Research Article

An Improved MMSE Amplitude Estimator under Generalized Gamma Distribution Based on Auditory Perception

Chabane Boubakir^{1,2} and Daoud Berkani²

¹ Department of Electronics, Lamel Laboratory, University of Jijel, P.O. Box 98, Ouled Aissa, 18000 Jijel, Algeria

² Department of Electronics, Signal and Communications Laboratory, National Polytechnic School,
10 Avenue Hassen Badi, P.O. Box 182, 16200 El-harrach, Algeria

Correspondence should be addressed to Chabane Boubakir; chboubakir@yahoo.fr

Received 7 January 2013; Revised 2 March 2013; Accepted 16 March 2013

Academic Editor: Matjaz Perc

Copyright © 2013 C. Boubakir and D. Berkani. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper describes a new speech enhancement approach which employs the minimum mean square error (MMSE) estimator based on the generalized gamma distribution of the short-time spectral amplitude (STSA) of a speech signal. In the proposed approach, the human perceptual auditory masking effect is incorporated into the speech enhancement system. The algorithm is based on a criterion by which the audible noise may be masked rather than being attenuated, thereby reducing the chance of speech distortion. Performance assessment is given to show that our proposal can achieve a more significant noise reduction as compared to the perceptual modification of Wiener filtering and the gamma based MMSE estimator.

1. Introduction

Speech enhancement is concerned with improving the quality and intelligibility of speech signal. The need to enhance speech signals arises in many situations in which the speech signal originates from a noisy location or is affected by noise over a communication channel.

Speech enhancement methods are employed more and more often in applications such as mobile telephony, speech recognition, and human-machine communication systems [1–10]. Speech enhancement algorithms can therefore be used as a preprocessor in speech-coding systems employed in cellular phones. In the case of the speech recognition system, the noisy speech signal can be preprocessed by a speech enhancement method before being fed to the speech recognizer. In an air-ground communication scenario, as well as in similar communication systems used by the military, it is more desirable to enhance the intelligibility rather than the quality of speech [11]. Besides, a further possible application is the enhancement of mating sounds and bioacoustic signals before their analysis. Therefore, algorithms based on Wiener filter or spectral subtraction can be used to eliminate background noise and other sounds of nature not inherent to that of the animal, followed, for example, by a nonlinear

time series analysis methods to analyze the dynamics of the sound-producing apparatus of the animal [12].

In this paper, single-microphone speech enhancement is studied. One of the main approaches of speech enhancement algorithms is to obtain the best possible estimates of the short-time spectral amplitude of a speech signal from a given noisy speech. The performance of enhanced speech is characterized by a tradeoff between the amount of noise reduction, the speech distortion, and the level of musical residual noise.

Several methods have been proposed to reduce the residual noise. Ephraim and Malah [1, 2] used the conventional hypothesis that, for speech enhancement in the discrete Fourier transform (DFT) domain, the distribution of the complex speech DFT coefficients is Gaussian. Nowadays, super-Gaussian models of the DFT coefficients are used because they lead to estimators with improved performance compared to those based on a Gaussian model [3–5]. For example, the minimum mean square error estimators for the amplitudes, assuming a one-sided generalized gamma distribution, are studied in [6–9]. Experimental results showed that the gamma-based estimator had higher preference scores compared to the Gaussian-based estimator for various types of noise and at different noise levels.

It is very difficult to suppress residual noise without decreasing intelligibility and without introducing speech distortion and musical residual noise [13]. Several methods [13–17] attempted to reduce the musical residual noise by emulating the human auditory system, based on the fact that the human ears cannot perceive the additive noise when the noise level falls below the auditory masking threshold (AMT). These methods are predominantly based on spectral subtraction and Wiener filtering, which have exploited the masking properties of the human auditory system. However, they do not perform well at very low varying signal-to-noise ratio (SNR) and introduce a perceptually disturbing musical noise especially with colored and nonstationary noise.

In this work, the human perceptual auditory masking effect is incorporated into the estimator based on the gamma model in order to obtain a more accurate estimate and achieve an effective suppression of noise as well as minimal musical tones in the residual signal. This study is followed by numerical simulations of these algorithms and an objective evaluation using a corpus of speech.

The rest of this paper is organized as follows. Section 2 describes the gamma-based short-time spectral amplitude estimator with some details. Section 3 presents our proposed enhancement method, and Section 4 demonstrates our implementations and results. Conclusions are finally drawn in Section 5.

2. Gamma-Based Short-Time Spectral Amplitude Estimator

The noisy speech signal is given by the following:

$$y(n) = x(n) + d(n), \quad (1)$$

where $x(n)$ is the clean speech signal which is assumed to be independent of the additive noise $d(n)$. Their representation in the short time fourier transform (STFT) domain is given by

$$Y(k, l) = X(k, l) + D(k, l), \quad (2)$$

where $Y(k, l)$, $X(k, l)$, and $D(k, l)$ are the samples of the noisy speech, the clean speech, and the noise signal's STFT correspondingly. The index k corresponds to the frequency bins and the index l to the time frames of the STFT. Since DFT coefficients from different time frames and frequency indices are assumed to be independent, the indices k and l will be sometimes omitted for simplicity. We can write $X = Ae^{j\Phi}$ and $Y = Re^{j\Theta}$, where random variables A and R represent the magnitude of the clean speech DFT coefficient and the noisy speech DFT coefficient, respectively, and Φ and Θ represent the corresponding phases values. We use upper case letters to denote random variables and the corresponding lower case letters to denote their realizations.

In this paper, we focus on the minimum mean square error estimation of the clean magnitude $A(k, l)$. The MMSE estimate of $A(k, l)$ is the expectation of the clean magnitude

conditional on the noisy magnitude $r(k, l)(E\{A/r\})$. With Bayes formula we can express $\hat{A}$ as follows:

$$\hat{A}(k, l) = E\{A/r\} = \frac{\int_0^\infty a f_{R/A}(r/a) f_A(a) da}{\int_0^\infty f_{R/A}(r/a) f_A(a) da}. \quad (3)$$

The estimation of the clean magnitude $A(k, l)$ requires some assumptions about the distribution of the speech and the noise. The speech has usually been assumed Gaussian, for example, [1, 2], but in recent times estimators based on super-Gaussian speech assumptions such as Laplacian or gamma distributions have been derived [8]. A similar development has been seen for the noise assumptions; most commonly, the noise is assumed Gaussian [8].

With the zero-mean Gaussian distribution assumption of the noise DFT coefficients, $f_{R/A}(r/a)$ can be written as follows [18]:

$$f_{R/A}(r/a) = \frac{2r}{\sigma_D^2(k, l)} \exp\left(-\frac{r^2 + a^2}{\sigma_D^2(k, l)}\right) I_0\left(\frac{2ar}{\sigma_D^2(k, l)}\right), \quad (4)$$

where I_0 is the 0th-order-modified Bessel function of the first kind, and $\sigma_D^2(k, l) = E\{|D(k, l)|^2\}$ is the noise spectral variance.

In the gamma-based MMSE estimators of the speech DFT magnitudes, we assume that the speech DFT magnitudes are distributed according to a one-sided generalized gamma prior density of the form

$$f_A(a) = \frac{\gamma \beta^\gamma}{\Gamma(\gamma)} a^{\gamma-1} \exp(-\beta a^\gamma), \quad a \geq 0, \quad (5)$$

where $\Gamma(\cdot)$ is the Gamma function and the random variable A represents the DFT magnitudes, with the constraints on the parameters $\beta > 0$, $\gamma > 0$, and $\nu > 0$.

The gamma-based MMSE magnitude estimators for the cases $\gamma = 1$ and $\gamma = 2$ have been derived in [6, 7, 19]. We will use the case $\gamma = 2$, as the related estimator can be derived without any approximations, and the maximum achievable performance for both cases is about the same.

Inserting (5) with $\gamma = 2$ and (4) into (3) give

$$\begin{aligned} \hat{A}(k, l) &= \frac{\int_0^\infty a^{2\gamma} \exp(-a^2/(\sigma_D^2(k, l)) - \beta a^2) I_0(2ar/(\sigma_D^2(k, l))) da}{\int_0^\infty a^{2\gamma-1} \exp(-a^2/(\sigma_D^2(k, l)) - \beta a^2) I_0(2ar/(\sigma_D^2(k, l))) da} \\ &= \frac{\int_0^\infty a^{2\gamma} \exp(-a^2/(\sigma_D^2(k, l)) - \beta a^2) I_0(2ar/(\sigma_D^2(k, l))) da}{\int_0^\infty a^{2\gamma-1} \exp(-a^2/(\sigma_D^2(k, l)) - \beta a^2) I_0(2ar/(\sigma_D^2(k, l))) da}. \end{aligned} \quad (6)$$

Using [20, Theorem 6.643.2], the integrals can be solved for $\nu > 0$. After inserting the relation between β and the second-moment $E\{A^2(k, l)\}$, which for this case is $\beta = \nu/\sigma_X^2(k, l)$, with $\sigma_X^2(k, l) = E\{|X(k, l)|^2\}$, the estimator is as follows [7]:

$$\hat{A}(k, l) = \frac{\Gamma(\nu + 0.5)}{\Gamma(\nu)} \frac{\sqrt{Q(k, l)}}{\gamma(k, l)} \frac{M_{(-\nu, 0)}(Q(k, l))}{M_{(-\nu+0.5, 0)}(Q(k, l))} r(k, l), \quad (7)$$

FIGURE 1: Block diagram of the proposed speech enhancement technique.

where $Q(k, l) = (\gamma(k, l)\eta(k, l))/(\nu + \eta(k, l))$; $\gamma(k, l) = R^2(k, l)/E\{|D(k, l)|^2\}$; $\eta(k, l) = E\{|X(k, l)|^2\}/E\{|D(k, l)|^2\}$ with $\eta(k, l)$ is called the a priori SNR, $\gamma(k, l)$ is the a posteriori SNR, and $M_{\nu, \mu}$ is recognized as the Whittaker function or in terms of confluent hypergeometric function ${}_1F_1(a; b; x)$ [20, Equations 9.210.1 and 9.220.2]:

$$\hat{A}(k, l) = \frac{\Gamma(\nu + 0.5)}{\Gamma(\nu)} \frac{\sqrt{Q(k, l)}}{\gamma(k, l)} \frac{{}_1F_1(\nu + 0.5; 1; Q(k, l))}{{}_1F_1(\nu; 1; Q(k, l))} r(k, l). \quad (8)$$

The special case $\nu = 1$ is the traditional MMSE-STSA estimator derived in [1].

In order to evaluate the previous gain functions, we must first estimate the noise power spectrum $\lambda_d(k, l) = \sigma_D^2(k, l) = E\{|D(k, l)|^2\}$. This is often done during periods of speech absence as determined by a voice activity detector (VAD), by using a noise-estimation algorithm such as the minimum statistics approach [21, 22], or by using a real noise in comparative studies.

The a posteriori SNR estimator $\gamma(k, l)$ is the ratio of the squared noisy magnitude $R^2(k, l)$ and the estimated noise power spectrum. Furthermore, we use the decision-directed approach for the estimation of the a priori SNR like in [1, 2, 23]. Thus, $\hat{\eta}_k$ is given by the following:

$$\hat{\eta}_k(k, l) = \max \left[\alpha \frac{\hat{A}^2(k, l-1)}{\lambda_d(k, l-1)} + (1-\alpha) [\gamma(k, l) - 1], \eta_{\min} \right], \quad (9)$$

where the smoothing factor is $0 \leq \alpha \leq 1$. A value of $\alpha = 0.98$ was used in the implementation and the lower limit $\eta_{\min}$ recommended in [23], is similar to the use of the spectral floor in the basic spectral subtraction method [24]. A lower limit of at least -15 dB is recommended.

3. Proposed Enhancement Technique

Figure 1 contains the flowchart of the proposed speech enhancement scheme, which consists of different steps as described below.

- (i) Spectral decomposition: windowing + fast fourier transform (FFT).
- (ii) Speech/noise detection and noise estimation.
- (iii) Rough estimation of the speech magnitude spectra $\hat{X}(k, l)$.
- (iv) Calculation of the auditory masking threshold $T(k, l)$.
- (v) Calculation of the enhanced speech magnitude spectra $\hat{A}(k, l)$ using (7)–(9).
- (vi) Calculation of the enhanced speech signal $\hat{x}(n, l)$ based on the following equation:

$$\hat{x}(n, l) = \text{IFFT} \left\{ \hat{A}(k, l) \exp \left[j\Theta_y(k, l) \right] \right\}. \quad (10)$$

3.1. Auditory Masking Threshold (AMT) Calculation. The auditory masking threshold is obtained through modeling the frequency selectivity of the human ear and its masking property. This paper only considers the simultaneous masking. Before computing the auditory masking threshold, the speech spectra must be estimated. A spectral subtraction or a Wiener filter is used for obtaining a rough estimate of the speech spectra. The speech magnitude spectra estimated by spectral subtraction is given by the following:

$$\hat{X}(k, l) = \max \left(Y(k, l) - \hat{D}(k, l), \varepsilon \right), \quad (11)$$

where ε is a small positive value.

Once $\hat{X}(k, l)$ is obtained, the auditory masking threshold $T(k, l)$ can be calculated based on the Johnston model [25]; then the gamma-based short-time spectral amplitude estimator can be applied.

4. Performance Evaluation

This section presents the performance evaluation of the proposed enhancement algorithm as well as a comparison with two other estimators. The first one is the gamma-based MMSE estimator presented in Section 2 which does not take the masking properties into account. The second one is the estimator based on a perceptual modification of the generalized Wiener filtering, which was proposed by Lin et al. in [15].

For evaluation purposes, we used the noisy speech signal taken from the Noizeus database [26], which consists of 30 speech signals sampled at 8 kHz, contaminated by eight different real-world noises at different SNRs (babble, car, exhibition hall, restaurant, street, airport, train station, train). The frame size is 256 samples, with an overlap of 50%, and the data window used was a Hanning window, while the total number of critical bands was $K = 18$. The enhanced signal was combined using the overlap and add method. The initial noise variance was estimated from 0.64 seconds of noise only, preceding speech activity.

MATLAB implementations available from [27] have been used to evaluate the confluent hypergeometric functions.

To measure the quality of the enhanced signal, we have used the segmental SNR, the weighted spectral slope measure (WSS), and the perceptual evaluation of speech quality (PESQ) [28–30]. All the measures show high correlation with subjective quality assessments.

The WSS measure is based on an auditory model and finds a weighted difference between the spectral slopes in each band. The magnitude of each weight reflects whether the band is near a spectral peak or valley and whether the peak is the largest in the spectrum. One implementation of the WSS measure can be defined as follows:

$$d_{\text{WSS}} = \frac{1}{M} \sum_{l=0}^{M-1} \frac{\sum_{j=1}^K W(j, l) (S_c(j, l) - S_p(j, l))^2}{\sum_{j=1}^K W(j, l)}, \quad (12)$$

where K is the number of bands, M is the total number of frames, and $S_c(j, l)$ and $S_p(j, l)$ are the spectral slopes (typically the spectral differences between neighboring bands) of the j th band in the l th frame for clean and processed speech signals, respectively. $W(j, l)$ are weights, which can be calculated as shown by Klatt in [28]. The highest 5% of the WSS measure values were discarded, as suggested in [29], to exclude unrealistically high spectral distance values. The lower the WSS measure for an enhanced speech, the better is its perceived quality.

Segmental SNR is based on the classical SNR and it is one of the most widely used methods for testing enhancement algorithms. Since the correlation of classical SNR with subjective quality is so poor. Instead, we choose the frame-based segmental SNR by averaging the frame level SNR estimates defined by [29]

$$\text{SNR}_{\text{seg}} = \frac{10}{M} \sum_{l=0}^{M-1} \log_{10} \frac{\sum_{n=Nl}^{Nl+N-1} x_{\phi}^2(n)}{\sum_{n=Nl}^{Nl+N-1} [x_p(n) - x_{\phi}(n)]^2}, \quad (13)$$

FIGURE 2: Performance in terms of mean segmental SNR, PESQ, and WSS measures as a function of input SNR (dB) for white gaussian noise of the noisy speech signal, the speech signal enhanced by the gamma-based MMSE estimator, the perceptual Wiener filtering, and the proposed estimator. Higher values of the segmental SNR and PESQ and lower WSS measures indicate better performances.

where $x_{\phi}(n)$ and $x_p(n)$ represent the clean and processed speech sample, respectively, N denotes the frame length, and M is the number of frames. The lower and upper thresholds are selected to be -10 dB and $+35$ dB, respectively.

The perceptual evaluation of speech quality (PESQ) measure, described in [30], was selected as the ITU-T recommendation P.862 [31]. PESQ measure is one of the most commonly used measures to predict the subjective opinion score of a degraded or enhanced speech. In the PESQ measure, a reference signal and the enhanced signal are first aligned in both time and level. This is followed by a range of perceptually significant transforms which include Bark spectral analysis, frequency equalization, gain variation equalization, and loudness mapping. The difference, termed the disturbance, between the loudness spectra is computed and averaged over time and frequency to produce the prediction of subjective MOS score. The PESQ score ranges from 1.0 (worst) to 4.5 (best), with higher scores indicating better quality [32].

Figures 2, 3, and 4 show plots of mean results in terms of segmental SNR, PESQ, and WSS measures, for 30 Noizeus sentences corrupted by white, babble, and car noise, respectively, at 0–15 dB SNR.

As can be seen, the proposed method outperforms the two estimators in terms of SNR segmental, PESQ, and WSS

FIGURE 3: Performance in terms of mean segmental SNR, PESQ, and WSS measures as a function of input SNR (dB) for Babble noise of the noisy speech signal, the speech signal enhanced by the gamma-based MMSE estimator, the perceptual Wiener filtering, and the proposed estimator. Higher values of the segmental SNR and PESQ and lower WSS measures indicate better performances.

measures for all SNR values. The improvement in the higher input SNRs is more noticeable because of the more accurate AMT calculation.

Table 1 presents an example of the objective results obtained for noisy speech and enhanced speech with the three estimators, in the case of white, babble and car noise at 0 dB SNR and 5 dB SNR.

From the objective results (Figures 2, 3, and 4, Table 1), it can be seen that the proposed estimator has higher preference scores compared to the two other estimators for all noise levels from 0 dB to 15 dB SNR. Furthermore, informal listening tests confirmed that the proposed estimator yields better quality with significantly lower noise distortion than the gamma-based estimator and comparable quality with the perceptual Wiener estimator.

5. Conclusions

In this paper, a gamma-based minimum mean square error estimator for speech enhancement incorporating masking properties was proposed. We showed an increase in the quality of the enhanced speech with different noise types.

Results, in terms of objective measures and listening tests, indicated that the proposed approach has a better tradeoff between the amount of noise reduction, the speech distortion,

FIGURE 4: Performance in terms of mean segmental SNR, PESQ, and WSS measures as a function of input SNR (dB) for car noise, of the noisy speech signal, the speech signal enhanced by the gamma-based MMSE estimator, the perceptual Wiener filtering, and the proposed estimator. Higher values of the segmental SNR and PESQ and lower WSS measures indicate better performances.

TABLE 1: Objective quality scores for various algorithms under white, babble and car noise, SNR = 0 dB and 5 dB.

Noise	Method	SNR = 0 dB			SNR = 5 dB		
		SNRseg	PESQ	WSS	SNRseg	PESQ	WSS
White	Noisy	-5.081	1.539	82.69	-2.327	1.799	69.97
	Gamma	2.769	2.545	56.86	5.135	2.901	46.49
	WienerP	3.4101	2.961	43.69	5.802	3.301	35.53
	Proposed	3.8214	2.981	41.03	6.435	3.382	33.92
Babble	Noisy	-4.632	1.705	70.35	-1.783	2.006	56.02
	Gamma	2.218	2.587	51.47	4.595	2.958	41.59
	WienerP	2.946	2.855	40.98	5.282	3.173	32.84
	Proposed	2.963	2.897	40.42	5.576	3.194	30.17
Car	Noisy	-4.959	1.634	66.94	-2.173	1.891	54.08
	Gamma	2.191	2.527	49.74	4.495	2.870	40.95
	WienerP	2.812	2.858	40.21	5.119	3.168	32.87
	Proposed	2.841	2.889	39.42	5.428	3.210	30.38

and the level of musical residual noise than perceptual Wiener filter and gamma-based estimator.

The implementation of the estimator based on gamma speech modeling requires the evaluation of the Gamma and the confluent hypergeometric functions, in addition to the AMT computation in the proposed estimator. In a real-time

implementation, these functions can be stored in a table. The computation of these estimators during runtime will not be then much more complex than that of the perceptual Wiener estimator.

Based on the previous findings, using noise masking properties to adapt a speech enhancement system was beneficial to improve the minimum mean square error amplitude estimator under generalized gamma distribution.

In the future, we plan to evaluate its possible application in preprocessing for new communication systems and hearing aids.

References

- [1] 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. 32, no. 6, pp. 1109–1121, 1984.
- [2] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error-log-spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 443–445, 1985.
- [3] R. Martin, "Speech enhancement based on minimum mean-square error estimation and supergaussian priors," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 845–856, 2005.
- [4] T. Lotter and P. Vary, "Speech enhancement by MAP spectral amplitude estimation using a super-Gaussian speech model," *EURASIP Journal on Applied Signal Processing*, vol. 2005, no. 7, pp. 1110–1126, 2005.
- [5] J. Jensen, R. C. Hendriks, J. S. Erkelens, and R. Heusdens, "MMSE estimation of complex-valued discrete fourier coefficients with generalized gamma priors," in *Proceedings of the 9th International Conference on Spoken Language Processing (INTERSPEECH '06)*, pp. 257–260, September 2006.
- [6] I. Andrianakis and P. R. White, "MMSE speech spectral amplitude estimators with Chi and Gamma speech priors," in *Proceedings of the 31st IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '06)*, pp. 1068–1071, Toulouse, France, May 2006.
- [7] R. C. Hendriks, J. S. Erkelens, J. Jensen, and R. Heusdens, "Minimum mean-square error amplitude estimators for speech enhancement under the generalized Gamma distribution," in *Proceeding of the International Workshop on Acoustic Echo and Noise Control (IWAENC '06)*, Paris, France, 2006.
- [8] T. Lotter and P. Vary, "Noise reduction by joint maximum a posteriori spectral amplitude and phase estimation with super-Gaussian speech modeling," in *Proceeding of the 12th European Signal Processing Conference*, pp. 1457–1460, Vienna, Austria, 2004.
- [9] C. Boubakir and D. Berkani, "Speech enhancement using minimum mean-square error amplitude estimators under normal and generalized gamma distribution," *Journal of Computer Science*, vol. 6, no. 7, pp. 700–705, 2010.
- [10] C. Boubakir and D. Berkani, "On the use of Kalman filter for enhancing speech corrupted by colored noise," *WSEAS Transactions on Signal Processing*, vol. 4, no. 12, pp. 657–666, 2008.
- [11] P. C. Loizou, *Speech Enhancement, Theory and Practice*, Taylor & Francis, New York, NY, USA, 1st edition, 2007.
- [12] T. P. Benko and M. Perc, "Nonlinearities in mating sounds of American crocodiles," *BioSystems*, vol. 97, no. 3, pp. 154–159, 2009.
- [13] N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 2, pp. 126–137, 1999.
- [14] D. E. Tsoukalas, J. N. Mourjopoulos, and G. Kokkinakis, "Speech enhancement based on audible noise suppression," *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 6, pp. 497–514, 1997.
- [15] L. Lin, W. H. Holmes, and E. Ambikairajah, "Speech denoising using perceptual modification of Wiener filtering," *Electronics Letters*, vol. 38, no. 23, pp. 1486–1487, 2002.
- [16] Y. Hu and P. C. Loizou, "Incorporating a psychoacoustical model in frequency domain speech enhancement," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 270–273, 2004.
- [17] L. Gong, C. Chen, Q. Chen, and H. Xu, "An Improved LSA-MMSE speech enhancement approach based on auditory perception," in *Proceedings of the International Seminar on Future Information Technology and Management Engineering (FITME '08)*, pp. 292–295, Leicestershire, UK, November 2008.
- [18] R. J. McAulay and M. L. 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, 1980.
- [19] J. Erkelens, J. Jensen, and R. Heusdens, "Improved speech spectral variance estimation under the generalized Gamma distribution," in *Proceedings of the SPS-DARTS of the third annual IEEE BENELUX/DSP valley signal processing symposium*, pp. 43–46, Antwerp, Belgium, 2007.
- [20] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series and Products*, Academic, New York, NY, USA, 6th edition, 2000.
- [21] R. Martin, "Spectral subtraction based on minimum statistics," in *Proceedings of the European Signal Processing Conference (EUSIPCO '94)*, pp. 1182–1185, Edinburgh, UK, 1994.
- [22] 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, 2001.
- [23] O. Cappé, "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, 1994.
- [24] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '79)*, pp. 208–211, Washington, DC, USA, 1979.
- [25] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE Journal on Selected Areas in Communications*, vol. 6, no. 2, pp. 314–323, 1988.
- [26] <http://www.utdallas.edu/~loizou/speech/noizeus/>.
- [27] Matlab routines for computation of special functions, <http://jin.ece.illinois.edu/routines/routines.html>.
- [28] D. H. Klatt, "Prediction of perceived phonetic distance from critical-band spectra: a first step," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '82)*, vol. 2, pp. 1278–1281, 1982.
- [29] J. Hansen and B. Pellom, "An effective quality evaluation protocol for speech enhancement algorithm," in *Proceeding of the 5th International Conference on Spoken Language Processing (ICSLP '98)*, vol. 7, pp. 2819–2822, Sydney, Australia, 1998.

- [30] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ)—a new method for speech quality assessment of telephone networks and codecs," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 749–752, Salt Lake City, Utah, USA, May 2001.
- [31] ITU P.862, "Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs," ITU-T Recommendation P.862, 2000.
- [32] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 1, pp. 229–238, 2008.