A Priori SNR Estimation Using Weibull Mixture Model

Aleksej Chinaev, Jens Heitkaemper, Reinhold Haeb-Umbach

Department of Communications Engineering, Paderborn University, 33100 Paderborn, Germany
Email: {chinaev,haeb}@nt.upb.de
Web: nt.upb.de

Abstract

This contribution introduces a novel causal *a priori* signal-to-noise ratio (SNR) estimator for single-channel speech enhancement. To exploit the advantages of the generalized spectral subtraction, a normalized α-order magnitude (NAOM) domain is introduced where an *a priori* SNR estimation is carried out. In this domain, the NAOM coefficients of noise and clean speech signals are modeled by a Weibull distribution and a Weibull mixture model (WMM), respectively. While the parameters of the noise model are calculated from the noise power spectral density estimates, the speech WMM parameters are estimated from the noisy signal by applying a causal Expectation-Maximization algorithm. Further a maximum *a posteriori* estimate of the *a priori* SNR is developed. The experiments in different noisy environments show the superiority of the proposed estimator compared to the well-known decision-directed approach in terms of estimation error, estimator variance and speech quality of the enhanced signals when used for speech enhancement.

1 Introduction

A single-channel speech spectral enhancement system based on statistical methods is usually composed of three modules: a noise tracker, an *a priori* SNR estimator and a gain function operator [1]. According to [2] the *a priori* SNR estimator provides a dominant parameter of a spectral gain function applied to short-time Fourier transform (STFT) coefficients of the noisy signal for noise suppression. A famous approach for *a priori* SNR estimation is the decision-directed (DD) approach [3]. It combines the information of the noise tracker and the gain function, and thus successfully reduces the musical noise in the enhanced signal. However, the DD approach suffers from slow response to an abrupt change of the instantaneous SNR taking place on speech onsets [4], because it uses the gain function of the previous frame.

Many different approaches for *a priori* SNR estimation were developed in recent years [5–13]. To reduce the estimation error during speech activity, a model-based approach was recently presented in [14]. There, the magnitudes of the clean speech STFT coefficients are modelled by a Gaussian mixture model (GMM) proposed in [15]. The GMM-based estimation helps to decouple the *a priori* SNR estimation and the gain function operator. A drawback of this approach, however, is the necessity of a prior GMM training phase.

Inspired by the advantages of the generalized spectral subtraction (GSS) regarding a reduction of musical noise investigated in [16], we propose to carry out the *a priori* SNR estimation in the NAOM domain by using a Weibull distribution [17]. In contrast to [17] we suggest to model the NAOM coefficients of the noise signal by a single Weibull distribution and of the clean speech signal by a Weibull mixture model. To reduce the number of mixture components we allow the WMM components to have a shape parameter β different from the spectral magnitude order α. The WMM parameters are determined from the noisy observations by an Expectation Maximization (EM) approach. Given the model parameters we derive a maximum *a posteriori* (MAP) estimator of the NAOM coefficients of the clean speech, which are used for the calculation of the *a priori* SNR. Further, a causal implementation of the proposed estimator is derived.

The remainder of this contribution is structured as follows: in Section 2 we develop a MAP-based estimator of the *a priori* SNR based on WMM. In Section 3 we optimize the parameters of the proposed estimator and describe the results of the experimental evaluation of the proposed approach compared to the DD approach, before conclusions are drawn in Section 4.

2 MAP estimation of the a priori SNR based on Weibull mixture models

We observe the STFT coefficients of a clean speech signal corrupted by uncorrelated additive noise denoted by $Y(k,\ell) = S(k,\ell) + N(k,\ell)$, where $S(k,\ell)$ and $N(k,\ell)$ represent the STFT coefficients of the clean speech and of the noise signal, respectively, with a frequency bin index $k \in [1;K]$ and a frame index $\ell \in [1;L]$. Motivated by a central limit theorem STFT coefficients $S(k,\ell)$ and $N(k,\ell)$ are modelled as non-stationary complex valued zero-mean Gaussian random processes with PSDs $\lambda_S(k,\ell) = \mathrm{E}\left[|S(k,\ell)|^2\right]$ and $\lambda_N(k,\ell) = \mathrm{E}\left[|N(k,\ell)|^2\right]$, where $E[\cdot]$ denotes the expectation operator [3, 18]. Then the *a priori* SNR is defined as:

$$\xi(k,\ell) = \frac{\lambda_S(k,\ell)}{\lambda_N(k,\ell)}. \tag{1}$$

Given an estimate of $\lambda_N(k,\ell)$, the *a priori* SNR estimator aims to estimate $\xi(k,\ell)$ from the periodogram $|Y(k,\ell)|^2$.

2.1 Normalized α-order magnitude domain

In the GSS performance improvements have been observed when carrying out the spectral subtraction in the so-called α-domain, where $\alpha = 2$ corresponds to the PSD domain [16]. To utilize the benefits of processing in such an alternative domain for the *a priori* SNR estimation, we consider the α-order magnitudes (AOM) of STFT coefficients and assume its additivity as in [19]

$$|Y(k,\ell)|^\alpha = |S(k,\ell)|^\alpha + |N(k,\ell)|^\alpha, \tag{2}$$

where $\alpha > 0$ is the spectral magnitude order. Furthermore, to obtain statistical models, which are independent of the signal energy, the AOM coefficients in (2) are normalized to the root of the frequency - dependent averaged power of

the clean speech AOM coefficients defined as

$$P_S(k) = \frac{1}{L}\sum_{\ell=1}^{L} |S(k,\ell)|^{2\alpha} \tag{3}$$

resulting in the normalized AOM (NAOM) coefficients

$$S_\alpha(k,\ell) = \frac{|S(k,\ell)|^\alpha}{\sqrt{P_S(k)}}, \tag{4}$$

$$N_\alpha(k,\ell) = \frac{|N(k,\ell)|^\alpha}{\sqrt{P_S(k)}} \tag{5}$$

with $E[S_\alpha^2(k,\ell)] = 1$. The noisy NAOM coefficients are

$$Y_\alpha(k,\ell) = \frac{|Y(k,\ell)|^\alpha}{\sqrt{P_S(k)}} = S_\alpha(k,\ell) + N_\alpha(k,\ell). \tag{6}$$

Note, that we will drop the indices k and ℓ in the following wherever possible without sacrificing clarity.

Under the stated statistical assumptions, the noise NAOM coefficients $N_\alpha(k,\ell)$ are Weibull distributed and follow the probability density function $p_{N_\alpha}(n) = \text{Weib}\,(n;\lambda_{N_\alpha},\alpha)$ with the Weibull distribution defined as

$$\text{Weib}\,(x;\lambda,\alpha) = \frac{2}{\alpha\cdot\lambda}\cdot x^{\frac{2}{\alpha}-1}\cdot\exp\left(-\frac{x^{\frac{2}{\alpha}}}{\lambda}\right)\cdot\varepsilon(x) \tag{7}$$

where the spectral magnitude order α is a shape parameter, $\lambda > 0$ a scale parameter, and $\varepsilon(x)$ the unit step function. For the κ-th order raw moment of a Weibull distributed random variable X one has

$$E[X^\kappa] = \lambda^{\kappa\cdot\frac{\alpha}{2}}\cdot\Gamma\left(\kappa\cdot\frac{\alpha}{2}+1\right). \tag{8}$$

In order to achieve a robust *a priori* SNR estimation we propose to model $N_\alpha(k,\ell)$ by a single Weibull distribution with a frequency-dependent time-invariant scale parameter $\lambda_{N_\alpha}(k)$ calculated from the noise PSD via

$$\lambda_N(k) = \frac{1}{L}\sum_{\ell=1}^{L}\lambda_N(k,\ell), \tag{9}$$

$$\lambda_{N_\alpha}(k) = \frac{\lambda_N(k)}{\sqrt[\alpha]{P_S(k)}}. \tag{10}$$

In this way, we model the noise NAOM coefficients by

$$p_{N_\alpha(k)}(n) = \text{Weib}\,(n;\lambda_{N_\alpha}(k),\alpha). \tag{11}$$

Based on the sparseness of the clean speech STFTs we suggest to model the clean speech NAOM coefficients $S_\alpha(k,\ell)$ in (4) by a Weibull mixture model (WMM)

$$p_{S_\alpha(k)}(s) = \sum_{m=1}^{M}\pi_m\cdot\text{Weib}\,(s,\lambda_m(k),\beta), \tag{12}$$

where M is the number of components, π_m the weight of the m-th component, and $\lambda_m(k)$ the component-specific frequency-dependent scale parameter. Furthermore, we take deviations from the probability model into account by using a shape parameter $\beta \neq \alpha$. This allows to model a variability of energy-rich clean speech NAOM coefficients by using more heavy-tailed Weibull distribution ($\beta > \alpha$) resulting in reducing the number of necessary mixture components. While the first WMM component $m = 1$ aims to model low-energy NAOM coefficients of the clean speech signal, the other WMM components should model energy-rich coefficients. According to (8), the parameters of (12) fulfill the condition $E[S_\alpha^2(k,\ell)] = 1$ if

$$\sum_{m=1}^{M}\pi_m\cdot\lambda_m^\beta(k) = \frac{1}{\Gamma(\beta+1)}. \tag{13}$$

In Fig. 1(a) a distribution of white noise NAOM coefficients and the estimated distribution $p_{N_\alpha}(n)$ are depicted for a single frequency bin. slope is The distribution is well represented here by the estimation. Fig. 1(b) shows a distribution of the clean speech NAOM coefficients (black) for a single frequency bin and the estimated WMM $p_{S_\alpha}(s)$ for $M = 2$ (red), where the first component (green) represents NOAM coefficients with low energy and the second component (blue) the more scattered high-energy coefficients. To capture the complexity of the clean speech NAOM coefficients, a mixture model with at least two-components is necessary.

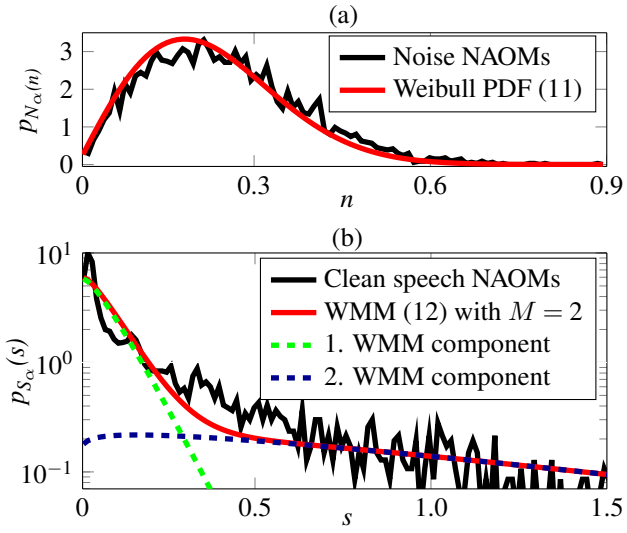

Figure 1: Histograms of NAOM coefficients of the frequency bin centered at 718.75 Hz and the estimated distributions: (a) for a white noise signal with $\alpha = 0.7$ and (b) for a clean speech signal with $\alpha = 0.7$ and $\beta = 2.5$.

2.2 Maximum a-priori estimation

Based on (6) and given $p_{N_\alpha(k)}(n)$ and $p_{S_\alpha(k)}(s)$, the maximum a-posteriori (MAP) estimate of the clean speech NAOM coefficient $S_\alpha(k,\ell)$ can be obtained from

$$\hat{S}_\alpha^{\text{MAP}}(k,\ell) = \underset{s}{\text{argmax}}\underbrace{p_{N_\alpha(k)}(y-s)\cdot p_{S_\alpha(k)}(s)}_{\propto\, p_{S_\alpha(k)|Y_\alpha(k,\ell)}(s|y)}, \tag{14}$$

where $y = Y_\alpha(k,\ell)$ is the current observation. Using the models (11) and (12), the MAP estimation turns out to be a maximum search in the interval $s\in(0;y)$ as follows

$$\hat{S}_\alpha^{\text{MAP}}(k,\ell) = \underset{s\in(0;y)}{\text{argmax}}\left[\frac{\varepsilon(s)\cdot\varepsilon(y-s)}{\beta\alpha\cdot\lambda_{N_\alpha}}\cdot s^{\frac{2}{\beta}-1}\cdot(y-s)^{\frac{2}{\alpha}-1}\cdot\right.$$
$$\left.\cdot\sum_{m=1}^{M}\frac{\pi_m}{\lambda_m}\exp\left(-\frac{1}{\lambda_m}s^{\frac{2}{\beta}}-\frac{1}{\lambda_{N_\alpha}}(y-s)^{\frac{2}{\alpha}}\right)\right]. \tag{15}$$

After derivation with respect to $s\in(0;y)$ we have to solve

$$\sum_{m=1}^{M}\frac{\pi_m}{\lambda_m}\left[\frac{2-\beta}{\beta s}-\frac{2-\alpha}{\alpha(y-s)}-\frac{2}{\beta\lambda_m}s^{\frac{2}{\beta}-1}+\right. \tag{16}$$
$$\left.+\frac{2}{\alpha\lambda_{N_\alpha}}(y-s)^{\frac{2}{\alpha}-1}\right]\cdot\exp\left(-\frac{(y-s)^{\frac{2}{\alpha}}}{\lambda_{N_\alpha}}-\frac{s^{\frac{2}{\beta}}}{\lambda_m}\right)\overset{!}{=}0.$$

To enable an efficient root finder for (16), we suggest two simplifications of (16), which lead to the approximative analytically calculated roots. First, to get rid of the exponential function we suggest to approximate it at $s = 0$ similar to [20]. And, second, we propose to use $\beta = \alpha = 1$ in (16), since we observed a stable behaviour of a root finder for this values. Both simplifications applied to (16) results in

$$\left(\frac{K_1}{\lambda_{N_\alpha}} + K_2\right) \cdot s^3 - y\left(\frac{2K_1}{\lambda_{N_\alpha}} + K_2\right) \cdot s^2 + K_1\left(\frac{y^2}{\lambda_{N_\alpha}} - 1\right) \cdot s + \frac{yK_1}{2} = 0 \quad (17)$$

with constants $K_1 = \sum_{m=1}^{M} \frac{\pi_m}{\lambda_m}$ and $K_2 = \sum_{m=1}^{M} \frac{\pi_m}{\lambda_m^2}$. Solving (17) using the Cardano's formulas delivers the desired $\hat{S}_\alpha^{\text{MAP}}(k,\ell)$ estimate [21], which is used for calculating the a-priori SNR estimate via

$$\hat{\xi}(k,\ell) = \frac{\left(\hat{S}_\alpha^{\text{MAP}}(k,\ell) \cdot \sqrt{P_S(k)}\right)^{\frac{2}{\alpha}}}{\lambda_N(k,\ell)}. \quad (18)$$

If more than one root of (17) is found in the interval $(0; y)$, the largest one is chosen as the solution.

2.3 Estimation of WMM parameters

For the MAP estimation (14), we assumed the knowledge of $p_{S_\alpha(k)}(s)$ from (12). In this section we discuss the estimation of the parameters of the two-component WMM $p_{S_\alpha(k)}(s)$.

Cappé discovered, that it is advantageous for *a priori* SNR estimation to use a minimum value of *a priori* SNR estimate ξ_{min} , which should be chosen according to 'the average *a priori* SNR' in the time-frequency slots 'containing noise only' [2]. Inspired by this, we suggest to integrate ξ_{min} in the clean speech model $p_{S_\alpha(k)}(s)$ by setting $\lambda_1(k)$ in a way that, the mean of the first WMM component $\mu_1(k)$ corresponds to the ξ_{min}. According to (1) and (4), $\mu_1(k)$ can then be calculated as

$$\mu_1(k) = \frac{(\lambda_N(k) \cdot \xi_{\text{min}})^{\frac{\alpha}{2}}}{\sqrt{\overline{P}_S}}, \quad (19)$$

where $\overline{P}_S = \frac{1}{K}\sum_{k=1}^{K} P_S(k)$ is the frequency-independent power of the clean speech AOM coefficients. Using $\overline{P}_S$ instead of $P_S(k)$ has proven to be advantageous for a robust causal *a priori* SNR estimation described below. With (8) and (19) we then get

$$\lambda_1(k) = \left(\frac{\mu_1(k)}{\Gamma\left(\frac{\beta}{2}+1\right)}\right)^{\frac{2}{\beta}}. \quad (20)$$

The parameter $\lambda_2(k)$ of $p_{S_\alpha(k)}(s)$ can be calculated by using the estimate obtained from the EM algorithm for the WMM presented in [22] with

$$\lambda_2(k) = \frac{1}{L_2(k)} \sum_{\ell=1}^{L} \gamma_2(k,\ell) \cdot \hat{S}_\alpha(k,\ell)^{\frac{2}{\beta}}, \quad (21)$$

where $L_2(k) = \sum_{\ell=1}^{L} \gamma_2(k,\ell)$, with the responsibility of the 2nd mixture component being computed in the E-step as

$$\gamma_2(k,\ell) = \frac{\pi_2 \cdot \text{Weib}\left(\hat{S}_\alpha(k,\ell); \lambda_2(k), \beta\right)}{\sum_{m=1}^{M} \pi_m \cdot \text{Weib}\left(\hat{S}_\alpha(k,\ell); \lambda_m(k), \beta\right)}. \quad (22)$$

Here, $\hat{S}_\alpha(k,\ell)$ denotes the denoised NAOM coefficients

$$\hat{S}_\alpha(k,\ell) = \max\left(Y_\alpha(k,\ell) - \frac{\lambda_N(k)^{\frac{\alpha}{2}}}{\sqrt{P_S(k)}} \cdot \Gamma\left(\frac{\alpha}{2}+1\right),\ \mu_1(k)\right), \quad (23)$$

taken into consideration that the second WMM component models the high-energy NAOM coefficients. The EM algorithm is initialized by $\pi_2 = 0.5$ and $\lambda_2(k) = 1$. To ensure $E[S_\alpha^2(k,\ell)] = 1$, the weights are adjusted after convergence of the EM algorithm with

$$\pi_1(k) = 1 - \pi_2(k), \quad \pi_2(k) = \frac{\frac{1}{\Gamma(\beta+1)} - \lambda_1(k)^\beta}{\lambda_2(k)^\beta - \lambda_1(k)^\beta}. \quad (24)$$

2.4 Causal implementation

In Sections 2.1, 2.2 and 2.3 a non-causal *a priori* SNR estimation was presented assuming the knowledge of the noisy NAOM coefficients $Y_\alpha(k,\ell)$ for $\forall \ell \in [1; L]$. However, a causal low-latency solution is required for most applications. To gain a causal estimation, the equations (3), (9) and (21) For this, every average over time has to be done recursively. Thus, the corresponding parameters become frame-dependent. The estimation of $P_S(k)$ in (3) turns to

$$P_S(k,\ell) = \max\left(P_Y(k,\ell) - P_N(k,\ell),\ \xi_{\text{min}}^{\alpha} \cdot P_N(k,\ell)\right) \quad (25)$$

where

$$P_Y(k,\ell) = \frac{\ell-1}{\ell} \cdot P_Y(k,\ell-1) + \frac{1}{\ell} \cdot |Y(k,\ell)|^{2\alpha}, \quad (26)$$

$$P_N(k,\ell) = \frac{\ell-1}{\ell} \cdot P_N(k,\ell-1) + \frac{1}{\ell} \cdot \lambda_N(k,\ell)^\alpha \cdot \Gamma(\alpha+1). \quad (27)$$

Further, the calculation of $\lambda_N(k)$ in (9) is modified to

$$\overline{\lambda}_N(k,\ell) = \frac{\ell-1}{\ell} \cdot \overline{\lambda}_N(k,\ell\text{-}1) + \frac{1}{\ell} \cdot \lambda_N(k,\ell). \quad (28)$$

Since $P_S(k,\ell)$ and $\overline{\lambda}_N(k,\ell)$ are current estimates of the time-independent parameters $P_S(k)$ and $\lambda_N(k)$ respectively, the values calculated in equations (26)-(28) lose their ability to react to local statistics with increasing ℓ. Finally, the estimation of $\lambda_2(k)$ in (21) is altered to

$$\lambda_2(k,\ell) = \frac{L_2(k,\ell\text{-}1)}{L_2(k,\ell)} \cdot \lambda_2(k,\ell\text{-}1) + \frac{\gamma_2(k,\ell)}{L_2(k,\ell)} \cdot \hat{S}_\alpha(k,\ell)^{\frac{2}{\beta}} \quad (29)$$

with $L_2(k,\ell) = L_2(k,\ell\text{-}1) + \gamma_2(k,\ell)$. To keep the computational effort low, we suggest to calculate $\gamma_2(k,\ell)$ just once for the current observation $Y_\alpha(k,\ell)$ and carry out only one iteration of the EM algorithm in each frame.

The proposed causal estimation procedure can be summarized in an algorithm with 6 steps, which are outlined in the Algorithm 1.

3 Experimental results

To investigate the performance of the proposed *a priori* SNR estimator in noisy environments, an experimental evaluation is carried out on the training subset of the single-channel clean speech signals of the CHiME database [23]. The isolated signals of one female and one male speaker at a sampling of $F_S = 16\,\text{kHz}$ are concatena-

Algorithm 1 The WMM-based *a priori* SNR estimator

Input: $|Y(k,\ell)|^2$, $\hat{\lambda}_N(k,\ell)$
Output: *a priori* SNR estimate $\hat{\xi}(k,\ell)$
Initialization: set $\pi_1 = \pi_2 = 0.5$, $\lambda_2 = 1$ and α, β
for all time frames ℓ **do**
1. Estimate $P_S(k,\ell)$ by utilizing (25)-(27)
2. Compute $Y_\alpha(k,\ell)$ and $\lambda_{N_\alpha}(k,\ell)$ by (6), (28), (10)
3. Calculate $\lambda_1(k,\ell)$ by (19)-(20) and $\lambda_2(k,\ell)$ with (29) by executing one EM-iteration (22)-(23)
4. Adjust $\pi_1(k,\ell)$ and $\pi_2(k,\ell)$ via (24)
5. Find $\hat{S}_\alpha^{\text{MAP}}(k,\ell)$ as a root of the polynomial (17)
6. Obtain *a priori* SNR estimate $\hat{\xi}(k,\ell)$ by (18)

ted to test signals of 2 minutes length each. The signals of 7 noise types are taken from the Noisex92 database: white, pink, f16, hfchannel, factory-1, factory-2 and babble [24]. The noise signals are artificially added to the clean speech signals at a global SNR ranging from -5 dB to 25 dB in steps of 5 dB. For STFT we used a 512 samples Hamming window and a shift factor 0.25.

The noise PSD $\lambda_N(k,\ell)$ is estimated by the minimum statistics method [25]. For comparison purposes the *a priori* SNR is also estimated by the DD method with a weighting factor of 0.98 [3]. Both the proposed WMM-based and the DD approach used the same lower bound for *a priori* SNR $\xi_{\min} = -18$ dB [2]. For the DD approach a log-spectral amplitude (LSA) gain function is used [26]. STFT coefficients of an enhanced signal are calculated here by employing the LSA gain with a gain floor $G_{\min} = -25$ dB [27]. For the proposed WMM-based approach the Wiener filter is applied [28], which does not need any additional gain floor as our experiments showed.

To find the optimal values α and β for the WMM-based approach, we carried out speech enhancement of female and male speech signals (other than test signals) of 1 minute length each distorted by all considered noise types for values $\alpha \in [0.3; 1.2]$ and $\beta \in [1; 5]$. Further we calculated the wide-band mean opinion score (MOS) listening quality objective scores of the enhanced signals [29]. Fig. 2 shows the positive MOS improvement

$$\Delta\text{MOS} = \max\left(\text{MOS}_{\text{WMM}} - \text{MOS}_{\text{DD}}\,,\, 0\right) \qquad (30)$$

averaged over all simulated values of global SNR as a function of α and β. It is obvious, that the proposed approach outperforms the DD approach for a wide range of α and β values. However, it achieves the best performance for $\alpha \in [0.55; 0.7]$ and $\beta \in [2; 5]$. Similar to [16] we observed less musical noise in enhanced signals for values of α smaller than 1. For further experiments $\alpha_{\text{opt}} = 0.64$ and $\beta_{\text{opt}} = 2.7$ are chosen.

To investigate the performance of the *a priori* SNR estimation we further calculated the Itakura-Saito Distance (ISD) [30] and the variance of the logarithmic error variance (LEV) [31]. While the ISD is the measure of the mean estimation error, the LEV represents the estimator variance. Smaller values of ISD and LEV correspond to better estimator performance. Since the true *a priori* SNR (1) is not known, the instantaneous *a priori* SNR values calculated from the available clean speech and pure noise signals are used as reference *a priori* SNR. Table 1 summarizes the resulting ISD, LEV and MOS values of the considered approaches for the test signals distorted by various noise types for different SNR values.

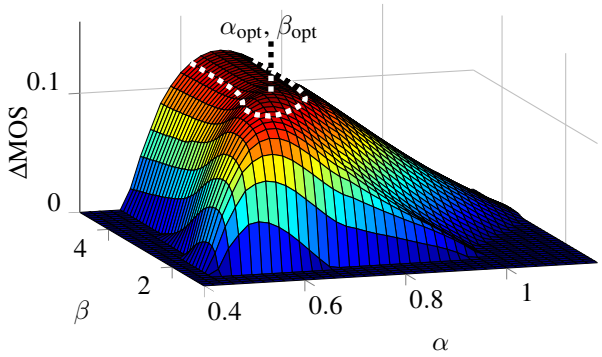

Figure 2: MOS improvement of the proposed algorithm over the DD method averaged over all considered noise types and all SNR values as a function of α and β.

All performance measure show poorer values with decreasing global SNR. In terms of all considered performance measures the proposed approach outperforms the DD method. The proposed approach does not seem to suffer from *a priori* SNR overestimation since this would be penalized by the ISD measure. The reduction of ISD achieved by the WMM-based approach increases with decreased global SNR and reaches 14-19 % on average. In contrast to the ISD measure, we boserved the best reduction of the LEV measure for the smallest and the highest global SNR values. In comparison to the DD approach the proposed *a priori* SNR estimator reduces the LEV measure on average by around 15 %. The reduction of the estimation error and of the variance of the estimator leads consequently to an improvement of the speech quality of the enhanced signals by about 4.3 %, as measured by MOS. In light of the achieved performance improvement it should be mentioned that the computational effort of the proposed approach is much higher than that of the DD approach.

SNR, dB		−5	0	5	10	15	20	25
ISD	DD	48.8	44.0	39.6	34.9	30.2	24.5	19.1
	WMM	42.6	38.1	34.1	30.4	27.3	23.0	18.9
LEV	DD	53.1	49.0	46.4	45.1	45.5	47.4	50.5
	WMM	45.6	43.9	42.6	41.1	39.0	37.0	35.9
MOS	DD	1.11	1.30	1.63	2.09	2.57	3.00	3.39
	WMM	1.18	1.46	1.77	2.13	2.62	3.16	3.61

Table 1: Resulting ISD, LEV and MOS values of the proposed WMM-based approach and of the DD method averaged over considered noise types for different SNR values.

4 Conclusions

In this contribution we proposed a novel causal *a priori* SNR estimator based on a Weibull mixture model for the normalized α-order magnitudes and optimized its parameters with respect to the achievable MOS values. The proposed estimator achieves superior results compared to the popular decision directed approach [3] for all examined cases and performance measures. The price to pay is, however, a higher computational cost, which should be reduced in our future work by developing a universal WMM.

References

[1] I. Cohen and S. Gannot, "Spectral Enhancement Methods," in *Springer Handbook of Speech Processing* (J. Benesty, M. M. Sondhi, and Y. A. Huang, eds.), pp. 873–902, Springer Berlin Heidelberg, 2008.

[2] O. Cappe, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," *IEEE Transactions on Speech and Audio Processing*, vol. 2, pp. 345–349, Apr. 1994.

[3] Y. Ephraim 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. ASSP-32, pp. 1109–1121, Dec. 1984.

[4] C. Plapous, C. Marro, and P. Scalart, "Improved signal-to-noise ratio estimation for speech enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, pp. 2098–2108, Nov. 2006.

[5] I. Y. Soon and S. N. Koh, "Low distortion speech enhancement," *IEE Proceedings - Vision, Image and Signal Processing*, vol. 147, pp. 247–253, June 2000.

[6] M. K. Hasan, S. Salahuddin, and M. R. Khan, "A modified a priori SNR for speech enhancement using spectral subtraction rules," *IEEE Signal Processing letters*, vol. 8, pp. 450–453, Apr. 2004.

[7] I. Cohen, "Speech enhancement using a noncausal a priori SNR estimator," *Signal Processing Letters, IEEE*, vol. 11, pp. 725–728, Sept. 2004.

[8] I. Cohen, "Relaxed statistical model for speech enhancement and a priori SNR estimation," *Speech and Audio Processing, IEEE Transactions on*, vol. 13, pp. 870–881, Sept. 2005.

[9] Y.-S. Park and J.-H. Chang, "A novel approach to a robust a priori SNR estimator in speech enhancement," *IEICE transactions on communications*, vol. 90, pp. 2182–2185, Aug. 2007.

[10] S. Suhadi, C. Last, and T. Fingscheidt, "A data-driven approach to a priori SNR estimation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, pp. 186–195, Jan. 2011.

[11] P. C. Yong, S. Nordholm, and H. H. Dam, "Optimization and evaluation of sigmoid function with a priori SNR estimate for real-time speech enhancement," *Speech Communication*, vol. 55, pp. 358 – 376, Sept. 2013.

[12] S. Lee, C. Lim, and J.-H. Chang, "A new a priori SNR estimator based on multiple linear regression technique for speech enhancement," *Digital Signal Processing*, vol. 30, pp. 154–164, Apr. 2014.

[13] A. Chinaev and R. Haeb-Umbach, "A priori SNR estimation using a generalized decision directed approach," in *Seventeenth Annual Interspeech Conference of the International Speech Communication Association*, Sept. 2016.

[14] S. Elshamy, N. Madhu, W. Tirry, and T. Fingscheidt, "An iterative speech model-based a priori SNR estimator," in *Sixteenth Annual Conference of the International Speech Communication Association*, pp. 1740–1744, Sept. 2015.

[15] P. Mowlaee and R. Saeidi, "Target speaker separation in a multisource environment using speaker-dependent post-filter and noise estimation," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 7254–7258, May 2013.

[16] T. Inoue, Y. Takahashi, H. Saruwatari, K. Shikano, and K. Rondo, "Theoretical analysis of musical noise in generalized spectral subtraction: Why should not use power/amplitude subtraction?," in *Signal Processing Conference, 2010 18th European*, pp. 994–998, Aug. 2010.

[17] I. Tashev and A. Acero, "Statistical modeling of the speech signal," in *International Workshop on Acoustic, Echo, and Noise Control (IWAENC)*, Aug 2010.

[18] J. Jensen, I. Batina, R. C. Hendriks, and R. Heusdens, "A study of the distribution of time-domain speech samples and discrete Fourier coefficients," in *Proc. SPS-DARTS*, vol. 1, pp. 155–158, 2005.

[19] B. L. Sim, Y. C. Tong, J. S. Chang, and C. T. Tan, "A parametric formulation of the generalized spectral subtraction method," *Speech and Audio Processing, IEEE Transactions on*, vol. 6, pp. 328–337, July 1998.

[20] S. Chehresa and M. Savoji, "Speech enhancement using Gaussian mixture models, explicit Bayesian estimation and Wiener filtering," *Iranian Journal of Electrical and Electronic Engineering*, vol. 10, pp. 168–175, Sept. 2014.

[21] E. W. Weisstein, "Cubic formula," *Omega*, vol. 86, p. 87, 2002.

[22] T. Bucar, M. Nagode, and M. Fajdiga, "Reliability approximation using finite Weibull mixture distributions," *Reliability Engineering & System Safety*, vol. 84, no. 3, pp. 241 – 251, 2004.

[23] J. Barker, R. Marxer, E. Vincent, and S. Watanabe, "The third "CHiME" speech separation and recognition challenge: Dataset, task and baselines," in *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, pp. 504–511, Dec. 2015.

[24] A. Varga and H. J. M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Communication*, vol. 12, pp. 247–251, July 1993.

[25] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *Speech and Audio Processing, IEEE Transactions on*, vol. 9, pp. 504–512, July 2001.

[26] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, pp. 443–445, Apr. 1985.

[27] I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," *Signal processing*, vol. 81, no. 11, pp. 2403–2418, 2001.

[28] R. McAulay and M. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 28, pp. 137–145, Apr. 1980.

[29] "Application guide for objective quality measurement based on recommendations P.862, P.862.1 and P.862.2." ITU-T Recommendation P.862.3, Nov. 2007.

[30] R. Gray, A. Buzo, A. Gray, and Y. Matsuyama, "Distortion measures for speech processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, pp. 367–376, Aug. 1980.

[31] J. Taghia, J. Taghia, N. Mohammadiha, J. Sang, V. Bouse, and R. Martin, "An evaluation of noise power spectral density estimation algorithms in adverse acoustic environments," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4640–4643, May 2011.