A Priori SNR Estimation Using a Generalized Decision Directed Approach

Aleksej Chinaev, Reinhold Haeb-Umbach

Department of Communications Engineering, Paderborn University, 33098 Paderborn, Germany
{chinaev,haeb}@nt.uni-paderborn.de

Abstract

In this contribution we investigate *a priori* signal-to-noise ratio (SNR) estimation, a crucial component of a single-channel speech enhancement system based on spectral subtraction. The majority of the state-of-the art *a priori* SNR estimators work in the power spectral domain, which is, however, not confirmed to be the optimal domain for the estimation. Motivated by the generalized spectral subtraction rule, we show how the estimation of the *a priori* SNR can be formulated in the so called generalized SNR domain. This formulation allows to generalize the widely used decision directed (DD) approach. An experimental investigation with different noise types reveals the superiority of the generalized DD approach over the conventional DD approach in terms of both the mean opinion score - listening quality objective measure and the output global SNR in the medium to high input SNR regime, while we show that the power spectrum is the optimal domain for low SNR. We further develop a parameterization which adjusts the domain of estimation automatically according to the estimated input global SNR.

Index Terms: single-channel speech enhancement, *a priori* SNR estimation, generalized spectral subtraction

1. Introduction

The *a posteriori* SNR and the *a priori* SNR estimation are two crucial tasks in any so called analysis-modification-synthesis (AMS) framework for single-channel spectral enhancement, often referred to as spectral subtraction (SS) systems. Based on their estimates the desired gain function can be calculated and in the modification step applied to the short-time Fourier transform (STFT) of the noisy signal. While the *a posteriori* SNR is considered as a correction parameter of the gain function, the *a priori* SNR has been advised to be used as its dominant parameter [1]. Since both SNR quantities are defined in the power spectral density (PSD) domain, they are usually calculated from PSD estimates of the noise signal and of the clean speech signal. For this first the *a posteriori* SNR is estimated by applying one of the many sophisticated noise PSD trackers, e.g., the minimum statistics (MS) approach [2, 3] or its version extended by a Bayesian postprocessor [4]. The following *a priori* SNR estimate is usually calculated as a weighted sum of two terms in the spirit of the decision directed (DD) approach [5]. The first is the *a priori* SNR estimate calculated from the spectral magnitude of the enhanced speech signal of the previous frame, and the second is the maximum likelihood (ML) estimate of the *a priori* SNR based on the current *a posteriori* SNR estimate. Thus the *a priori* SNR estimation exploits information of both the noise PSD tracker and the used gain function, and it can be considered a central component of any spectral AMS framework.

Without exaggeration one can assert that the DD approach is the most popular one for the *a priori* SNR estimation. Essentially, its simplicity and good performance contributed to its tremendous acceptance. However, the DD approach suffers from one well known drawback – slow response to an abrupt change in the instantaneous SNR known also as the reverberation effect [6]. The main reason for this is a constant weighting factor, which should be set to about 0.98 to avoid the so called musical noise in the enhanced signal, while simultaneously achieving a reasonable noise suppression.

Many approaches have been developed to overcome this shortcoming. The authors in [7] and [8] propose the use of a time-variant weighting factor, which is adapted by the energy change of consecutive frames. In [9] a noncausal *a priori* SNR estimator is developed, which is able to distinguish between speech onsets and noise irregularities. In [10] the DD approach is considered as a 'Kalman filter'-like estimator consisting of a propagation- and an update-step. Beside the non-causal estimator, a causal approach is proposed, which is more relevant in a real-world application. An improved version of the DD approach was developed in [6], where the *a priori* SNR estimation is refined by applying the gain function of the previous time step to the current observation instead of the previous one. In [11] an adaptive weighting factor incorporating a sigmoid-type control function calculated on the transient of the *a posteriori* SNR is proposed. In [12] a data-driven approach is presented by using two neural networks, one trained for speech absence and another for speech presence in an elaborated preprocessing phase. In [13] a modified sigmoid gain function is proposed, which is mapped with the *a priori* SNR estimates instead of *a posteriori* SNR in contrast to [11]. In [14] the conventional equation of the DD approach is considered as a linear regression model, whose coefficients are calculated by least squares optimization. To improve the tracking ability, the DD equation is extended in [15] by an additional momentum term with a time-invariant factor chosen in the MMSE sense. To reduce the estimation errors in speech presence, the use of Gaussian mixture models of clean speech has been lately proposed in [16]. The models had to be estimated in an extra training phase.

Following its definition in the PSD domain, all mentioned approaches also estimate the *a priori* SNR in the PSD domain [17]. The majority of them make use of the ML estimate of the *a priori* SNR, which exploits the additivity assumption of the PSDs of the clean speech and noise signals. On the other hand, a recently published investigation [18] showed that the additivity of power spectra is not an optimal assumption in terms of the quality of the speech signal enhanced by spectral subtraction. Earlier publications report, that the processed signals enhanced by using a so-called generalized spectral subtraction (GSS) sound 'less noisy' [19] and of 'higher quality' [20] compared to the signals obtained by the conventional SS rule. The GSS gain functions developed in [21] outperform slightly the gain function of the conventional SS approach in the majority of tests. Motivated by these observations we are going to investigate, whether the estimation of *a priori* SNR should best be

performed also in a domain other than the PSD domain.

The remainder of this contribution is structured as follows: The DD approach is briefly revisited in Section 2. Next the statistical modelling in the generalized spectral domain is described in Section 3, followed by the derivation of the generalized decision directed approach and its parametrization in the experiments with white noise. The experimental comparison of the proposed GDD approach with the conventional DD approach is given in the Section 4, and the conclusions are drawn in Section 5.

2. Decision directed approach

The spectral enhancement aims to estimate the clean speech STFT coefficients $S(k,\ell)$ from the noisy STFT coefficients:

$$Y(k,\ell) = S(k,\ell) + D(k,\ell) \tag{1}$$

distorted by an additive noise signal with the STFT coefficients $D(k,\ell)$, where k and ℓ are the frequency bin and frame indices, respectively. $S(k,\ell)$ and $D(k,\ell)$ are modelled as two uncorrelated non-stationary complex valued zero-mean Gaussian distributed random processes with time-variant PSD parameters

$$\lambda_S(k,\ell) \triangleq E[|S(k,\ell)|^2], \quad (2) \qquad \lambda_D(k,\ell) \triangleq E[|D(k,\ell)|^2], \quad (3)$$

where $E[\cdot]$ denotes the expectation operator. The *a posteriori* SNR $\gamma(k,\ell)$ and *a priori* SNR $\xi(k,\ell)$ are defined as in [5]:

$$\gamma(k,\ell) \triangleq \frac{|Y(k,\ell)|^2}{\lambda_D(k,\ell)}, \quad (4) \qquad \xi(k,\ell) \triangleq \frac{\lambda_S(k,\ell)}{\lambda_D(k,\ell)}. \quad (5)$$

Similarly, using the additivity assumption of the power spectra:

$$|Y(k,\ell)|^{2\rho} = |S(k,\ell)|^{2\rho} + |D(k,\ell)|^{2\rho}. \tag{6}$$

For $\rho = 1$ one can define the problem formulation for *a priori* SNR estimation in the domain of the SNR quantities as follows: estimate the *a priori* SNR $\xi(k,\ell)$ from the *a posteriori* SNR

$$\gamma(k,\ell) = \zeta(k,\ell) + \eta(k,\ell), \tag{7}$$

where the instantaneous *a priori* SNR $\zeta(k,\ell)$ and the ratio of the instantaneous noise PSD to the (long-term) noise PSD parameter (NPR) $\eta(k,\ell)$ are defined as follows:

$$\zeta(k,\ell) \triangleq \frac{|S(k,\ell)|^2}{\lambda_D(k,\ell)}, \quad (8) \qquad \eta(k,\ell) \triangleq \frac{|D(k,\ell)|^2}{\lambda_D(k,\ell)}. \quad (9)$$

Note, that in [22], where the *a priori* SNR is introduced for the first time, it is defined as a random variable (8) and not as a parameter (5) according to [5]. From (5) with (7) we get

$$\xi(k,\ell) = E[\zeta(k,\ell)] = E[\gamma(k,\ell)] - 1, \tag{10}$$

a groundbreaking equation in sense of using $\xi(k,\ell)$ for calculation of gain functions instead of *a posteriori* SNR estimates $\hat{\gamma}(k,\ell)$ [23]. (10) is the power subtraction rule based on the additivity assumption (6) for $\rho = 1$. Since $\gamma(k,\ell)$ is exponentially distributed, (10) appears in the ML estimation of the *a priori* SNR based on a single observation $\hat{\gamma}(k,\ell)$ as in [5]:

$$\hat{\xi}^{\mathrm{ML}}(k,\ell) = \max(\hat{\gamma}(k,\ell) - 1, 0), \tag{11}$$

where a maximum operator $\max(\cdot)$ ensures the positiveness of $\hat{\xi}^{\mathrm{ML}}(k,\ell)$. According to [5] $\hat{\xi}^{\mathrm{ML}}(k,\ell)$ appears as the second of two terms in the DD estimate:

$$\hat{\xi}^{\mathrm{DD}}(k,\ell) = \alpha \cdot \tilde{\xi}(k,\ell-1) + (1-\alpha) \cdot \hat{\xi}^{\mathrm{ML}}(k,\ell), \tag{12}$$

where α is a weighting factor and $\tilde{\xi}(k,\ell-1)$ a propagated *a priori* SNR estimate of previous processing step defined as

$$\tilde{\xi}(k,\ell-1) = \frac{|\hat{S}(k,\ell-1)|^2}{\hat{\lambda}_D(k,\ell-1)} = G^2(k,\ell-1)\cdot\hat{\gamma}(k,\ell-1), \tag{13}$$

where $G(k,\ell-1)$ is a gain function calculated by using the previous *a priori* SNR estimate $\hat{\xi}(k,\ell-1)$. Note, that (12) is not a recursive averaging but just a weighted sum of two terms, because $\tilde{\xi}(k,\ell-1) \neq \hat{\xi}^{\mathrm{DD}}(k,\ell-1)$. In [5], $\tilde{\xi}(k,\ell-1)$ is introduced by dropping the expectation operator in (2), which is the numerator of (5), because $S(k,\ell)$ is highly non-stationary. Or in our notation, the instantaneous *a priori* SNR (8) is used as a definition for the second term of the DD approach instead of (5). Thus the DD approach makes use of both definitions of the *a priori* SNR, as a parameter (5) and as a random variable (8).

To reduce a 'low-level' musical noise, in [1] it is suggested to introduce a lower bound $\xi_{\min}$ for the *a priori* SNR and to calculate the resulting *a priori* SNR estimate via

$$\hat{\xi}^{\mathrm{DD}}(k,\ell) = \max\left(\hat{\xi}^{\mathrm{DD}}(k,\ell),\ \xi_{\min}\right). \tag{14}$$

The investigations in [24] showed that it is advantageous to use in (13) the gain function in speech activity $G_{H_1}(k,\ell)$ instead of the gain function $G(k,\ell)$, which is applied to get the enhanced STFTs $\hat{S}(k,\ell) = G(k,\ell)\cdot Y(k,\ell)$ and is calculated via

$$G(k,\ell) = \max\left(G_{H_1}(k,\ell),\ G_{\min}\right), \tag{15}$$

where $G_{\min}$ is a gain floor introduced in [25] to hide the residual noise in the enhanced signal. In our experiments we used the minimum mean-square error (MMSE) log-spectral amplitude (LSA) gain function $G_{H_1}(k,\ell)$ as defined in [24].

3. Generalized decision directed approach

3.1. Derivation

Motivated by the GSS rule, which is given in [19] as

$$|\hat{S}(k,\ell)| = \left(\,|Y(k,\ell)|^{\rho} - E[\,|D(k,\ell)|^{\rho}\,]\,\right)^{\frac{1}{\rho}}, \tag{16}$$

where $\rho > 0$ is an arbitrary constant. We define the *a priori* SNR estimation task in the generalized SNR domain as follows: estimate the *a priori* SNR $\xi(k,\ell)$ from the generalized *a posteriori* SNR

$$\gamma_\rho(k,\ell) = \frac{|Y(k,\ell)|^{2\rho}}{E[|D(k,\ell)|^{2\rho}]} = \zeta_\rho(k,\ell) + \eta_\rho(k,\ell), \tag{17}$$

where the generalized instantaneous SNR and NPR defined as:

$$\zeta_\rho(k,\ell) \triangleq \frac{|S(k,\ell)|^{2\rho}}{E[|D(k,\ell)|^{2\rho}]}, \tag{18}$$

$$\eta_\rho(k,\ell) \triangleq \frac{|D(k,\ell)|^{2\rho}}{E[|D(k,\ell)|^{2\rho}]}. \tag{19}$$

Under the assumptions made for $S(k,\ell)$ and $D(k,\ell)$, the $|S(k,\ell)|^{2\rho}$ and $|D(k,\ell)|^{2\rho}$ are uncorrelated non-stationary

real-valued Weibull-distributed random processes with the probability density functions (PDF)

$$p_{|S(k,\ell)|^{2\rho}}(s) = \text{Weib}(s;\, \lambda_S(k,\ell),\ \rho), \tag{20}$$

$$p_{|D(k,\ell)|^{2\rho}}(d) = \text{Weib}(d;\, \lambda_D(k,\ell),\ \rho), \tag{21}$$

where the Weibull PDF $p_X(x) = \text{Weib}(x;\, \lambda, \rho)$ is defined as:

$$\text{Weib}(x;\, \lambda, \rho) \triangleq \frac{1}{\rho\lambda} \cdot x^{\frac{1}{\rho}-1} \cdot e^{-\frac{1}{\lambda}\cdot x^{\frac{1}{\rho}}} \cdot \epsilon(x), \tag{22}$$

where ρ, λ and $\epsilon(x)$ are the shape parameter, the scale parameter and the unit step function, respectively. Note that, for $\rho = 1$ the generalized SNR becomes the conventional SNR and the Weibull PDF simplifies to the exponential distribution, often used to model the conventional SNR quantities.

Using the additivity (1), the PDF of $|Y(k,\ell)|^{2\rho}$ is given by:

$$p_{|Y(k,\ell)|^{2\rho}}(y) = \text{Weib}(y;\, \lambda_S(k,\ell) + \lambda_D(k,\ell),\ \rho). \tag{23}$$

From (20), (21) and (23) one can easily obtain the PDFs of $\eta_\rho(k,\ell)$, $\zeta_\rho(k,\ell)$ and $\gamma_\rho(k,\ell)$ by using the raw moment of κ-th order of the Weibull-distributed random variable:

$$E[X^\kappa] = \lambda^{\kappa\cdot\rho} \cdot \Gamma(\kappa\cdot\rho + 1), \tag{24}$$

where $\Gamma(x)$ is the gamma function. From (24) with $\kappa = 1$ we get the following PDFs

$$p(\eta_\rho) = \text{Weib}(\eta_\rho;\, \Gamma^{-1/\rho}(\rho+1),\ \rho), \tag{25}$$

$$p(\zeta_\rho) = \text{Weib}(\zeta_\rho;\, \xi(k,\ell)\cdot\Gamma^{-1/\rho}(\rho+1),\ \rho), \tag{26}$$

$$p(\gamma_\rho) = \text{Weib}(\gamma_\rho;\, (\xi(k,\ell)+1)\cdot\Gamma^{-1/\rho}(\rho+1),\ \rho). \tag{27}$$

The *a priori* SNR $\xi(k,\ell)$ from (5) appears here as a parameter.

With (17) and (24) one can get the relationship between the conventional and the generalized *a posteriori* SNR

$$\gamma_\rho(k,\ell) = \frac{\gamma^\rho(k,\ell)}{\Gamma(\rho+1)}. \tag{28}$$

Now we can derive the GDD approach in the generalized SNR domain in the spirit of the DD approach, again by using the weighted sum

$$\hat{\xi}_\rho^{\text{GDD}}(k,\ell) = \alpha_\rho \cdot \tilde{\xi}_\rho(k,\ell-1) + (1-\alpha_\rho)\cdot\hat{\xi}_\rho^{\text{ML}}(k,\ell), \tag{29}$$

where α_ρ is a weighting factor of GDD approach. The first term

$$\tilde{\xi}_\rho(k,\ell-1) = \frac{|\hat{S}(k,\ell)|^{2\rho}}{\lambda_D^\rho(k,\ell)\cdot\Gamma(\rho+1)} = G_{H_1}^{2\rho}(k,\ell-1)\cdot\hat{\gamma}_\rho(k,\ell-1) \tag{30}$$

is a generalized propagated *a priori* SNR estimate defined as a random variable similar to (8), (13) and (18), and the second

$$\hat{\xi}_\rho^{\text{ML}}(k,\ell) = (\,\max(\,\hat{\gamma}(k,\ell) - 1\,,\, 0)\,)^\rho \tag{31}$$

is a ML estimate of the generalized *a priori* SNR $\xi_\rho(k,\ell)$ from the Weibull distributed observation $\gamma_\rho(k,\ell)$ with

$$\xi_\rho(k,\ell) = E[\zeta_\rho(k,\ell)] = \xi^\rho(k,\ell) \tag{32}$$

defined as a parameter similar to (5) and (10).

In the spirit of the DD approach, the desired *a priori* SNR of GDD approach in the PSD domain can be calculated from $\hat{\xi}_\rho(k,\ell)$ using (32) as follows

$$\hat{\xi}^{\text{GDD}}(k,\ell) = \max\left(\left(\hat{\xi}_\rho^{\text{GDD}}(k,\ell)\right)^{\frac{1}{\rho}},\ \xi_{\min}\right), \tag{33}$$

defining the output *a priori* SNR as a parameter. However, the output *a priori* SNR considered as a random variable becomes

$$\hat{\xi}^{\text{GDD}}(k,\ell) = \max\left(\left(\hat{\xi}_\rho^{\text{GDD}}(k,\ell)\cdot\Gamma(\rho+1)\right)^{\frac{1}{\rho}},\ \xi_{\min}\right). \tag{34}$$

Note, that for $\rho = 1$ the equations from (29) to (34) of the proposed GDD approach become the equations from (11) to (14) of the conventional DD approach. For initialization it is sufficient to set $\hat{\xi}_\rho^{\text{GDD}}(k,1) = (\max(\hat{\gamma}(k,1)-1, 0))^\rho$.

While the GDD approach can use the same parameter $\xi_{\min}$ as the DD approach, the parameters ρ and α_ρ have to be set appropriately. To do this and to get a deeper insight in the behaviour of the proposed GDD approach experiments with speech signals distorted by the white noise are carried out.

3.2. Parameterization

In our experiments the clean speech signals are generated by concatenating utterances of different male and female speakers from the TIMIT database [26], having a total length of 1 minute each. The white noise signal is taken from the NOISEX-92 database [27] and is added artificially to the clean speech signals at global SNR values varied from 0 dB to 30 dB in steps of 10 dB and denoted in the following as Υ. All signals are sampled at 16 kHz. The STFT spectral analysis used a Hanning window of 512 samples length with a frame overlap of 25%. The *a posteriori* SNR estimates are calculated from the noise PSD estimates $\hat{\lambda}_D(k,\ell)$ of the MS approach [2]. The length of the MS window for minimum search is set to $D = U\cdot V = 96$ frames divided into $U = 8$ subwindows of length of $V = 12$ frames. Further we set $\xi_{\min} = -25$ dB as in [9] and $G_{\min} = -25$ dB as in [24]. In every experiment the parameters of the GDD approach ρ and α_ρ are set to constant values from the ranges of $[0.05; 2]$ and $[0; 1]$, respectively.

Our simulations with white noise have shown, that the optimal weighting factor $\alpha_\rho^{\text{opt}}(\Upsilon)$, which maximize the mean opinion score - listening quality objective (MOS-LQO) measure of the enhanced signal, depends on both the input global SNR Υ

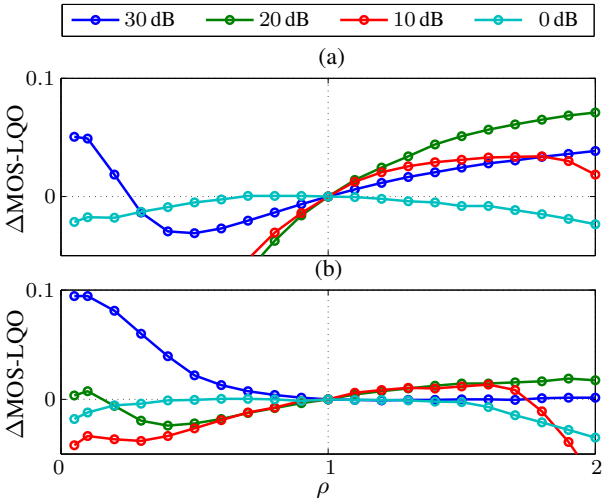

Figure 1: The MOS-LQO improvement ΔMOS-LQO of the proposed GDD approach over the DD approach in enhancing speech signals distorted by white noise at different input global SNR as a function of ρ: (a) by using (33), (b) by using (34).

and the value of ρ. By averaging over all SNR values a SNR-independent weighting factor α_ρ^{opt} can be calculated. For $\rho = 1$ we got $\alpha_\rho^{\text{opt}} = 0.975$, which is very close to the expected 0.98, since in this case the GDD approach reduces to the DD approach. While α_ρ^{opt} decreases for $\rho < 1$, it increases for $\rho > 1$, still remaining smaller than 1. Note, smaller values of α_ρ are preferable, since they deliver a smaller delay of the *a priori* SNR estimates contributing to reduction of reverberation effect.

In Fig. 1, the achievable MOS-LQO improvement of the proposed GDD approach over the conventional DD approach

$$\Delta\text{MOS-LQO} = \text{MOS-LQO}_{\text{GDD}} - \text{MOS-LQO}_{\text{DD}} \qquad (35)$$

averaged over male and female speech signals is depicted as a function of ρ at different input global SNR either by using (33) in (a) and by using (34) in (b). As we can see (33) and (34) show slightly different behaviour.

According to the Fig. 1(a), the GDD approach with (33) and $\rho > 1$ achieves a positive MOS-LQO gain for all input global SNR > 0 dB. For the global SNR of 20 dB the MOS-LQO improvement is particularly high. The Fig. 1(b) shows that the MOS-LQO gain of the proposed GDD approach by using (34) is especially high for the global SNR of 30 dB and for $\rho \approx 0.1$. Here the trajectories of the *a priori* SNR estimates contain less fluctuations in speech absence because the estimator works on highly compressed magnitudes. At the same time the estimator's tracking ability to the rising and decaying speech power remains high, since the optimal weighting factor has small values. Both Figs. 1 (a) and (b) show that for low global SNR of 0 dB $\rho = 1$ is the best choice for the *a priori* SNR estimation which corresponds to the conventional DD approach.

To attain the highest possible MOS-LQO improvement for arbitrary input global SNR Υ we suggest to adapt ρ for every frame ℓ using the following estimate of the input global SNR

$$\hat{\Upsilon}_0(\ell) = \frac{\ell - 1}{\ell} \cdot \hat{\Upsilon}_0(\ell - 1) + \frac{1}{\ell} \cdot \sum_{k=1}^{K} \hat{\xi}^{\text{GDD}}(k, \ell), \qquad (36)$$

where K is a number of frequency bins until the Nyquist frequency. For initialization we used $\hat{\Upsilon}(1) = 15$ dB with the relation $\hat{\Upsilon} = 10 \cdot \log_{10} \hat{\Upsilon}_0$.

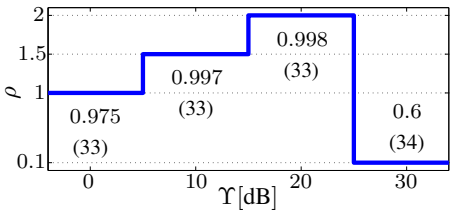

Figure 2: Used adaptation function $\rho(\Upsilon)$ with corresponding α_ρ^{opt} values and the applied final GDD equations in parentheses.

In Fig. 2 the chosen adaptation function $\rho(\Upsilon)$ used in our experiments for adaptation is depicted together with the corresponding values of α_ρ^{opt} and the applied final GDD equations.

4. Experimental results

To investigate the performance of the proposed GDD approach with the suggested adaptation scheme, we extended our previous experiments in two aspects. First, we used longer signals of the TIMIT database with a total length of 3 minutes and, second, we additionally employed 14 remaining noise types from the NOISEX-92 database listed in the legend of Fig. 3. To show that the quality improvement of the enhanced signals measured by ΔMOS-LQO is not achieved at costs of the output global SNR (SNR-OUT), we calculated the SNR-OUT improvement of the proposed GDD approach over the DD approach

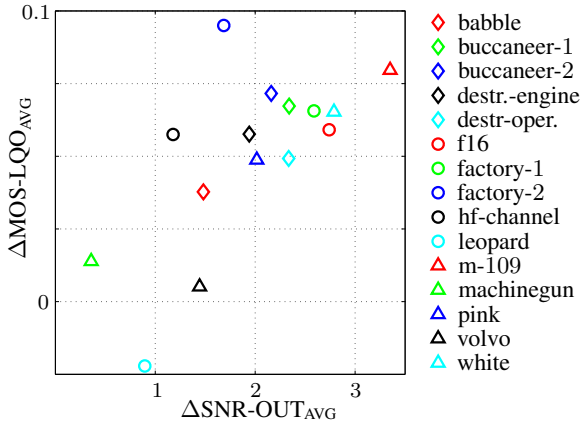

Figure 3: Overall improvement of the proposed GDD approach compared to the DD approach in terms of $\Delta\text{MOS-LQO}_{\text{AVG}}$ and $\Delta\text{SNR-OUT}_{\text{AVG}}$ values for NOISEX-92 database.

$$\Delta\text{SNR-OUT} = \text{SNR-OUT}_{\text{GDD}} - \text{SNR-OUT}_{\text{DD}} \qquad (37)$$

averaged over male and female speech signals and measured in dB. To show the overall improvement, the values of ΔMOS-LQO and ΔSNR-OUT are additionally averaged over all simulated input global SNR values resulting in $\Delta\text{MOS-LQO}_{\text{AVG}}$ and $\Delta\text{SNR-OUT}_{\text{AVG}}$ values, which are depicted in Fig. 3 for all noise types of the NOISEX-92 database.

Our experimental results show, that the proposed GDD approach compared to the conventional DD approach improves both the quality and the output global SNR of the processed signals for almost all considered noise types. Averaged over all noise types we observed a moderate improvement of speech quality of $\Delta\text{MOS-LQO} = 0.05$ score points (consistent to the ΔMOS-LQO values in Fig. 1) and a remarkable increase in output global SNR of $\Delta\text{SNR-OUT} = 2$ dB (i.e. from 18 dB for DD approach up to 20 dB for the proposed GDD approach for averaged global input SNR of 15 dB). Only for leopard noise an improvement in the output global SNR of almost 1 dB causes a slight loss in speech quality of output signals. Note, that the performance improvement comes mostly from signals generated in the medium to high input global SNR regime. In general it looks as if the proposed generalization of the conventional DD approach helps to find a slightly better tradeoff of noise suppression against speech distortion.

5. Conclusions

In this contribution we described a statistical modeling of the *a priori* SNR estimation task in the generalized SNR domain defined for an arbitrary power exponent of the spectral magnitude. We developed a generalized decision directed approach, took deeper insight in its behaviour and parameterized it for usage in arbitrary noisy conditions. In the experiments we showed, first, that the proposed approach is able to improve the performance of the decision directed approach in high input global SNR and, second, that the power spectral domain is an optimal choice for the *a priori* SNR estimation for low input global SNR.

6. References

[1] O. Cappe, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 2, pp. 345–349, Apr. 1994.

[2] 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, July 2001.

[3] A. Chinaev and R. Haeb-Umbach, "On optimal smoothing in minimum statistics based noise tracking," in *16th Annual Interspeech Conference of the International Speech Communication Association (ISCA)*, Sept. 2015, pp. 1785–1789.

[4] A. Chinaev, R. Haeb-Umbach, J. Taghia, and R. Martin, "Improved single-channel nonstationary noise tracking by an optimized MAP-based postprocessor," in *38th International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013, pp. 7477–7481.

[5] 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, no. 6, pp. 1109–1121, Dec. 1984.

[6] 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, no. 6, pp. 2098–2108, Nov. 2006.

[7] I. Y. Soon and S. N. Koh, "Low distortion speech enhancement," *IEE Proceedings - Vision, Image and Signal Processing*, vol. 147, no. 3, pp. 247–253, June 2000.

[8] 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, no. 4, pp. 450–453, Apr. 2004.

[9] I. Cohen, "Speech enhancement using a noncausal a priori SNR estimator," *IEEE Signal Processing Letters*, vol. 11, no. 9, pp. 725–728, Sept. 2004.

[10] ——, "Relaxed statistical model for speech enhancement and a priori SNR estimation," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 870–881, Sept. 2005.

[11] Y. Park and J. Chang, "A novel approach to a robust a priori SNR estimator in speech enhancement," *IEICE Transactions on Communications*, vol. 90, no. 8, pp. 2182–2185, Aug. 2007.

[12] 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, no. 1, pp. 186–195, Jan. 2011.

[13] 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, no. 2, pp. 358 – 376, Sept. 2013.

[14] 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.

[15] H. Sun, S. Ou, R. Liu, and Y. Gao, "A variable momentum factor algorithm for a priori SNR estimation in speech enhancement," in *7th International Congress on Image and Signal Processing (CISP)*, Oct. 2014, pp. 888–892.

[16] S. Elshamy, N. Madhu, W. Tirry, and T. Fingscheidt, "An iterative speech model-based a priori snr estimator," in *16th Annual Interspeech Conference of the International Speech Communication Association (ISCA)*, Sept. 2015, pp. 1740–1744.

[17] 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 *18th European Signal Processing Conference*, Aug. 2010, pp. 994–998.

[18] S. Voran, "Exploration of the additivity approximation for spectral magnitudes," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct. 2015, pp. 1–5.

[19] J. Lim, "Evaluation of a correlation subtraction method for enhancing speech degraded by additive white noise," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 26, no. 5, pp. 471–472, Oct. 1978.

[20] J. Lim and A. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proceedings of the IEEE*, vol. 67, no. 12, pp. 1586–1604, Dec. 1979.

[21] B. L. Sim, Y. C. Tong, J. S. Chang, and C. T. Tan, "A parametric formulation of the generalized spectral subtraction method," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 4, pp. 328–337, July 1998.

[22] R. McAulay and M. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 28, no. 2, pp. 137–145, Apr. 1980.

[23] P. Scalart and J. Filho, "Speech enhancement based on a priori signal to noise estimation," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2, May 1996, pp. 629–632.

[24] I. Cohen, "On speech enhancement under signal presence uncertainty," *In Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1, pp. 661–664, May 2001.

[25] J. Yang, "Frequency domain noise suppression approaches in mobile telephone systems," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2, Apr. 1993, pp. 363–366.

[26] "TIMIT, Acoustic-Phonetic Continuous Speech Corpus," *DARPA, NIST Speech Disc 1-1.1*, Oct. 1990.

[27] 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, no. 3, pp. 247–251, July 1993.