

# A GENERALIZED LOG-SPECTRAL AMPLITUDE ESTIMATOR FOR SINGLE-CHANNEL SPEECH ENHANCEMENT

*Aleksej Chinaev, Reinhold Haeb-Umbach*

Department of Communications Engineering, Paderborn University, 33098 Paderborn, Germany

{chinaev,haeb}@nt.uni-paderborn.de

## ABSTRACT

The benefits of both a logarithmic spectral amplitude (LSA) estimation and a modeling in a generalized spectral domain (where short-time amplitudes are raised to a generalized power exponent, not restricted to magnitude or power spectrum) are combined in this contribution to achieve a better tradeoff between speech quality and noise suppression in single-channel speech enhancement. A novel gain function is derived to enhance the logarithmic generalized spectral amplitudes of noisy speech. Experiments on the CHiME-3 dataset show that it outperforms the famous minimum mean squared error (MMSE) LSA gain function of Ephraim and Malah in terms of noise suppression by 1.4 dB, while the good speech quality of the MMSE-LSA estimator is maintained.

**Index Terms**— single-channel spectral speech enhancement, generalized statistical-model based algorithms

## 1. INTRODUCTION

Despite the recent success of neural networks for speech enhancement, there is still an interest in parametric single-channel spectral speech enhancement algorithms, since they tend to require less computational and memory resources than neural networks and since they do not need a training phase. Starting with the seminal paper [1] introducing the spectral subtraction algorithm for noise suppression of short-time spectral amplitudes of noisy speech signal much research has been devoted to finding an optimal tradeoff between high noise suppression and low speech distortion. One line of research was concerned with finding better spectral gain functions. Thus the MMSE-LSA estimator derived in [2] was shown to successfully reduce the musical noise phenomenon as reported in [3]. However a closer look at the shapes of the MMSE-LSA gain curves revealed that the price to pay for the good quality of the enhanced speech signals was a weaker noise suppression in regions with low speech energy [4].

Further it was proposed to carry out the enhancement in domains other than the magnitude or power spectral domain [4–10]. The generalized spectral subtraction (GSS) gain functions in [4] were derived, e.g., in the domain of the spectral amplitudes raised to a generalized power exponent  $\alpha \in \mathbb{R}_{>0}$

denoted further as generalized spectral amplitude (GSA) domain, where  $\alpha = 1$  and  $\alpha = 2$  correspond to the magnitude and the power spectral domain, respectively. According to [4] the constrained parametric MMSE-GSS estimator results in a respectable ability to suppress noise.

Recently we applied the spectral speech enhancement with a generalized power exponent for the a priori SNR estimation and discovered that this is beneficial for noise suppression without any loss in speech quality [11]. Motivated by this observation the goal of this contribution is to combine the advantages of the LSA estimation with those of GSA domain processing. Indeed, it will be shown that the logarithmic GSA (LGSA) gain function derived in this paper achieves high noise suppression and good speech quality at the same time, thus improving the noise suppression of the MMSE-LSA method while maintaining its good speech quality, which is better than that of the MMSE-GSS estimator.

In the next section we introduce a statistical modeling in the GSA domain, and derive a maximum a posteriori (MAP) LGSA estimator of the spectral amplitude of the clean speech. In Section 3 we introduce an additional parameter to achieve more modeling freedom and thus a better approximation to the true distributions. A parameterization of the proposed gain function and experimental results are presented in Sections 4 and 5, while Section 6 offers some conclusions.

## 2. DERIVATION

Observing a speech signal distorted by an additive uncorrelated noise results in the short-time Fourier transform (STFT) coefficients  $Y(k, \ell)$  of the noisy signal according to

$$Y(k, \ell) = S(k, \ell) + D(k, \ell), \quad (1)$$

where  $S(k, \ell)$  and  $D(k, \ell)$  are the STFT coefficients of the clean speech and noise signal, respectively, with a frequency bin index  $k$  and a frame index  $\ell$ . Motivated by the Central Limit Theorem, the STFT coefficients  $S(k, \ell)$  and  $D(k, \ell)$  are modelled as non-stationary complex-valued zero-mean Gaussian random processes with power spectral densities  $\lambda_S(k, \ell) = \mathbb{E}[|S(k, \ell)|^2]$  and  $\lambda_D(k, \ell) = \mathbb{E}[|D(k, \ell)|^2]$ , where  $\mathbb{E}[\cdot]$  denotes the expectation operator [12, 13].

## 2.1. Generalized spectral amplitudes

The notion of GSA domain refers to raising the involved spectral amplitudes to the power of an arbitrary constant  $\alpha \in \mathbb{R}_{>0}$ :

$$X_\alpha(k, \ell) = |X(k, \ell)|^\alpha \text{ for } X \in \{Y, S, D\}. \quad (2)$$

In consideration of (1) and under the made statistical assumptions, the GSAs of the involved processes are non-stationary real-valued Weibull-distributed random processes with probability density function (PDF)

$$p_{X_\alpha(k, \ell)}(x) = \text{Weib}(x; \lambda_X(k, \ell), \alpha) \quad (3)$$

where the Weibull PDF introduced in [14] is defined here as

$$\text{Weib}(x; \lambda_X, \alpha) \triangleq \frac{2}{\alpha \lambda_X} \cdot x^{\frac{2}{\alpha}-1} \cdot \exp\left(-\frac{x^{\frac{2}{\alpha}}}{\lambda_X}\right) \cdot \epsilon(x), \quad (4)$$

and where  $\lambda_X \in \mathbb{R}_{>0}$ ,  $\alpha$  and  $\epsilon(x)$  are a scale parameter, a shape parameter and the unit step function, respectively. The raw moment of  $\kappa$ -th order is given by

$$\mathbb{E}[X_\alpha^\kappa] = \Gamma\left(\kappa \cdot \frac{\alpha}{2} + 1\right) \cdot \lambda_X^{\kappa \cdot \frac{\alpha}{2}}, \quad (5)$$

where  $\Gamma(x)$  is the gamma function. Note that the Weibull PDF simplifies to the Rayleigh distribution for  $\alpha = 1$  and to the exponential distribution for  $\alpha = 2$ . The additivity of Eq. (1) results in  $\lambda_Y(k, \ell) = \lambda_S(k, \ell) + \lambda_D(k, \ell)$ .

## 2.2. Approximation by consistent Gaussian

As a computationally efficient LGSA estimator is pursued as our ultimate goal, that is analytically intractable for Weibull-distributed GSAs, we suggest to approximate the Weibull PDFs of involved GSAs by a Gaussian distribution

$$p_{X_\alpha(k, \ell)}(x) = \text{Weib}(x; \lambda_X, \alpha) \approx \mathcal{N}(x; \mu_X, \sigma_X^2) \quad (6)$$

using moment matching for mean and variance, resulting in

$$\mu_X \triangleq \mathbb{E}[X_\alpha] = \Gamma\left(\frac{\alpha}{2} + 1\right) \cdot \lambda_X^{\frac{\alpha}{2}} \quad (7)$$

$$\sigma_X^2 = \underbrace{\left[\Gamma(\alpha + 1) - \Gamma^2\left(\frac{\alpha}{2} + 1\right)\right]}_{c_\alpha} \cdot \lambda_X^\alpha = \frac{c_\alpha \cdot \mu_X^2}{\Gamma^2\left(\frac{\alpha}{2} + 1\right)}. \quad (8)$$

Admittedly the reasons for such an approximation are not obvious *prima facie*. But taking a closer look at the Weibull PDF reveals, that at least for a certain range of  $\alpha \in (0.5; 0.6)$ , where the skewness of the Weibull PDF is around zero, such an approximation is indeed well justified.

Note that the Gaussian distribution introduced in (6) exhibits some specific properties. First, the mean from (7) has to be positive  $\mu_X \in \mathbb{R}_{>0}$ , and second,  $\mu_X$  and  $\sigma_X^2$  are not two independent parameters, since they are connected via (8) with  $c_\alpha > 0$ . As a consequence larger values of  $\mu_X$  are accompanied with larger values of  $\sigma_X^2$ . Normal distributions with this linkage between mean and variance are sometimes referred to as consistent Gaussian distributions [15].

## 2.3. MMSE estimator of GSA

Before deriving a desired LGSA estimator let us first consider a MMSE-GSA estimator denoted further as a GSS estimator as named by its developers in [4]. Based on the introduced approximation by the consistent normal distributions and assuming similar to [4, 9] the additivity

$$Y_\alpha(k, \ell) = S_\alpha(k, \ell) + D_\alpha(k, \ell) \quad (9)$$

a MMSE-GSA estimator can be derived in contrast to [4] via the following conditional expectation

$$\hat{S}_\alpha^{\text{GSS}}(k, \ell) = \mathbb{E}[S_\alpha | Y_\alpha] = G_\alpha^{\text{GSS}}(k, \ell) \cdot Y_\alpha(k, \ell). \quad (10)$$

Since all involved GSAs are approximated by Gaussian distributions, one can easily obtain  $\mathbb{E}[S_\alpha | Y_\alpha]$  using the moments given in eqs. (7) and (8) resulting in the GSS gain function

$$G_\alpha^{\text{GSS}}(k, \ell) = \frac{\xi^\alpha}{\xi^\alpha + 1} \left(1 - \frac{\Gamma\left(\frac{\alpha}{2} + 1\right)}{\gamma^{\frac{\alpha}{2}}} (1 - \xi^{-\frac{\alpha}{2}})\right), \quad (11)$$

where  $\xi \rightarrow \xi(k, \ell)$  and  $\gamma \rightarrow \gamma(k, \ell)$  are the *a priori* SNR and the *a posteriori* SNR, respectively, defined as in [12]

$$\gamma(k, \ell) \triangleq \frac{|Y(k, \ell)|^2}{\lambda_D(k, \ell)}, \quad (12) \quad \xi(k, \ell) \triangleq \frac{\lambda_S(k, \ell)}{\lambda_D(k, \ell)}. \quad (13)$$

Note, that (10) describes the denoising of the generalized  $\alpha$ -order spectral magnitudes of the noisy signal by a gain function  $G_\alpha^{\text{GSS}}(k, \ell)$ , which depends on the parameter  $\alpha$ . As we discovered, the gain function from (11) rewritten as  $G_{\text{GSS}}(k, \ell) = [G_\alpha^{\text{GSS}}(k, \ell)]^{\frac{1}{\alpha}}$  to be applied to the noisy spectral amplitudes was already derived in [4], however using another problem formulation. There, a parametric GSS estimator defined as  $\hat{S}_\alpha(k, \ell) = a \cdot Y_\alpha(k, \ell) - b \cdot \mathbb{E}[D_\alpha(k, \ell)]$  was derived by minimizing the mean squared error cost function  $\mathbb{E}\{[S_\alpha(k, \ell) - \hat{S}_\alpha(k, \ell)]^2\}$  w.r.t. the parameters  $a$  and  $b$ . This accordance provides another justification for the approximation (6) leading to the Gaussian conditional PDF

$$p_{S_\alpha|Y_\alpha}(s|y) = \mathcal{N}(s; \mu_{S|Y}, \sigma_{S|Y}^2), \quad (14)$$

$$\mu_{S|Y} = G_\alpha^{\text{GSS}} \cdot Y_\alpha, \quad (15) \quad \sigma_{S|Y}^2 = \frac{c_\alpha}{\gamma^\alpha} \cdot \frac{\xi^\alpha}{\xi^\alpha + 1} \cdot Y_\alpha^2. \quad (16)$$

Note, that in contrast to [4] the approximation (6) allows us to get a closed form conditional PDF (14), which now can be used to derive a desired estimator of LGSA.

## 2.4. MAP estimator of logarithmic GSA

To derive an estimator of LGSA denoted as  $Z_\alpha = \ln S_\alpha$ ,  $p_{S_\alpha|Y_\alpha}(s|y)$  from (14) has to be modified in a way that it is defined only for  $s > 0$ , as a prerequisite for going to the logarithmic domain. Since all realizations of  $S_\alpha$  are positive according to the definition (2) we suggest to approximate (14)

by a normal distribution truncated at  $s = 0$  while maintaining the mode of distribution, resulting in

$$\tilde{p}_{S_\alpha|Y_\alpha}(s|y) = \frac{\epsilon(s)}{Q\left(-\frac{\mu_{S|Y}}{\sigma_{S|Y}}\right)} \cdot \mathcal{N}(s; \mu_{S|Y}, \sigma_{S|Y}^2), \quad (17)$$

where  $Q(x)$  is the complementary cumulative distribution of the standard normal density. A change of variable leads to the following PDF of the LGSA

$$p_{Z_\alpha|Y_\alpha}(z|y) = \frac{e^z \cdot \mathcal{N}(e^z; \mu_{S|Y}, \sigma_{S|Y}^2)}{Q\left(-\frac{\mu_{S|Y}}{\sigma_{S|Y}}\right)} \propto e^{\frac{1}{2} \cdot f(z)} \quad (18)$$

with  $f(z) = 2 \cdot z - (e^z - \mu_{S|Y})^2 / \sigma_{S|Y}^2$ . Since the derivation of the MMSE-LGSA estimator  $\hat{S}_\alpha = \exp(\mathbb{E}[Z_\alpha|Y_\alpha])$  is analytically intractable, we suggest to employ the maximum a posteriori based LGSA estimator defined as

$$\hat{S}_\alpha^{\text{LGSA}} = \exp\left(\arg \max_z p_{Z_\alpha|Y_\alpha}(z|y)\right). \quad (19)$$

Finding a maximum of  $p_{Z_\alpha|Y_\alpha}(z|y)$  and using it in (19) results in the desired simple MAP-based LGSA estimator

$$\hat{S}_\alpha^{\text{LGSA}} = \frac{\mu_{S|Y}}{2} + \sqrt{\left(\frac{\mu_{S|Y}}{2}\right)^2 + \sigma_{S|Y}^2}. \quad (20)$$

Using (15) and (16) in (20) provides the resulting MAP-LGSA gain function  $G_{\text{LGSA}}(k, \ell) = [G_\alpha^{\text{LGSA}}(k, \ell)]^{\frac{1}{\alpha}}$  with

$$G_\alpha^{\text{LGSA}}(k, \ell) = \frac{G_\alpha^{\text{GSS}}}{2} + \sqrt{\left(\frac{G_\alpha^{\text{GSS}}}{2}\right)^2 + \frac{c_\alpha}{\gamma^\alpha} \cdot \frac{\xi^\alpha}{\xi^\alpha + 1}}. \quad (21)$$

Note,  $G_\alpha^{\text{LGSA}}(k, \ell) > G_\alpha^{\text{GSS}}(k, \ell)$  holds always for a given  $\alpha$ . To our knowledge  $G_{\text{LGSA}}(k, \ell)$  is a first gain function in logarithmic domain among the MAP-based gain functions [16].

### 3. ADDITIONAL MODELING FREEDOM

Following George E. P. Box' statement that 'Essentially, all models are wrong, but some are useful' [17], we suggest a mechanism to increase the flexibility of our modeling similar to [18] and allow the models in (3) to have a shape parameter  $\beta$  different from the used power exponent  $\alpha$  as follows

$$p_{X_\alpha(k, \ell)}(x) = \text{Weib}(x; \lambda_X(k, \ell), \beta). \quad (22)$$

Thus we model spectral amplitudes raised to the power  $\alpha$  with a Weibull PDF by using a shape parameter  $\beta$  not necessarily equal to  $\alpha$ . Such modeling causes in (11) and (21) a substitution of  $\alpha$  by  $\beta$ . With this additional parameter  $\beta$  it is possible to better approximate the true distributions of GSAs and to increase usefulness of introduced statistical models.

Thus, in contrast to (10) we suggest to denote the noisy GSAs  $Y_\alpha(k, \ell)$  by the gain functions  $G_\beta^{\text{GSS}}(k, \ell)$  or  $G_\beta^{\text{LGSA}}(k, \ell)$ , which are dependent on  $\beta$  resulting in

$$G_{\text{EST}}(k, \ell) = [G_\beta^{\text{EST}}]^{\frac{1}{\alpha}} \text{ for } \text{EST} \in \{\text{GSS}, \text{LGSA}\}. \quad (23)$$

## 4. PARAMETERIZATION

In order to use the gain functions (23) for speech enhancement a power exponent  $\alpha$  and a shape parameter  $\beta$  have to be set appropriately. For this some experiments are conducted with speech signals distorted by white noise. Clean speech signals for male and female speakers are taken from the TIMIT database [19] and are concatenated to a total length of 1 minutes each. These are distorted by a white noise signal taken from the signal processing information base (SPIB) data [20] at global SNR values  $\text{SNR}_{\text{IN}} \in \{-5, 0, 5, 10, 15\}$  dB. To obtain a frequency representation of signals sampled at 16 kHz, a STFT transformation with a Hamming analysis window of 512 samples length with a shift factor of 0.25 is used. As a noise power spectral density estimator  $\hat{\lambda}_D(k, \ell)$ , the minimum statistics (MS) approach is applied with a length of the MS window for minimum search of 96 frames divided into  $U_{\text{MS}} = 8$  sub-windows of length of  $V_{\text{MS}} = 12$  frames [21]. A minimum value of the MS smoothing parameter is set to a constant value  $\alpha_{\text{MS}, \text{min}} = 0.3$  For the *a priori* SNR estimation the decision-directed (DD) approach [12] is applied with a weighting factor of 0.975 and a minimum *a priori* SNR of  $\xi_{\text{min}} = -25$  dB [22]. The gain functions are delimited by an upper bound of 1 and a lower bound of  $G_{\text{min}} = -25$  dB [23]. Values of  $\alpha$  and  $\beta$  are varied in the ranges of  $[0.01; 3]$  and  $[0.2; 5]$ , respectively. As an objective performance measure, the wide-band mean opinion score - listening quality objective (MOS-LQO) measure is used [24]. Note, higher MOS-LQO values indicate better performance.

Fig. 1 shows the resulting MOS-LQO values averaged over signals of male and female speakers at  $\text{SNR}_{\text{IN}} = 5$  dB for the GSS and LGSA gain functions entitled by the values 'MOS-LQO<sub>max</sub>( $\alpha_{\text{opt}}, \beta_{\text{opt}}$ )', where the parameters ( $\alpha_{\text{opt}}, \beta_{\text{opt}}$ ) depicted by big black points maximize the MOS-LQO scores. The experiments show, that both gain functions achieve similar MOS-LQO<sub>max</sub> values but for different optima ( $\alpha_{\text{opt}}, \beta_{\text{opt}}$ ). Further, the GSS gain function provides high MOS-LQO values for a larger range of ( $\alpha, \beta$ ) values than the LGSA gain

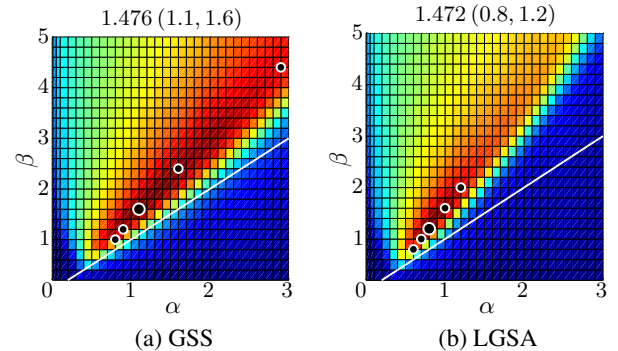


Fig. 1. Averaged MOS-LQO scores for white noise at 5 dB.

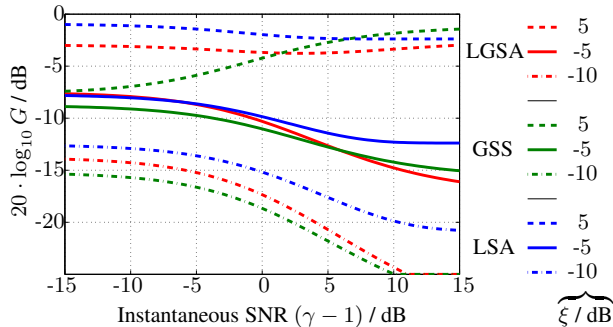


Fig. 2. The proposed LGSA, GSS and LSA gain functions.

function. However, the optimal values  $(\alpha_{\text{opt}}, \beta_{\text{opt}})$  for different  $\text{SNR}_{\text{IN}}$  values depicted by small black points scatter for the GSS more than for the LGSA. In general the points with smaller values of  $\alpha_{\text{opt}}$  and  $\beta_{\text{opt}}$  correspond to higher  $\text{SNR}_{\text{IN}}$  values and vice versa. None of the points  $(\alpha_{\text{opt}}, \beta_{\text{opt}})$  lies on a constraint  $\beta = \alpha$  depicted by a white line justifying usefulness of the additional modeling freedom proposed in Section 3. It is preferable to choose the shape parameter  $\beta_{\text{opt}}$  of the Weibull distribution higher than the power exponent  $\alpha_{\text{opt}}$  of the GSAs, which means that distributions with higher kurtosis are favoured as for  $\beta = \alpha$ .

The curves of the gain functions from (23) for  $(\alpha_{\text{opt}}, \beta_{\text{opt}})$  at  $\text{SNR}_{\text{in}} = 5$  dB are depicted in the Fig. 2 over the instantaneous SNR  $\gamma - 1$  at *a priori* SNR  $\xi \in \{-10, -5, 5\}$  dB together with the curves of the MMSE-LSA gain from [2] denoted by LSA. A desired ability of the LSA gain concerning reducing the musical tones is decreasing of its curves with increasing  $\gamma$  values [3, 4]. In contrast to the GSS gain, the proposed LGSA gain approximately maintains this desired behavior even for the higher region of  $\xi$  values (e.g., for  $\xi = 5$  dB). As the gain curves of the LSA gain for  $\xi < -5$  dB show, the price to pay for good speech signal quality is a poor noise suppression. On the contrary, using  $\beta \neq \alpha$  in (23) causes a higher noise attenuation for both generalized gain functions.

## 5. EXPERIMENTAL RESULTS

In order to evaluate the performance of the gain functions, we carried out single-channel speech enhancement experiments on the development dataset of the third *computational hearing in multisource environments* (CHiME-3) challenge [25], where signals are sampled at 16 kHz and represent in total about 2.88 hours of audio data. The simulated isolated data consist of 410 utterances in every of 4 different noise environments: on the bus (bus), in a cafe (caf), in a pedestrian area (ped) and on a street junction (str). We used recordings of the 5<sup>th</sup> tablet microphone with an averaged global input SNR of  $\text{SNR}_{\text{in}} \approx 5.8$  dB and denoised them by the same enhancement system as in the experiments with white noise. In the gener-

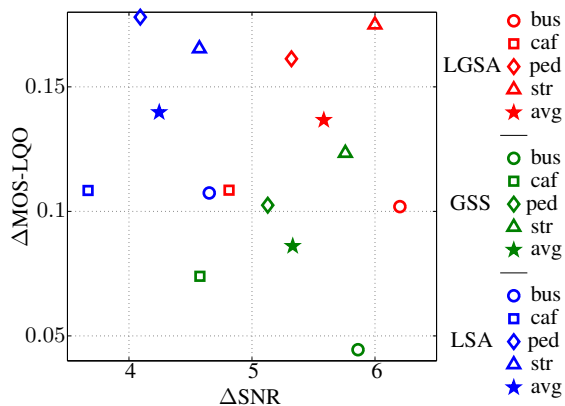


Fig. 3. Average improvement in terms of  $\Delta\text{MOS-LQO}$  and  $\Delta\text{SNR}$  for development set of the CHiME-3 database [25].

alized gain functions the fixed parameters  $(\alpha_{\text{opt}}, \beta_{\text{opt}})$  given in Fig. 1 are used with resulting gain curves as in Fig. 2. Beside the speech quality improvement measured in terms of  $\Delta\text{MOS-LQO} = \text{MOS-LQO}_{\text{out}} - \text{MOS-LQO}_{\text{in}}$  calculated for every enhanced output and noisy input signal, we evaluated the increase in global SNR measured on the output of the system regarding to its input via  $\Delta\text{SNR} = \text{SNR}_{\text{out}} - \text{SNR}_{\text{in}}$  to show the ability of the estimators to suppress noise.

The resulting  $\Delta\text{MOS-LQO}$  values averaged over all utterances of a certain noise type are depicted in Fig. 3 over the averaged  $\Delta\text{SNR}$  values for the LGSA, GSS and LSA estimators. Additionally,  $\Delta\text{MOS-LQO}$  and  $\Delta\text{SNR}$  values averaged (avg) over all noise types are pointed out. As expected the LSA estimator delivers enhanced signals with a good speech quality but poor noise suppression. On the contrary the GSS estimator achieves good noise suppression, however at the cost of poorer signal quality. Amazingly, the proposed LGSA gain function almost achieves the speech quality of the LSA estimator and at the same time outperforms the GSS estimator in terms of noise suppression. Compared to the LSA, the proposed LGSA estimator improves noise suppression by approximately 1.4 dB on average (from 4.2 dB to 5.6 dB) almost without loss in speech quality. Thus, the proposed LGSA gain function provides a better tradeoff between speech quality and noise suppression than both other estimators.

## 6. CONCLUSIONS

A novel short-time spectral gain function is derived in this work in the domain of logarithmic generalized spectral amplitudes. Using the MAP criterion here leads to a computationally efficient estimator which achieves a better tradeoff between speech quality and noise suppression compared to the famous MMSE-LSA estimator from [2] and to the MMSE-GSS estimator proposed in [4]. The achieved improvement comes at virtually no increased computational cost.

## 7. REFERENCES

- [1] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. on Acoustics, Speech and Signal Processing (ASSP)*, vol. 27, no. 2, pp. 113–120, Apr. 1979.
- [2] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator," *IEEE Trans. on ASSP*, vol. 33, no. 2, pp. 443–445, Apr. 1985.
- [3] O. Cappe, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," *IEEE Trans. on Speech and Audio Processing (SAP)*, vol. 2, no. 2, pp. 345–349, Apr. 1994.
- [4] B.L. Sim, Y.C. Tong, J.S. Chang, and C.T. Tan, "A parametric formulation of the generalized spectral subtraction method," *IEEE Trans. on SAP*, vol. 6, no. 4, pp. 328–337, July 1998.
- [5] J.S. Lim and A.V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. of the IEEE*, vol. 67, no. 12, pp. 1586–1604, Dec. 1979.
- [6] C.H. You, S.N. Koh, and S. Rahardja, " $\beta$ -order MMSE spectral amplitude estimation for speech enhancement," *IEEE Trans. on SAP*, vol. 13, no. 4, pp. 475–486, July 2005.
- [7] J. Li, S. Sakamoto, S. Hongo, M. Akagi, and Y. Suzuki, "Adaptive  $\beta$ -order generalized spectral subtraction for speech enhancement," *Signal Processing*, vol. 88, no. 11, pp. 2764–2776, June 2008.
- [8] T. Inoue, H. Saruwatari, Y. Takahashi, K. Shikano, and K. Kondo, "Theoretical Analysis of Musical Noise in Generalized Spectral Subtraction Based on Higher Order Statistics," *IEEE Trans. on ASLP*, vol. 19, no. 6, pp. 1770–1779, Aug. 2011.
- [9] S. Voran, "Exploration of the additivity approximation for spectral magnitudes," in *IEEE Workshop on ASPAA*, Oct. 2015, pp. 1–5.
- [10] Y. Tsao and Y. Lai, "Generalized maximum a posteriori spectral amplitude estimation for speech enhancement," *Speech Comm.*, vol. 76, pp. 112–126, Feb. 2016.
- [11] A. Chinaev and R. Haeb-Umbach, "A Priori SNR Estimation Using a Generalized Decision Directed Approach," in *17-th Annual INTERSPEECH Conf. of the ISCA*, Sept. 2016, pp. 3758–3762.
- [12] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator," *IEEE Trans. on ASSP*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.
- [13] D. Brillinger, *Time series: data analysis and theory*, vol. 36, Siam, 2001.
- [14] W. Weibull, "A statistical distribution function of wide applicability," *Journal of Applied Mechanics*, vol. 18, pp. 293–297, Sept. 1951.
- [15] T. Richardson, A. Shokrollahi, and R. Urbanke, "Design of provably good low-density parity check codes," in *IEEE Int'l Symp. on Information Theory*, June 2000, p. 199.
- [16] Y. Lu and P. C. Loizou, "Estimators of the Magnitude-Squared Spectrum and Methods for Incorporating SNR Uncertainty," *IEEE Trans. on Audio, Speech, and Language Processing (ASLP)*, vol. 19, no. 5, pp. 1123–1137, July 2011.
- [17] George EP Box, "Robustness in the strategy of scientific model building," *Robustness in statistics*, vol. 1, pp. 201–236, May 1979.
- [18] A. Chinaev, J. Heitkaemper, and R. Haeb-Umbach, "A Priori SNR Estimation Using Weibull Mixture Model," in *12th ITG Symposium on Speech Communication*, Oct. 2016, pp. 297–301.
- [19] "TIMIT, Acoustic-Phonetic Continuous Speech Corpus," *DARPA, NIST Speech Disc 1-1.1*, Oct. 1990.
- [20] D. Johnson and P. N. Shami, "The signal processing information base," in *IEEE Signal Processing Magazine*, Oct. 1993, vol. 10, pp. 36–43.
- [21] R. Martin, "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics," *IEEE Trans. on SAP*, vol. 9, no. 5, pp. 504–512, July 2001.
- [22] I. Cohen, "Speech enhancement using a noncausal a priori SNR estimator," *IEEE Signal Processing Letters*, vol. 11, no. 9, pp. 725–728, Sept. 2004.
- [23] I. Cohen, "On speech enhancement under signal presence uncertainty," in *Proc. IEEE Int'l Conf. on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 661–664, May 2001.
- [24] "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.
- [25] 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*, Dec. 2015, pp. 504–511.