22. Konferenz Elektronische Sprachsignalverarbeitung (ESSV), September 2011, Aachen, Germany (TuDPress, ISBN 978-3-942710-37-4)

Evaluation of Single- and Dual-channel Noise Power Spectral Density Estimation Algorithms for Mobile Phones

Christian Herglotz, Marco Jeub, 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,nelke,vary}@ind.rwth-aachen.de
christophe.beaugeant@intel.com

Abstract: Noise suppression has become a standard option for high end mobile phones. One essential component for most of these algorithms is the estimation of the noise power spectral density (PSD) of the unwanted background noise. A large number of single and dual-channel approaches that exploit a large variety of different signal properties have been presented in literature. So far, a comprehensive evaluation of these approaches in realistic and reproducible situations does not exist. In this contribution, the performance of well-known as well as recently proposed noise PSD estimators is analyzed with respect to an integration in dual-microphone mobile phones. Additionally, a new algorithm is proposed which exploits explicitly the dual-microphone configuration of state-of-the-art mobile phones and smartphones.

1 Introduction

In recent years background noise reduction in mobile communication devices has been subject to extensive research. As manufacturing costs drop and technology develops, nowadays it is more and more common to integrate more than one microphone in the mobile device. As a consequence, new algorithms are developed which benefit from the additional information captured by a second microphone.

A key part of many existing noise reduction algorithms is the noise estimation. The objective is to find an accurate and reliable estimate of the current PSD of the background noise. Common single-channel algorithms often reach their limits in difficult situations like non-stationary noise or low input signal-to-noise-ratios (SNR). The use of multiple microphones provides the possibility to exploit further properties of the input signals, e.g., the coherence or the direction-of-arrival of the multichannel input signal. Thus, typical single-channel drawbacks can be overcome.

Recent developments concentrate on the use of two microphones. In this dual-channel approach, two alignments are of special interest, cf. Figure 1: The first possibility is to place one microphone at the bottom as in classic mobile phones and a second microphone placed on the top backside of the mobile which will be referred to as bottom-top (BT) in the remainder of the paper. In the second considered alignment, two closely-spaced microphones are located at the bottom of the mobile which will be referred to as bottom-bottom (BB). When single-channel algorithms are presented in this contribution, only the signal of microphone B1 is taken into account. Only the so-called handset mode of mobile phones is taken into account where the loudspeaker of the mobile device is pressed on the ear and the bottom microphones are held close to the mouth of the speaker. Similar performance evaluations of dual-channel noise PSD estimation algorithms have been published in [14] and for single-channel algorithms recently in [15]. However, in this contribution we give a more elaborate evaluation and extend the experiments to recently proposed dual-channel methods.Real recordings with mock-up phones were conducted to analyse the noise estimation algorithms in realistic conditions.

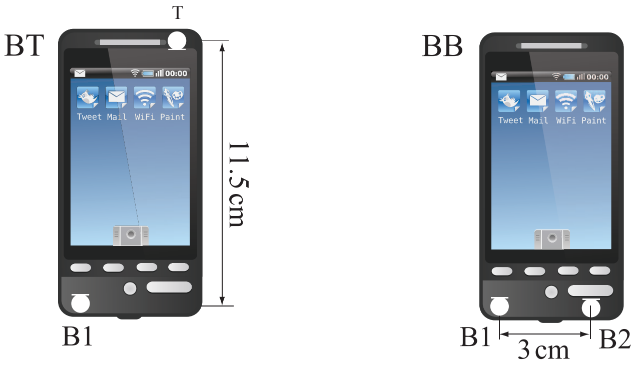


Figure 1 - Considered microphone alignments for dual-microphone mobile phones. The left plot shows the bottom-top (BT) alignment and the right plot the bottom-bottom (BB) alignment.

2 Noise Reduction Framework

A block diagram of the considered noise reduction system for dual-microphone mobile phones is depicted in Figure 2. Since state-of-the-art mobile communication is constrained to single-channel transmission, only a single-channel output has to be computed from the dual-channel input signals. We assume the input signal $x_i(m)$ to be the sum of the desired speech $s_i(m)$ and the noise $n_i(m)$, where m is the sample and i the microphone index ($x_i(m) = s_i(m) + n_i(m)$). The signals are segmented, windowed and then transformed to the frequency domain. We obtain the signals $X_i(\lambda, \mu)$ in the frequency domain, where λ is the frame and μ the frequency

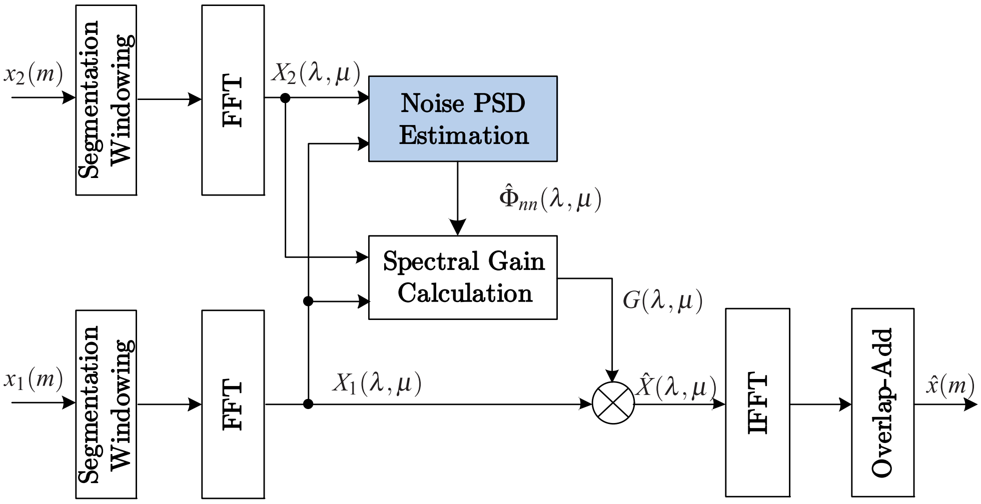


Figure 2 - Block diagram of the dual-channel noise reduction system where the different noise PSD estimators are employed.

Table 1 - Main simulation parameters.

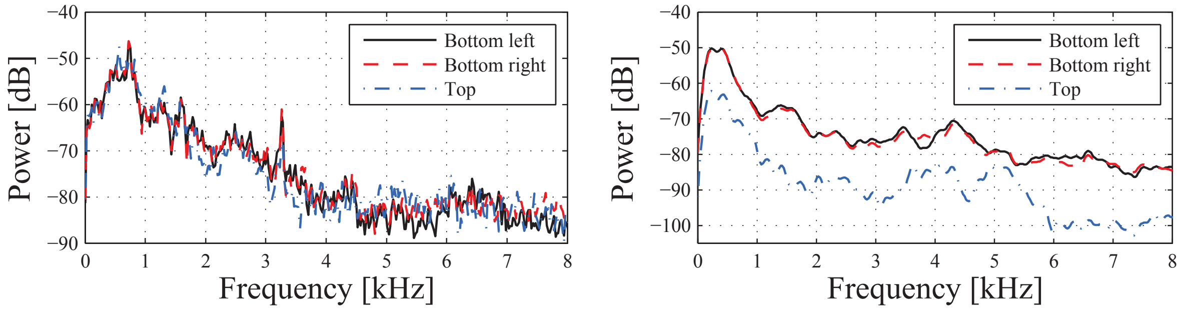
Parameter	Settings
Sampling frequency	$f_s = 16\,\mathrm{kHz}$
Frame length	$L = 320$ (20 ms)
FFT length	$M = 512$ (including zero-padding)
Frame overlap	50% (Hann window)

bin index. $\hat{\Phi}_{nn}(\lambda,\mu)$ is the estimated noise PSD and $G(\lambda,\mu)$ contains the output spectral weighting gains. $\hat{X}(\lambda,\mu)$ and $\hat{x}(m)$ are the enhanced output signals in the frequency and time domain, respectively. The used simulation parameters are given in Table 1.

In this paper we focus on the noise estimation module. Thus, only the output of the noise PSD Estimation $\hat{\Phi}_{nn}(\lambda,\mu)$ will be examined. Furthermore, a delay compensation of the useful speech between the microphones is presumed to have already been performed.

3 Measurement Setup

The background noise analysis is based on measurements inside an acoustic chamber using the standardized procedure described in [4] to generate realistic, i.e., diffuse noise fields. Here, we restrict the analysis to two commonly used noise types: car and babble noise from [4]. Exemplary, the average periodograms of babble noise for the three microphones are depicted in Figure 3(a). The recording system consists of a HEAD acoustics HMS II.3 artificial head which includes a mouth simulator. Two mock-up phones were build by integrating two omnidirectional Beyerdynamic MM1 measurement microphones in a 6x12x3 cm^3 plastic housing.



(a) Pub noise in diffuse sound field

(b) Speech generated by the the artificial mouth of the dummy head

Figure 3 - Average periodograms of the recorded input in the bottom left, bottom right and top microphone. The averaged signals were taken from measurements. Differences can be explained by the distance and the mounting direction of the specific microphones.

The phones were mounted on the artificial head in the flat handset position. This procedure allows to record speech and noise separately which is usually not possible in real acoustic environments. As speech input, the set of English sentences spoken by a woman and a man from the speech database presented in [8] was used. The average periodograms are depicted in Figure 3(b). The attenuation of the speech signal in the top microphone with respect to the bottom microphones is found to be nearly constant, usually higher than 10dB. We chose a signal length of 16.6s. As input SNR we chose four different values: −5dB, 5dB, 15dB and 25dB.

As the signals are recorded in a realistic environment, the coherence of the speech is lower than one. The effective, average magnitude squared coherence (MSC) of the measured speech is depicted in Figure 4. It can be seen that in both alignments, no perfect coherence of the desired speech signal (a coherence close to one) is achieved. Especially in BT one can find some significant notches which are caused by shadowing effects, microphone mismatch and due to the different orientation of the microphones.

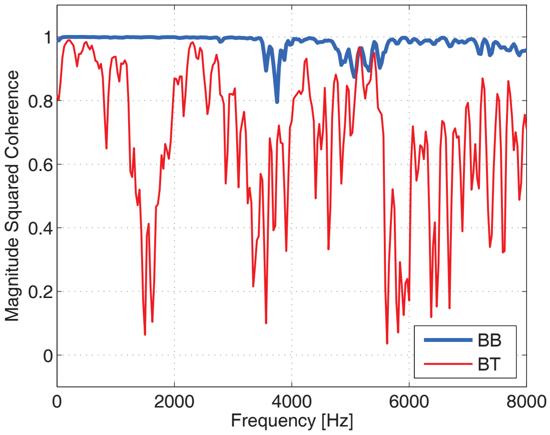
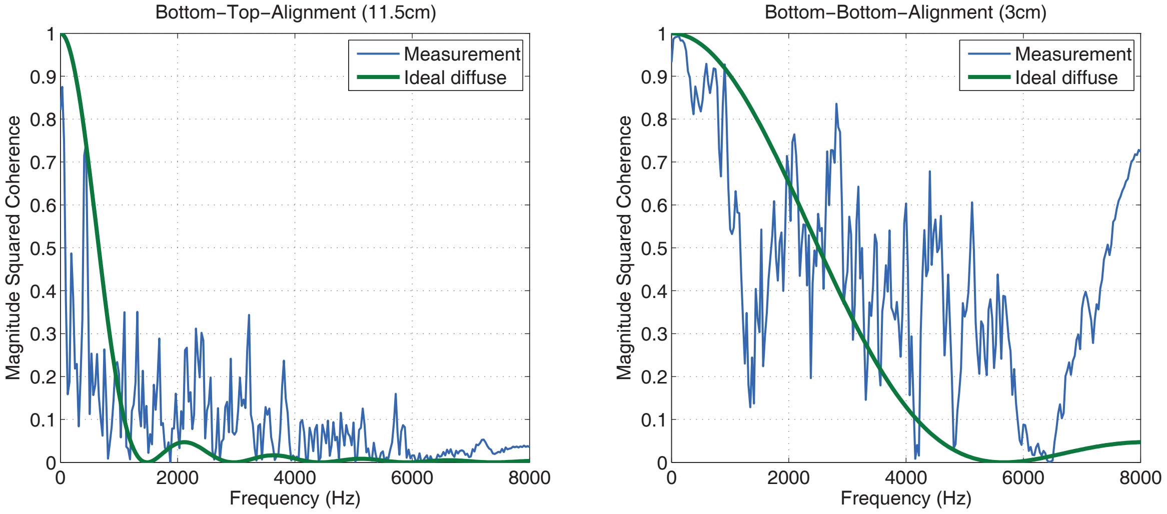


Figure 4 - Magnitude squared coherence of speech signals recorded in an acoustic chamber for both investigated alignments. The BT alignment shows considerable notches.

The MSC of the considered background noise is shown in Figure 5 where differences to the theoretical diffuse noise field models can be observed. However, the curves are always below the speech MSC and the overall shape of the curves corresponds to the diffuse coherence models.



(a) Bottom-Top Alignment

(b) Bottom-Bottom Alignment

Figure 5 - Magnitude squared coherence of measured noise signals in BT (a) and BB (b) alignment. High coherence in low and low coherence in high frequencies can be explained by the diffusity of the noise field.

4 Noise PSD Estimation Algorithms

Since we restrict the algorithms to a possible application in mobile phones, only a limited number of single- and dual-channel algorithms have been selected that show most promising effectiveness. Due to the limitation of space, the twelve analyzed algorithms are only briefly described in this section. For further details, the reader is referred to the corresponding publications. The required tuning parameters are selected as proposed in the publications.

1. The *ideal noise estimate* (Perf) takes the pure, unaltered noise input and is used as a reference. To obtain ”fair” results it is recursively smoothed over time using a constant smoothing factor ($\alpha = 0.9$). This also explains the error which is unequal to zero in the evaluation part.

2. The *First-T-Frames* (FTF) algorithm is a single-channel algorithm and uses the first T frames (here: T = 20 frames) of the input signal assuming no speech activity as well as stationary noise. During these first frames, the noise PSD is calculated using recursive smoothing ($\alpha = 0.9$). The resulting noise PSD is kept and used throughout the entire length of the input signal.

3. The *Minimum Statistics* (MS) algorithm is a single-channel method explained in [13]. The algorithm tracks the noise PSDs by searching the minima in every frequency bin of the recent input frames from a given time span (which is usually set to $1.5\text{s} - 2\text{s}$). The noise PSD is smoothed over time using adaptive smoothing parameters. Furthermore, a bias compensation and a maximum noise-slope is applied which only allows slow noise changes to be tracked.

4. The *MMSE-Tracker* is an algorithm proposed by Hendriks et al. [6]. It is developed for single-channel applications and is based on a statistical approach to determine an expected value of the noise PSD. It estimates the expected noise under the condition of the input signal X and an estimated a-priori-SNR. As probability density function (PDF) for speech and noise, a Gaussian distribution in the frequency domain is assumed for both. The a-priori-SNR is estimated using the spectral smoothing proposed by Ephraim and Malah in the Decision Directed Approach [2]. Furthermore a bias correction is performed. The MMSE-Tracker requires an estimation of the speech-PSD using an algorithm that calculates spectral gains for speech enhancement. Here, the approach presented in [3] is used.

5. The *coherence-based* algorithm (Coh) by Dörbecker [1] estimates the background noise by exploiting the coherence between the signals arriving at both microphones. If the arriving signal is coherent it is considered as speech. If it is not coherent it is considered as noise. The averaged power level of the incoherent part of both signals is used to estimate the noise PSD. This method is applicable for both microphone alignments.

6. The *enhanced coherence-based* algorithm (ECoh) [7] is a generalized dual-channel noise PSD estimator which exploits a-priori-knowledge about the noise field coherence. The approach (Coh) [1] can be seen as a special case for an uncorrelated background noise assumption. Here, an ideal diffuse noise field as depicted in Figure 5 is assumed.

7. The *binaural approach* by Kamkar-Parsi (KKP) et al. [9] exploits a similar approach as the ECoh. The authors also assume that the coherence function of the background noise is known. However, as a preprocessing step, a simple coherence based Wiener Filter as

presented in [16] is applied on the input signal to extract the coherent parts of the signal. This algorithm is applicable for both alignments.

8. The *Channel 2* approach (Ch2) assumes that in the top microphone of the BT-alignment the speech signal arrives with much lower sound pressure level than in the bottom microphone. Assuming that the noise power is equal at both microphones (which holds, e.g., in diffuse noise fields), the smoothed signal from the top microphone is directly used as a noise estimate.

9. The ninth algorithm is the *Dual Microphone Spectral Subtraction* approach (DMSS) by Gustaffson et al. in [5]. It is applicable for the BT-alignment and derives the estimated noise PSD in two steps: At first a rough speech estimate is calculated by spectrally subtracting the noise estimate of the last frame from the current input frame of the bottom microphone. Afterwards, this speech estimate is subtracted from the current input frame of the top microphone. Thus, only the noise part of the top microphone should remain.

10. The *phase-based* approach (Phase) by Kim et al. is a dual-channel approach for the BB allignment presented in [12]. It evaluates the phase difference of the input signal and derives a speech presence probability. If the probability is below a given threshold the input is assumed to be noise. It is used to update the estimated noise PSD. Otherwise the noise PSD of the last frame is kept.

11. The *weighted noise estimator* (WNE) is a single-channel approach taken from Kawamura et al. [11], originally taken from Kato et al. [10]. It uses the estimated noise PSD of the last frame to estimate a current SNR. If the SNR is negative the input is assumed to be pure noise and used as a noise estimate. If the SNR is very high the estimate of the last frame is kept. In between these extremes a weighted update of the noise estimate is performed.

12. The new algorithm (Proposed) is explicitly designed for the BT microphone alignment. It exploits the difference of the power levels in both channels in order to distinguish between speech and noise and generates a noise estimate using this information. Investigations showed that results are comparable to the MMSE-Tracker by Hendriks. At the same time this new approach shows a much lower computational complexity.

To get a rough insight into the computational complexity of the algorithms, the computing times in the used MATLAB simulations can be considered. Figure 6 depicts the normalized computation times, where Minimum Statistics has been used as a reference. As a first impression one

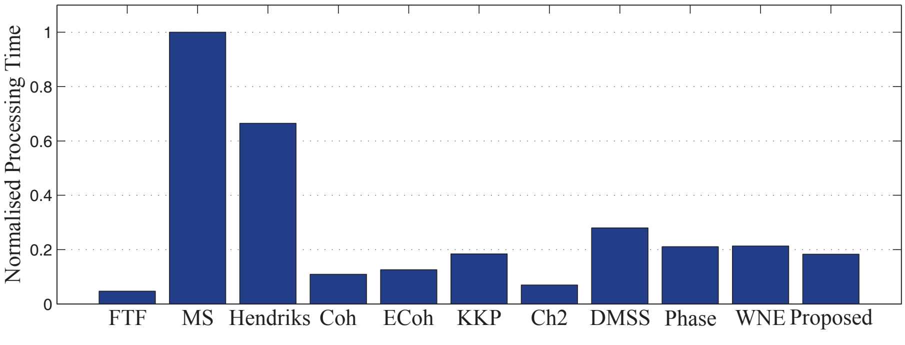


Figure 6 - Normalized processing times of the investigated algorithms in Matlab-Environment.

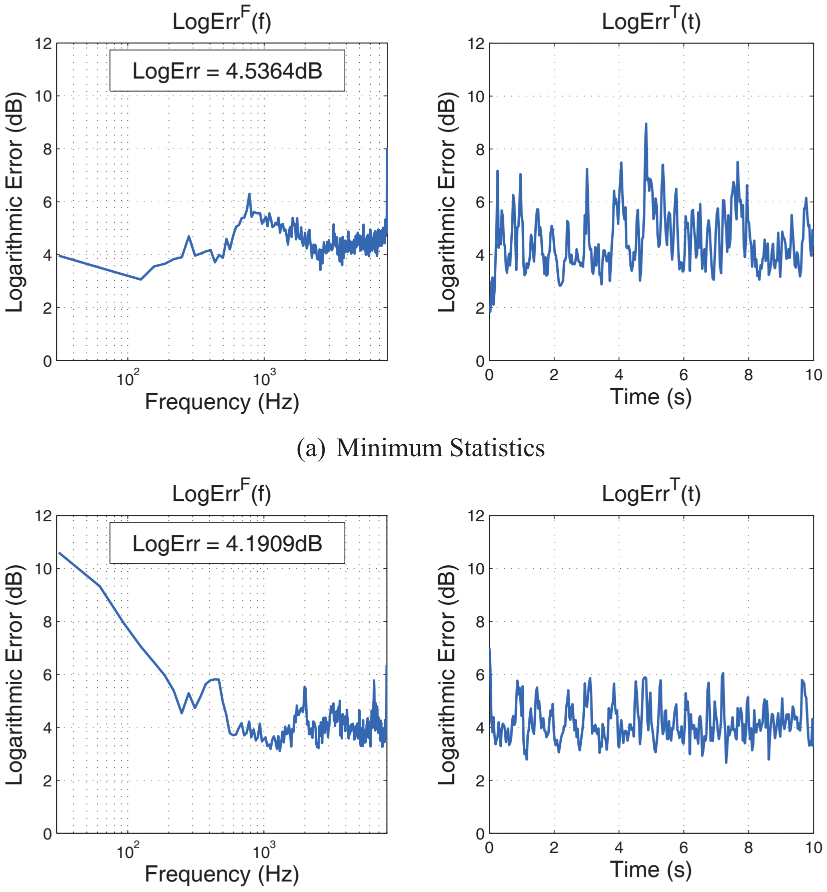
can see that the investigated dual-channel approaches usually show much less complexity than the single-channel-based methods.

5 Evaluation

The objective measure used to determine the accuracy of a given noise estimate is the so-called logarithmic error (LogErr) as used in [6, 7]. It is defined as a distance measure between the estimated noise power $\hat{\Phi}_{nn}$ and the ideal noise power Φ_{nn}. As a consequence, the LogErr can only be calculated when the ideal pure noise is known, which is the case in our simulations. The corresponding equation is given by

$$\text{LogErr}[dB] = \frac{1}{\Lambda M} \sum_{\lambda=1}^{\Lambda} \sum_{\mu=1}^{M} \left| 10 \cdot \log_{10} \left(\frac{\Phi_{nn}(\lambda,\mu)}{\hat{\Phi}_{nn}(\lambda,\mu)} \right) \right| . \tag{1}$$

A more detailed analysis is obtained when investigating the LogErr over time and frequency separately. As an example, Figure 7(a) and 7(b) show the LogErr of the MS and the Coh algorithm. The error is calculated using a noisy speech signal at 0 dB SNR (babble noise) in the BT-alignment. In the time-domain plot one can see that the estimate of Coh shows less variation.



(a) Minimum Statistics

(b) Coherence based approach

Figure 7 - Logarithmic errors over time and frequency. (a) depicts the estimation error for Minimum Statistics [13] and (b) for the coherence based approach [1].

MS shows deficiencies especially when the background noise changes which results in peaks in the time plot. On the other hand, in the frequency domain evaluation, one can see that MS shows a rather constant performance over all frequencies. In contrast, the performance of Coh diminishes strongly below 300 Hz. This is caused by the high coherence of low frequency noise in a diffuse field. As a consequence we decided to evaluate the performance of the algorithms in two separate frequency bands: 0..1 kHz and 1..8 kHz.

6 Results

The results of the simulations are depicted in Figure 8. As a first impression we can see that the MMSE-Tracker by Hendriks and the proposed approach return the best results. In terms of single-channel performance, the results are in accordance with the evaluation paper by Taghia et al. [15], where, e.g., the MMSE-Tracker was found to be superior to the Minimum Statistics approach. One can also see that many algorithms show their best performance in medium SNR-cases. In high SNRs the results are often unsatisfying because the dominant speech prevents an accurate estimate. The high LogErr of the Coh-algorithm in low frequencies is caused by the high coherence of the background noise in this frequency region. The high value of the FTF can be explained by the short period of pure noise before speech starts. As a consequence the first speech parts are part of the noise estimate.

7 Conclusions

We have developed a framework that allows to evaluate and compare different noise PSD estimation algorithms which are applicable for noise reduction in dual-microphone mobile phones. When only one microphone is available, the single-channel MMSE noise tracker by Hendriks should be employed. With respect to noise reduction, the performance gain due to a secondary microphone can only be fully exploited in the bottom-top configuration which should be the design target for any future mobile communication device. The proposed dual-channel algorithm outperforms all related algorithms in terms of estimation accuracy and computational complexity, which was confirmed by experiments.

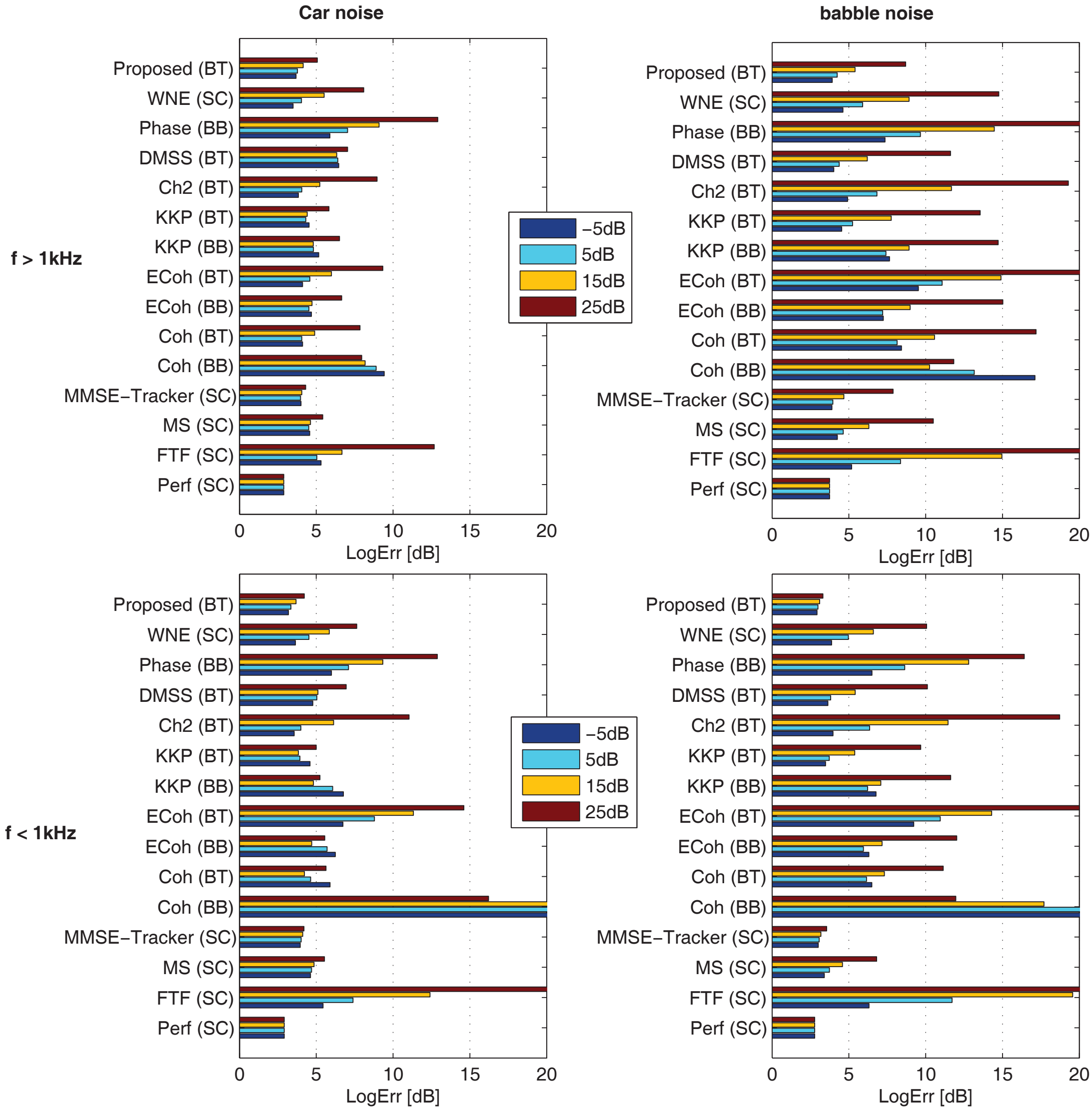


Figure 8 - Results of investigated noise PSD estimation algorithms in terms of the logarithmic error (LogErr). The left diagrams show the results for car noise and the right diagrams for babble noise. The top diagrams depict the performance above and the bottom diagrams the performance below 1kHz.

References

[1] Dörbecker, M. 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.

[2] Ephraim, Y. and D. Malah: *Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator*. IEEE Trans. on Acoustics, Speech and Signal Process., 32(6):1109–1121, 1984.

[3] Erkelens, J. S., R. C. Hendriks, R. . Heusdens and J. . Jensen: *Minimum Mean-Square Error Estimation of Discrete Fourier Coefficients With Generalized Gamma Priors*. IEEE Transactions on Audio, Speech, and Language Processing, 15(6):1741–1752, 2007.

[4] 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.

[5] Gustaffson, H., I. Claesson, S. Nordholm and U. Lindgren: *Dual Microphone Spectral Subtraction*. Techn. Rep., Department of Telecommunications and Signal Processing, University of Karlskrona/Ronneby, Sweden, 2000.

[6] Hendriks, R. C., 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)*, pp. 4266–4269, Dallas, TX, USA, 2010.

[7] Jeub, M., 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] Kabal, P.: *TSP Speech Database*. Techn. Rep., Department of Electrical & Computer Engineering, McGill University, Montreal, Quebec, Canada, 2002.

[9] Kamkar-Parsi, A. H. 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., 17(4):521–533, 2009.

[10] Kato, M., A. Sugiyama and M. Serizawa: *Noise Suppression with High Speech Quality BAsed on Weighted Noise Estimation and MMSE STSA*. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Darmstadt, Germany, 2001.

[11] Kawamura, A., W. Thanhikam and Y. Iihuni: *A Speech Spectral Estimator Using Adaptive Speech Probability Density Function*. In *Proc. European Signal Processing Conference (EUSIPCO)*, Aalborg, Denmark, 2010.

[12] Kim, K., S.-Y. Jeong, J.-H. Jeong, K.-C. Oh and J. Kim: *Dual channel noise reduction method using phase difference-based spectral amplitude estimation*. In *Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 217–220, Dallas, TX, USA, 2010.

[13] Martin, R.: *Noise power spectral density estimation based on optimal smoothing and minimum statistics*. IEEE Trans.on Speech and Audio Process., 9(5):504–512, 2001.

[14] Meyer, J., K. Simmer and K. Kammeyer: *Comparison Of One- And Two-Channel Noise-Estimation Techniques*. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, London, UK, 1997.

[15] Taghia, J., 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. IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, 2011.

[16] Vary, P. and R. Martin: *Digital Speech Transmission. Enhancement, Coding and Error Concealment*. Wiley&Sons, Chichester, 2006.