BIAS CORRECTION METHODS FOR ADAPTIVE RECURSIVE SMOOTHING WITH APPLICATIONS IN NOISE PSD ESTIMATION

Robert Rehr, Timo Gerkmann

Speech Signal Processing Group, Department of Medical Physics and Acoustics
Cluster of Excellence Hearing4all, University of Oldenburg, Germany
robert.rehr@uni-oldenburg.de, timo.gerkmann@uni-oldenburg.de

ABSTRACT

Due to the low computational complexity and the low memory consumption, first-order recursive smoothing is a technique often applied to estimate the mean of a random process. For instance, recursive smoothing is used in noise power estimators where adaptively changing smoothing factors are used instead of fixed ones to prevent the speech power from leaking into the noise estimate. However, in general, the usage of adaptive smoothing factors leads to a biased estimate of the mean. In this paper, we propose a novel method to correct the bias evoked by adaptive smoothing factors. We compare this method to a recently proposed compensation method in terms of the log-error distortion using real world signals for two noise power estimators. We show that both corrections reduce the distortion measure in noisy speech while the novel method has the advantage that no iteration is required for determining the correction factor.

Index Terms— Error correction, adaptive estimation, smoothing methods, IIR filters, speech enhancement

1. INTRODUCTION

Mobile telephones and hearing aids are used to facilitate communication and are often employed in noisy conditions. The background noise usually reduces the perceived speech quality and speech intelligibility. Thus, to counteract the detrimental effects of noise, enhancement algorithms are employed. If the noisy speech signal is recorded using a single microphone, the noise is commonly suppressed by attenuating specific frequency bands when they are dominated by the disturbance, e.g., by using the Wiener filter. Here, the attenuation is controlled by the background noise power spectral density (PSD) and speech PSD which have to be estimated from the noisy input signal, e.g, using [1–4], [5, Section 14] or [6]. These PSDs can be understood as the mean of the speech and noise periodograms, respectively. To estimate these means, first-order recursive smoothing filters are often used due to their low computational complexity and low memory demands. These filters can be interpreted as the application of an exponentially decaying smoothing window. This allows first-order recursive smoothing filters also to track the mean of nonstationary signals.

For example, the noise PSD estimators proposed in [5, Section 14.1.3] and [6] reveal a first-order recursive filter structure where the fixed smoothing constant has been replaced by an adaptive smoothing factor [7]. The filter equation is then given by

$$\overline{x}_\ell = [1-\alpha(x_\ell,\overline{x}_{\ell-1})]x_\ell + \alpha(x_\ell,\overline{x}_{\ell-1})\overline{x}_{\ell-1}. \tag{1}$$

Here, x_ℓ is the observation of the random process at time ℓ and describes the periodogram of the noisy input signal. The estimated mean, i.e., the PSD, is denoted by $\overline{x}_\ell$. The choice of $0 \leq \alpha(x_\ell,\overline{x}_{\ell-1}) \leq 1$ is a trade-off in the speed with which a changing mean $\mathbb{E}\{x_\ell\}$ can be tracked on the one hand and the variance of the estimate on the other. The noise estimators in [5, Section 14.1.3] and [6] use the adaptive smoothing factors where the smoothing factor increases with an increasing *a posteriori* signal-to-noise ratio (SNR). In [5, Section 14.1.3], $\alpha(x_\ell,\overline{x}_{\ell-1})$ can take two different values which are selected based on a threshold while in [6], the smoothing factor is implicitly adapted using the speech presence probability (SPP).

In [7], it has been shown that the application of adaptive smoothing factors in general leads to a biased estimate of the mean. An iterative method has been presented which allows to quantify the bias evoked by adaptive smoothing factors. The estimated bias has been used to correct the deviation from the mean by applying a correction factor to the filter *output* $\overline{x}_\ell$. In this paper, we show that the correction factor can also be applied to the filter *input* if the smoothing factor depends only on the ratio $x_\ell/\overline{x}_{\ell-1}$. Further, we propose a second compensation method where no iteration is required to determine the correction factor. By means of experiments, we show that both correction methods can lead to a better estimation of the noise PSD in terms of the log-error distortion [8]. For this, we use the noise PSD estimators described in [5, 6] as examples.

After introducing the signal model and its relationship to (1), we describe the noise estimators proposed in [5, 6] in the context of adaptive smoothing. We continue with the two proposed correction methods and describe the procedures to estimate the correction factors. After that, we evaluate the precision of the methods proposed to estimate the correction factors using Monte-Carlo simulations. Further, we conduct experiments using real world signals and show that the correction methods reduce the log-error distortion.

2. SIGNAL MODEL

The considered adaptive smoothing functions are used in noise PSD estimators that operate in the short-time Fourier transform (STFT) domain. We now introduce the employed signal model and explain the relationship to the quantities used in (1).

We assume that the input signal is represented in the STFT domain which is obtained by splitting the signal into overlapping frames and transforming each frame using the discrete Fourier transform. Usually, a spectral analysis window, e.g., a Hann window, is applied before the Fourier transform is performed. Here, we assume that the speech spectral coefficients $S[k,\ell]$ and the noise spectral coefficients $D[k,\ell]$ are additive as

$$X[k,\ell] = S[k,\ell] + D[k,\ell]. \tag{2}$$

Correspondingly, $X[k,\ell]$ are the spectral coefficients of the noisy input signal while k and ℓ denote the frequency and time index, respectively. We omit the frequency index k for better readability if the considered expression does not depend on this quantity. Additionally, we assume that the periodogram of the noisy input signal $|X[\ell]|^2$ follows an exponential distribution as

$$f(|X[\ell]|^2) = \begin{cases} (1/\mu)\exp(-|X[\ell]|^2/\mu), & \text{if } |X[\ell]|^2 > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

This work was funded by the PhD program "Signals and Cognition", and the DFG Research Grant GE2538/2-1.

Here, the mean of the distribution is given by the scaling parameter μ as $\mathbb{E}\{|X[\ell]|^2\}=\mu$ where $\mathbb{E}\{\cdot\}$ is the expectation operator. This model holds if the speech coefficients $X[\ell]$ and noise coefficients $D[\ell]$ follow circular complex Gaussian distributions. The considered noise PSD estimators [5, 6] determine the noise PSD by adaptively smoothing the noise periodogram $|X[\ell]|^2$ across time ℓ. Consequently, the filter input x_ℓ in (1) can be understood as the periodogram $|X[\ell]|^2$ while the filter output $\overline{x}_\ell$ corresponds to the estimated noise PSD $\hat{\sigma}_d^2[\ell]$.

3. NOISE PSD ESTIMATORS USING ADAPTIVE SMOOTHING FACTORS

In this section, we present the noise PSD estimators proposed in [5, 6] in the context of adaptive smoothing.

3.1. Two different smoothing factors

In [5, Section 14.1.3], a simple approach has been proposed for estimating the noise PSD. Here, one out of two fixed smoothing constants is selected based on a threshold. A larger constant is used if the *a posteriori* SNR, i.e., the ratio $|X[\ell]|^2/\hat{\sigma}_d^2[\ell-1]$, is larger than one and a smaller one is used otherwise. The goal of this procedure is to avoid speech to leak into the noise PSD estimate $\hat{\sigma}_d^2[\ell]$ by excluding spectral coefficients which are likely to contain speech, i.e., the spectral coefficients having high energy relative to the estimated noise PSD. Mathematically, this method is given by

$$\alpha(x_\ell,\overline{x}_{\ell-1})=\begin{cases}\alpha^{\uparrow}, & \text{if } x_\ell/\overline{x}_{\ell-1}>1,\\ \alpha^{\downarrow}, & \text{otherwise}\end{cases} \tag{4}$$

where $\alpha^{\uparrow}$ and $\alpha^{\downarrow}$ denote the two fixed smoothing constants which are chosen between zero and one where $\alpha^{\uparrow}$ is larger than $\alpha^{\downarrow}$.

3.2. Speech presence probability based noise PSD estimation

In the noise PSD estimator proposed in [6], an estimate of the SPP is used to avoid speech leakage. Even though the noise estimator is not explicitly described by adaptive smoothing, it could be shown in [7] that this algorithm can be described as an adaptive smoothing factor $\alpha(x_\ell,\overline{x}_{\ell-1})$. In [6], two hypothesis are introduced: one for speech presence H_1, where $X[\ell]=S[\ell]+D[\ell]$, and one for speech absence H_0, where $X[\ell]=D[\ell]$. Under the assumption that $S[\ell]$ and $D[\ell]$ follow complex Gaussian distributions and that the prior probabilities are identical, i.e., $P(H_0)=P(H_1)$, the SPP is obtained as

$$P(H_1|X[\ell])=\left(1+(1+\xi)\exp\left(-\frac{|X[\ell]|^2}{\hat{\sigma}_d^2[\ell-1]}\frac{\xi}{\xi+1}\right)\right)^{-1}. \tag{5}$$

The quantity ξ denotes the SNR that is expected if a spectral coefficient contains speech. In [6], ξ is not adaptively estimated but set to a fixed value which is optimized such that the Bayesian risk, i.e., the misclassification between speech presence and speech absence, is minimized [9]. In [7], the corresponding adaptive smoothing factor has been derived as

$$\alpha(x_\ell,\overline{x}_{\ell-1})=\beta+\frac{1-\beta}{1+(1+\xi)\exp(-x_\ell\xi/\{\overline{x}_{\ell-1}(1+\xi)\})} \tag{6}$$

where β is a fixed recursive smoothing constant. Here, the term weighted by $1-\beta$ is the SPP given in (5) with $x_\ell/\overline{x}_{\ell-1}=|X[\ell]|^2/\hat{\sigma}_d^2[\ell-1]$. The behavior of this adaptive smoothing function is similar to the factor proposed in [5, Section 14.1.3] as the expression in (6) is close to one for high *a posteriori* SNRs and approaches the fixed smoothing constant β for low *a posteriori* SNRs.

4. CORRECTION METHODS

If the adaptive smoothing factor depends only on the ratio $x_\ell/\overline{x}_{\ell-1}$, the adaptive smoothing factor is scale-invariant, i.e., if the filter input x_ℓ is scaled by some $r\in\mathbb{R}$, the filter output $\overline{x}_\ell$ is scaled by the same factor r.

Algorithm 1 Bias compensation by scaling the input signal by c_1 [7].

1: Initialize algorithm and compensate bias:
$\breve{\overline{x}}_0\leftarrow c_1x_0$, with c_1 obtained from Algorithm 3.
2: **for all** all remaining frames ℓ **do**
3: Perform smoothing and correct bias:
$\breve{\overline{x}}_\ell=(1-\alpha(c_1x_\ell,\breve{\overline{x}}_{\ell-1}))c_1x_\ell+\alpha(c_1x_\ell,\breve{\overline{x}}_{\ell-1})\breve{\overline{x}}_{\ell-1}$
4: **end for**

Algorithm 2 Proposed bias compensation by scaling only the x_ℓ not occurring in $\alpha(x_\ell,\overline{x}_{\ell-1})$ by c_2.

1: Initialize algorithm and compensate bias:
$\breve{\overline{x}}_0\leftarrow c_2x_0$, with c_2 obtained from Algorithm 4.
2: **for all** all remaining frames ℓ **do**
3: Perform smoothing and correct bias:
$\breve{\overline{x}}_\ell=(1-\alpha(x_\ell,\breve{\overline{x}}_{\ell-1}))c_2x_\ell+\alpha(x_\ell,\breve{\overline{x}}_{\ell-1})\breve{\overline{x}}_{\ell-1}$
4: **end for**

This can be proven using the method of induction. For recursive filters it is often assumed that the system is initially at rest, i.e., $\overline{x}_\ell=0$ for $\ell<0$. Instead, as this would lead to a division by zero, we assume that the first filter output is given by the first input as $\overline{x}_0=x_0$. Consequently, multiplying the filter input by r leads to $r\overline{x}_0=rx_0$. Then, for the following input samples of the smoothing filter in (1), it can be shown that

$$\left[1-\alpha\left(\frac{rx_\ell}{r\overline{x}_{\ell-1}}\right)\right]rx_\ell+\alpha\left(\frac{rx_\ell}{r\overline{x}_{\ell-1}}\right)r\overline{x}_{\ell-1} \tag{7}$$

$$=r\left(\left[1-\alpha\left(\frac{x_\ell}{\overline{x}_{\ell-1}}\right)\right]x_\ell+\alpha\left(\frac{x_\ell}{\overline{x}_{\ell-1}}\right)\overline{x}_{\ell-1}\right) \tag{8}$$

$$=r\overline{x}_\ell. \tag{9}$$

It thus follows that, the adaptive smoothing factors of the considered noise PSD estimators [5, 6] are indeed scale-invariant.

Because of the scale-invariance the bias can be corrected by applying a correction factor c_1 either to the filter input or to the filter output. For obtaining an unbiased estimate of the mean, this correction factor has to be set to $c_1=\mathbb{E}\{x_\ell\}/\mathbb{E}\{\overline{x}_\ell\}$. As this ratio does not depend on the scaling of x_ℓ or $\overline{x}_\ell$, it is sufficient to know the bias for any given mean of the signal, e.g., $\mathbb{E}\{x_\ell\}=1$. Algorithm 1 summarizes the resulting correction method. Here, $\breve{\overline{x}}_\ell$ denotes the corrected filter output. The same correction can also be achieved by applying c_1 to the filter output. For this, the input signal is smoothed with (1) and corrected via $\breve{\overline{x}}_\ell=c_1\overline{x}_\ell$. For obtaining the correct compensation, the corrected filter output must not be recursively fed back.

In this paper, we show that instead of multiplying the filter input or output by a correction factor, the bias can also be corrected by applying the correction factor only to the x_ℓ not occurring in the adaptive smoothing function $\alpha(x_\ell,\overline{x}_{\ell-1})$ as shown in Algorithm 2. Here, the correction factor c_2 is used, whose value is in general different to c_1 used in Algorithm 1. Also for this case, the scale-invariance can be exploited meaning that it is sufficient to obtain a correction factor c_2 for a known mean value of the input, e.g., $\mathbb{E}\{x_\ell\}=1$. In contrast to Algorithm 1, this method has the advantage that also the values of the adaptive smoothing factor $\alpha(x_\ell,\overline{x}_{\ell-1})$ are corrected. For smoothing constants that depend on the ratio $x_\ell/\overline{x}_{\ell-1}$, this cannot be achieved with Algorithm 1 as the input x_ℓ is scaled with the same factor as the corrected output $\breve{\overline{x}}_{\ell-1}$. As the correction factor c_1 cancels out, the smoothing factors remain the same independent of c_1. In contrast, in Algorithm 2, only the corrected output $\breve{\overline{x}}_{\ell-1}$ is used to determine the adaptive smoothing factor while the input x_ℓ remains unscaled. Consequently, the smoothing factor is determined based on a corrected version of the *a posteriori* SNR. Considering the noise PSD estimator in [6], for example, this also leads to a correction of the estimated SPP. Another advantage is that for determining c_2, the iteration described in [7] is not required. Depending on the used adaptive smoothing factor,

Algorithm 3 Iterative estimation of the correction factor c_1 for adaptive functions depending on $\overline{x}_{\ell-1}$.

$i \leftarrow 0$, $\rho_0 \leftarrow 1$, $\mu \leftarrow 1$.
while convergence criterion for ρ_i is not met **do**
  Obtain ρ_{i+1} using (12). The solutions for the adaptive functions in [5,6] are given in (15) and (16).
  $i \leftarrow i+1$
end while
Compute compensation factor: $c_1 = \mu/\rho_i$.

this correction can lead to a lower overestimation of the mean compared to Algorithm 1 proposed in [7] as we will show in our experiments.

5. DETERMINATION OF THE CORRECTION FACTOR

In this section, we propose two methods to determine the correction factors c_1 and c_2, respectively. In [7], an analytical solution has been described to determine $\mathbb{E}\{\overline{x}_\ell\}$ for adaptive smoothing factors that depend on the filter input x_ℓ, but not on the previous filter output $\overline{x}_{\ell-1}$. Based on this solution, an iterative method has been proposed that can determine the bias for smoothing factors depending on both x_ℓ and $\overline{x}_{\ell-1}$ in good approximation. Here, we extend the methods proposed in [7] to estimate the correction factor c_2 used in Algorithm 2.

5.1. Determination of the correction factor c_1

First, we address the correction factor c_1. If adaptive smoothing factors $\alpha(x_\ell,\overline{x}_{\ell-1})$ are considered that do not depend on the previous filter output $\overline{x}_{\ell-1}$, equation (1) simplifies to

$$\overline{x}_\ell = [1-\alpha(x_\ell)]x_\ell + \alpha(x_\ell)\overline{x}_{\ell-1}. \tag{10}$$

For the derivations, we assume that all x_ℓ are identically distributed and uncorrelated. Further, we assume that the filter output $\overline{x}_\ell$ will remain stationary if the filter input x_ℓ is stationary. Experiments indicate that this property is sufficiently fulfilled. From this it follows that $\mathbb{E}\{\overline{x}_n\} = \mathbb{E}\{\overline{x}_m\}$. With the first assumption, the expected value $\mathbb{E}\{x_\ell \overline{x}_{\ell-1}\}$ can be written as $\mathbb{E}\{x_\ell\}\mathbb{E}\{\overline{x}_{\ell-1}\}$. Thus, by applying $\mathbb{E}\{\cdot\}$ to (10) and rearranging the terms

$$\mathbb{E}\{\overline{x}_\ell\} = \frac{\mathbb{E}\{x_\ell\} - \mathbb{E}\{x_\ell \alpha(x_\ell)\}}{1-\mathbb{E}\{\alpha(x_\ell)\}} \tag{11}$$

is obtained [7]. The result given in (11) depends only on the adaptive function $\alpha(x_\ell)$ and the probability density function of x_ℓ. Based on the solution in (11), an iteration has been derived in [7]. For this, the adaptive smoothing factors are simplified by replacing $\overline{x}_{\ell-1}$ by a fixed value ρ. Then, the equation for the iteration is given by

$$\rho_i = \frac{\mathbb{E}\{x_\ell\} - \mathbb{E}\{x_\ell \alpha(x_\ell,\rho_{i-1})\}}{1-\mathbb{E}\{\alpha(x_\ell,\rho_{i-1})\}}. \tag{12}$$

Here, ρ_i is the estimate of $\mathbb{E}\{\overline{x}_\ell\}$ obtained for the ith iteration step whereas the initial condition is denoted by ρ_0. For determining the final estimate of $\mathbb{E}\{\overline{x}_\ell\}$, the iteration is continued until it converges. In [7], it has been shown that the parameter ρ_0 does not influence the convergence of the iterative approach for the considered noise PSD estimators [5,6]. With the converged ρ_i, the correction factor is determined by $c_1 = \mu/\rho_i$. This procedure is summarized in Algorithm 3.

5.2. Determination of the proposed correction factor c_2

For determining the correction factor c_2, the expected value of line three in Algorithm 2 is considered. First, as in Section 5.1, we deal with the case where the adaptive smoothing factor does not depend on the previous estimate $\overline{x}_{\ell-1}$. This means that $\mathbb{E}\{\overline{x}_\ell\}$ is now determined for $\overline{x}_\ell = (1-\alpha(x_\ell))c_2 x_\ell + \alpha(x_\ell)\overline{x}_{\ell-1}$. The result is similar to (11) and differs only by the multiplication by c_2 as

$$\mathbb{E}\{\overline{x}_\ell\} = c_2 \frac{\mathbb{E}\{x_\ell\} - \mathbb{E}\{x_\ell \alpha(x_\ell)\}}{1-\mathbb{E}\{\alpha(x_\ell)\}}. \tag{13}$$

Algorithm 4 Estimation of the correction parameter c_2 for adaptive functions depending on $\overline{x}_{\ell-1}$.

$\rho_0 \leftarrow 1$, $\mu \leftarrow 1$.
Obtain ρ_1 using (12). The solutions for the adaptive functions in [5,6] are given in (15) and (16).
Compute compensation factor: $c_2 = \mu/\rho_1$.

Smoothing factor	Monte-Carlo	Sec. 5
(4), [5, Section 14.1.3]	10.42	10.33
(6), [6]	1.31	1.23

Table 1: Correction factor c_1 as proposed in Algorithm 1 for the adaptive smoothing functions in (4) and (6).

Here, the same assumptions have been made as for (11). Interestingly, the compensation method in Algorithm 2 leads to a scaling of the filter output by c_2 for adaptive smoothing functions that depend only on x_ℓ but not $\overline{x}_{\ell-1}$. Consequently, setting $c_2 = \mathbb{E}\{x_\ell\}/\mathbb{E}\{\overline{x}_\ell\}$ leads to a compensation of the bias for such smoothing factors. Analogously, an iteration can be derived for the smoothing in Algorithm 2 for factors $\alpha(x_\ell,\overline{x}_{\ell-1})$ depending on x_ℓ and $\overline{x}_{\ell-1}$. Also here, $\overline{x}_{\ell-1}$ is replaced by a fixed constant and we obtain

$$\tilde{\rho}_i = c_2 \frac{\mathbb{E}\{x_\ell\} - \mathbb{E}\{x_\ell \alpha(x_\ell,\tilde{\rho}_{i-1})\}}{1-\mathbb{E}\{\alpha(x_\ell,\tilde{\rho}_{i-1})\}}, \tag{14}$$

which is (12) multiplied by c_2. To distinguish the iteration steps in (12) from the ones used in (14), $\tilde{\rho}_i$ is used here. To obtain an unbiased estimate, the constant c_2 has to be set such that the iteration converges to $\mathbb{E}\{x_\ell\}$. To achieve this, ρ_1 is determined with (12) where ρ_0 is set to the mean of the input signal $\mathbb{E}\{x_\ell\}$. With that, the correction factor c_2 is determined as $c_2 = \mathbb{E}\{x_\ell\}/\rho_1$. Using the resulting c_2 and $\tilde{\rho}_0 = \mathbb{E}\{x_\ell\}$ in (14), it can be seen that this correction factor enforces the same result, namely $\mathbb{E}\{x_\ell\}$, for each iteration step $\tilde{\rho}_i$. Similar to [7], it can be shown experimentally that for other initializations $\tilde{\rho}_0$ the iteration in (14) converges to the same value, i.e., $\mathbb{E}\{x_\ell\}$. This indicates that the determined value for c_2, which can be obtained without an iteration, compensates the bias in Algorithm 2. Due to the scale-invariance, this procedure leads to the same c_2 for any given mean $\mathbb{E}\{x_\ell\}$, e.g., $\mathbb{E}\{x_\ell\} = \rho_0 = 1$. A summary of this procedure is shown in Algorithm 4.

5.3. Analytic solutions for specific adaptive smoothing factors

Finally, we show the solutions to (11) for the adaptive smoothing factors given in (4) and (6). Here, it is assumed that x_ℓ follows an exponential distribution as in (3) and that $\overline{x}_{\ell-1}$ is replaced by ρ. For the noise PSD estimator proposed in [5, Section 14.1.3], the expected value $\mathbb{E}\{\overline{x}_\ell\}$, i.e., the solution to (11) given (4), results in

$$\mathbb{E}\{\overline{x}_\ell\} = \mu \frac{(\alpha^{\downarrow}-1)\exp(\lambda) + (\alpha^{\uparrow}-\alpha^{\downarrow})(1+\lambda)}{(\alpha^{\downarrow}-1)\exp(\lambda) + \alpha^{\uparrow} - \alpha^{\downarrow}}, \tag{15}$$

with $\lambda = \rho/\mu$. The expected value $\mathbb{E}\{\overline{x}_\ell\}$ for the expression in (6) can be derived using the property of the geometric series [10, 1.112.1] and the analytic continuation property of the hypergeometric series [10, 9.130]. The result is

$$\mathbb{E}\{\overline{x}_\ell\} = \mu \frac{1 - {}_3F_2[1,\zeta,\zeta;\zeta+1,\zeta+1;-(1+\xi)]}{1 - {}_2F_1[1,\zeta;\zeta+1;-(1+\xi)]}, \tag{16}$$

where ${}_pF_q$ is the generalized hypergeometric function and ζ is

$$\zeta = \lambda \frac{\xi+1}{\xi}. \tag{17}$$

6. EVALUATION

First, we analyze the precision of the estimation methods described in Section 5, which can be used to determine the correction factors c_1 and

Smoothing factor	Monte-Carlo	Sec. 5
(4), [5, Section 14.1.3]	2.44	2.37
(6), [6]	1.20	1.16

Table 2: Correction factor c_2 as proposed in Algorithm 2 for the adaptive smoothing functions in (4) and (6).

c_2. We employ Monte-Carlo simulations to obtain a ground truth. We generate 10^6 independent realizations for x_ℓ that follow an exponential distribution as given in (3) where the mean μ is kept fixed. These samples are used to compute the output $\overline{x}_\ell$ of the adaptive first-order recursive smoothing filter which is used to estimate the expected value $\mathbb{E}\{\overline{x}_\ell\}$ by temporal averaging. The estimated mean is denoted by $\hat{\mathbb{E}}\{\overline{x}_\ell\}$ and is used to determine a Monte-Carlo estimate c_1^{MC} of the correction factor c_1 by computing $c_1^{\mathrm{MC}} = \mu/\hat{\mathbb{E}}\{\overline{x}_\ell\}$. The Monte-Carlo reference c_2^{MC} for the factor c_2 used in Algorithm 2 is determined as the factor that minimizes the difference between the mean of the input signal $\mathbb{E}\{x_\ell\}$ and the mean of the corrected filter output $\hat{\mathbb{E}}\{\breve{\overline{x}}_\ell\}$, i.e.,

$$c_2^{\mathrm{MC}} = \arg\min_{c_2} |\mathbb{E}\{x_\ell\} - \hat{\mathbb{E}}\{\breve{\overline{x}}_\ell\}| \tag{18}$$

where $\hat{\mathbb{E}}\{\breve{\overline{x}}_\ell\}$ is determined as the temporal average of the filter output $\breve{\overline{x}}_\ell$ which is computed as in Algorithm 2.

In the evaluations, we use the default values of the noise PSD estimators as described in the literature. Correspondingly, $\alpha^{\uparrow}$ and $\alpha^{\downarrow}$ are set to 0.9995 and 0.9, respectively [5, Section 14.1.3]. For the SPP based noise estimator $\xi = 15$ dB and $\beta = 0.8$ are used [6].

For the considered noise estimators described in [5,6], 10 to 15 iteration step are required in Algorithm 3 to determine correction factor c_1. In accordance to the findings in [7], the results in Table 1 show that the correction factor c_1 is slightly underestimated by the iterative method given in Algorithm 3. The ratio between the reference and the estimated correction factor is -0.04 dB and -0.27 dB for the smoothing factors in (4) and (6), respectively. This means that the mean will still be underestimated by this amount if the correction factors obtained with Algorithm 3 are used.

Table 2 shows the correction factors for c_2 obtained with Algorithm 4. Compared to the correction factor c_1 used in Algorithm 1, the respective values of c_2 are smaller. The difference is due to the different influence of Algorithm 2 on the recursion. Further, also the factors c_2 determined with Algorithm 4 are smaller than the compensation factors obtained from the Monte-Carlo simulations. With these correction factors an underestimation of 0.6 dB and 0.2 dB remains for the smoothing factors proposed in [5, Section 14.1.3] and [6], respectively. In general, a good approximation of the true correction factors can be achieved using the methods described in Section 5.

The following experiment is conducted using real world signals. We use 192 sentences from the TIMIT corpus [11] which are artificially corrupted by background noises at SNRs of -10 dB, 10 dB and 10 dB. A variety of synthetic and natural noise types is employed namely the pink and the babble noise taken from the Noisex-92 corpus [12] and a passing car noise. The considered signals have a sample rate of 16 kHz. For the STFT, we use 32 ms frames with an overlap of 50 % and a square-root Hann window for spectral analysis. The correction methods are compared in terms of the log-error distortion proposed in [8]. Similar to [6], we consider the contribution of the underestimation $\mathrm{LogErr}_{\downarrow}$ and the overestimation $\mathrm{LogErr}_{\uparrow}$ separately as

$$\mathrm{LogErr}_{\uparrow} = \frac{1}{KL}\sum_{\ell=0}^{L-1}\sum_{k=0}^{K-1}\max\left(0, 10\log_{10}\frac{\hat{\sigma}_d^2[k,\ell]}{\sigma_d^2[k,\ell]}\right), \tag{19}$$

$$\mathrm{LogErr}_{\downarrow} = -\frac{1}{KL}\sum_{\ell=0}^{L-1}\sum_{k=0}^{K-1}\min\left(0, 10\log_{10}\frac{\hat{\sigma}_d^2[k,\ell]}{\sigma_d^2[k,\ell]}\right), \tag{20}$$

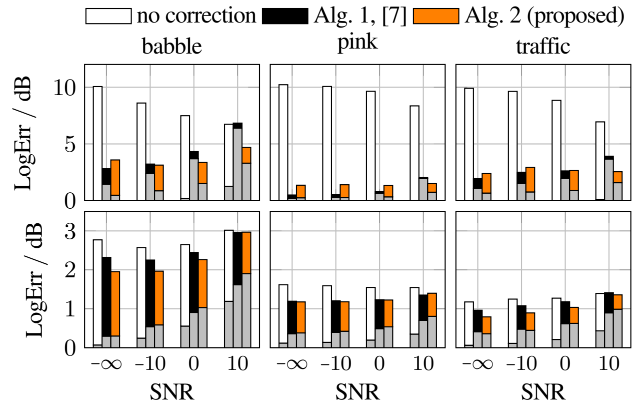

Fig. 1: Comparison of the proposed correction methods in terms of the log-error distortion with respect to the noise PSD estimators proposed in (4), [5] (top) and (6), [6] (bottom), for noisy speech. The lower part of the bars (gray) is the overestimation whereas the upper part (colored) is the underestimation.

where the total log-error distortion is given by $\mathrm{LogErr} = \mathrm{LogErr}_{\downarrow} + \mathrm{LogErr}_{\uparrow}$. For the evaluation, we ensure that the first 5 s are excluded to allow the noise estimators to adapt to the background noise. Further, we exclude the 0 Hz bin and the Nyquist frequency as the periodogram cannot be considered exponentially distributed here. As the pink noise is stationary, the reference noise PSD $\sigma_d^2[\ell]$ is obtained by temporal averaging of the noise periodogram. For the nonstationary noises, we employ a slightly smoothed version of the noise periodogram, where a fixed smoothing constant with $\alpha = 0.73$ is used. The results are shown in Figure 1 for the smoothing factors in (4), [5, Section 14.1.3], and (6), [6], respectively.

Considering the noise only case, i.e., SNR $= -\infty$, both proposed correction methods lead to lower log-error distortions at the cost of a slightly increased noise overestimation in comparison to the case where no correction is applied. As a consequence, the total log-error distortion is considerably smaller compared to the case where no correction is employed. For [5, Section 14.1.3], Algorithm 1 gives a slightly lower distortion while Algorithm 2 appears to be the better solution for the SPP based noise estimator. In general, the speech leakage increases with increasing SNR. Thus, in high SNRs, the correction factors that are required to obtain an unbiased estimate amplify this overestimation. This is especially true for the noise estimator proposed in [5, Section 14.1.3] which always allows to track the noise albeit slowly. This issue is smaller if the proposed Algorithm 2 is used for correcting the bias induced by the noise PSD estimator in [5, Section 14.1.3]. With this method, the log-error distortion can be reduced for all noise types and SNRs. For the SPP based noise PSD estimator, the scaling used in Algorithm 1 leads to a lower overestimation in high SNRs. For the SPP based noise estimator, however, the total log-error distortion is in general rather small so that the speech distortion caused by the overestimation in high SNRs is most likely negligible while in low SNRs estimation errors are reduced.

7. CONCLUSIONS

In this paper, we proposed a novel method to compensate the bias caused by adaptive smoothing. In contrast to the method in [7], the correction factor can be determined without any iteration. We analyzed both compensation methods and the procedures to determine the respective correction terms. We showed that the precision of the correction factors can be estimated with sufficient accuracy with the proposed methods. Further, experiments on real world signals were conducted. While the multiplication with a compensation factor larger than one necessarily results in an increase of overestimation errors, we showed that the overall estimation error is decreased for the critical SNR range below 10 dB.

8. REFERENCES

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

[2] C. Breithaupt, T. Gerkmann, and R. Martin, "A novel a priori SNR estimation approach based on selective cepstro-temporal smoothing," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas, Nevada, USA, Apr. 2008, pp. 4897–4900.

[3] 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, Jul. 2001.

[4] F. Heese and P. Vary, "Noise PSD estimation by logarithmic baseline tracing," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2015, pp. 4405–4409.

[5] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control – A Practical Approach*, ser. Adaptive and Learning Systems for Signal Processing, Communication and Control, S. Haykin, Ed. Wiley & Sons, 2004.

[6] 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)*, New Paltz, New York, USA, 2011, pp. 145–148.

[7] R. Rehr and T. Gerkmann, "On The Bias Of Adaptive First-Order Recursive Smoothing," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, New York, USA, Oct. 2015.

[8] R. C. Hendriks, J. Jensen, and R. Heusdens, "DFT Domain Subspace Based Noise Tracking for Speech Enhancement," in *Interspeech*, Aug. 2007, pp. 830–833.

[9] T. Gerkmann, C. Breithaupt, and R. Martin, "Improved A Posteriori Speech Presence Probability Estimation Based on a Likelihood Ratio With Fixed Priors," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 5, pp. 910–919, 2008.

[10] I. S. Gradshteyn and I. W. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed., D. Zwillinger and V. Moll, Eds. Academic Press, Feb. 2007.

[11] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren, and V. Zue, "TIMIT Acoustic-Phonetic Continuous Speech Corpus," Linguistic Data Consortium, Philadelphia, Tech. Rep., 1993.

[12] H. J. M. Steeneken and F. W. M. Geurtsen, "Description of the RSG.10 noise database," TNO Institute for perception, Technical Report IZF 1988-3, 1988.