Parchami *et al. EURASIP Journal on Advances in Signal Processing* (2015) 2015:87
DOI 10.1186/s13634-015-0270-6

EURASIP Journal on Advances in Signal Processing
a SpringerOpen Journal

RESEARCH Open Access

Bayesian STSA estimation using masking properties and generalized Gamma prior for speech enhancement

Mahdi Parchami[1]*, Wei-Ping Zhu[1], Benoit Champagne[2] and Eric Plourde[3]

Abstract

We consider the estimation of the speech short-time spectral amplitude (STSA) using a parametric Bayesian cost function and speech prior distribution. First, new schemes are proposed for the estimation of the cost function parameters, using an initial estimate of the speech STSA along with the noise masking feature of the human auditory system. This information is further employed to derive a new technique for the gain flooring of the STSA estimator. Next, to achieve better compliance with the noisy speech in the estimator's gain function, we take advantage of the generalized Gamma distribution in order to model the STSA prior and propose an SNR-based scheme for the estimation of its corresponding parameters. It is shown that in Bayesian STSA estimators, the exploitation of a rough STSA estimate in the parameter selection for the cost function and the speech prior leads to more efficient control on the gain function values. Performance evaluation in different noisy scenarios demonstrates the superiority of the proposed methods over the existing parametric STSA estimators in terms of the achieved noise reduction and introduced speech distortion.

Keywords: Generalized Gamma distribution (GGD); Masking; Noise reduction; Short-time spectral amplitude (STSA); Speech enhancement

1 Introduction

Speech enhancement aims at the reduction of corrupting noise in speech signals while keeping the introduced speech distortion at the minimum possible level. In this respect, considerable interest has been directed toward the estimation of the speech spectral amplitude, due to its perceptual importance in the frequency domain approaches [1, 2].

Within this framework, the general goal is to provide an estimate of the short-time spectral amplitude (STSA) of the clean speech using statistical models for the noise and speech spectral components. In [3], Ephraim and Malah proposed to estimate the speech signal amplitude through the minimization of a Bayesian cost function which measures the mean square error between the clean and estimated STSA; accordingly, the resulting estimator was called the minimum mean square error (MMSE) spectral amplitude estimator. Later in [4], a logarithmic version of the proposed estimator, i.e., the Log-MMSE, was introduced by considering that the logarithm of the STSA is perceptually more relevant to the human auditory system. Even though alternatives to the Bayesian STSA estimators were proposed, e.g., in [5], due to the satisfying performance of the latter, they are still found to be appealing in the literature. More recently, further modifications to the STSA Bayesian cost functions were suggested by Loizou in [6] by taking advantage of the psycho-acoustical models initially employed for speech enhancement purposes in [7, 8]. Therein, it was shown that the estimator emphasizing the spectral valleys (minima) of the speech STSA, namely the weighted Euclidean (WE) estimator, achieves the best overall performance. Along the same line of thought, You et al. [9] proposed to use the β power of the STSA term in the Bayesian cost function, in order to obtain further flexibility in the corresponding STSA gain function. These authors investigated the performance of

*Correspondence: m_parch@ece.concordia.ca
The first two authors contributed equally.
The third and fourth authors contributed equally, as well.
[1] Department of Electrical and Computer Engineering, Concordia University, 1455 De Maisonneuve Blvd. West, H3G 1M8 Montreal, Canada
Full list of author information is available at the end of the article

the so-called β-order MMSE estimator for different values of β and found that it is moderately better than the MMSE and Log-MMSE estimators proposed earlier. In that work, an adaptive scheme based on the frame SNR was also suggested to determine β.

Plourde and Champagne in [10] suggested to take advantage of STSA power weightings (as used in the WE estimator) in the β-order MMSE cost function and introduced the parameter α as the power of their new weighting. They further proposed to select the two estimator parameters as functions of frequency, according to the psycho-acoustical properties of the human auditory system and showed a better quality in the enhanced speech in most of the input SNR range. Yet, at high input SNRs, the performance of the developed estimator may not be appealing due to the undesired distortion in the enhanced speech. Further in [11], the same authors introduced a generalized version of the Wβ-SA estimator by including a new weighting term in the Bayesian cost function which provides additional flexibility in the estimator's gain. However, apart from the mathematically tedious solution for the gain function, the corresponding estimator does not provide further noticeable improvement in the enhanced speech quality.

Overall, the parametric Bayesian cost functions as those in [6, 9, 10] can provide further noise reduction over the previous estimators, thanks to the additional gain control obtained by the appropriate choice of the cost function parameters. In [6], fixed values were used for the STSA weighting parameter, whereas in [9], an experimental scheme was proposed in order to adapt β to the estimated frame SNR. In the latter, the adaptive selection of the cost function parameters has been proved to be advantageous over fixed parameter settings in most of the tested scenarios. To make use of the noise masking properties as in [8], it was suggested in [12] to select the power β as a linear combination of both the frame SNR and the noise masking threshold; subsequently, improvements with respect to the previous schemes were reported. In [10], rather than an adaptive scheme, the values of the estimator parameters are chosen only based on the perceptual properties of human auditory system. Whereas this scheme is in accordance with the spectral psycho-acoustical models of the hearing system in neural science [13], it does not take into account the noisy speech features in updating the parameters.

In the aforementioned works, since the complex Gaussian probability distribution function (PDF) is considered for the speech short-time Fourier transform (STFT) coefficients, the speech STSA actually takes the Rayleigh PDF. However, as it was indicated in [14], parametric non-Gaussian (super-Gaussian) PDFs are able to better model the speech STSA prior. In [15], the Chi PDF with fixed parameter settings was used as the speech STSA prior for a group of perceptually motivated STSA estimators. Use of Chi and Gamma speech priors was further studied in [16] and training-based procedures using the histograms of clean speech data were proposed for the estimation of the speech STSA prior parameters. Yet, apart from being computationally tedious, training-based methods depend largely on the test data, and unless a very lengthy set of training data is used, their performance may not be reliable. Within the same line of work, the generalized Gamma distribution (GGD) has also been taken into account, which includes some other non-Gaussian PDFs as a special case. In [17, 18], it was confirmed that the most suitable PDF for the modeling of speech STSA priors is the GGD, given that the corresponding parameters are estimated properly. Two mathematical approaches, i.e., the maximum likelihood and the method-of-moments, have been used in [18] for the estimation of the GGD parameters. However, as the evaluations showed in [19] and our experiments proved, these two approaches do not lead to acceptable results, due to the coarse approximations involved in their derivation. Other major studies within this field such as those in [20, 21], use either fixed or experimentally set values for the GGD model parameters, lacking the adaptation with the noisy speech data. Hence, an adaptive scheme to estimate the STSA prior parameters with moderate computational burden and fast adaptability with the noisy speech samples is further needed.

In this work, by taking into account the parametric Wβ-SA estimator, we first propose novel schemes for the parameter selection of the cost function as well as the gain flooring. The new schemes make use of the prior information available through a preliminary estimate of the speech STSA, noise masking threshold, and the compression property of the human auditory system. Next, a generalization of this estimator by employing the GGD prior model is derived and an efficient yet low-complexity scheme is introduced for the estimation of its parameters. We assess the performance of the proposed methods in terms of speech quality and the amount of noise reduction and demonstrate their advantage with respect to the previous STSA estimators. In particular, through a series of controlled experiments, we demonstrate the incremental advantages brought about by each one of the newly proposed modifications to the original Wβ-SA estimator.

The remainder of this paper is organized as follows. In Section 2, a brief overview of the auditory-based Wβ-SA estimator is presented. Section 3 proposes new schemes for the parameter selection of the Bayesian cost function as well as a new gain flooring scheme for STSA estimators. Section 4 exploits the application of the GGD prior to the proposed STSA estimator and discusses an efficient method for the estimation of its parameters. Performance of the proposed STSA estimation schemes is evaluated in Section 5 in terms of objective performance measures. Conclusions are drawn in Section 6.

2 Background: parametric STSA estimation

In this section, a brief overview of a generic STSA estimation method, namely the Wβ-SA estimator, is presented. This estimator will be used as a basis for further developments in the next sections. Suppose that the noisy speech signal, $y(t)$, consists of a clean speech, $x(t)$, and the additive noise signal, $v(t)$, that is statistically independent of $x(t)$. After sampling and taking STFT with analysis window of length K, by denoting the frequency bin and time frame indices as $k \in \{0, 1, \ldots, K-1\}$ and $l \in \mathbb{N}$, respectively, it follows that

$$Y(k,l) = X(k,l) + V(k,l) \tag{1}$$

where $Y(k,l)$, $X(k,l)$, and $V(k,l)$ are the STFTs of the noisy observation, clean speech and noise, respectively. Expressing the complex-valued speech coefficients, $X(k,l)$, as $\chi(k,l)e^{j\Omega(k,l)}$ with χ and Ω as the amplitude and phase in respect, the purpose of speech STSA estimation is to estimate the speech amplitude, $\chi(k,l)$, given the noisy observations, $Y(k,l)$. The estimated amplitude will then be combined with the noisy phase of $Y(k,l)$ to provide an estimate of the speech Fourier coefficients. For sake of brevity, we may discard the indices k and l in the following.

The Bayesian STSA estimation problem can be formulated as the minimization of the expectation of a cost function that represents a measure of distance between the true and estimated speech STSAs, denoted respectively by χ and $\hat{\chi}$. This problem can be expressed as

$$\hat{\chi}^{(o)} = \underset{\hat{\chi}}{\operatorname{argmin}}\, E\left\{C(\chi, \hat{\chi})\right\} \tag{2}$$

where $C(.)$ is the Bayesian cost function, $E\{.\}$ denotes statistical expectation and $\hat{\chi}^{(o)}$ is the optimal speech STSA estimate in a Bayesian sense. Following a Bayesian framework, the expected value of the cost function in (2) can be written as [6]

$$\begin{aligned} E\left\{C(\chi,\hat{\chi})\right\} &= \int\int C(\chi,\hat{\chi})p\left(\chi,Y\right)d\chi d_{\mathrm{Y}} \\ &= \int\left[\int C(\chi,\hat{\chi})p\left(\chi|Y\right)d\chi\right]p\left(Y\right)d_{\mathrm{Y}} \end{aligned} \tag{3}$$

with $p\left(\chi, Y\right)$ and $p\left(\chi|Y\right)$ being the joint and conditional PDFs of the speech STSA and observation Y respectively. Note that in order to derive the optimum STSA estimator in (2), it suffices to minimize the inner integral in (3) with respect to $\hat{\chi}$. As discussed in Section 1, the weighted version of the β-SA, i.e., the Wβ-SA estimator, has been found to be advantageous with respect to the other Bayesian estimators. In fact, previously proposed Bayesian cost functions can be expressed as a special case of the underlying Wβ-SA cost function, which is defined as [10]

$$C(\chi,\hat{\chi}) = \chi^{\alpha}\left(\chi^{\beta} - \hat{\chi}^{\beta}\right)^2 \tag{4}$$

with α and β being the corresponding cost function parameters. Note that, for notational ease, compared to [10], a slight modification is done in (4) by replacing -2α in Eq. (10) in [10] by α. Substituting (4) into (3) and minimizing the expectation results in [10]

$$\hat{\chi}^{(W\beta-SA)} = \left(\frac{E\{\chi^{\beta+\alpha}|Y\}}{E\{\chi^{\alpha}|Y\}}\right)^{1/\beta} \tag{5}$$

The conditional moments of the form $E\{\chi^m|Y\}$ appearing in (5) can be obtained as

$$E\{\chi^m|Y\} = \frac{\int_0^{\infty}\int_0^{2\pi}\chi^m p(Y|\chi,\Omega)p(\chi,\Omega)d\Omega d\chi}{\int_0^{\infty}\int_0^{2\pi}p(Y|\chi,\Omega)p(\chi,\Omega)d\Omega d\chi} \tag{6}$$

with $p(Y|\chi,\Omega)$ and $p(\chi,\Omega)$ being respectively the conditional PDF of the noisy observation given the clean speech and the joint PDF for the speech amplitude and phase. In the relevant literature, the speech phase Ω is considered to be independent of the speech amplitude χ and also uniformly distributed over $[0,2\pi)$. Also, due to complex zero-mean Gaussian PDF assumed for the noise spectral coefficients, the conditional PDF $p(Y|\chi,\Omega)$ takes a Gaussian form. Thus, it follows that

$$\begin{aligned} p(Y|\chi,\Omega) &= \frac{1}{\pi\sigma_{\nu}^2}\exp\left(-\frac{\left|Y-\chi e^{j\Omega}\right|^2}{\sigma_{\nu}^2}\right) \\ p(\chi,\Omega) &= p(\chi)\,p(\Omega) = \frac{\chi}{\pi\sigma_{\chi}^2}\exp\left(-\frac{\chi^2}{\sigma_{\chi}^2}\right) \end{aligned} \tag{7}$$

where σ_{ν}^2 and σ_{χ}^2, respectively, denote the noise and speech spectral variances and a Rayleigh PDF has been assumed for the speech STSA PDF $p(\chi)$ as in [3]. By insertion of (7) into (6), the STSA moment $E\{\chi^m|Y\}$ is obtained and using this moment in (5) leads to the following gain function for the Wβ-SA estimator [10]

$$\begin{aligned} G^{(W\beta-SA)} &\triangleq \frac{\hat{\chi}^{(W\beta-SA)}}{|Y|} \\ &= \frac{\sqrt{\nu}}{\gamma}\left(\frac{\Gamma\left(\frac{\alpha+\beta}{2}+1\right)\mathrm{M}\left(-\frac{\alpha+\beta}{2},1;-\nu\right)}{\Gamma\left(\frac{\alpha}{2}+1\right)\mathrm{M}\left(-\frac{\alpha}{2},1;-\nu\right)}\right)^{1/\beta} \end{aligned} \tag{8}$$

where $\Gamma(.)$ and $\mathrm{M}(.,.;.)$ denote the Gamma and confluent hypergeometric functions [22], respectively, and the gain parameters γ and ν are defined as

$$\gamma = \frac{|Y|^2}{\sigma_{\nu}^2}, \nu = \frac{\zeta}{1+\zeta}\gamma, \zeta = \frac{\sigma_{\chi}^2}{\sigma_{\nu}^2} \tag{9}$$

where ζ and γ are called the *a priori* and *a posteriori* SNRs, respectively. Figure 1 shows theoretical gain curves of the estimator in (8) for different values of the parameters α and β. Herein, the fixed values $\zeta = 0$ dB and $\gamma = 0$ dB are considered to account for a highly noisy scenario. It is observed that the STSA gain function can be controlled by the selection of its two parameters and an increment in either of the two, especially α, would result in an increment in the gain function values. This realization will be used in the following sections to propose new schemes for the choice of these parameters.

3 Proposed noise masking-based STSA estimator

In this section, we propose a new parametric STSA estimator with a focus on its parameter selection and gain flooring using the noise masking property of the human auditory system. Figure 2 shows a block diagram of the proposed algorithm for the STSA estimation that is based on the noise masking threshold. As indicated, an initial estimate of the speech STSA is first obtained to calculate the noise masking threshold and the estimator parameters. This preliminary estimate can be obtained through a basic STSA estimator, e.g., the MMSE estimator in [3], as only a rough estimate of the speech STSA is needed at this step. As the experiments revealed, use of more accurate estimates of the speech STSA, either in the calculation of the noise masking threshold or in the parameters of the STSA estimator, do not result in any considerable improvements in the performance of the entire algorithm. Next, the STSA estimator parameters, α and β, are estimated using both the noise masking threshold and the available initial estimate of the speech STSA. These two parameters along with the noisy speech are fed into the STSA gain calculation block. Note that noise-related parameters, i.e., the noise spectral variance and the *a priori* SNR, should be estimated within this block in order to achieve the gain function value. This gain function is further thresholded and modified by the proposed gain flooring scheme. This modified gain is the ultimate form of the gain function being applied on the STSA of the noisy speech and leading to the enhanced STSA in the output. The enhanced STSA is to be combined with the phase of the noisy speech to generate the STFT of the enhanced speech. The following subsections describe the proposed block diagrams for the STSA estimation method in detail.

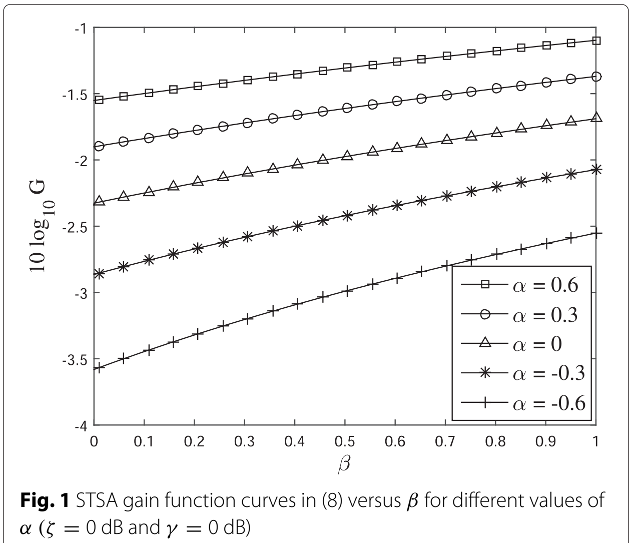

Fig. 1 STSA gain function curves in (8) versus β for different values of α ($\zeta = 0$ dB and $\gamma = 0$ dB)

3.1 Selection of parameter α

In the original proposition of the Wβ-SA estimator [10], the parameter α was selected as an increasing piecewise-linear function of frequency, in order to increase the contribution of high-frequency components of the speech STSA in the Bayesian cost function. This is because these frequencies often include small speech STSAs that can be easily masked by stronger noise components. However, increasing the values of this parameter monotonically with the frequency without considering the estimated speech STSA values results in over-amplification of high-frequency components, and therefore, large amount of distortion may appear in the enhanced speech. This will be further investigated in Section 5. We here employ the available initial estimate for the speech STSA, denoted by $\hat{\chi}_0(k,l)$ (the one used to calculate the noise masking threshold), to propose a new scheme for the selection of α. Specifically, we propose to select α according to the following scheme

$$\alpha(k,l) = \begin{cases} c_\alpha \frac{\hat{\chi}_0(k,l)}{\hat{\chi}_{0,\max}(l)}, & \text{if } \hat{\chi}_0(k,l) \geq \frac{\hat{\chi}_{0,\max}(l)}{4} \\ 0\,, & \text{otherwise} \end{cases} \qquad (10)$$

where $\hat{\chi}_{0,\max}(l)$ is the maximum value of the initial STSA estimate over the frequency bins at frame l and c_α, which determines the maximum value taken by α, is experimentally fixed at 0.55 to avoid excessively large α values. The major reasoning for the proposed frequency-based selection of the parameter α is to emphasize the weighting term χ^α in (10) for larger speech spectral components, while avoiding the use of such weighting for smaller components within each time frame. This further helps to distinguish the speech STSA components from the noise components of the same frequency at each time frame. In fact, if the speech STSA, $\hat{\chi}_0(k,l)$, falls above the threshold $\hat{\chi}_{0,\max}(l)/4$, increasing α results in the magnification of the weight χ^α in (4), provided that the speech STSA, χ, is large enough to be greater than unity. In contrast, for the speech STSA values smaller than the threshold, α is simply set to zero implying no further emphasis on the speech STSA component. In this case, the Wβ-SA estimator actually turns into the β-SA estimator in [9]. It should be noted

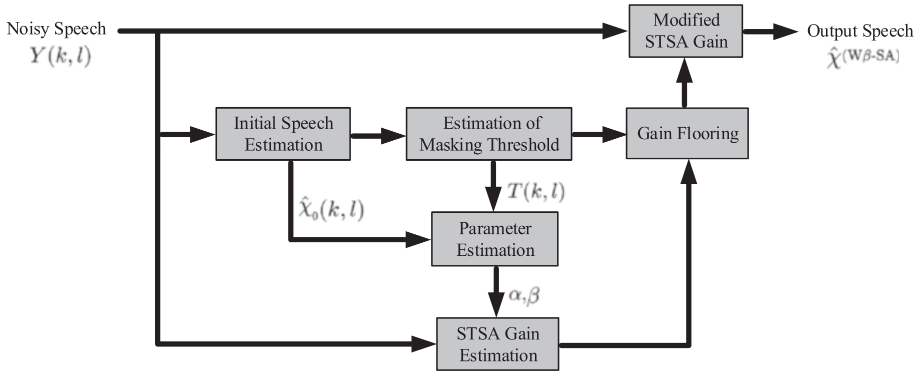

Fig. 2 Block diagram of the proposed algorithm for speech STSA estimation

that the threshold $\hat{\chi}_{0,\max}(l)/4$ was selected as a means to compare the relative intensity of the speech STSA components within the same time frame. Also, the normalization with respect to $\hat{\chi}_{0,\max}$ ensures that the resulting value of α will not be increased excessively in time frames where many frequency bins reach large values. Note that the magnification of strong speech components through the suggested selection of α can also be justified by considering the increment of the gain function through increasing α in the gain curves plotted in Fig. 1. In Fig. 3, the choice of the parameter α versus lower frequency bins for one sample time frame of the noisy speech along with the corresponding initial estimate of the speech STSA have been illustrated. In Section 5, it will be shown that the undesirable distortion resulting from the original selection of α as in [10] is compensated by using the proposed scheme.

3.2 Selection of parameter β

The adaptive selection of parameter β was primarily suggested in [9] as a linear function of frame SNR. Later in [12], it was suggested to choose this parameter as a linear function of both the frame SNR and noise masking threshold, as

$$\beta^{(1)}(k,l) = d_0 + d_1\mathrm{SNR}(l) + d_2T(k,l) + d_3\ \max\{\mathrm{SNR}(l) - d_4, 0\}T(k,l) \tag{11}$$

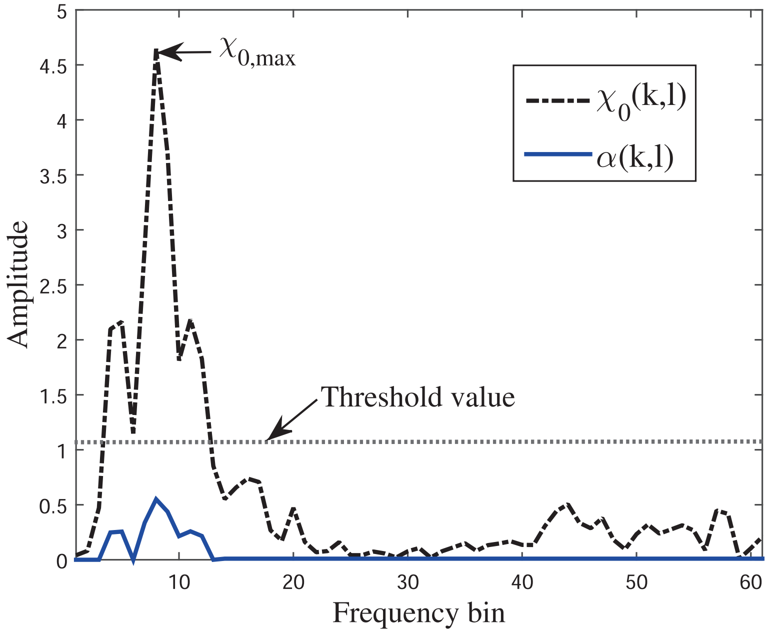

Fig. 3 Variation of the proposed choice of α versus frequency compared to that of initial speech STSA estimate for one sample time frame of the noisy speech

where SNR(l) is the frame SNR in dB, $T(k,l)$ is the normalized noise masking threshold [8] and d_i's are empirical numerical values. $T(k,l)$ represents the threshold below which the human auditory system cannot recognize the noise component and its calculation, which requires an initial estimate of the speech STSA, say $\hat{\chi}_0(k,l)$, involves a multiple-step algorithm detailed in [12]. The motivation for the choice of β in (11) is to increase the gain function values in frames/frequencies with higher frame SNRs or noise masking thresholds, given that the β-SA gain is a monotonically increasing function of β. The corresponding observations $Y(k,l)$ are dominated by strong speech components and it is hence desirable to employ a larger gain value in the enhancement process. In [10], however, from a psycho-acoustical point of view, it was suggested to choose β based on the compression rate between the sound intensity and perceptual loudness in the human ear. The suggested β therein takes the following form

$$\beta^{(2)}(k) = \frac{\log_{10}\left(g_1 k + g_2\right)}{\log_{10}\left(g_1 \frac{K}{2} + g_2\right)}\left(\beta_{\max} - \beta_{\min}\right) + \beta_{\min} \tag{12}$$

where K is the number of STFT frequency bins, g_1 and g_2 are two constants depending on the physiology of human ear [23] and $\beta_{\max}$ and $\beta_{\min}$ are set to 1 and 0.2, respectively. However, since β is chosen only as a function of the frequency, it is not adapted to the noisy speech. Furthermore, as experiments show, there may appear excessive distortion in the enhanced speech using the STSA estimator with this parameter choice, especially at high SNRs. Hence, we propose to use the adaptive approach in (11) as the basis for the selection of β, but to further apply the scheme in (12) as a form of frequency weighting to take into account the psycho-acoustics of the human auditory system within each time frame. Specifically, the following approach is proposed for the selection of β:

$$\beta(k,l) = C_\beta\ \beta^{(1)}(k,l)\ \beta^{(2)}(k) \tag{13}$$

where the purpose of the constant $C_\beta = 1/0.6$ is to scale up to one the median value of the frequency weighting parameter $\beta^{(2)}(k)$ in (12).

3.3 Proposed gain flooring scheme

In frequency bins characterized by weak speech components, the gain function of STSA estimators often approaches very small, near zero values, implying too much attenuation on the speech signal. To avoid the resulting speech distortion, various flooring schemes have been applied on the gain function values in these estimators. In [12], it is suggested to make use of the noise masking threshold in the spectral flooring scheme by employing a modification of the generalized spectral subtraction method in [8], namely,

$$G_M(k,l) = \begin{cases} G(k,l), & \text{if } \gamma(k,l) > \rho_1(k,l) \\ \sqrt{\frac{\rho_2(k,l)}{\gamma(k,l)}}, & \text{otherwise} \end{cases} \tag{14}$$

where $G(k,l)$ and $G_M(k,l)$ are the original and modified (thresholded) gain functions, respectively, and $\rho_1(k,l)$ and $\rho_2(k,l)$ are given by [12]

$$\begin{aligned} \rho_1(k,l) &= 5.28\ \frac{T(k,l) - T_{\min}(l)}{T_{\max}(l) - T_{min}(l)} + 1 \\ \rho_2(k,l) &= 0.015\ \frac{T(k,l) - T_{\min}(l)}{T_{\max}(l) - T_{min}(l)} \end{aligned} \tag{15}$$

with $T_{\min}(l)$ and $T_{\max}(l)$ denoting the minimum and maximum of $T(k,l)$ at the lth time frame. The *a posteriori* SNR, $\gamma(k,l)$, is used in the top branch of (14) as an indicator of the speech signal intensity while the term $\sqrt{\frac{\rho_2(k,l)}{\gamma(k,l)}}$ in the bottom branch determines the thresholded value of the gain function. Still, (14) is characterized by a number of limitations. As originally proposed by Cohen in [24], the gain function itself is a more relevant indicator of speech signal intensity and is therefore more appropriate for use in the thresholding test than $\gamma(k,l)$. Another problem with (14) is that the thresholded value may increase uncontrollably at very low values of $\gamma(k,l)$. Rather than relying on $\gamma(k,l)$, it was suggested in [25] to make use of the estimated speech STSA in the thresholded value, as in the following

$$G'_M(k,l) = \begin{cases} G(k,l), & \text{if } G(k,l) > \mu_0 \\ \frac{1}{2}\frac{\mu_0|Y(k,l)| + \hat{\chi}(k,l-1)}{|Y(k,l)|}, & \text{otherwise} \end{cases} \tag{16}$$

where μ_0 is a fixed threshold taken between 0.05 and 0.22. Our experimentations, however, provided different proper values for μ_0 in various noise scenarios and input SNRs. Hence, considering the wide range of values for the gain function and also the variations in speech STSA, it is appropriate for the threshold μ_0 to be selected as a function of the time frame and frequency bin. Herein, by employing the adaptive threshold $\rho_1(k,l)$ in (15) and using a variable recursive smoothing for the thresholded value, we propose the following alternative flooring scheme

$$G''_M(k,l) = \begin{cases} G(k,l), & \text{if } G(k,l) > \rho_1(k,l) \\ \frac{p(k,l)\hat{\chi}_0(k,l) + [1-p(k,l)]\hat{\chi}(k,l-1)}{|Y(k,l)|}, & \text{otherwise} \end{cases} \tag{17}$$

where $p(k,l)$ is the speech presence probability which can be estimated through a soft-decision noise PSD estimation method. Using the popular improved minima controlled recursive averaging (IMCRA) in [26] provides enough precision for the estimation of this parameter in the proposed gain flooring scheme. According to (17), for higher speech presence probabilities or equivalently in frames/frequencies with stronger speech components, the contribution of the current frame in the recursive smoothing through the term $\hat{\chi}_0(k,l)$ will be larger than that of the previous frame $\hat{\chi}(k,l-1)$. Conversely, in case of a weak speech component in the current frame, the smoothing gives more weight to the previous frame. Hence, this choice of the flooring value favors the speech component over the noise component in adverse noisy conditions where the gain function is mainly determined by the second branch in (17).

4 Incorporation of GGD as speech prior

As mentioned in Section 1, use of the parametric GGD model as the STSA prior, due to providing further flexibility in the resulting gain function, is advantageous compared to the conventional Rayleigh prior. In this section, we first derive an extended Wβ-SA estimator under the GGD speech prior and then propose an efficient method to estimate its corresponding parameters.

4.1 Extended Wβ-SA estimator with GGD prior

The GGD model can be expressed as

$$p(\chi) = \frac{ab^c}{\Gamma(c)}\chi^{ac-1}\exp(-b\chi^a);\ \chi \geq 0,\ \ a,b,c > 0 \tag{18}$$

with a and c as the shape parameters and b as the scaling parameter [21]. To obtain a solution to the Wβ-SA estimator as in (5), we consider the moment term $E\{\chi^m|Y\}$ based on the above PDF for the speech STSA. In view of the comprehensive experimental results in [14, 20] for different values of a and in order to arrive at a closed-form solution in the Bayesian sense, we choose $a = 2$ in our work. For this choice of a, the GGD prior is actually simplified into a generalized form of the Chi distribution with $2c$ degrees of freedom and $1/\sqrt{2b}$ as the scale parameter [27]. Based on the second moment of the derived Chi distribution, it can be deduced that the two parameters b and c satisfy the relation $c/b = \sigma_\chi^2$ [28]. Therefore, the scale parameter b has to be chosen as c/σ_χ^2, given an estimate of the speech STSA variance, σ_χ^2, and the shape parameter c. Using an estimate of the noise variance, σ_ν^2, and the *a priori* SNR, ζ, we can obtain an estimate of the speech STSA variance as $\sigma_\chi^2 = \zeta\sigma_\nu^2$. The selection of the shape parameter c will be discussed in the next subsection. Taking this into consideration, the following expression for the STSA moment can be derived (see Appendix for details):

$$E\{\chi^m|Y\} = \frac{\Gamma\left(\frac{m+2c}{2}\right)\mathrm{M}\left(\frac{2-m-2c}{2},1;-\nu'\right)}{\Gamma(c)\lambda^{m/2}\mathrm{M}\left(1-c,1;-\nu'\right)} \tag{19}$$

where

$$\lambda = \frac{c}{\sigma_\chi^2} + \frac{1}{\sigma_\nu^2}\ ,\ \ \nu' = \frac{\zeta}{c+\zeta}\gamma \tag{20}$$

Now, by using (19) into (5) we can derive

$$G^{(\mathrm{MW}\beta-\mathrm{SA})} = \frac{\sqrt{\nu'}}{\gamma}\left(\frac{\Gamma\left(\frac{\alpha+\beta+2c}{2}\right)\mathrm{M}\left(\frac{2-\alpha-\beta-2c}{2},1;-\nu'\right)}{\Gamma\left(\frac{\alpha}{2}+c\right)\mathrm{M}\left(\frac{2-\alpha-2c}{2},1;-\nu'\right)}\right)^{1/\beta} \tag{21}$$

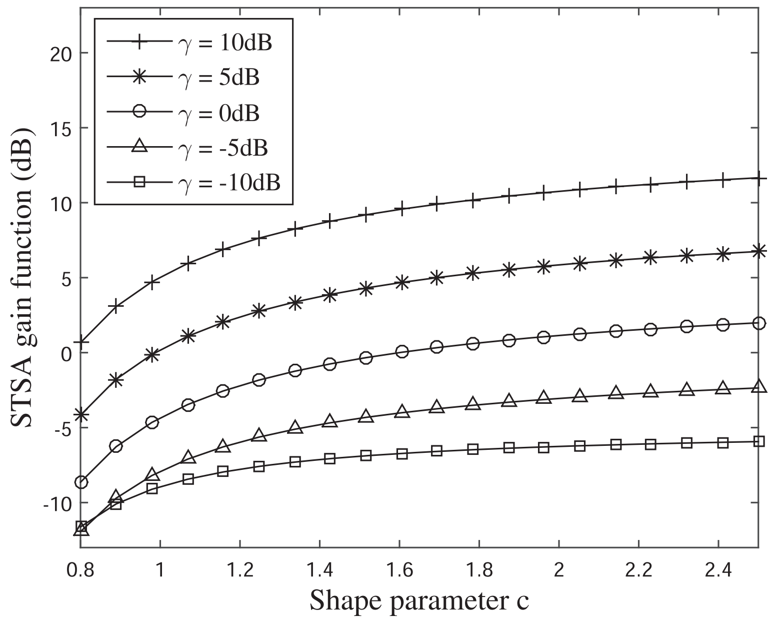

Fig. 4 Gain function of the modified Wβ-SA estimator in (21) versus the GGD shape parameter c for different values of γ ($\zeta = -5$ dB)

where the notation MWβ-SA is used to denote the modified Wβ-SA estimator. It is obvious that, for $c = 1$ where the Rayleigh prior is obtained as a special case, (21) degenerates to the original Wβ-SA. In the following, we present a simple approach for the selection of the GGD parameter c for the proposed STSA estimator.

4.2 Selection of the GGD shape parameter

In [14, 20], experimental fixed values in the range of [0,2] have been used for the GGD shape parameter c in different noisy scenarios. Rather than using experimental values, we here take advantage of the behavior of the proposed gain function in (21) with respect to the shape parameter

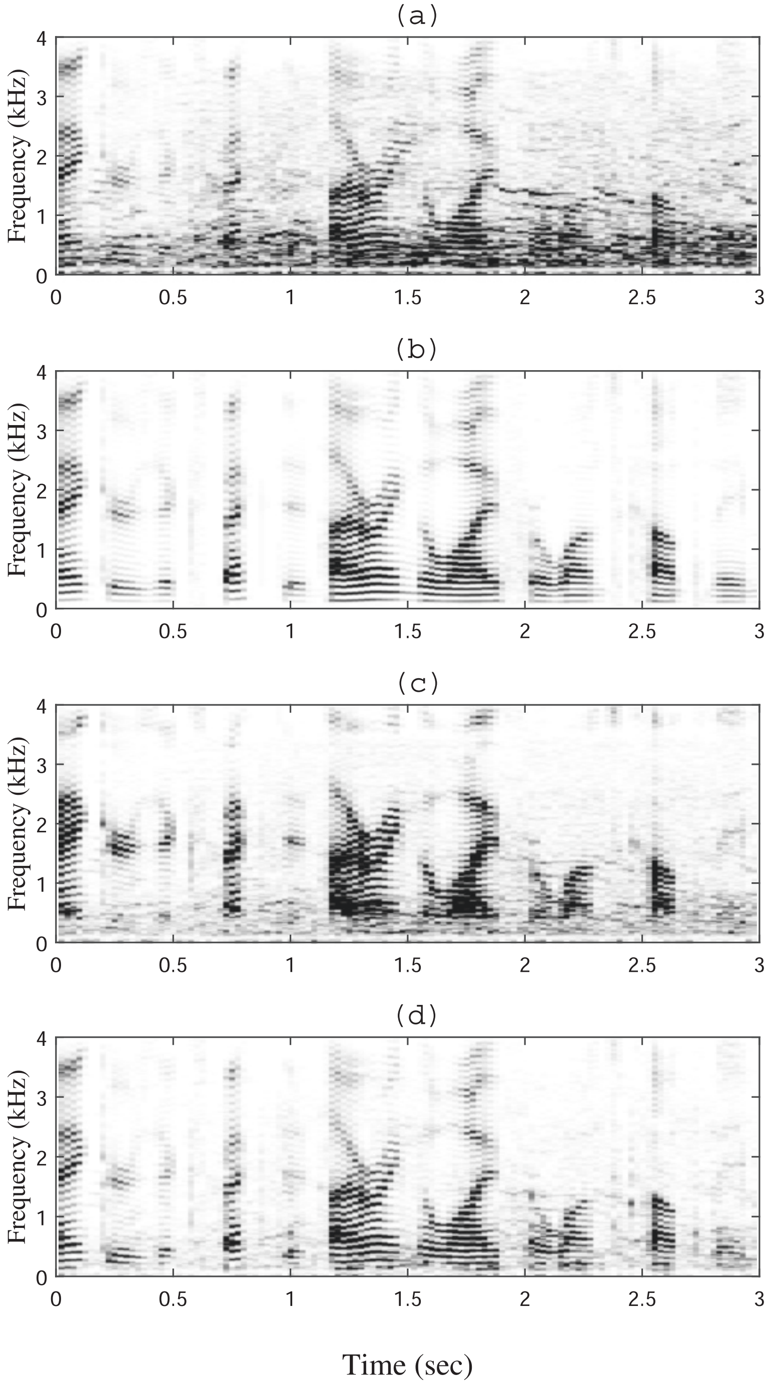

Fig. 5 Spectrograms of (**a**) input noisy speech, (**b**) clean speech, (**c**) enhanced speech by the original Wβ-SA estimator, and (**d**) enhanced speech by the proposed Wβ-SA estimator, in case of babble noise (Input SNR = 5 dB)

c and propose an adaptive scheme for the determination of this parameter. Figure 4 depicts curves of the proposed gain function in (21) versus the shape parameter *c* for different *a posteriori* SNRs. As observed, increasing the shape parameter leads to a monotonic increase of the gain function for all considered values of SNR. Note that for stronger speech STSA components (or equivalently weaker noise components) a larger gain function value is desirable in general. Therefore, we suggest to choose the shape parameter as a linear function of the SNR values at each time frame, namely,

$$c(l) = c_{\min} + (c_{max} - c_{min})\, \zeta_{norm}(l) \qquad (22)$$

where, based on the comprehensive experimentations in [21], $c_{\min}$ and $c_{\max}$ are chosen as 1 and 3, respectively and

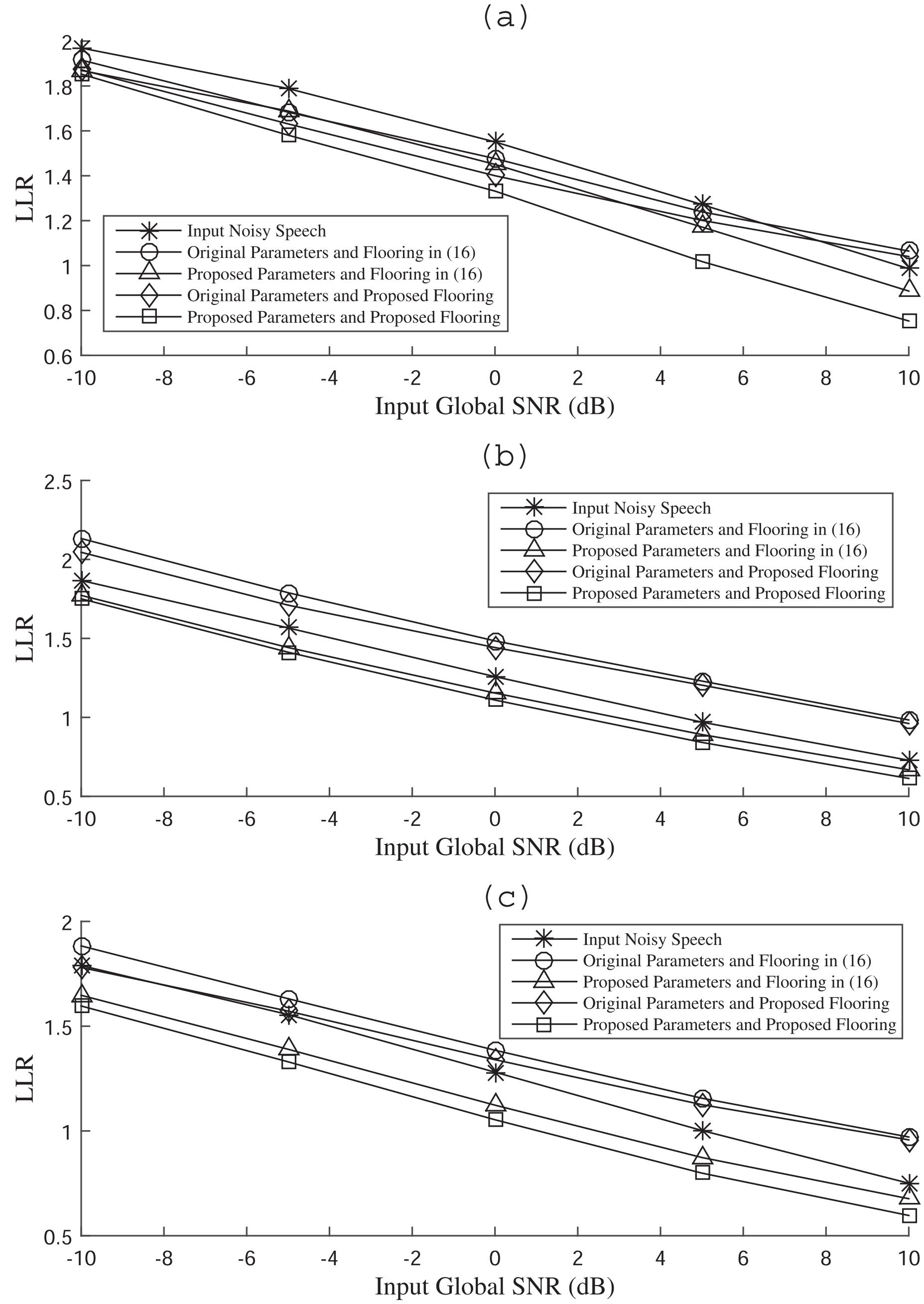

Fig. 6 LLR versus global SNR for different Wβ-SA estimators, (**a**) white noise, (**b**) babble noise, and (**c**) car noise

$0 < \zeta_{\mathrm{norm}}(l) < 1$ is the normalized *a priori* SNR. The latter is obtained as

$$\zeta_{\mathrm{norm}}(l) = \frac{\zeta_{av}(l) - \zeta_{min}(l)}{\zeta_{max}(l) - \zeta_{min}(l)} \tag{23}$$

with $\zeta_{\mathrm{av}}(l)$ as the *a priori* SNR being averaged over the frequency bins of the lth frame, and $\zeta_{\mathrm{min}}(l)$ and $\zeta_{\mathrm{max}}(l)$ as the minimum and maximum of the *a priori* SNR at the same time frame, respectively. According to (22), the shape parameter c takes on its values as a linearly increasing function of the SNR in its possible range between c_{min} and c_{max}, leading to the appropriate adjustment of the estimator gain function based on the average power of the speech STSA components at each frame.

5 Performance evaluation

In this section, we evaluate the performance of the proposed STSA estimation methods using objective speech

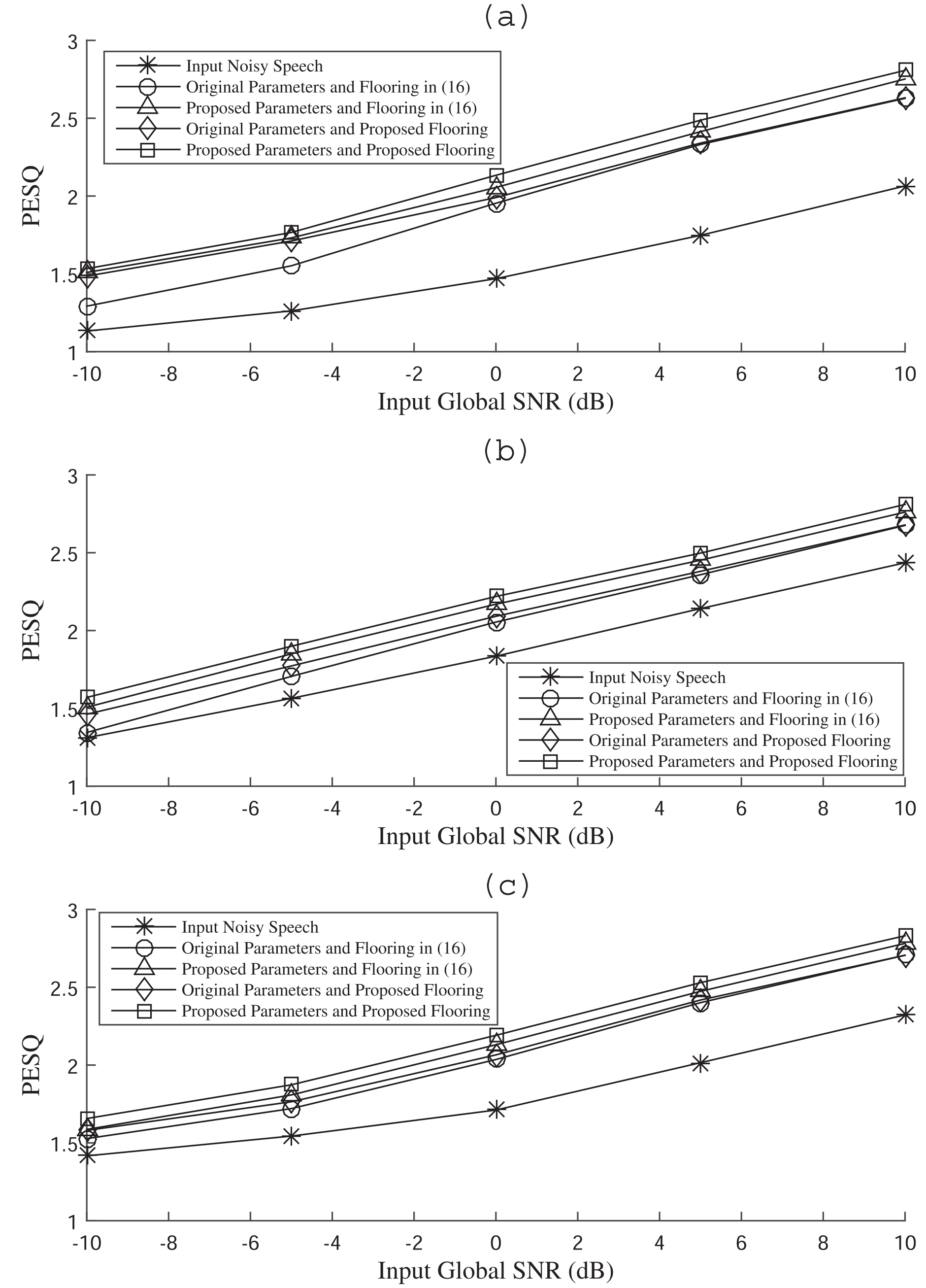

Fig. 7 PESQ versus global SNR for different Wβ-SA estimators, (**a**) white noise, (**b**) babble noise, and (**c**) car noise

quality measures. First, the performance of the proposed STSA parameter selection and gain flooring schemes are compared to the previous methods. Next, the proposed GGD-based estimator is compared to the estimators using the conventional Rayleigh prior. Due to the performance advantage of the generic Wβ-SA estimator over the previous versions of STSA estimators, it is used throughout the following simulations.

Various types of noise from NOISEX-92 database [29] were considered for the evaluations, out of which, the results are presented for three noise types, i.e., white, babble, and car noises. Speech utterances including 10 male and 10 female speakers are used from the TIMIT speech database [30]. The sampling rate is set to 16 kHz and a Hamming window with length 20 ms and overlap of 75 % between consecutive frames is used for STFT analysis and overlap-add synthesis. In all simulations, the noise variance is estimated by the soft-decision IMCRA method [26] eliminating the need to use a hard-decision voice activity detector (VAD). Also, the decision-directed (DD)

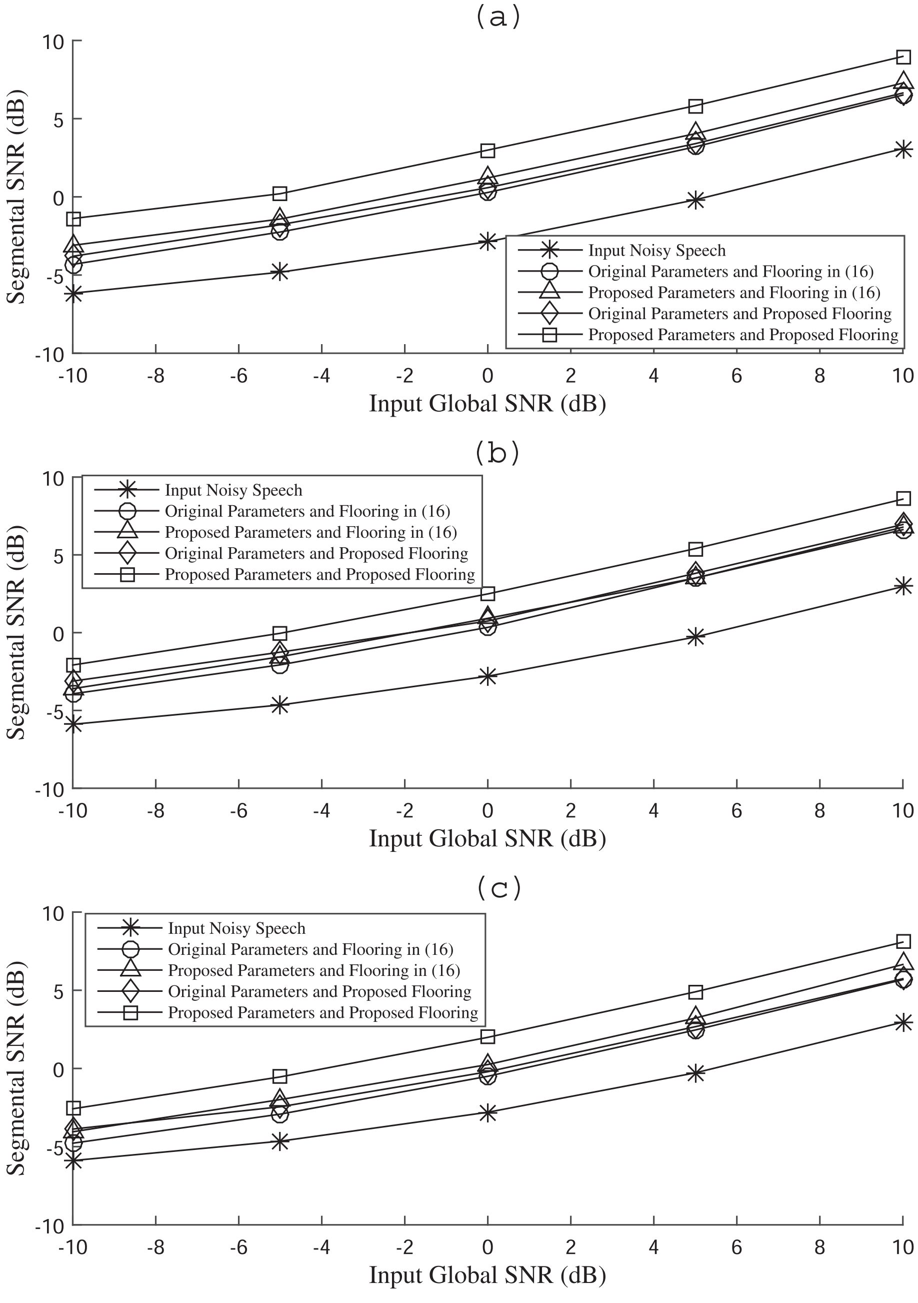

Fig. 8 Segmental SNR versus global SNR for different Wβ-SA estimators, (**a**) white noise, (**b**) babble noise, and (**c**) car noise

approach [3] is used to estimate the *a priori* SNR. Even though more accurate methods of noise and SNR estimation exist, use of the aforementioned approaches provided enough accuracy for our purpose.

As for the assessment of the enhanced speech quality, various objective measures have been employed in the literature. In order to obtain a measure of the overall quality of the enhanced speech, we use the Perceptual Evaluation of Speech Quality (PESQ) scores. Nowadays, PESQ is a widely accepted industrial standard for objective voice quality evaluation and is standardized as ITU-T recommendation P.862 [31]. Since PESQ measurements principally model the Mean Opinion Scores (MOS), it has a close connection to subjective performance tests performed by a human. On the other hand, the log-likelihood ratio (LLR) score which measures a logarithmic distance between the linear prediction coefficients (LPC) of the enhanced and clean speech utterances, is more related to the introduced distortion in the clean speech signal [32]. Whereas PESQ takes values between 1 (worst) and 4.5 (best), the lower the LLR the less distorted the speech signal. To have a more complete evaluation of the noise reduction performance, we also consider the segmental SNR which correlates well with the level of noise reduction regardless of the existing distortion in the speech [32].

To illustrate graphically the advantage achieved by the proposed parameter selection scheme, first we plot the speech spectrograms for the noisy, clean, and enhanced speech signals for the case of babble noise in Fig. 5. We considered the original frequency-based scheme in [10] and compared it to the suggested scheme in Section 3 where, for both schemes, the gain flooring in (17) is used. It is observed that, particularly at low frequencies, the estimator with the original scheme cannot preserve the clean speech component satisfactorily, whereas it over-amplifies other parts of the speech spectrum. The disappearance of the very low-frequency portion of the spectral content is mainly due to the too small values of the parameter α given by this scheme. However, the proposed parameter selection scheme is capable of retaining most of the strong components of the clean speech spectrum, especially in the low frequencies. Further noise reduction can also be observed through the use of the proposed selection schemes for α and β.

To evaluate the efficiency of the proposed selection of the estimator parameters as well as the proposed gain flooring scheme, we herein present the performance measures for Wβ-SA estimator with the parameter scheme in [10], Wβ-SA estimator using the proposed parameter selection in Section 3 and also the same estimators with the proposed gain flooring in (17). We employed the gain flooring scheme in (16) in cases where the proposed gain flooring is not used, since the closest results to the proposed flooring were obtained under this scheme. The LLR results for the three noise types in the range of input global SNR between −10 and 10 dB are presented in Fig. 6. As stated in Section 3, the original choice of the parameters of Wβ-SA estimator results in an excessive distortion in the enhanced speech, which is observable through the LLR values in Fig. 6. Yet, the suggested adaptive parameter selection completely resolves this problem and is also able to yield further improvement. Moreover, the use of the recursive smoothing-based gain flooring in (17) is able to remove further speech distortion compared to the gain flooring scheme in [25] as given by (16), especially at higher SNRs. This is due to the incorporation of the estimated speech, that is strongly present at high SNRs, in the flooring value instead of using the noise masking threshold-based method. The result is that the gain floor is kept at more moderate levels in order not to distort the existing speech components. Similar trends can be observed in Figs. 7 and 8 in terms of the speech quality determined by PESQ and noise reduction evaluated by segmental SNR measurements, respectively. As it is observed, in cases where the proposed parameter setting is able to provide only minor improvements over the original method, the combination of the proposed parameters with the gain flooring improves the performance to a considerable degree.

Table 1 PESQ values for the Wβ-SA estimator with different schemes of parameter α, case of white noise

Input SNR (dB)	−10	−5	0	5	10
Input noisy speech	1.13	1.26	1.47	1.75	2.06
Choice of $\alpha = 0$	1.49	1.70	2.03	2.39	2.72
Choice of $\alpha = 0.22$	1.49	1.73	2.06	2.41	2.76
Original choice of α	1.50	1.73	2.08	2.44	2.78
Proposed choice of α	1.54	1.77	2.14	2.49	2.81

To have a more detailed evaluation of each of the suggested schemes, we present the results obtained by individually applying each of them to the Wβ-SA estimator. In Tables 1, 2, and 3, PESQ results for the Wβ-SA estimator considering $\alpha = 0$ (corresponding to the β-SA estimator), $\alpha = 0.22$ (an empirically fixed choice of α), original

Table 2 PESQ values for the Wβ-SA estimator with different schemes of parameter α, case of babble noise

Input SNR (dB)	−10	−5	0	5	10
Input noisy speech	1.31	1.56	1.83	2.14	2.43
Choice of $\alpha = 0$	1.48	1.71	2.03	2.40	2.73
Choice of $\alpha = 0.22$	1.51	1.82	2.14	2.42	2.77
Original choice of α	1.54	1.86	2.16	2.45	2.79
Proposed choice of α	1.58	1.91	2.23	2.51	2.82

Table 3 PESQ values for the Wβ-SA estimator with different schemes of parameter α, case of car noise

Input SNR (dB)	−10	−5	0	5	10
Input noisy speech	1.41	1.54	1.71	2.01	2.32
Choice of $\alpha = 0$	1.57	1.76	2.06	2.40	2.75
Choice of $\alpha = 0.22$	1.58	1.78	2.11	2.46	2.77
Original choice of α	1.60	1.81	2.15	2.50	2.79
Proposed choice of α	1.66	1.88	2.20	2.54	2.84

Table 5 PESQ values for the Wβ-SA estimator with different schemes of parameter β, case of babble noise

Input SNR (dB)	−10	−5	0	5	10
Input noisy speech	1.31	1.56	1.83	2.14	2.43
Choice of $\beta = 1.82$	1.49	1.73	2.04	2.42	2.73
Choice of β by (11)	1.55	1.88	2.18	2.46	2.76
Choice of β by (12)	1.55	1.88	2.17	2.47	2.79
Proposed choice of β	1.58	1.91	2.23	2.51	2.82

scheme for α as in [10] and the proposed scheme for α in (10). In all cases, the proposed scheme for β and so for the gain flooring have been employed. It is observed that, whereas the employment of the STSA weighting through the parameter α results in a considerable improvement compared to the β-SA estimator, the suggested scheme represented in the last row attains the best results. Within the same line, Tables 4, 5, and 6 are representative of the evaluations performed on the Wβ-SA estimator by using $\beta = 1.82$ (an empirically fixed value), β given by (11), β given by (12), and the proposed choice of β as in (13). In all cases, we employed α as proposed in (10) and the gain flooring proposed in (17). It can be deduced that, apart from the benefit obtained by the frequency-dependent choices of β through (11) and (12) over the fixed choice of this parameter, the suggested scheme in (13) is able to achieve notable improvements compared to the others.

To investigate the performance improvement attained by the proposed gain flooring scheme in (17) individually, we implemented the Wβ-SA estimator in Section 3 using different gain flooring schemes. In Fig. 9, PESQ results have been shown for this estimator using the developed gain flooring in (17), those given by (14) and (16), as well as a fixed gain thresholding with $\mu_0 = 0.08$. It is observed that, whereas the gain flooring in (16) leads to improvements with respect to the conventional fixed thresholding, the one in (14) only slightly outperforms the employed fixed flooring. This shows that the gain function itself, as used in (16), is a better measure for gain flooring compared to the *a posteriori* SNR used in (14). This is the reason we based our gain flooring scheme on (16) but further employed the noise masking concept to threshold the gain function values. As illustrated, the proposed gain flooring outperforms the scheme in (16) considerably even in the higher range of the input SNR. This is due to the fact that, even at such SNRs, there are frequencies in which the gain function decays abruptly below the threshold value, requiring an appropriate flooring value to keep the speech components.

Next, we investigated the performance advantage obtained by the proposed GGD-based estimator in Section 4 over the original Rayleigh-based estimator [10]. Also, to illustrate the superiority of the proposed scheme for the selection of the GGD parameter c in Section 4.2 with respect to the employed fixed values as in [21], we considered the same GGD-based estimator with different choices of the parameter c. In Fig. 10, PESQ results are plotted for the original and suggested Wβ-SA estimators as well as two fixed choices of the parameter c in the range of $[\,c_{\min}, c_{\max}]$ as in Section 4.2. As it is observed, whereas the use of GGD speech prior with fixed choices of c results in improvements with respect to the Rayleigh speech prior in most of the cases, the suggested SNR-based scheme for choosing c is capable of providing further enhancement compared to different fixed c choices. Other choices of the parameter c did not result in further improvements than those considered herein.

To evaluate the performance of the proposed GGD-based Wβ-SA estimator in Section 4 with respect to the recent STSA estimators using super-Gaussian priors, we considered the STSA estimation methods proposed in [33, 34]. In [33], the GGD model with a few choices of fixed parameters is applied as the STSA prior using the Log-MMSE estimator, whereas in [34], WE and WCOSH

Table 4 PESQ values for the Wβ-SA estimator with different schemes of parameter β, case of white noise

Input SNR (dB)	−10	−5	0	5	10
Input noisy speech	1.13	1.26	1.47	1.75	2.06
Choice of $\beta = 1.82$	1.48	1.69	2.00	2.32	2.68
Choice of β by (11)	1.53	1.74	2.08	2.39	2.72
Choice of β by (12)	1.52	1.74	2.06	2.42	2.75
Proposed choice of β	1.54	1.77	2.14	2.49	2.81

Table 6 PESQ values for the Wβ-SA estimator with different schemes of parameter β, case of car noise

Input SNR (dB)	−10	−5	0	5	10
Input noisy speech	1.13	1.26	1.47	1.75	2.06
Choice of $\beta = 1.82$	1.60	1.81	2.09	2.43	2.76
Choice of β by (11)	1.63	1.84	2.14	2.49	2.78
Choice of β by (12)	1.62	1.83	2.14	2.51	2.80
Proposed Choice of β	1.66	1.88	2.20	2.54	2.84

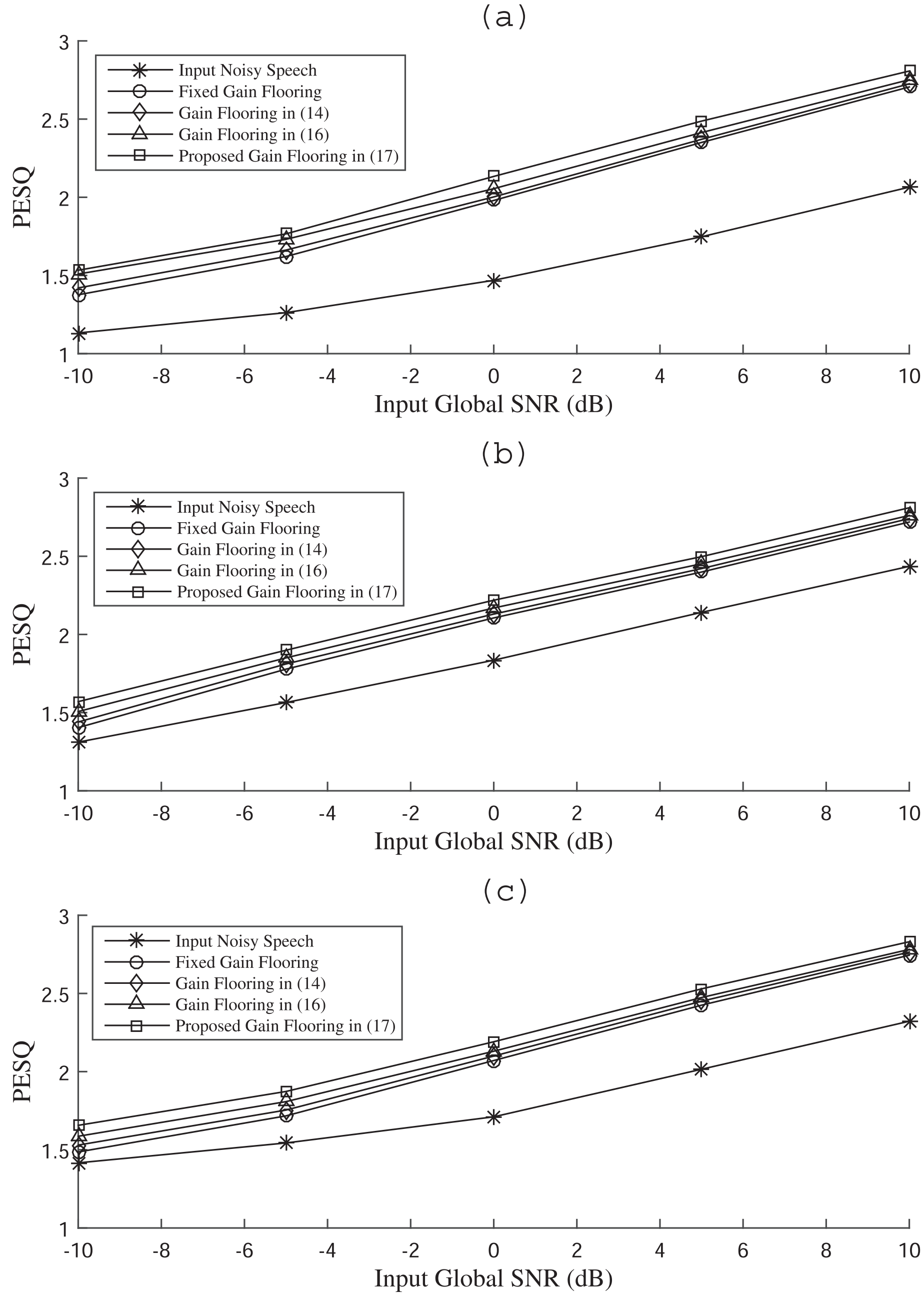

Fig. 9 PESQ versus global SNR for Wβ-SA estimator with the proposed parameters in Section 3 using different gain flooring schemes, (**a**) white noise, (**b**) babble noise, and (**c**) car noise

estimators (originally introduced in [6]) are developed exploiting Chi PDF with fixed parameters as the STSA prior. Figure 11 illustrates speech spectrograms for the aforementioned STSA estimators in case of babble noise. Through careful inspection of the speech spectrograms, it is observed that the proposed estimator is capable of maintaining clean speech components at least as much as the other estimators whereas further noise reduction, especially in the lower frequency range, is clearly obtained by using the proposed estimator. In Figs. 12, 13, and 14, performance comparisons for the same estimators are depicted in terms of LLR, PESQ, and segmental SNR, respectively. We used the gain flooring scheme proposed in Section 3.3 for all of the estimators. It is observed that, while the estimators suggested in [34] perform better than the one in [33] in most of the cases, the proposed STSA

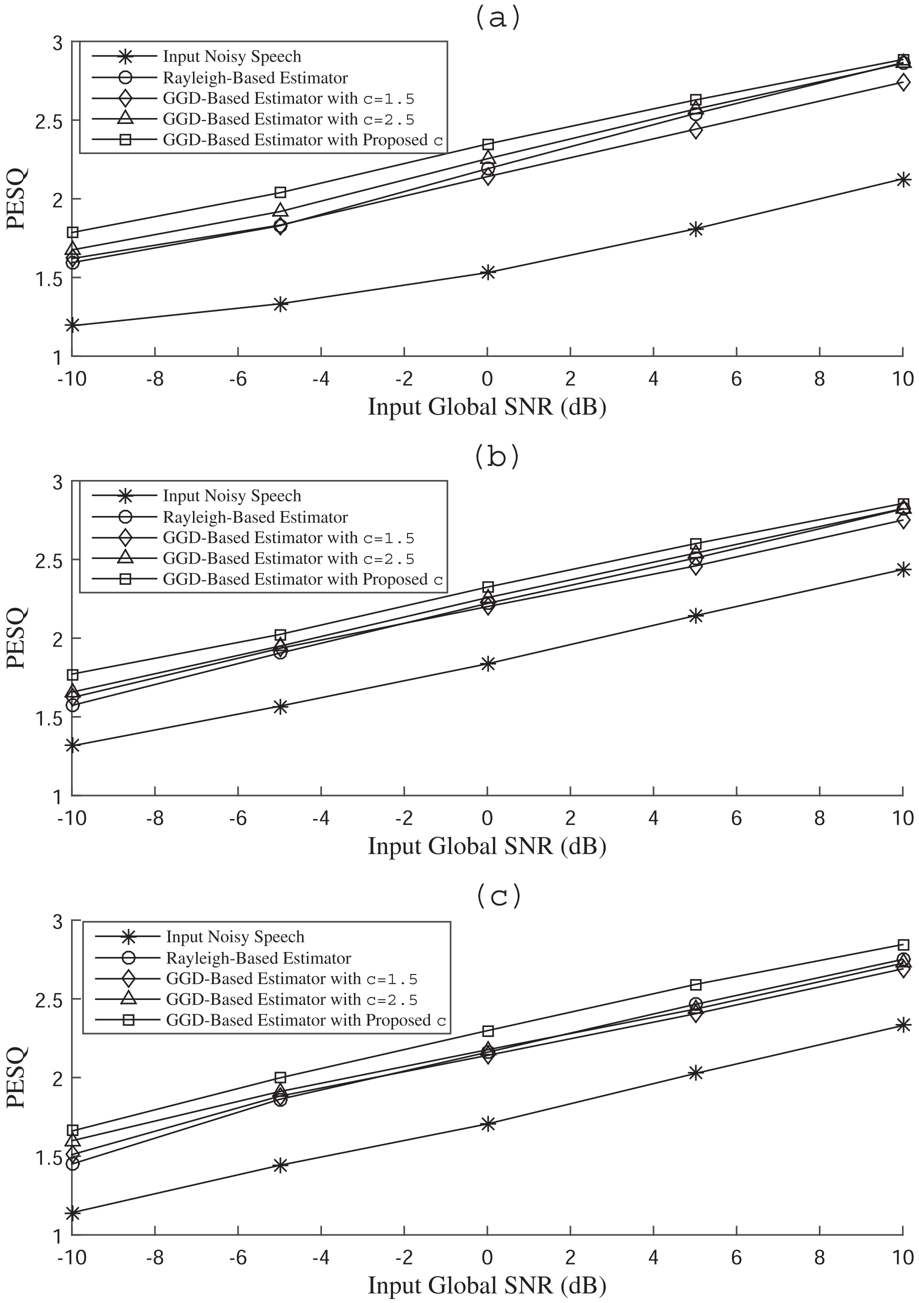

Fig. 10 PESQ versus global SNR for the Rayleigh-based estimator in Section 3, the GGD-based estimator in Section 4 with $c = 1.5, 2.5$ and the proposed choice of c in Section 4.2, (**a**) white noise, (**b**) babble noise, and (**c**) car noise

estimator in Section 4 is able to achieve superior performance especially at the lower SNR. This is mainly due to the further contribution of the speech STSA in the Bayesian cost function parameters through (10) as well as properly selecting the STSA prior shape parameter using (22) to adjust the gain function values. Whereas the latter is assigned a fixed value in the two previous STSA estimation methods, careful selection of this parameter based on the estimated *a priori* SNR leads to a more accurate model for the speech STSA prior.

6 Conclusions

In this work, we presented new schemes for the selection of Bayesian cost function parameters in parametric STSA

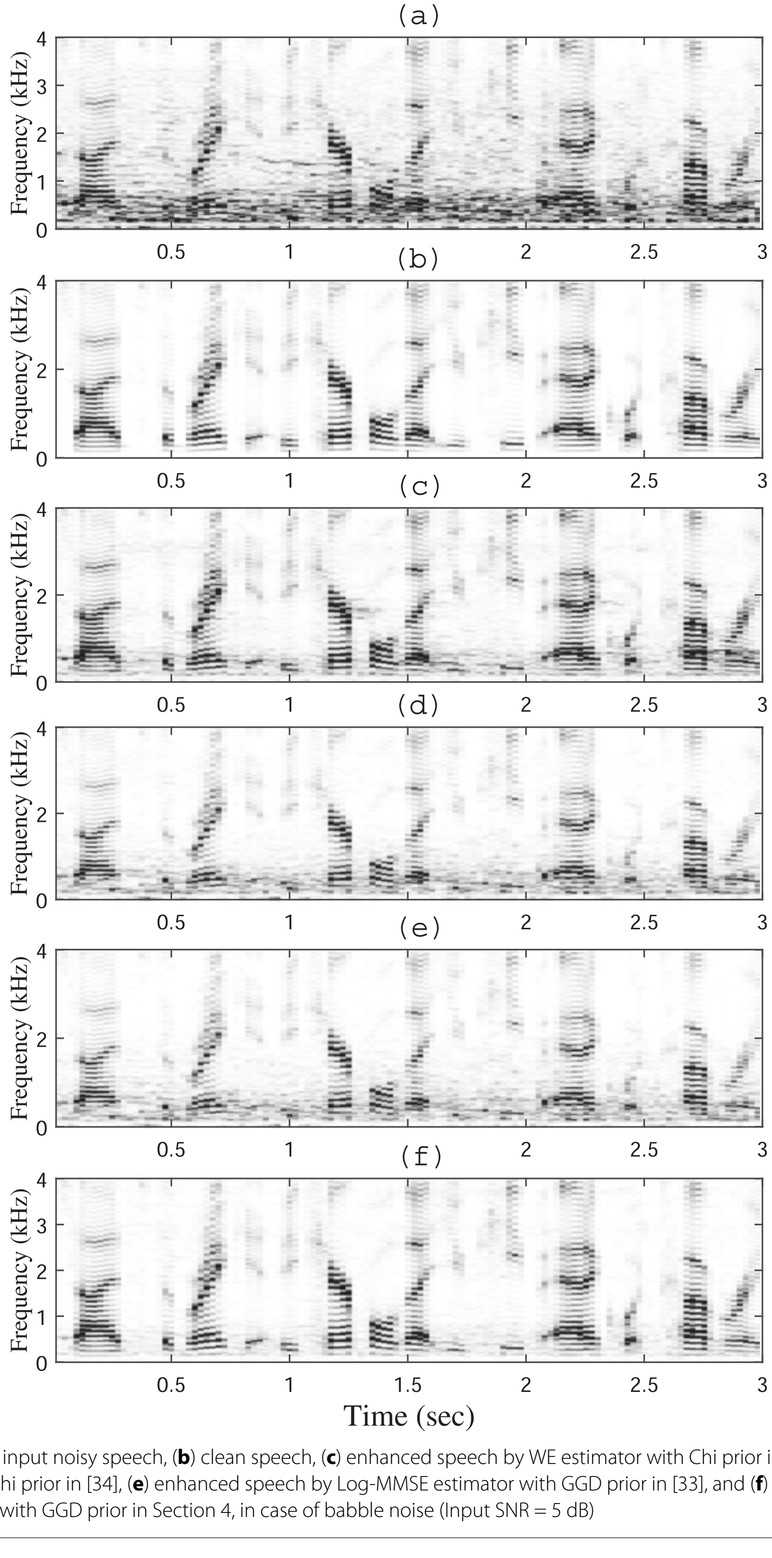

Fig. 11 Spectrograms of (**a**) input noisy speech, (**b**) clean speech, (**c**) enhanced speech by WE estimator with Chi prior in [34], (**d**) enhanced speech by WCOSH estimator with Chi prior in [34], (**e**) enhanced speech by Log-MMSE estimator with GGD prior in [33], and (**f**) enhanced speech by the proposed Wβ-SA estimator with GGD prior in Section 4, in case of babble noise (Input SNR = 5 dB)

estimators, based on an initial estimate of the speech and the properties of human audition. We further used these quantities to design an efficient flooring scheme for the estimator's gain function, which employs recursive smoothing of the speech initial estimate. Next, we applied the GGD model as the speech STSA prior to the

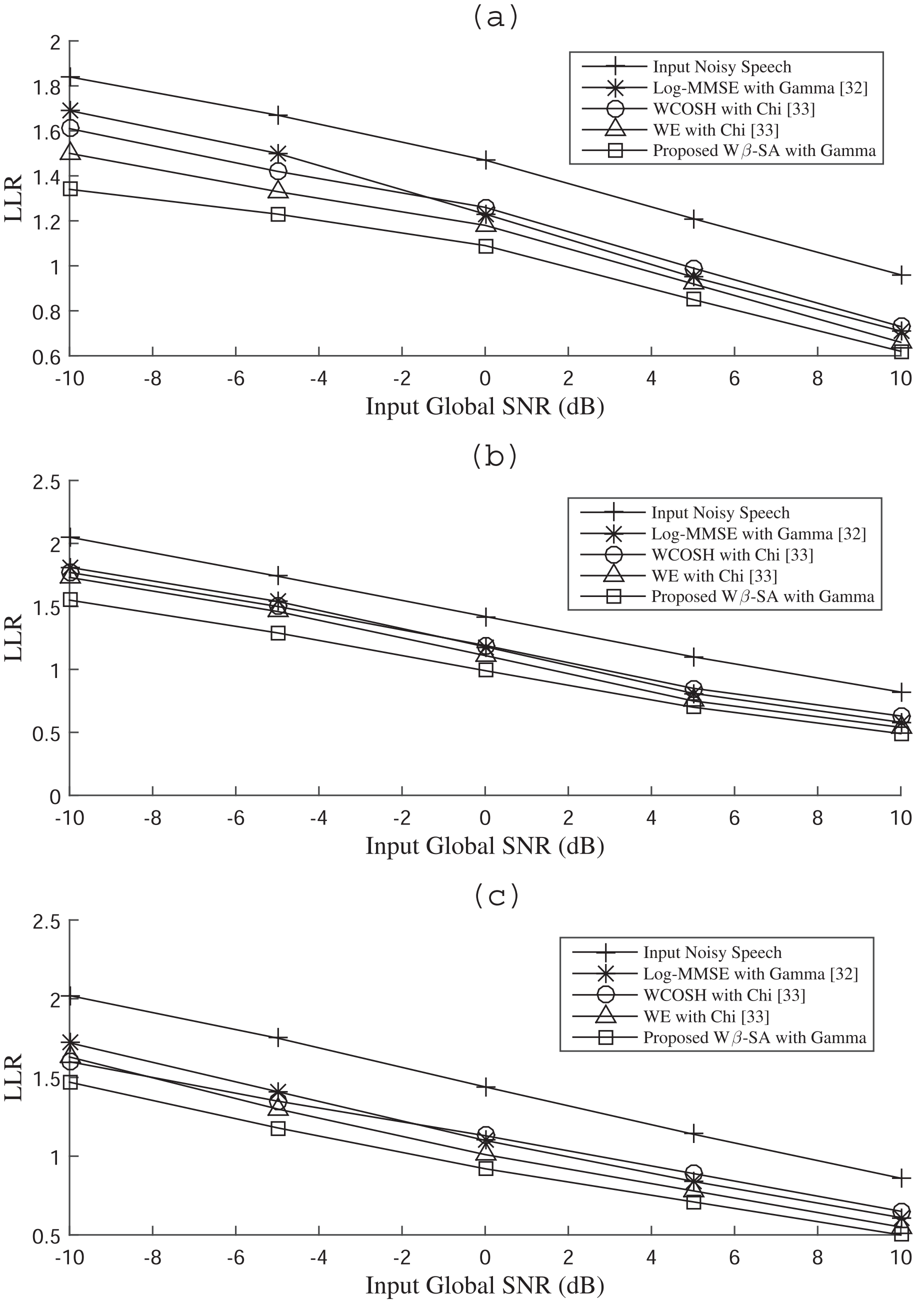

Fig. 12 LLR versus global SNR for the STSA estimators in [33, 34] and the proposed STSA estimator in Section 4, (**a**) white noise, (**b**) babble noise, and (**c**) car noise

Wβ-SA estimator and proposed to choose its parameters using the noise spectral variance and the *a priori* SNR. Due to the more efficient adjustment of the estimator's gain function by the suggested parameter choice and also further keeping the speech strong components from being distorted through the gain flooring scheme, our STSA estimation schemes are able to provide better noise reduction as well as less speech distortion compared to the previous methods. Also, by taking into account a more precise modeling of the speech STSA prior through using the GGD function with the suggested adaptive parameter selection, improvements were achieved with respect to the recent speech STSA estimators. Quality and noise reduction performance evaluations indicated the superiority of the proposed speech STSA estimation with respect to the previous estimators.

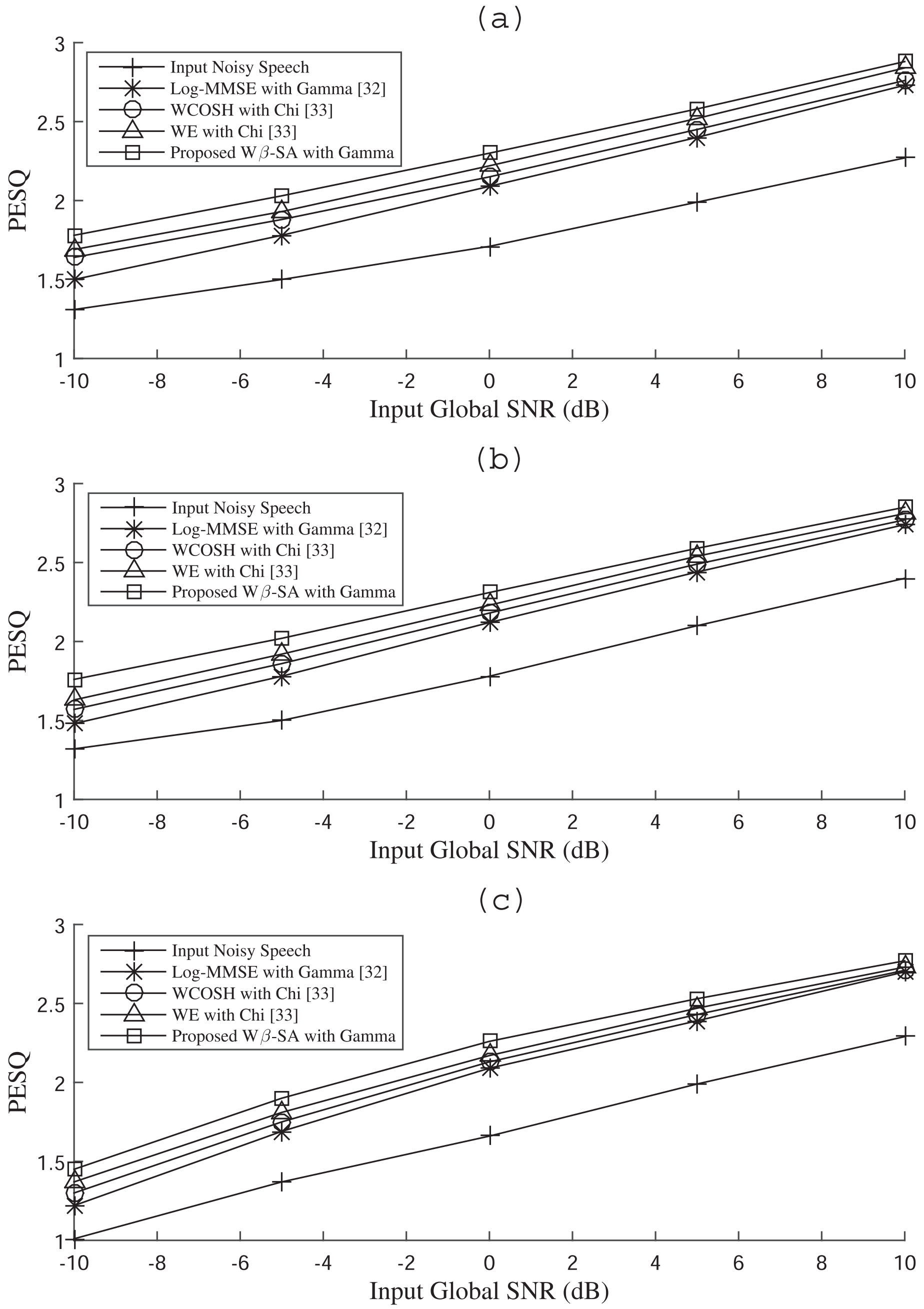

Fig. 13 PESQ versus global SNR for the STSA estimators in [33, 34] and the proposed STSA estimator in Section 4, (**a**) white noise, (**b**) babble noise, and (**c**) car noise

Appendix: Derivation of Eq. (19)

Based on (6), we obtain

$$E\{\chi^m|Y\} = \frac{\int_0^\infty \int_0^{2\pi} \chi^m p(Y|\chi,\Omega)p(\chi,\Omega)d\Omega d\chi}{\int_0^\infty \int_0^{2\pi} p(Y|\chi,\Omega)p(\chi,\Omega)d\Omega d\chi} \triangleq \frac{\mathrm{NUM}}{\mathrm{DEN}} \tag{24}$$

Obviously, it suffices to derive the numerator in (24) and then obtain the denominator as a special case where $m = 0$. Using the GGD model in (18) with $a = 2$ for the speech STSA and the uniform PDF for the speech phase, it follows

$$p(\chi,\Omega) = \frac{1}{2\pi}\frac{2b^c}{\Gamma(c)}\chi^{2c-1}\exp(-b\chi^2) \tag{25}$$

Substitution of (25) and also $p(Y|\chi,\Omega)$ from (7) into the numerator of (24) results in

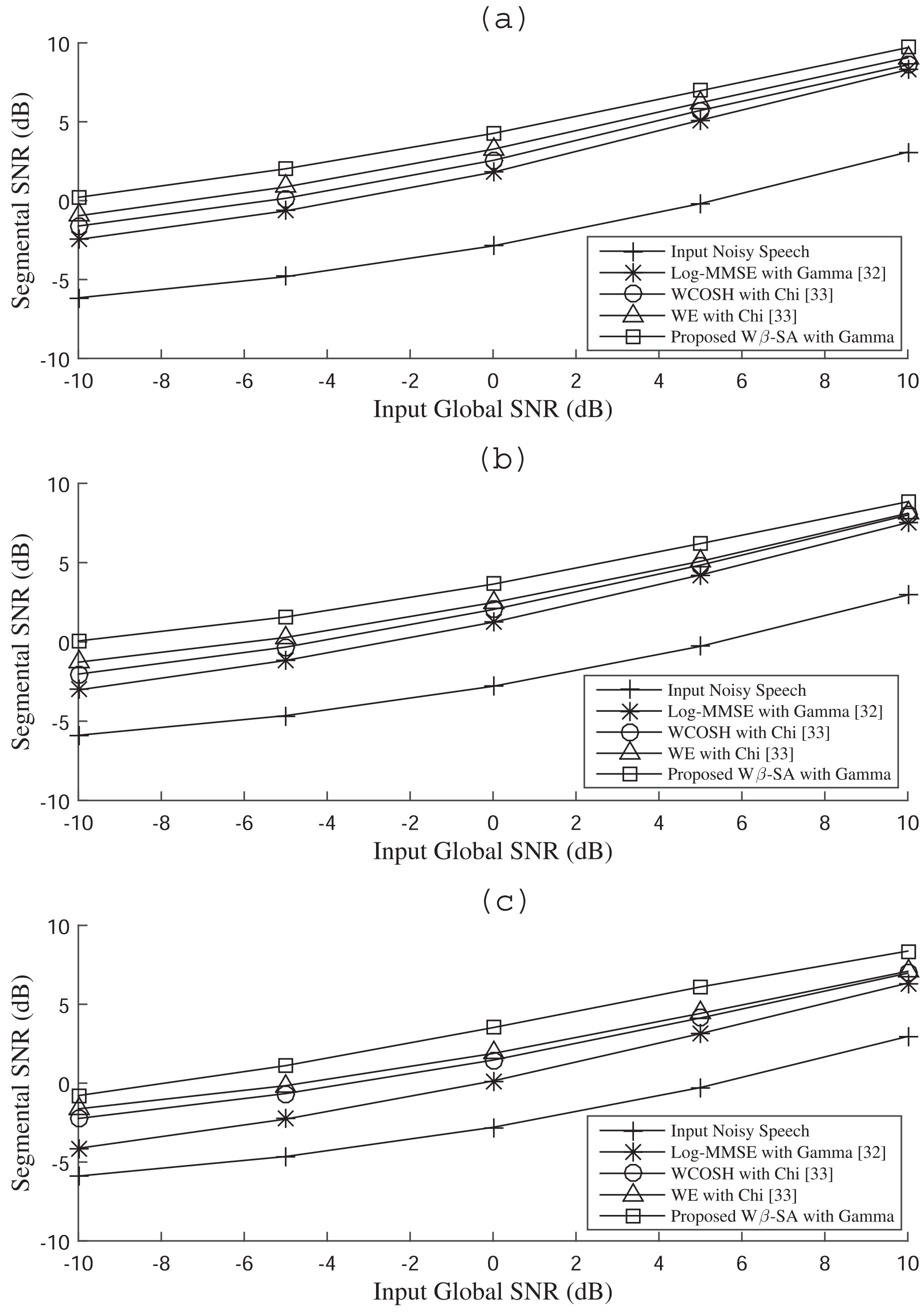

Fig. 14 Segmental SNR versus global SNR for the STSA estimators in [33, 34] and the proposed STSA estimator in Section 4, (**a**) white noise, (**b**) babble noise, and (**c**) car noise

$$\mathrm{NUM} = \underbrace{\frac{2b^{c}}{2\pi\,\Gamma(c)}\frac{1}{\pi\sigma_{v}^{2}}}_{K_1}\int_{0}^{\infty}\int_{0}^{2\pi}\chi^{m+2c-1}\exp\left(-b\chi^{2}\right) \times \exp\left(\frac{1}{\sigma_{v}^{2}}\left(|Y|^{2}+\chi^{2}-2|Y|\chi\cos\left(\psi-\Omega\right)\right)\right)d\Omega d\chi \tag{26}$$

with ψ as the phase of the complex observation Y. To further progress with (26), the integration with respect to Ω should be performed first. To this end, we may write

$$\mathrm{NUM} = \underbrace{K_1\exp\left(-\frac{|Y|^{2}}{\sigma_{v}^{2}}\right)}_{K_2}\int_{0}^{\infty}\chi^{m+2c-1}\exp\left(-b\chi^{2}\right) \times \exp\left(-\frac{\chi^{2}}{\sigma_{v}^{2}}\right)\Delta_{1}d\chi \tag{27}$$

with

$$\Delta_1 = \int_0^{2\pi} \exp\left(\frac{\chi |Y| \cos(\psi - \Omega)}{\sigma_\nu^2}\right) d\Omega \tag{28}$$

Further manipulation of Δ_1 results in

$$\Delta_1 = \pi \mathrm{I}_0\left(\frac{2\chi |Y|}{\sigma_\nu^2}\right) \tag{29}$$

with $\mathrm{I}_0(.)$ as the zero-order modified Bessel function of the first kind [22]. Now, by inserting (29) into (27) and using Equation (6.631-1) in [22] to solve the resulting integral, it follows

$$\mathrm{NUM} = \pi K_2 \frac{\Gamma\left(\frac{m+2c}{2}\right)}{\left(b + \frac{1}{\sigma_\nu^2}\right)^{\frac{m+2c}{2}}} \mathrm{M}\left(\frac{m+2c}{2}, 1; \nu'\right) \tag{30}$$

with ν' as defined in (20). Using the following property of the confluent hypergeometric function,

$$\mathrm{M}(x, y; z) = e^z \mathrm{M}(y - x, y; -z) \tag{31}$$

we further obtain

$$\mathrm{NUM} = \pi K_2 e^{\nu'} \frac{\Gamma\left(\frac{m+2c}{2}\right)}{\left(b + \frac{1}{\sigma_\nu^2}\right)^{\frac{m+2c}{2}}} \mathrm{M}\left(\frac{2-m-2c}{2}, 1; -\nu'\right) \tag{32}$$

where, according to Section 4.1, we have $b = c/\sigma_\chi^2$. Now, by considering $m = 0$ in the above, a similar expression is derived for DEN in (24). Division of the obtained expression of NUM by that of DEN results in Eq. (19).

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Electrical and Computer Engineering, Concordia University, 1455 De Maisonneuve Blvd. West, H3G 1M8 Montreal, Canada. [2]Department of Electrical and Computer Engineering, McGill University, 3480 University Street, H3A 0E9 Montreal, Canada. [3]Department of Electrical and Computer Engineering, Université de Sherbrooke, 2500 boul. de l'Université, J1K 2R1 Sherbrooke, Quebec, Canada.

Received: 21 October 2014 Accepted: 9 September 2015
Published online: 06 October 2015

References

1. PC Loizou, *Speech Enhancement: Theory and Practice*. (CRC Press, Boca Raton, FL, USA, 2007)
2. J Benesty, Y Huang, *Springer Handbook of Speech Processing*. (Springer, Secaucus, NJ, USA, 2008)
3. Y Ephraim, D Malah, Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. IEEE Trans. Acous. Speech and Sig. Process. **32**(6), 1109–1121 (1984)
4. Y Ephraim, D Malah, Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. IEEE Trans. Acous. Speech and Sig. Process. **33**(2), 443–445 (1985)
5. PJ Wolfe, SJ Godsill, Efficient alternatives to the Ephraim and Malah suppression rule for audio signal enhancement. EURASIP J. Adv. Sig. Process. **2003**, 1043–1051 (2003)
6. PC Loizou, Speech enhancement based on perceptually motivated Bayesian estimators of the magnitude spectrum. IEEE Trans. Speech and Audio Process. **13**(5), 857–869 (2005)
7. DE Tsoukalas, JN Mourjopoulos, G Kokkinakis, Speech enhancement based on audible noise suppression. IEEE Trans. Speech and Audio Process. **5**(6), 497–514 (1997)
8. N Virag, Single channel speech enhancement based on masking properties of the human auditory system. IEEE Trans. Speech and Audio Process. **7**(2), 126–137 (1999)
9. CH You, SN Koh, S Rahardja, β-order MMSE spectral amplitude estimation for speech enhancement. IEEE Trans. Speech and Audio Process. **13**(4), 475–486 (2005)
10. E Plourde, B Champagne, Auditory-based spectral amplitude estimators for speech enhancement. IEEE Trans. Audio, Speech, Lang. Process. **16**(8), 1614–1623 (2008)
11. E Plourde, B Champagne, Generalized Bayesian estimators of the spectral amplitude for speech enhancement. IEEE Signal Process. Letters. **16**(6), 485–488 (2009)
12. CH You, SN Koh, S Rahardja, Masking-based β-order MMSE speech enhancement. Speech Comm. **48**(1), 57–70 (2006)
13. E Kandel, J Schwartz, *Principles of Neural Science, Fifth Edition*. (McGraw-Hill Education, Secaucus, NJ, USA, 2013)
14. R Martin, Speech enhancement based on minimum mean-square error estimation and superGaussian priors. IEEE Trans. Speech and Audio Process. **13**(5), 845–856 (2005)
15. MB Trawicki, MT Johnson, Speech enhancement using Bayesian estimators of the perceptually-motivated short-time spectral amplitude (STSA) with Chi speech priors. Speech Comm. **57**(0), 101–113 (2014)
16. I Andrianakis, PR White, Speech spectral amplitude estimators using optimally shaped Gamma and Chi priors. Speech Comm. **51**(1), 1–14 (2009)
17. T Lotter, P Vary, Speech enhancement by MAP spectral amplitude estimation using a super-Gaussian speech model. EURASIP J. Appl. Sig. Process. **2005**, 1110–1126 (2005)
18. R Prasad, H Saruwatari, K Shikano, Probability distribution of time-series of speech spectral components. IEICE rans. Fundam. Electron. Commun. Comput. Sci. **E87-A**(3), 584–597 (2004)
19. I Andrianakis, Bayesian Algorithms for Speech Enhancement. PhD thesis, University of Southampton (2007). http://eprints.soton.ac.uk/66244/1.hasCoversheetVersion/P2515.pdf
20. JS Erkelens, RC Hendriks, R Heusdens, J Jensen, Minimum mean-square error estimation of discrete fourier coefficients with generalized Gamma priors. IEEE Trans. Audio, Speech, Lang. Process. **15**(6), 1741–1752 (2007)
21. BJ Borgstrom, A Alwan, A unified framework for designing optimal STSA estimators assuming maximum likelihood phase equivalence of speech and noise. IEEE Trans. Audio, Speech, Lang. Process. **19**(8), 2579–2590 (2011)
22. A Jeffrey, D Zwillinger, *Table of Integrals, Series, and Products*. (Elsevier Science, Boston, 2007)
23. DD Greenwood, A cochlear frequency-position function for several species–29 years later. The J. Acoust. Soc. Am. **87**(6), 2592–2605 (1990)
24. I Cohen, B Berdugo, Speech enhancement for non-stationary noise environments. Sig. Process. **81**(11), 2403–2418 (2001)
25. BL Sim, YC Tong, JS Chang, CT Tan, A parametric formulation of the generalized spectral subtraction method. IEEE Trans. Speech and Audio Process. **6**(4), 328–337 (1998)
26. I Cohen, Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging. IEEE Trans. Speech and Audio Process. **11**(5), 466–475 (2003)
27. NL Johnson, S Kotz, N Balakrishnan, *Continuous Univariate Distributions*. (Wiley & Sons, New York, 1995)
28. O Gomes, C Combes, A Dussauchoy, Parameter estimation of the generalized Gamma distribution. Math. Comput. Simul. **79**(4), 955–963 (2008)
29. Noisex-92 database. Speech at CMU, Carnegie Mellon University, available at: http://www.speech.cs.cmu.edu/comp.speech/Section1/Data/noisex.html. Accessed date Sept 2014
30. JS Garofolo, DARPA TIMIT acoustic-phonetic speech database. National Institute of Standards and Technology (NIST) (1988). https://catalog.ldc.upenn.edu/LDC93S1
31. Recommendation P.862: Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. ITU-T (2001). http://www.itu.int/rec/T-REC-P.862

32. Y Hu, PC Loizou, Evaluation of objective quality measures for speech enhancement. IEEE Trans. Audio, Speech, Lang. Process. **16**(1), 229–238 (2008)
33. BJ Borgstrom, A Alwan, in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. Log-spectral amplitude estimation with Generalized Gamma distributions for speech enhancement, (2011), pp. 4756–4759. doi:10.1109/ICASSP.2011.5947418
34. MB Trawicki, MT Johnson, Speech enhancement using Bayesian estimators of the perceptually-motivated short-time spectral amplitude (STSA) with Chi speech priors. Speech Comm. **57**(0), 101–113 (2014)