NOISE REDUCTION FOR DUAL-MICROPHONE MOBILE PHONES EXPLOITING POWER LEVEL DIFFERENCES

Marco Jeub, Christian Herglotz, Christoph Nelke, Christophe Beaugeant⋆*, and Peter Vary*

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

ABSTRACT

This paper discusses the application of noise reduction algorithms for dual-microphone mobile phones. An analysis of the acoustical environment based on recordings with a dual-microphone mock-up phone mounted on a dummy head is given. Motivated by the recordings, a novel dual-channel noise reduction algorithm is proposed. The key components are a noise PSD estimator and an improved spectral weighting rule which both explicitly exploit the Power Level Differences (PLD) of the desired speech signal between the microphones. Experiments with recorded data show that this low complexity system has a good performance and is beneficial for an integration into future mobile communication devices.

Index Terms— Noise reduction, noise estimation, speech enhancement, dual-channel, power level difference.

1. INTRODUCTION

Mobile phone conversations can take place in nearly every acoustical situation. Since the listener at the far-end usually suffers from unwanted background noise if the talker is located in an adverse acoustical situation, most mobile phones have integrated algorithms to enhance the speech quality, cf. [1]. The algorithms aim to reduce unwanted background noise while ensuring that the occurring speech distortions are inaudible to the greatest possible extent. For such algorithms, the computational complexity and algorithmic delay is of significant importance. Besides, the algorithm should be able to converge fast in changing noise conditions.

In this contribution, we discuss the application of noise reduction algorithms for dual-microphone mobile phones. In order to employ such algorithms, a secondary microphone can be placed either next to the common primary microphone on the bottom of the device or on top of the device (see Fig.1). In the first part of this paper, an analysis of the acoustical environment is given, which is entirely based on recordings taken with a dual-microphone mock-up phone in typical acoustical situations. Based on these observations, in the second part, a novel algorithm is proposed which exploits the *Power Level Differences* (PLD) of the different signal components and has a very low computational complexity.

2. ANALYSIS OF THE ACOUSTICAL ENVIRONMENT

Common mobile phones use a single microphone for capturing the speech signal. This primary microphone is usually mounted on the bottom of the device in order to allow for a short acoustic path between mouth and microphone, which ensures a high direct path energy and less reverberation. Depending on the phone design, a secondary microphone can be placed either on the bottom next to the primary microphone, or on top of the device in order to capture the speech signal with a lower sound pressure level (SPL).

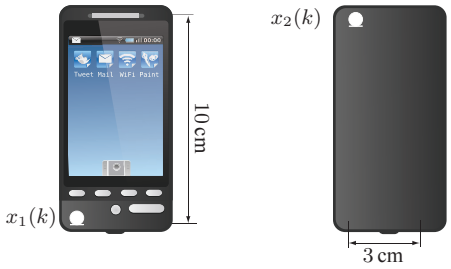

Fig. 1. Illustration of mobile phone with the considered microphone position. (left) front side , (right) rear side.

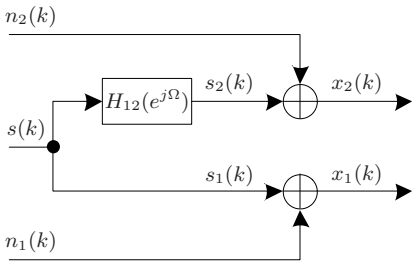

Fig. 2. Dual-channel signal model.

In the remainder of this paper, the dual-channel microphone configuration according to Fig.1 is considered. A primary microphone is placed at the bottom and a secondary microphone on the top rear side of the device. The two microphone signals $x_1(k)$ and $x_2(k)$ are related to clean speech $s(k)$ and additive background noise signals $n_m(k)$ by the signal model shown in Fig 2, with $m = 1, 2$ and discrete time index k. The acoustic transfer function of the desired speech signal between the two microphones is denoted by $H_{12}(e^{j\Omega})$.

The following background noise analysis is based on measurements inside an acoustic chamber using the standardized multi-loudspeaker procedure described in [2] to generate realistic noise fields. Here, we restrict the analysis to two important noise types: car and babble noise from [2]. The recording system consists of a HEAD acoustics HMS II.3 artificial head which includes a mouth simulator. A mock-up phone was mounted on the artificial head in the flat handset position. This procedure allows to record speech (taken from [3]) and noise separately which is usually not possible in real acoustic environments.

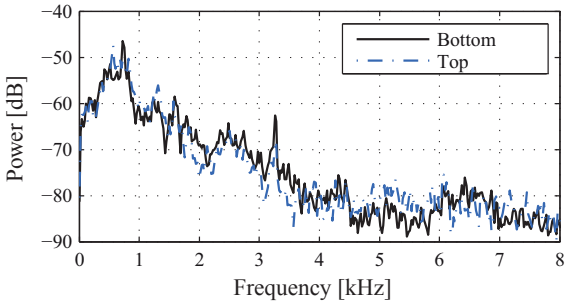

Fig. 3. PSD of babble noise captured by the two microphones.

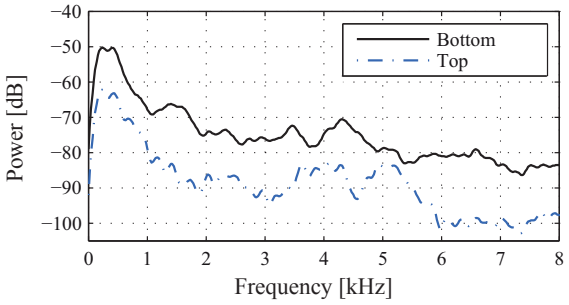

Fig. 4. PSD of speech signal from artificial mouth captured by the two microphones.

2.1. Analysis of Background Noise

Important acoustical quantities are the power spectral densities (PSD) recorded at the positions of the two microphones for both, speech and noise. Figure 3 shows exemplarily the PSD of babble noise for the two microphones. It can be seen that both signals have roughly the same PSD and hence, a homogeneous noise field exists as confirmed by the investigation of further noise types.

A further coherence evaluation of the background noise showed a good match between the theoretical coherence using the free-field diffuse model, cf. [4], with the corresponding inter-microphone distances and the recorded data. All experiments with noise-only conditions have also been verified with the same mock-up phone, which was placed outside in crowded places and a busy street.

2.2. Analysis of Speech

The attenuation of the desired speech signal from the mouth to the possible microphone locations is of significant importance. Figure 4 shows the PSD of the speech signals picked up by the two microphones (noise-free case) where a power level difference of $\approx 10\,\mathrm{dB}$ is measured between the bottom and top microphone for all frequencies.

3. NOISE REDUCTION SYSTEM

The novel speech enhancement system which operates in the short-time Fourier domain is depicted in Fig. 5. The system can be divided into two novel components: a dual-channel noise PSD estimator as well as a dual-channel spectral weighting rule. Each of the two components works independently and can be incorporated in any related speech enhancement system. The enhanced spectrum $\hat{S}(\lambda,\mu)$ is given by multiplying the primary input $X_1(\lambda,\mu)$ with the spectral weighting gains $G(\lambda,\mu)$. Discrete frequency bin and frame index are denoted by μ and λ. The required estimate of the noise PSD is

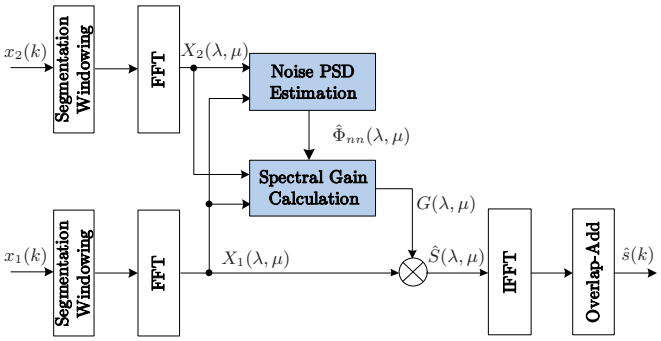

Fig. 5. Block diagram of the proposed dual-channel noise reduction system.

denoted by $\hat{\Phi}_{nn}(\lambda,\mu)$. The enhanced time domain signal $\hat{s}(k)$ is obtained by using the IFFT and overlap-add.

3.1. Noise PSD Estimation (PLDNE Algorithm)

The motivation for the novel PLD-based noise PSD estimator, which is termed as *Power Level Difference Noise Estimator* (PLDNE), is given by the preceding measurements. Two important assumptions are the existence of a homogeneous diffuse noise field, i.e., $\Phi_{n_1n_1}(\lambda,\mu) = \Phi_{n_2n_2}(\lambda,\mu) = \Phi_{nn}(\lambda,\mu)$, as well as a sufficient attenuation of the desired speech signal between the two microphones of, e.g., 10 dB.

In a first step, the normalized difference of the power spectral density $0 \leq \Delta\Phi_{\mathrm{PLDNE}}(\lambda,\mu) \leq 1$ of the noisy input is calculated for every frequency bin μ by

$$\Delta\Phi_{\mathrm{PLDNE}}(\lambda,\mu) = \left| \frac{\Phi_{x_1x_1}(\lambda,\mu) - \Phi_{x_2x_2}(\lambda,\mu)}{\Phi_{x_1x_1}(\lambda,\mu) + \Phi_{x_2x_2}(\lambda,\mu)} \right|, \qquad (1)$$

where $\Phi_{x_1x_1}(\lambda,\mu)$ and $\Phi_{x_2x_2}(\lambda,\mu)$ represent the auto-PSD of $x_1(k)$ and $x_2(k)$ respectively. The cross-PSD is denoted by $\Phi_{x_1x_2}(\lambda,\mu)$. All PSD values are calculated by recursive smoothing over time with constant α_1.

The idea behind the subsequent noise PSD estimation is as follows. In case of background noise-only periods, $\Delta\Phi_{\mathrm{PLDNE}}(\lambda,\mu)$ will be close to zero as the input power levels are almost equal. If the value lies below a threshold $\phi_{\min}$, the noise PSD estimate is determined directly from the input signal $x_1(k)$ by

$$\hat{\Phi}_{nn}(\lambda,\mu) = \alpha_2 \cdot \hat{\Phi}_{nn}(\lambda-1,\mu) + (1-\alpha_2)\cdot |X_1(\lambda,\mu)|^2, \quad \text{if } \Delta\Phi_{\mathrm{PLDNE}}(\lambda,\mu) < \phi_{\min}. \qquad (2)$$

Regarding the noise-free case, the auto-PSD at $x_1(k)$ will be larger than at $x_2(k)$ according to Fig. 4 and thus, the value of $\Delta\Phi_{\mathrm{PLDNE}}(\lambda,\mu)$ will be close to one. As a consequence, the updating of the noise estimate will be stopped if the difference is larger than a threshold $\phi_{\max}$, i.e.,

$$\hat{\Phi}_{nn}(\lambda,\mu) = \hat{\Phi}_{nn}(\lambda-1,\mu) \quad \text{if } \Delta\Phi_{\mathrm{PLDNE}}(\lambda,\mu) > \phi_{\max}. \qquad (3)$$

In between these two extremes, a noise estimation using $x_2(k)$ is used as approximation according to

$$\hat{\Phi}_{nn}(\lambda,\mu) = \alpha_3 \cdot \hat{\Phi}_{nn}(\lambda-1,\mu) + (1-\alpha_3)\cdot |X_2(\lambda,\mu)|^2, \qquad (4)$$

since the highly attenuated speech components in $x_2(k)$ can be neglected. In situations with babble noise, it is beneficial to combine the PLDNE algorithm with further single- or dual-channel noise PSD estimators, e.g., [5, 6, 7] instead of keeping the last estimate in Eq.(3).

3.2. Noise Reduction (PLD Algorithm)

The second component of the novel noise reduction system comprises the calculation of the spectral weighting gains $G(\lambda, \mu)$. The method is motivated by the PLD algorithm initially proposed in [8]. Here, an alternative calculation of the spectral gains and an additional smoothing is proposed.

It is again assumed that the power levels are equal for noise whereas speech results in a higher PSD at microphone $x_1(k)$. The auto-PSDs of the inputs are given by

$$\Phi_{x_1x_1}(\lambda, \mu) = \Phi_{s_1s_1}(\lambda, \mu) + \Phi_{n_1n_1}(\lambda, \mu), \quad (5)$$

$$\Phi_{x_2x_2}(\lambda, \mu) = \Phi_{s_2s_2}(\lambda, \mu) + \Phi_{n_2n_2}(\lambda, \mu). \quad (6)$$

By introducing a transfer function of the desired speech signal between the microphones (see Fig. 2), the auto-PSD at the secondary microphone can be expressed by

$$\Phi_{x_2x_2}(\lambda, \mu) = |H_{12}(\lambda, \mu)|^2 \cdot \Phi_{s_1s_1}(\lambda, \mu) + \Phi_{n_2n_2}(\lambda, \mu). \quad (7)$$

Two difference equations for the auto-PSD of the noisy input and the noise-only signals are introduced as

$$\Delta\Phi_{\mathrm{PLD}}(\lambda, \mu) = \Phi_{x_1x_1}(\lambda, \mu) - \Phi_{x_2x_2}(\lambda, \mu), \quad (8)$$

$$\Delta\Phi_{nn}(\lambda, \mu) = \Phi_{n_1n_1}(\lambda, \mu) - \Phi_{n_2n_2}(\lambda, \mu). \quad (9)$$

The power level difference of the noisy input signal can thus be expressed as

$$\Delta\Phi_{\mathrm{PLD}}(\lambda, \mu) = \Phi_{s_1s_1}(\lambda, \mu)(1 - |H_{12}(\lambda, \mu)|^2) + \Delta\Phi_{nn}(\lambda, \mu). \quad (10)$$

Due to the assumption of a homogeneous noise field the difference $\Delta\Phi_{nn}(\lambda, \mu)$ can be neglected, i.e., $\Delta\Phi_{nn}(\lambda, \mu) \approx 0$. Hence, the equation for the PLD reads

$$\Delta\Phi_{\mathrm{PLD}}(\lambda, \mu) = (1 - |H_{12}(\lambda, \mu)|^2) \cdot \Phi_{s_1s_1}(\lambda, \mu). \quad (11)$$

The final spectral weighing rule is the Wiener filter equation

$$G(\lambda, \mu) = \frac{\Phi_{s_1s_1}(\lambda, \mu)}{\Phi_{s_1s_1}(\lambda, \mu) + \Phi_{n_1n_1}(\lambda, \mu)}. \quad (12)$$

By expanding both nominator and denominator by $1 - |H_{12}(\lambda, \mu)|^2$ as in [8] and by inserting Eq.(11), the weighting rule reads

$$G(\lambda, \mu) = \frac{\Delta\Phi_{\mathrm{PLD}}(\lambda, \mu)}{\Delta\Phi_{\mathrm{PLD}}(\lambda, \mu) + \gamma(1 - |H_{12}(\lambda, \mu)|^2) \cdot \Phi_{nn}(\lambda, \mu)}, \quad (13)$$

with a noise overestimation factor denoted by γ. In the case of speech absence $\Delta\Phi_{\mathrm{PLD}}(\lambda, \mu)$ will be zero and hence, the gains will be zero, too. When there is pure speech the right part of the denominator of Eq.(13) will be zero. Thus the gains $G(\lambda, \mu)$ will turn to one. The required transfer function $H_{12}(\lambda, \mu)$ is derived from the cross-PSD of the noisy input $\Phi_{x_1x_2}(\lambda, \mu)$. In [8], the cross-PSD is expressed by

$$\Phi_{x_1x_2}(\lambda, \mu) = H_{12}(\lambda, \mu) \cdot \Phi_{x_1x_1}(\lambda, \mu) + \Phi_{n_1n_2}(\lambda, \mu), \quad (14)$$

and the transfer function is given by

$$H_{12}(\lambda, \mu) = \frac{\Phi_{x_1x_2}(\lambda, \mu) - \Phi_{n_1n_2}(\lambda, \mu)}{\Phi_{x_1x_1}(\lambda, \mu) - \Phi_{nn}(\lambda, \mu)}. \quad (15)$$

Table 1. Main simulation parameters.

Sampling frequency	$f_s = 16$ kHz
Frame length	$L = 320$ (20 ms)
FFT length	$M = 512$ (including zero-padding)
Frame overlap	50% (Hann window)
Smoothing factors	$\alpha_1 = 0.9, \alpha_2 = 0.9, \alpha_1 = 0.8, \alpha_{nn} = 0.9$
Smoothing threshold	$f_0 = 1$ kHz
PLDNE thresholds	$\phi_{\min} = 0.2, \phi_{\max} = 0.8$
Overestimation factor	$\gamma = 4$

The required cross-PSD of the background noise $\Phi_{n_1n_2}(\lambda, \mu)$ is calculated in [8] from the first 400 ms where no speech activity is assumed. In contrast to Eq.(14), in our implementation the cross-PSD is correctly expressed by

$$\Phi_{x_1x_2}(\lambda, \mu) = H_{12}(\lambda, \mu) \cdot \Phi_{s_1s_1}(\lambda, \mu) + \Phi_{n_1n_2}(\lambda, \mu). \quad (16)$$

By incorporating the coherence of the noise field $\Gamma_{n_1n_2}(\mu)$, the cross-PSD reads with $\Phi_{s_1s_1}(\lambda, \mu) = \Phi_{x_1x_1}(\lambda, \mu) - \Phi_{nn}(\lambda, \mu)$

$$\begin{aligned}\Phi_{x_1x_2}(\lambda, \mu) = H_{12}(\lambda, \mu) \cdot (\Phi_{x_1x_1}(\lambda, \mu) - \Phi_{nn}(\lambda, \mu)) \\ + \Gamma_{n_1n_2}(\mu) \cdot \Phi_{nn}(\lambda, \mu).\end{aligned} \quad (17)$$

Hence, the proposed transfer function is given by

$$H_{12}(\lambda, \mu) = \frac{\Phi_{x_1x_2}(\lambda, \mu) - \Gamma_{n_1n_2}(\mu) \cdot \Phi_{nn}(\lambda, \mu)}{\Phi_{x_1x_1}(\lambda, \mu) - \Phi_{nn}(\lambda, \mu)}. \quad (18)$$

With Eq.(18), the computation of the transfer function does not require an additional calculation of the noise cross-PSD anymore and allows the algorithm to cope with non-stationary noise and changing SNR conditions compared to [8]. In the practical implementation, the power level difference is proposed to be calculated by

$$\Delta\Phi_{\mathrm{PLD}}(\lambda, \mu) = \max\left(\Phi_{\mathrm{x_1x_1}}(\lambda, \mu) - \Phi_{\mathrm{x_2x_2}}(\lambda, \mu), 0\right), \quad (19)$$

which prevents speech distortions if the assumption of a homogeneous noise field is violated, e.g., due to an interfering talker.

In order to reduce the amount of musical tones, a smoothing over frequency using the approach of [9] is employed for frequencies above f_0.

4. PERFORMANCE EVALUATION

The experiment section is separated into an evaluation of the proposed noise estimator as well as the complete noise reduction system using PLDNE and the PLD-based weighting rule. The input signals are taken from recordings using the same experimental setup with a dual-microphone mock-up phone used for the acoustical analysis carried out in Section 2. Speech and noise were recorded separately and mixed together with different SNR conditions, ensuring the same power level difference of the speech signal. We investigate the PLD algorithm assuming an ideal diffuse noise field and use the following coherence model in Eq.(18)

$$\Gamma_{n_1n_2}(f) = \mathrm{sinc}\left(2\pi f d_{\mathrm{mic}}/c\right), \quad (20)$$

with distance $d_{\mathrm{mic}} = 0.1$ m and sound velocity $c = 340$ m/s. Further simulation parameters are listed in Tbl.1.

4.1. Noise Estimation Accuracy

The PLDNE algorithm (Proposed) is compared to the generalized dual-channel coherence-based noise PSD estimator [7] (GCoh). It has to be mentioned that the estimator presented in [7] was mainly developed for binaural hearing aids with a larger inter-microphone

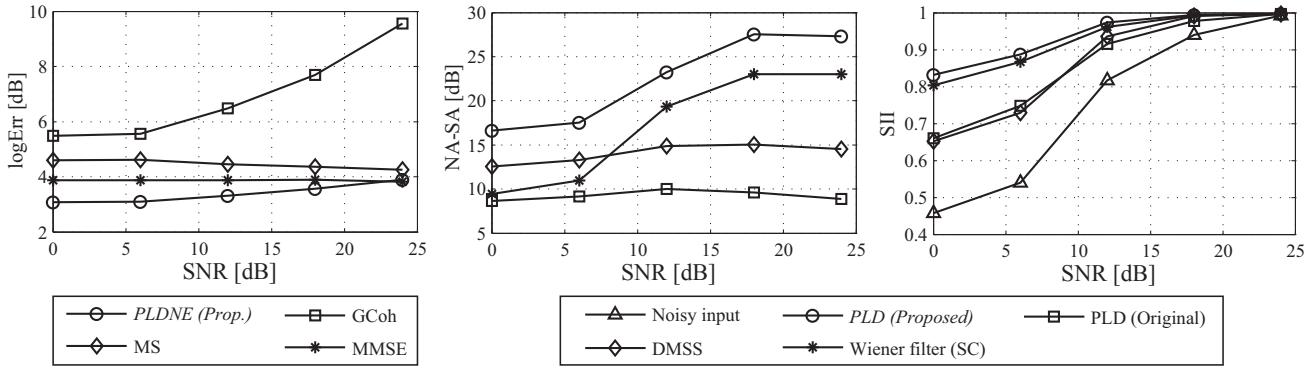

Fig. 6. Simulation results: (left) noise estimation accuracy, (middle) noise suppression performance, (right) influence on intelligibility. NA-SA: noise attenuation minus speech attenuation, SII: speech intelligibility index.

spacing of $0.15 - 0.2\,\mathrm{m}$. Besides, the two single-channel approaches Minimum Statistics (MS) [5] and MMSE-based noise tracker (MMSE) [6], which work on the primary signal $x_1(k)$ only, are used as state-of-the art references. The performance is rated in terms of the symmetric segmental logarithmic estimation error between the ideal noise PSD $\Phi_{nn}(\lambda, \mu)$ and the estimated noise PSD $\hat{\Phi}_{nn}(\lambda, \mu)$ by

$$\mathrm{logErr} = \frac{1}{K\,M} \sum_{\lambda=1}^{K} \sum_{\mu=1}^{M} \left| 10 \log_{10} \left[\frac{\Phi_{\mathrm{nn}}(\lambda, \mu)}{\hat{\Phi}_{\mathrm{nn}}(\lambda, \mu)} \right] \right|, \qquad (21)$$

with total number of frames K. The ideal noise PSD is obtained using the true noise periodograms smoothed over time λ with smoothing factor α_{nn}. The averaged results for babble and traffic noise are depicted in Fig.6 (left). It can be seen that the novel algorithm outperforms all related approaches and is nearly independent of the input SNR.

4.2. Noise Reduction Performance

The performance of the PLD weighting rule (Proposed) using Eq.(18) is compared with the original implementation by [8] (Original) using Eq.(15) and a single-channel (SC) Wiener filter with decision-directed approach for the a priori SNR calculation. All algorithms use the PLDNE noise PSD estimator. Besides, a dual-channel spectral subtraction algorithm [10] (DMSS) is evaluated. In a first step, two common spectral subtraction approaches provide a rough speech and noise estimate for each channel by using the other channel respectively. In a following step these estimates are used by a third spectral subtraction stage which results in the enhanced output. The noise reduction performance is determined by means of the noise attenuation minus speech attenuation (NA-SA) measure, where higher values indicate an improvement. Besides, the speech intelligibility index (SII) [11] was calculated for the noisy as well as the enhanced signal. An SII higher than 0.75 indicates a good communication system and values below 0.45 correspond to a poor system. The averaged results for babble and traffic are shown in Figs. 6 (middle/right). From the plots, we can conclude that the proposed noise reduction system outperforms related approaches in terms of noise suppression performance and increase in speech intelligibility. The modifications on the original PLD implementation also result in a high performance gain. All results are consistent with the subjective listening impression where the highest amount of musical tones was observed for the DMSS algorithm. Since for babble noise the major frequency components of the noise signal lie in the same regions as those of the desired speech signal, this scenario can be seen as the most difficult one. However, all experiments have also been conducted with train station noise where the same tendency has been observed.

5. CONCLUSIONS

We propose a noise reduction system which is suitable for speech enhancement in dual-microphone mobile phones. A novel noise PSD estimator as well as a modified spectral weighting rule are presented, which both exploit the power level differences of the desired speech signal between the microphones. The algorithms require a low computational complexity and can efficiently be implemented using first order IIR filters for the auto- and cross-PSD estimation. Experiments have shown that the novel system is capable of reducing unwanted background noise and increase the intelligibility in terms of the SII measure.

6. REFERENCES

[1] L. Watts, "Advanced noise reduction for mobile telephony," *IEEE Computer Magazine*, vol. 41, no. 8, pp. 90–92, 2008.

[2] 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.

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

[4] H. Kuttruff, *Room Acoustics*, Spon Press, Oxon, 2009.

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

[6] 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, USA, 2010.

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

[8] N. Yousefian, A. Akbari, and M. Rahmani, "Using power level difference for near field dual-microphone speech enhancement," *Applied Acoustics*, vol. 70, pp. 1412 – 1421, 2009.

[9] T. Esch and P. Vary, "Efficient musical noise suppression for speech enhancement systems," in *Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Taipei, Taiwan, 2009.

[10] H. Gustaffson, I. Claesson, S. Nordholm, and U. Lindgren, "Dual microphone spectral subtraction," Tech. Rep., Department of Telecommunications and Signal Processing, University of Karlskrona/Ronneby, Sweden, 2000.

[11] ANSI S3.5-1997, *Methods for the Calculation of the Speech Intelligibility Index*, ANSI, r2007 edition, 2007.