NOISE PSD ESTIMATION BY LOGARITHMIC BASELINE TRACING

Florian Heese and Peter Vary

Institute of Communication Systems and Data Processing (ind)
RWTH Aachen University, Germany
{heese,vary}@ind.rwth-aachen.de

ABSTRACT

A novel noise power spectral density (PSD) estimator for disturbed speech signals which operates in the short-time Fourier domain is presented. A noise PSD estimate is provided by constrained tracing with time of the noisy observation separately for each frequency bin. The constraint is a limitation of the logarithmic magnitude change between successive time frames. Since speech onset is assumed as sudden rises in the noisy observation, a fixed and adaptive tracing parameter β has been derived to track the contained noise while preventing speech leakage to the noise PSD estimate. The experimental evaluation and comparison with state-of-the-art algorithms, *SPP* and *Minimum Statistics*, confirms a lower logarithmic noise estimation error and superior speech enhancement rated in a standard noise reduction system. The proposed concept has extremely low computational complexity and memory usage. Thus, it is well suited for applications where processing power and memory is limited.

***Index Terms*—** Noise power estimation, speech enhancement, noise reduction, low complexity, low memory

1. INTRODUCTION

In mobile communication voice is often captured in acoustically disturbed environments. A noisy *near end* signal, e.g., captured by a microphone, is usually enhanced for the *far end* by reducing the noise while preserving the target speech signal as much as possible, e.g., [1, 2, 3, 4, 5, 6, 7]. The intelligibility of a clean *far end* signal perceived in strong *near end* environmental noise can be enhanced by a pre-processing of the *far end* signal, e.g., [8, 9]. All mentioned algorithms rely on an estimate of the noise power spectral density (PSD). Thus, the noise PSD estimation is one of the most important prerequisite for speech enhancement.

1.1. Relation to prior work

If the noise is stationary or only slowly varying with time, a noise PSD estimate can either be obtained during speech pauses or by continuously tracking the magnitude minima in the short-time Fourier domain. Further processing and updating over time is necessary. Several methods have been proposed for the estimation of noise PSD by tracking and post-processing the magnitude minima in the short-time Fourier domain, e.g., [2, 3, 4, 10, 11, 12, 13].

In [2] the noise spectrum is estimated for each frequency bin based on a smoothed periodogram over time of the noisy observation by nonlinear temporal minima tracking. If the last noise PSD estimate is smaller than the current noisy observation the tracking is realized by a weighted average of the last and current noisy frame. In the other case the current noisy observation serves as new noise PSD estimate.

The *Minimum Statistics* [3, 4] method is based on two assumptions: speech and noise are statistically independent and the power of the noisy signal often decays to the power level of the noise. Using a smoothed periodogram of the noisy signal it is possible to track a minimum separately for each frequency within a certain time window to obtain a noise PSD estimate. The duration of the time window for the minimum search states a trade-off between fast noise tracking and speech portions in the noise PSD estimate.

The *SPP* algorithm [12] (a further development of [11]) estimates the noise PSD for each frequency by a smoothed linear combination of the current observed noisy PSD and the last estimate of the noise PSD weighted by the speech presence and speech absence probability, respectively. The determination of speech presence probability depends on the observed noisy PSD, the last noise PSD estimate and a threshold parameter.

These approaches take a quite significant portion of the memory capacity and the computational power of the whole enhancement algorithm. The application of speech enhancement in hearing aids or low cost mobile phones require low complexity and low memory algorithms. In this contribution a new noise PSD estimator operating in the short-time Fourier domain is presented and evaluated in comparison with [2], *Minimum Statistics* [3, 4] and *SPP* [12].

2. SIGNAL MODEL

The noisy input signal $x(k)$ consists of a clean speech signal $s(k)$ additively degraded by a noise component $n(k)$ according to:

$$x(k) = s(k) + n(k), \tag{1}$$

where k is the discrete time index.

Since the noise PSD estimation is performed in the short-time Fourier domain, $x(k)$ is segmented into overlapping frames of length L_F with frame advance L_A, followed by windowing with a square root Hann-window and zero-padding. Subsequently, each frame is transformed by applying the fast Fourier transform (FFT) of length M_F. The spectral coefficients of the input signal $x(k)$ at frequency bin μ and frame λ are given by:

$$X(\lambda, \mu) = S(\lambda, \mu) + N(\lambda, \mu), \tag{2}$$

where $S(\lambda, \mu)$ and $N(\lambda, \mu)$ correspond to the spectral coefficients of the speech and noise signal, respectively.

3. PROPOSED NOISE PSD BASELINE TRACING

The noise estimation problem is formulated in the logarithmic amplitude domain, while the actual processing is carried out with linear amplitudes. This procedure is beneficial for the following reasons:

- the linear domain processing is computationally less complex than in the log domain,
- the log domain estimator is inherently unbiased and does not need correction terms like *Minimum Statistics* [3, 4],

978-1-4673-6997-8/15/$31.00 ©2015 IEEE

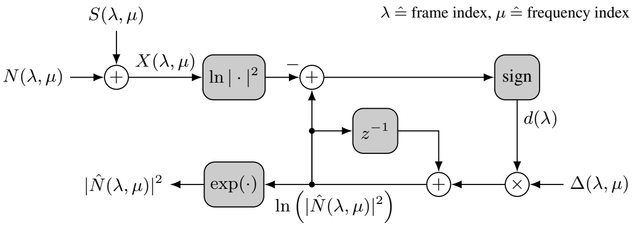
$S(\lambda,\mu)$
$\lambda \hat{=}$ frame index, $\mu \hat{=}$ frequency index
$N(\lambda,\mu)$
$X(\lambda,\mu)$
$\ln|\cdot|^2$
$-$
sign
$d(\lambda)$
z^{-1}
$|\hat{N}(\lambda,\mu)|^2$
$\exp(\cdot)$
$\ln\left(|\hat{N}(\lambda,\mu)|^2\right)$
$\Delta(\lambda,\mu)$

Fig. 1. Equivalent block diagram of proposed noise PSD estimator

- the log domain formulation of the estimator does not need explicit amplitude normalization.

The equivalent log domain block diagram of the proposed noise PSD estimator is depicted in Fig. 1. The estimator can be explained in terms of delta modulation with an adaptive step size $\Delta(\lambda,\mu)$. For each fixed frequency bin μ, the variable step size is deliberately adjusted such that the estimate $\ln|\hat{N}(\lambda,\mu)|^2$ follows the base line of the log noisy sub-band (*Baseline Tracing*).

In a first order delta modulator, the input signal is traced by an estimate which increases or decreases with a linear slope, which is determined by the step size Δ and the sign of the error between the input and the estimate. By adaptive control of the step size, the delta modulator is operated here in the slope overload mode [14] such that the estimate follows the base line, which is determined by the noise. Due to the additive noise, the magnitudes of the speech component frequently decay to the level of the noise component. This is also exploited by *SPP* [12] and *Minimum Statistics* [3, 4]. By means of a stationary noise component it can be seen, that the signum series $d(\lambda) \in \{-1, 0, 1\}$ alternates with time step λ and is zero mean on average. Thus, the proposed estimator is unbiased expect of the granular noise known from delta modulation. In contrast to delta modulation $d(\lambda) = 0$ is allowed, which is favorable as the noise estimation may exactly match the, e.g., constant input.

For complexity reasons, the logarithmic noise PSD estimator is implemented in the linear amplitude domain. The resulting equations (3) and (4) are partly similar to [13]. However, the adaptation mechanism is significantly different and the control is effective in the log amplitude domain. Given a noise estimate $\hat{N}(\lambda-1,\mu)^2$ from the last frame, the current estimate $\hat{N}(\lambda,\mu)^2$ is calculated by stretching or compressing the last estimate with the tracing factor $\beta(\mu)$ in each frequency bin. The tracing factor β is equivalent to $\exp\left(\Delta(\lambda,\mu)\right)$ and can be realized frequency dependent or independent. A further option is to use a time varying $\beta(\lambda,\mu)$ in analogy to the adaptive step size control in delta modulation [14, 15]. As criterion for stretching or compressing, the signum function is used. If the difference between the current noisy observation $X(\lambda,\mu)$ and the last estimate $\hat{N}(\lambda-1,\mu)$ is greater than zero, $\hat{N}(\lambda-1,\mu)$ will be stretched by β and compressed by $1/\beta$ in the other case. The estimation step, which is equivalent to the "Delta Modulation Algorithm" in the log amplitude domain of Fig. 1, is described by the following equations:

$$|\hat{N}(\lambda,\mu)|^2 = |\hat{N}(\lambda-1,\mu)|^2 \cdot \beta(\lambda,\mu)^{D(\lambda,\mu)}, \tag{3}$$

$$D(\lambda,\mu) = \operatorname{sign}\left(\ln|X(\lambda,\mu)|^2 - \ln|\hat{N}(\lambda-1,\mu)|^2\right) \tag{4}$$

$$= \operatorname{sign}\left(|X(\lambda,\mu)|^2 - |\hat{N}(\lambda-1,\mu)|^2\right), \tag{5}$$

with the initialization of the first estimate $|\hat{N}(1,\mu)|^2 = |X(1,\mu)|^2$.

A proof of concept example for a frequency bin corresponding to a frequency of 1816 Hz is depicted in Fig. 2. Therefore, a noisy signal consisting of factory1 noise [16] and a female speaker randomly taken from the NTT database [17] at 5 dB SNR was processed with a frequency independent $\beta(\lambda,\mu) = 1.05$, which corresponds to approx. 5 % change in $|\hat{N}(\lambda,\mu)|^2$ from frame to frame. In the lower plot the clean speech and noise signal can be seen, while in the upper plot the noisy mixture and the noise PSD estimate are depicted. It is visible that the simple concept of the new estimator is able to track the noise.

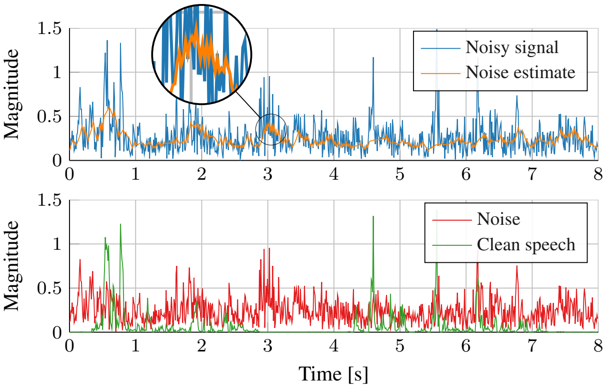

Fig. 2. Magnitude over frames for bin 59 (1816 Hz)

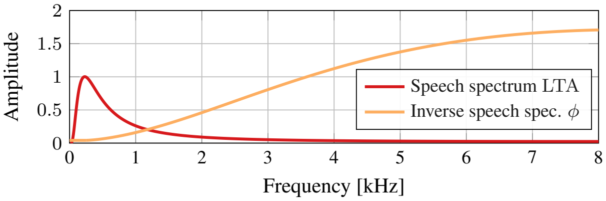

Fig. 3. Long-term speech spectrum $\mathrm{LTA}(f)$ plotted in the linear domain normalized for clarity to a max of one and its inverse $\phi(\mu)$

4. TRACING FACTOR β

Although the choice of $\beta = 1.05$ in the previous example (Fig. 2) works properly, it seems reasonable to define a frequency and time (frame) dependent scaling factor β:

$$\beta(\lambda,\mu) = 1 + \alpha(\lambda)\phi(\mu), \tag{6}$$

where α represents the time and ϕ the frequency dependent component. Since compression or stretching is realized by multiplication and division, β has to be greater than one.

4.1. Speech dependent scaling $\phi(\mu)$ over the frequency

If β is too large, $|\hat{N}(\lambda,\mu)|^2$ follows unintentionally also the speech signal and the noise PSD estimate thus contains parts of speech. In order to prevent that speech contributes to the noise PSD estimate, the tracking speed for speech relevant frequencies is decreased while allowing faster tracking at the remaining frequencies. Therefore, $\phi(\mu)$ is chosen proportional to the inverse of the long-term speech spectrum average (LTA) as shown in Fig. 3 with the definition of the LTA [18]

$$\begin{aligned}\mathrm{LTA}(f)|_{\mathrm{dB}} = &-376.44 + 465.439\log_{10}(f) \\ &- 157.745\log_{10}^2(f) + 16.7124\log_{10}^3(f),\end{aligned} \tag{7}$$

where f is the frequency in Hz. A piece-wise approximation of the inverse long-term speech spectrum average $\mathrm{INV}_{\mathrm{LTA}}(\mu)$ is introduced,

$$\mathrm{INV}_{\mathrm{LTA}}(\mu) = \begin{cases} \left(10^{\mathrm{LTA}\left(\frac{f_s}{M_F}\mu\right)/20}\right)^{-1} & \text{if } \frac{f_s}{M_F}\mu \geq 230\,\mathrm{Hz} \\ \left(10^{\mathrm{LTA}(230\,\mathrm{Hz}/20)}\right)^{-1} & \text{if } \frac{f_s}{M_F}\mu < 230\,\mathrm{Hz}, \end{cases} \tag{8}$$

which ensures a smooth transition at low frequencies. Hence, the new speech dependent $\phi(\mu)$ is specified as:

$$\phi(\mu) = \frac{M_F \cdot \mathrm{INV}_{\mathrm{LTA}}(\mu)}{\sum_{i=0}^{M_F-1} \mathrm{INV}_{\mathrm{LTA}}(i)}. \tag{9}$$

Note, $\phi(\mu)$ is normalized to a mean of one. Both, the long-term speech spectrum and its inverse $\phi(\mu)$ are depicted in Fig. 3.

4.2. Fixed scaling α with the time

As mentioned above, a large β leads to an erroneous noise PSD estimate including also speech. As $\phi(\mu)$ is one on average, $\beta(\lambda, \mu)$ may be too large in many cases and $|\hat{N}(\lambda, \mu)|^2$ changes excessive in successive frames, which can be solved by an appropriate choice of $\alpha(\lambda)$. According to Fig. 3 the main part of speech energy is distributed up to approx. 3.4 kHz. Allowing a change in 10 ms of about 5 % on average at this frequency range (as in the presented example Fig. 2), yields to a fixed $\alpha(\lambda)$ of:

$$\alpha(\lambda) = \frac{5 \cdot L_A\left(\left\lfloor\frac{3.4\,\mathrm{kHz}\cdot M_F}{f_s}\right\rfloor + 1\right)}{f_s \cdot \sum_{i=0}^{\left\lfloor\frac{3.4\,\mathrm{kHz}\cdot M_F}{f_s}\right\rfloor} \phi(i)} \approx 0.13 \mathrel{\hat{\approx}} 0.4\,\mathrm{dB}/10\,\mathrm{ms}. \tag{10}$$

4.3. Adaptive scaling $\alpha(\lambda)$ with the time

A further option is an adaptive $\alpha(\lambda)$ as a function of the frame *a posteriori* SNR. If the *a posteriori* SNR is extremely high, the adaptive $\alpha(\lambda)$ should be very small, resulting in small changes of $|\hat{N}(\lambda, \mu)|^2$ with the frames. Whereas with decreasing SNR, $\alpha(\lambda)$ should grow, allowing a faster tracking of the noise. In order to prevent error propagation, the adaptive $\alpha(\lambda)$ is chosen as a function of the segmental internal SNR with an upper limit of $\mathrm{SNR}_{\mathrm{max}}$ defined as

$$\mathrm{SNR}_{\mathrm{int}}(\lambda) = \min\left(\frac{1}{M_F}\sum_{\mu=0}^{M_F-1}\frac{|X(\lambda-1,\mu)|^2}{|\hat{N}(\lambda-1,\mu)|^2}, \mathrm{SNR}_{\mathrm{max}}\right), \tag{11}$$

controlled by a second independent *a posteriori* SNR estimate,

$$\mathrm{SNR}_{\mathrm{2nd}}(\lambda) = \frac{\sum_{\mu=0}^{M_F-1}|X(\lambda,\mu)|^2}{\sum_{\mu=0}^{M_F-1}|\hat{N}_{\mathrm{2nd}}(\lambda,\mu)|^2}, \tag{12}$$

Parameter	Settings
Sampling frequency f_s	16 kHz
Frame length L_{F}	320 ($\hat{=}$ 20 ms)
FFT length M_{F}	512 (including zero-padding)
Frame overlap	50% ($\sqrt{\mathrm{Hann}}$ – window)

Table 1. Simulation system settings

where $\hat{N}_{\mathrm{2nd}}(\lambda, \mu)$ is provided by a second *Baseline Tracer* with a large fixed α_{2nd}, resulting in a fast but rough noise tracking. Reasoning behind $\mathrm{SNR}_{\mathrm{2nd}}$ is to reduce the tracking speed in case of sudden increase of the speech component. Combining both SNR estimates, the adaptive $\alpha(\lambda)$ is now specified as,

$$\alpha(\lambda) = \frac{1 - \mathrm{SNR}_{\mathrm{int}}(\lambda)/\mathrm{SNR}_{\mathrm{max}}}{\mathrm{SNR}_{\mathrm{2nd}}(\lambda)}, \tag{13}$$

where the denominator provides fast and robust scaling of $\alpha(\lambda)$ which is refined by the nominator and $\mathrm{SNR}_{\mathrm{max}}$ defines the upper limit for noise tracking.

5. EVALUATION

A benchmark is carried out to compare the proposed noise PSD estimator *Baseline Tracing* in two different configurations for $\beta(\lambda, \mu)$ with three state-of-the-art methods: *Minimum Tracking* [2], *Minimum Statistics* [4] and the *SPP* noise tracker [12]. The first configuration employs a frequency dependent $\phi(\mu)$ according to the inverse long-term speech average spectrum (Sec. 4.1) and a fixed $\alpha(\lambda) = 0.4\,\mathrm{dB}/10\,\mathrm{ms}$, while in the second configuration $\alpha(\lambda)$ is *a posteriori* SNR dependent (Sec. 4.3) with an $\mathrm{SNR}_{\mathrm{max}} \hat{=} 15\,\mathrm{dB}$ and $\alpha_{\mathrm{2nd}} = 1.6\,\mathrm{dB}/10\,\mathrm{ms}$. The parameters of the *Minimum Tracking*, *Minimum Statistics* and *SPP* algorithm are chosen as suggested in [2, 4, 12], respectively. In the following, a standard speech enhancement system which is depicted in Fig. 5 serves as benchmark platform. The simulation parameters are summarized in Tab. 1.

The comparison is performed for all permutations of the following parameters: the input SNR varies from -15 to 25 dB in 5 dB steps and 15 male and female english speakers (randomly taken from the NTT database) are mixed with seven different stationary and non-stationary noise types (f16, factory1, babble, buccaneer1 [16], modulated Gaussian noise, vacuum cleaner, passing cars). The Gaussian noise is modulated with $f_{\mathrm{mod}} = 0.5\,\mathrm{Hz}$ according to $f(k) = 1 + 0.5\sin(2\pi k f_{\mathrm{mod}}/f_s)$. The evaluation is carried out by the logarithmic noise PSD estimation error. In addition, the performance is rated using a speech enhancement system by the objective scores segmental speech (SA) and noise attenuation (NA) as well as the cepstral distance (CD).

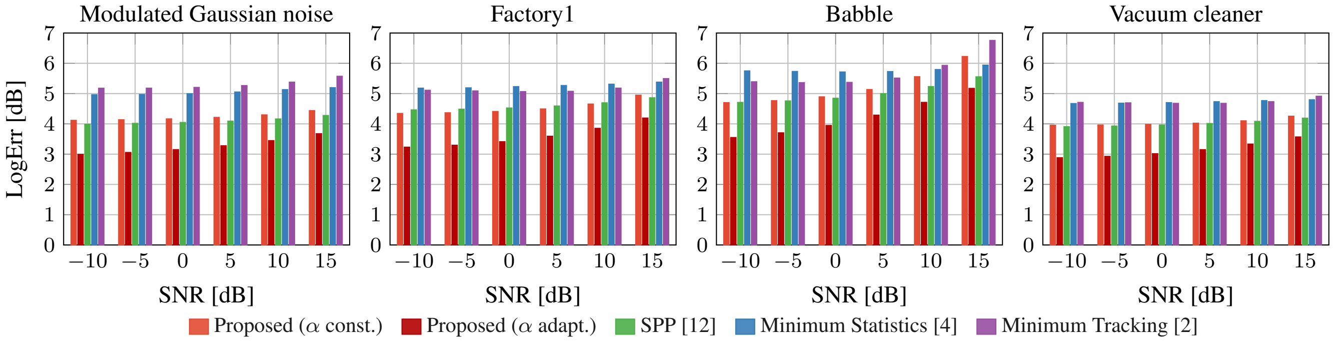

Fig. 4. Logarithmic error measure averaged over 30 speakers taken from the NTT database at various SNRs for selected noise types.

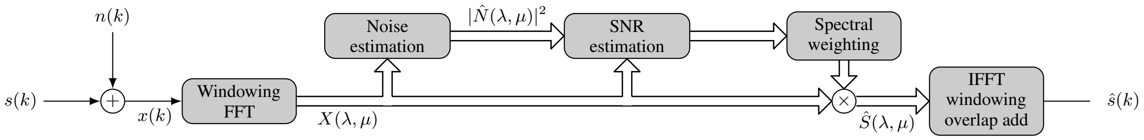

Fig. 5. Block diagram of standard noise reduction system

5.1. Noise PSD estimation performance

The logarithmic error measure between the estimated and the real noise PSD is defined as

$$\mathrm{LogErr} = \frac{1}{LM_F}\sum_{\lambda=1}^{L}\sum_{\mu=1}^{M_F}\left|10\log_{10}\left(\frac{|N(\lambda,\mu)|^2}{\left|\hat{N}(\lambda,\mu)\right|^2}\right)\right|, \qquad (14)$$

where, lower values indicate a better performance. In applications such as speech enhancement an overestimation of the true noise power likely results in an attenuation of the speech and thus in speech distortions. On the other hand, a noise power underestimation causes probable lower noise attenuation.

In Fig. 4 the averaged results are summarized for selected noise types at various SNRs. Comparing the proposed *Baseline Tracing* with fixed α (orange) to the best state-of-the-art algorithm, i.e., *SPP* (green), the performance is quite similar for all noise types and SNR conditions, except for babble noise at 10 and 15 dB, where *SPP* performs slightly better. The *Minimum Statistics* (blue) and *Minimum Tracking* (purple) have a comparable performance regarding the LogErr measure and perform 0.59 dB worse on average compared to *SPP* and the proposed estimator with fixed α. In contrast to *Minimum Statistics*, the LogErr analysis of *Minimum Tracking* confirmed a dominant underestimation of the noise PSD, indicating lower performance in terms of noise reduction. For all noises and SNR conditions, the proposed estimator *Baseline Tracing* with adaptive $\alpha(\lambda)$ (red) holds the best performance in all error measures with a projection up to 1.1 dB and 0.71 dB on average.

5.2. Noise reduction performance

The performance of the different noise estimators is also measured in terms of the cepstral distance (CD), segmental noise attenuation (NA) and speech attenuation (SA) [19] using them in a standard noise reduction system depicted in Fig. 5. Regarding the cepstral distance, lower values indicate a lower speech distortion. The difference between NA and SA corresponds to the noise reduction performance. In the following, it will be presented normalized to the NA-SA difference of a reference estimator using the real noise PSD, which is available in the simulation environment. Hence, lower values indicate better performance. The estimate of the *a priori* SNR and *a posteriori* SNR is provided by the decision-directed approach [1]. For the spectral gains, the Wiener filter is utilized which depends on the SNR estimate. The enhanced time domain signal $\hat{s}(k)$ is obtained by applying an Inverse Fast Fourier Transform (IFFT), windowing (square root Hann-window) and overlap-add.

Fig. 6 shows the results. As indicated in the previous section, the *Minimum Tracking* has the highest distance from the reference NA-SA measure over the SNR. Since the noise is underestimated significantly, the speech distortion should be low, which is confirmed by the CD measure up to 10 dB. While the *Minimum Statistics* and the proposed system with fixed and adaptive α perform in the NA-SA measure similar over the complete SNR, the *SPP* method has a higher distance of approx. 3.5 dB at -10 dB SNR reaching a similar performance starting with 10 dB SNR. Except the *Minimum Tracking* for high SNR, the *SPP* has a slightly higher CD over the SNR, where the proposed estimator with adaptive α and *Minimum Statistics* perform similar with the best scores on average. Up to 10 dB SNR, the *Baseline Tracing* with fixed α performs also likewise. This confirms the great LogErr performance also in the noise reduction task for both new *Baseline Tracing* estimators, as they provide a high noise attenuation at simultaneously low speech distortion.

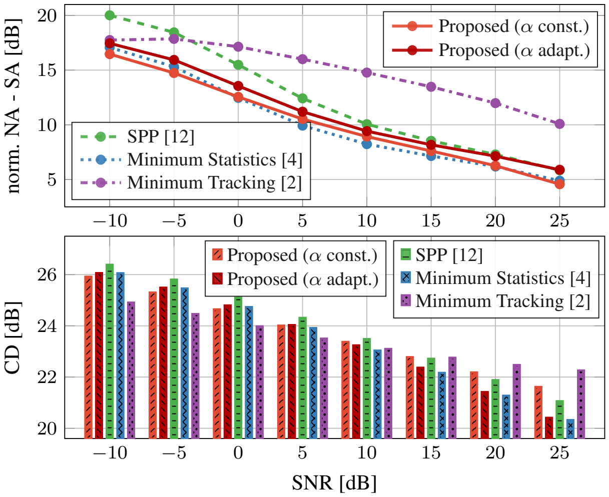

Fig. 6. The upper plot shows the normalized difference between noise attenuation (NA) and speech attenuation (SA) while the lower plot depicts the cepstral distance over the input SNR

6. CONCLUSIONS

A novel noise PSD estimator *Baseline Tracing* is presented which operates in the short-time Fourier domain. The basic idea consists of a constrained logarithmic magnitude tracing of the noisy observation separately for each frequency bin μ. The estimator can be explained in terms of delta modulation with an adaptive step size, operated in the slope overload mode. In the linear domain, the noise PSD of the current frame is calculated by a simple scaling of the last noise estimate with a certain frequency and time dependent β. Stretching or compressing is decided according to the sign of the difference between the last noise PSD estimate and the current noisy frame. Doing so, the estimator aims to follow the noisy observation. Since speech onset is assumed as sudden rises in the noisy observation, β has to be selected to only follow the noise. A fixed as well as an adaptive $\beta(\lambda,\mu)$ have been presented which consider the long-term speech spectrum and frame SNR. Compared to state-of-the-art systems, the new *Baseline Tracing* algorithm with adaptive $\beta(\lambda,\mu)$ has a superior performance with respect to the noise PSD error measure while performing similar to the *SPP* using a fixed $\beta(\mu)$. The noise reduction performance is characterized by a low cepstral distance, i.e., low speech distortion and strong NA-SA measures resulting in a high noise attenuation.

7. REFERENCES

[1] Yariv Ephraim and David Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 32, no. 6, pp. 1109–1121, 1984.

[2] Gerhard Doblinger, "Computationally efficient speech enhancement by spectral minima tracking in subbands," in *Proc. Eurospeech*, 1995, pp. 1513–1516.

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

[4] Rainer Martin, "Bias compensation methods for minimum statistics noise power spectral density estimation," *Signal Processing*, vol. 86, no. 6, pp. 1215–1229, June 2006.

[5] Thomas Esch and Peter Vary, "Model-based speech enhancement using SNR dependent MMSE estimation," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Piscataway, NJ, USA, May 2011, pp. 4652–4655, IEEE.

[6] Florian Heese, Thomas Esch, Bernd Geiser, and Peter Vary, "Noise reduction for wideband speech exploiting spectral dependencies based on conditional estimation," in *ITG-Fachtagung Sprachkommunikation*, Berlin, Oct. 2010, VDE Verlag GmbH.

[7] Florian Heese, Christoph Matthias Nelke, Markus Niermann, and Peter Vary, "Selflearning codebook speech enhancement," in *ITG Fachtagung Sprachkommunikation*. Sept. 2014, VDE Verlag GmbH.

[8] Pinaki Shankar Chanda and Sungjin Park, "Speech intelligibility enhancement using tunable equalization filter," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 2007, vol. 4, pp. IV–613, IEEE.

[9] Bastian Sauert and Peter Vary, "Recursive closed-form optimization of spectral audio power allocation for near end listening enhancement," in *ITG-Fachtagung Sprachkommunikation*, Berlin, Germany, Oct. 2010, VDE Verlag GmbH.

[10] Israel Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *Speech and Audio Processing, IEEE Transactions on*, vol. 11, no. 5, pp. 466–475, 2003.

[11] R. C. Hendriks, R. Heusdens, and J. Jensen, "MMSE based noise PSD tracking with low complexity," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2010, pp. 4266–4269.

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

[13] Christin Baasch, Vasudev Kandade Rajan, Mohamed Krini, and Gerhard Schmidt, "Low-complexity noise power spectral density estimation for harsh automobile environments," in *International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2014, pp. 219–223.

[14] N. S. Jayant and Peter Noll, "Digital coding of waveforms, principles and applications to speech and video," p. 688. Prentice-Hall, Englewood Cliffs NJ, USA, 1984.

[15] John G. Proakis and Masoud Salehi, *Communication Systems Engineering*, Prentice Hall, Upper Saddle River, N.J, 2 edition edition, Aug. 2001.

[16] A Varga, HJM Steeneken, and D Jones, "The noisex-92 study on the effect of additive noise on automatic speech recognition system," *Reports of NATO Research Study Group (RSG. 10)*, 1992.

[17] "Multi-lingual speech database for telephonometry," 1994, NTT-Corporation.

[18] ITU, "Artificial voices (ITU-t recommendation p.50)," Tech. Rep., International Telecommunication Union, Sept. 1999.

[19] Schuyler R. Quackenbush, Thomas Pinkney Barnwell, and Mark A. Clements, *Objective measures of speech quality*, Prentice Hall Englewood Cliffs, NJ, 1988.