{opt} is the fixed optimal *a priori* SNR. The SPP can then be used to update the noise estimate as a weighted sum of the noisy input and the noise estimate from the previous frame as

$$\widehat{\Phi}_{N,SPP}(\lambda,\mu) = \rho\cdot\widehat{\Phi}_{N,SPP}(\lambda-1,\mu) + (1-\rho)\cdot|X(\lambda,\mu)|^2. \tag{4}$$

More details on the derivations of this method can be found in [3] and the references therein.

3.2. Coherence Based Noise Estimation

For the auto- and cross-PSDs which are needed in the following the short-term estimates are calculated by recursive smoothing of the input signals as

$$\Phi_{x_i x_j}(\lambda,\mu) = \alpha_s \Phi_{x_i x_j}(\lambda-1,\mu) + (1-\alpha_s)X_i(\lambda,\mu)\cdot X_j(\lambda,\mu)^*. \tag{5}$$

The $\{\}^*$ operation denotes the complex conjugate of the signal and the smoothing factor α_s is chosen to 0.75. As initially mentioned, the coherence properties of the input signals $X_{1|2}(\lambda, \mu)$ can be used for the noise estimation in the higher frequency range. For the sake of brevity the frame and frequency indices (λ and μ) are omitted in the following equations. We assume that speech and noise signals are uncorrelated. Then, the auto- and cross PSDs of the input signals read

$$\Phi_{x_1x_1} = \Phi_{s_1s_1} + \Phi_{n_1n_1} \quad (6)$$
$$\Phi_{x_2x_2} = \Phi_{s_2s_2} + \Phi_{n_2n_2} \quad (7)$$
$$\Phi_{x_1x_2} = \Phi_{s_1s_2} + \Phi_{n_1n_2}. \quad (8)$$

Furthermore, we assume a homogeneous speech and noise field in both microphone signals of the system, i.e.

$$\Phi_{s_1s_1} = \Phi_{s_2s_2} = \Phi_s \quad (9)$$
$$\Phi_{n_1n_1} = \Phi_{n_2n2} = \Phi_n. \quad (10)$$

In [4] we assumed ideal coherent speech ($\Gamma_s = 1$) which is, however, not always fulfilled in real situation as it was shown in Sec. 2. In the following we skip this assumption and thus, the cross PSD in (8) can be rewritten with (1) and (9, 10) as

$$\Phi_{x_1x_2} = \Gamma_s \cdot \Phi_s + \Gamma_n \cdot \Phi_n, \quad (11)$$

where Γ_s and Γ_n are the coherence functions of the speech and noise signals, respectively. Inserting (9), (10) in (6) and (7) and using the geometric mean of two auto PSDs

$$\sqrt{\Phi_{x_1x_1}\Phi_{x_2x_2}} = \Phi_s + \Phi_n \quad (12)$$

(11) can be reordered to get the noise PSD estimate

$$\widehat{\Phi}_{N,theor} = \frac{\sqrt{\Phi_{x_1x_1}\Phi_{x_2x_2}} - \frac{\Phi_{x_1x_2}}{\Gamma_s}}{1 - \frac{\Gamma_n}{\Gamma_s}}. \quad (13)$$

Due to the fact that the noise PSD must be real-valued, the absolute values of $\frac{\Phi_{x_1x_2}}{\Gamma_s}$ and $\frac{\Gamma_n}{\Gamma_s}$ are used in the implementation of the algorithm. Furthermore, in (13) it has to be ensured that the denominator is greater than zero, e.g., by an upper threshold of the coherence ratio $|\frac{\Gamma_n}{\Gamma_s}| < 0.99$. Note, that if $\Gamma_s = 1$ is assumed this leads to the noise estimate provided by [4].

In periods, where speech is not predominant, it turned out that a weighted average with the noisy input signal (e.g., from the first microphone) is more accurate than the estimate from (13). Therefore, the final noise PSD estimate of coherence based stage is given by

$$\widehat{\Phi}_{N,coh} = \rho_{coh}\, \widehat{\Phi}_{N,theor} + (1 - \rho_{coh})|X_1|^2. \quad (14)$$

The weighting factor

$$\rho_{coh} = \frac{\Gamma_{x,\lambda} - \Gamma_n}{\Gamma_s - \Gamma_n} \quad (15)$$

is a function of the short-term coherence $\Gamma_{x,\lambda}$ in the current signal frame which is calculated by using the short-term estimates of the auto- and cross-PSDs (5) in (1). ρ_{coh} can be seen as a dual microphone voice detector equivalent to the one of (3).

3.3. Coherence Function Update

The coherence based noise estimate given in (13) - (15) requires the coherence functions. This can be constant functions as described in [4]. As in practice, Γ_s and Γ_n are not known, we propose to update Γ_s in periods where speech is predominant and Γ_n in periods where speech is absent. The speech presence probability ρ from (3) determines these periods by applying a simple threshold. The update rule is based on the short-term coherence $\Gamma_{x,\lambda}$ and reads for the noise coherence function

$$\widehat{\Gamma}_n = \alpha_\Gamma \widehat{\Gamma}_{n,last} + (1 - \alpha_\Gamma)\Gamma_{x,\lambda} \quad \text{, if } \rho < 0.1 \quad (16)$$

where $\widehat{\Gamma}_{n,last}$ is the last updated coherence function. This rule uses the speech pauses to update the noise coherence. The same rule can not be applied directly for the update of the speech coherence function because a high SPP value ρ does not indicate a noise-free signal segment. Hence, the influence of the noise must be taken into account. Using (6) - (8) and assuming again that noise and speech signals are uncorrelated the coherence function of $X_{1|2}$ can be expressed as

$$\begin{aligned}\Gamma_x &= \frac{\Phi_{s_1s_2} + \Phi_{n_1n_2}}{\sqrt{\Phi_{x_1x_1}\Phi_{x_2x_2}}} = \frac{\Phi_{s_1s_2} + \Phi_{n_1n_2}}{\Phi_s + \Phi_n} \\ &= \frac{\Phi_{s_1s_2}}{\Phi_s}\left(1 + \frac{\Phi_n}{\Phi_s}\right)^{-1} + \frac{\Phi_{n_1n_2}}{\Phi_n}\left(1 + \frac{\Phi_s}{\Phi_n}\right)^{-1}. \quad (17)\end{aligned}$$

With the definition of the *a posteriori* SNR

$$\gamma = \frac{\Phi_x}{\Phi_n} = \frac{\Phi_s + \Phi_n}{\Phi_n} \quad (18)$$

and inserting the coherence function (1) for speech and noise in (17) the coherence can be rewritten as

$$\Gamma_x = \Gamma_s \frac{\gamma - 1}{\gamma} + \Gamma_n \frac{1}{\gamma}. \quad (19)$$

For the *a posteriori* SNR the current noise estimate and the noisy input are used to compute Φ_n and Φ_x. Now (19) can be rearranged to finally lead to the corrected speech coherence function

$$\Gamma_{s,cor} = \Gamma_{x,\lambda} \frac{\gamma}{\gamma - 1} - \Gamma_n \frac{1}{\gamma - 1}. \quad (20)$$

The update of the speech coherence can be carried out during periods where speech is active:

$$\widehat{\Gamma}_s = \alpha_\Gamma \widehat{\Gamma}_{s,last} + (1 - \alpha_\Gamma)\Gamma_{s,cor} \quad \text{, if } \rho > 0.9. \quad (21)$$

It should be mentioned, that ρ is frequency dependent which means that the coherence functions (16) and (21) are updated in time-frequency ranges where the SPP exceeds the corresponding thresholds. The smoothing constants in (16) and (21) are chosen to $\alpha_\Gamma = 0.95$ and the coherence functions are initialized as $\Gamma_s = 1$ for the speech and ideal diffuse (2) for the noise signals.

The whole algorithm can be summarized as shown in the table below, where μ_s represents the split-frequency between the single microphone and dual microphone noise estimate. Here, we propose to use the frequency where the MSC of the ideal diffuse coherence in (2) takes the value 0.5 (see Fig. 2). All parameters for the SPP based component of the system are chosen as proposed in [3].

	Proposed noise PSD estimation (repeat for all signal frames)
1	Compute auto- and cross-PSDs (5)
2	Obtain single microphone estimate of SPP ρ and noise PSD $\Phi_{N,SPP}$ (3, 4)
3	Obtain dual microphone noise PSD estimation $\Phi_{N,coh}$ (13,14)
4	Combine noise PSD estimates, as $\widehat{\Phi}_N(\lambda,\mu) = \begin{cases}\widehat{\Phi}_{N,SPP}(\lambda,\mu) & \text{, for } \mu < \mu_s \\ \widehat{\Phi}_{N,coh}(\lambda,\mu) & \text{, for } \mu > \mu_s\end{cases}$
5	Update speech and noise coherence functions (16, 20, 21)
6	Apply temporal smoothing of $\widehat{\Phi}_N(\lambda,\mu)$ with $\alpha_N = 0.9$ to obtain the noise power estimate

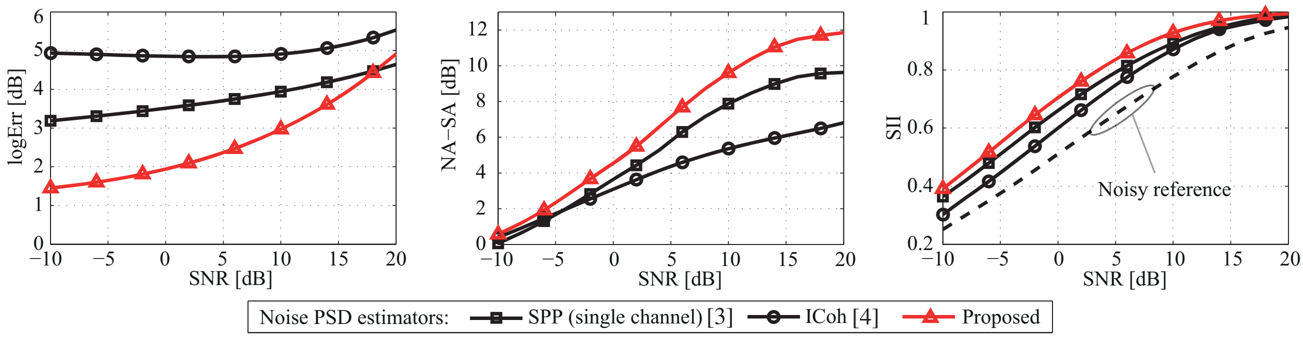

Fig. 4. Evaluation results: *left:* noise PSD estimation accuracy (logErr: logarithmic error); *middle:* noise reduction performance (NA-SA: noise attenuation minus speech attenuation); *right:* intelligibility enhancement (SII: speech intelligibility index)

4. EXPERIMENTS AND RESULTS

To rate the performance, the proposed system was evaluated with recorded audio signals. Thereby, we compared our method with two other algorithms: The SPP based method which turned out to give the best results among the single microphone algorithms (c.f., [3], [7]) and the dual microphone method in [4]. The latter was as well designed for the use in mobile phones and will be termed as ICoh. For all schemes the spectral gain calculation from [9] was applied and to prevent musical tones in the processed speech signal the gain smoothing from [10] was used. Besides, the frame length was 20 ms using a Hann window with 50 % overlap and the FFT-size was 512 (including zero-padding), while the sampling frequency was 16 kHz.

4.1. Measurement Set-up

For the evaluation measurements with a dual-microphone mock-up phone were carried out. The microphones were arranged as shown in Fig. 1 with $d_{mic} = 10\,\text{cm}$. The desired speech signal was produced by an artificial head (HEAD acoustics HMS II.3) in a low reverberant room (reverberation time $\approx 100\,\text{ms}$), while the mock-up phone was located at the hands-free reference point 50 cm in front of the artificial head [11]. A diffuse noise field was generated using four loudspeakers as defined in [12]. For the speech signals 60 samples were taken from the TSP database [13]. The background noise simulation was carried out by using three noise types from [12] ('Mensa','FullsizeCar','WorknoiseJackhammer') in addition to white noise and modulated white noise. The speech and noise signals were recorded separately and can be superposed to simulate different SNR conditions.

4.2. Evaluation

Estimation Accuracy: The accuracy of the investigated methods is measured with the segmental logarithmic error (logErr) between the estimated noise PSD $\widehat{\Phi}_N(\lambda, \mu)$ and a smoothed version of the true noise signal yielding the reference noise PSD $\Phi_{N,ref}(\lambda, \mu)$, where K and M are the number of frames and discrete frequency bins:

$$\text{logErr} = \frac{1}{KM} \sum_{\lambda=1}^{K} \sum_{\mu=1}^{M} \left| 10 \log_{10} \left(\frac{\widehat{\Phi}_N(\lambda, \mu)}{\Phi_{N,ref}(\lambda, \mu)} \right) \right| . \tag{22}$$

The smoothing constant for the reference noise PSD was chosen to 0.9 (c.f., [14]). A low logarithmic error indicates an accurate noise PSD estimation. The results of this measure are depicted in the left plot of Fig. 4 averaged over all noise types. For the whole SNR range the ICoh method shows the highest estimation error. This is caused by the non-ideal coherence properties of the audio data. In this case the single microphone SPP estimator yields to a better accuracy. Except for the 20 dB SNR condition, the proposed method gives the lowest logarithmic error and thus the most accurate noise PSD estimate.

Speech Enhancement Performance: The whole system is also evaluated in terms of the speech enhancement. The noise reduction performance is determined by means of the noise attenuation minus speech attenuation (NA-SA) measure (e.g., [15]), where an improvement results in higher values. In addition the Speech Intelligibility Index (SII) [16] is applied as measure. The SII provides a value between 0 and 1 where a SII higher than 0.75 indicates a good communication system and values below 0.45 correspond to a poor system. The results for both measures are shown in Fig. 4 in the middle and right plot. The SII of the noisy input signal is shown by the black dashed line. Both plots show the same ranking of the investigated methods as the logErr measure. It can be concluded, that the proposed algorithm outperforms the two other methods in terms of noise reduction and enhances the speech intelligibility. Informal listing tests confirm these results, where the proposed system shows less speech distortion and better reduction of low frequency noise then the previous coherence based method [4]. In addition, the SPP based method leaves some audible residual noise during changes in the background noise.

5. RELATION TO PRIOR WORK

The SPP based component of the proposed system can be ranked in the evolution of single microphone noise estimators such as [1], [2], [17] or [3]. For the coherence based part of the system the algorithms presented in [18] and [4] can be seen as special cases of the proposed noise estimation scheme, where both methods assume constant coherence properties of the speech and noise signals. In [19] and [20] coherence based techniques were presented which are designed small microphone distances as in hearing aids to suppress interfering coherent noise sources. A procedure to estimate the coherence properties in order to classify the background noise was shown in [8] and was advanced in [21]. In our system both the noise and speech coherence are estimated by a new update rule.

6. CONCLUSIONS

In this contribution a new noise PSD estimator is presented for mobile phones in *hands-free* position. The proposed system is a very effective combination of a single and dual microphone method. For the higher frequency range the coherence properties of speech and noise are exploited, while for the lower frequency range the SPP based noise estimate is directly used. Besides, a technique was introduced to estimate the coherence of the background noise and the desired speech signal. An evaluation with real audio recordings shows that the proposed method outperforms the single microphone SPP estimator and a previously proposed dual microphone method.

7. REFERENCES

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

[2] R.C. Hendriks, R. Heusdens, and J. Jensen, "MMSE based noise PSD tracking with low complexity," in *Proc. of IEEE Intern. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, Dallas, Texas, USA, 2010.

[3] T. Gerkmann and R. Hendriks, "Noise power estimation based on the probability of speech presence," in *Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, USA, 2011.

[4] M. Jeub, C.M. Nelke, H. Krüger, C. Beaugeant, and P. Vary, "Robust dual-channel noise power spectral density estimation," in *Proc. of European Signal Processing Conf. (EUSIPCO)*, Barcelona, Spain, 2011.

[5] B.F. Cron and C. H. Sherman, "Spatial-correlation functions for various noise models," *Journal of the Acoustical Society of America*, vol. 34, no. 11, pp. 1732–1736, 1962.

[6] M. Brandstein and D. Ward, *Microphone Arrays - Signal Processing Techniques and Applications*, Springer Verlag, 2001.

[7] M. Jeub, *Joint Dereverberation and Noise Reduction for Binaural Hearing Aids and Mobile Phones*, PhD Thesis, IND, RWTH Aachen, Aug. 2012.

[8] 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. of European Signal Processing Conf. (EUSIPCO)*, Barcelona, Spain, 2011.

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

[10] T. Esch and P. Vary, "Efficient musical noise suppression for speech enhancement system," in *Proc. of IEEE Intern. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, Taipei, Taiwan, 2009.

[11] ETSI EG 201 377-2, "Speech processing, transmission and quality aspects (STQ); Specification and measurement of speech transmission quality; Part 2: Mouth-to-ear speech transmission quality including terminals," April 2004.

[12] ETSI EG 202 396-1, "Speech and multimedia transmission quality (STQ); Part 1: Background noise simulation technique and background noise database," March 2009.

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

[14] J. Taghia, J. Taghia, N. Mohammadiha, J. Sang, V. Bouse, and R. Martin, "An evaluation of noise power spectral density estimation algorithms in adverse acoustic environments," in *Proc. of IEEE Intern. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, Prague, Czech Republic, 2011.

[15] S. Gustafsson, R. Martin, P. Jax, and P. Vary, "A psychoacoustic approach to combined acoustic echo cancellation and noise reduction," vol. 10, no. 5, pp. 245–256, July 2002.

[16] ANSI S3.5-1997, "Methods for the calculation of the speech intelligibility index," 1997.

[17] T. Gerkmann and R.C. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 20, no. 4, pp. 1383 – 1393, 2011.

[18] 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)*, Trieste, Italy, 1996.

[19] N. Yousefian and P. Loizou, "A dual-microphone speech enhancement algorithm based on the coherence function," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 20, no. 2, pp. 599–609, 2012.

[20] N. Yousefian and P. C. Loizou, "A dual-microphone algorithm that can cope with competing-talker scenarios," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 21, no. 1, pp. 143 –153, 2013.

[21] O. Thiergart, G. Del Galdo, and E.A.P. Habets, "Signal-to-reverberant ratio estimation based on the complex spatial coherence between omnidirectional microphones," in *Proc. of IEEE Intern. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, Kyoto, Japan, 2012.