Bin ZHOU, Xiongwei ZHANG, Xia ZOU, Gaihua ZHAO

PLA University of Science and Technology

Speech Enhancement by Short-Time Spectrum Estimation with Multivariate Laplace Speech Model

***Abstract**. The paper presents a new short-time spectrum estimation algorithm for speech enhancement. A novel multivariate Laplace speech model is utilized to characterize the dependencies between adjacent DFT coefficients of speech, based on which a minimum mean-square error (MMSE) estimator of speech spectral components is derived. Moreover, the speech presence uncertainty is incorporated to modify the MMSE estimator. Experimental results show that the developed algorithm achieves better noise suppression and lower speech distortion compared to the existing speech enhancement methods.*

***Streszczenie.** W artykule przedstawiono nowy algorytm estymacji krótkookresowego spektrum głosu do poprawy dźwięku mowy. Wykorzystano wieloczynnikowy model Laplace'a w celu scharakteryzowania zależności pomiędzy składnikami DFT dźwięku mowy. Na tej podstawie obliczane jest minimum błędu średnio-kwadratowego dla estymatora. Wyniki eksperymentalne potwierdzają ulepszoną skuteczność eliminacji zakłóceń mowy, w porównaniu ze stosowanymi metodami. (**Wieloczynnikowy model mowy Laplace'a w estymatorze spektrum krótkookresowego, na potrzeby polepszenia dźwięku**)*

Keywords: speech enhancement; minimum mean-square error (MMSE); multivariate Laplace distribution.
Słowa kluczowe: polepszenie dźwięku mowy, minimum błędu średniokwadratowego, wieloczynnikowy rozkład Laplace'a.

1 Introduction

Speech enhancement plays an important role in speech processing systems and is very useful in many applications such as speech communication and automatic speech recognition. As speech signal is often corrupted by noise during acquisition or transmission, the underlying goal of speech enhancement is to obtain the noise-free speech from the corrupted speech. Although various approaches to this problem have been adopted during the last decades, it is still one of the most fundamental, widely studied, and largely unsolved problems in speech processing.

Single channel speech enhancement methods based on short-time spectrum estimation have received significant interest due to the low complexity and relatively good performance. Fig. 1 presents the block diagram of a typical speech enhancement system based on short-time spectrum estimation. The clean speech signal $s(n)$ is mixed with additive background noise $d(n)$ to give the noise-corrupted speech signal $y(n)$. After segmentation and windowing with a function $h(n)$, e.g., Hamming window, the noisy speech is enhanced in the short-time Fourier transform (STFT) domain using a spectral gain function. The enhanced speech is then reconstructed from the inverse transformed frames using overlap-add (OLA) synthesis.

In the statistically motivated short-time spectrum estimation, the recovery of the underlying noise-free speech spectral coefficients from the noisy speech is generally treated as a Bayesian problem, where the statistical priors of speech and noise are modeled appropriately. For instance, in the well-known minimum mean square-error (MMSE) estimator derived in [1], the short-time spectral amplitudes of speech are restored assuming that the speech DFT coefficients are Gaussian distributed. Further results show that super-Gaussian priors are much better models for speech spectral components than the Gaussian priors [2]. Therefore, a number of speech enhancement algorithms with super-Gaussian priors have been developed, e.g., [3], [4], [5]. More recently, the generalized Gamma distribution is utilized to model speech spectral components, leading to more flexible spectrum estimators, e.g., [6], [7], [8].

Although those aforementioned methods have been successfully applied to improve the performance of speech enhancement, most of them are still based on the traditional assumption of independence, i.e., they assume that the speech spectral components are independent with each other. However, this assumption is inexact and there is some correlation between speech spectral coefficients in practice, mainly due to the effect of short-time window and the harmonicity of voiced speech [9]. Several methods have been proposed to solve this problem. In [9], a block-based linear MMSE estimator was developed to exploit the mutual correlations between spectral coefficients. Instead of assuming the spectral components to be independent, the method takes the spectral and temporal correlations into account by ways of an improved model for signal covariance matrix. Besides, a multidimensional short-time spectral amplitude estimator was derived in [10], in which the correlated spectral components were estimated jointly using a novel multidimensional Bayesian estimator. However, the closed-form solution was not given in the paper due to the complexity. Taking these factors into consideration, we propose a new speech enhancement method with a multivariate Laplace speech prior. Base on the assumption that the adjacent speech spectral coefficients can be modeled approximately by multivariate Laplace distribution, a new MMSE estimator that is able to exploit the spectral dependencies is derived analytically. Moreover, the speech presence uncertainty is also involved to further improve the performance.

The remainder of the paper is organized as follows. Section 2 gives an overview of the statistical framework of short-time spectrum estimation. Section 3 introduces the multivariate Laplace distribution as a new speech model. The proposed speech enhancement algorithm is presented in section 4, including the MMSE estimator and the speech presence possibility, both of which are derived based on the multivariate Laplace speech model. The experimental results and performance evaluations are given in section 5 and finally, section 6 presents our conclusion.

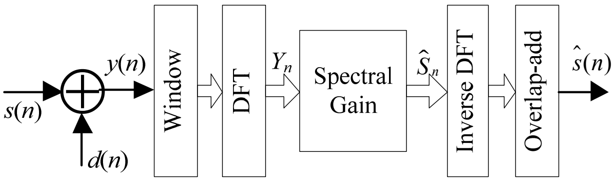

Fig.1. Block diagram of short-time spectrum estimation for speech enhancement.

2 Short-time spectrum estimation

Let $y(n)=s(n)+d(n)$ be the noisy speech signal consisting of the clean speech $s(n)$ and the additive noise $d(n)$. Taking the short-time Fourier transform (STFT) of $y(n)$, the DFT coefficients of the noisy speech at frequency bin k and frame l are given by

$$(1) \qquad Y_k(l) = S_k(l) + D_k(l)$$

where $S_k(l)$, $D_k(l)$ represent the DFT coefficients of the clean speech and the noise respectively. It is assumed conventionally that $S_k(l)$ and $D_k(l)$ are statistically independent across time and frequency, which allows to dropping the time and frequency indices l, k for the sake of readability, i.e.,

$$(2) \qquad Y = S + D$$

The objective of speech enhancement is to estimate the noise-free speech coefficients S from the noisy coefficients Y. The optimal estimate of S, in the sense of MMSE, is given as

$$(3) \qquad \hat{S} = E\{S \mid Y\} = \int_{-\infty}^{+\infty} S p_{S|Y}(S \mid Y) dS$$

where $p_{S|Y}(S|Y)$ is the probability distribution function (PDF) of clean speech coefficients S conditioned on noisy speech coefficients Y. Furthermore, given the assumed independence of the real and the imaginary parts of the complex DFT coefficients, the MMSE estimator in (3) may be split into the estimators for the real and the imaginary parts[3],

$$(4) \qquad \hat{S} = E\{S \mid Y\} = E\{S_R \mid Y_R\} + jE\{S_I \mid Y_I\}$$

where the subscripts R and I denote the real and imaginary parts of a complex variable respectively. Since the two parts in (4) can be treated independently in a similar procedure, we will only focus on the estimation of real part, $E\{S_R|Y_R\}$ in the following.

To derive the MMSE estimator in the Bayesian framework, it is required to assume appropriate statistical priors for the clean speech and the noise. Without loss of generality, the noise is often modeled as Gaussian priors, i.e., $D\sim N(D;\, 0,\, \sigma_D^2)$. Various priors have been developed to model the clean speech. For example, in the well-known Wiener estimator, the Gaussian prior is utilized to model speech spectral components motivated by the central limit theorem [1]. Thus,

$$(5) \qquad \widehat{S_R} = \frac{\sigma_S^2}{\sigma_S^2 + \sigma_D^2} Y_R$$

where σ_S^2 and σ_D^2 are the variance of the spectral components of the speech and the noise, respectively.

Considering the fact that the distribution of speech spectral coefficients is more super-Gaussian rather than Gaussian, a Laplace speech prior is applied in spectrum estimation, that is

$$(6) \qquad p_S(S_R) = \frac{1}{\sigma_S} \exp\left(-\frac{2|S_R|}{\sigma_S}\right)$$

and then under the additive white Gaussian noise assumption, the MMSE estimation of S_R is

$$(7)\ \widehat{S_R} = Y_R + \frac{\sigma_D}{\sqrt{\xi}} \frac{\exp\left(\frac{2Y_R}{\sigma_D\sqrt{\xi}}\right)\mathrm{erfc}(\lambda_+) - \exp\left(-\frac{2Y_R}{\sigma_D\sqrt{\xi}}\right)\mathrm{erfc}(\lambda_-)}{\exp\left(\frac{2Y_R}{\sigma_D\sqrt{\xi}}\right)\mathrm{erfc}(\lambda_+) + \exp\left(-\frac{2Y_R}{\sigma_D\sqrt{\xi}}\right)\mathrm{erfc}(\lambda_-)}$$

where $\lambda_+=\sigma_D/\sigma_S$, $\lambda_-=\sigma_D/\sigma_S - Y_R/\sigma_D$, $\xi = \sigma_S^2 / \sigma_D^2$ denotes *a prior* signal-to-noise ratio (SNR) , and $\mathrm{erfc}(\cdot)$ denotes the complementary error function [3].

3 Multivariate Laplace distribution

The PDF of a Laplace distributed random vector $\mathbf{x}$ is given as follows [11],

$$(8) \qquad p_{\mathbf{x}}(\mathbf{x}) = \frac{1}{\pi\sigma^2}\left(\frac{1}{\sqrt{2\pi\sigma}\|\mathbf{x}\|}\right)^{d/2-1} K_{d/2-1}\left(\frac{\sqrt{2}}{\sigma}\|\mathbf{x}\|\right)$$

where d is the dimension of $\mathbf{x}$, σ^2 is the variance of the Laplace marginal distribution of $\mathbf{x}$, and $K_\lambda(t)$ denotes the modified Bessel function of the second kind, which is defined as

$$(9) \qquad K_\lambda(t) = \frac{1}{2}\left(\frac{t}{2}\right)^\lambda \int_0^\infty a^{-\lambda-1} \exp\left(-a - \frac{t^2}{4t}\right) da$$

In order for derivation, the Gaussian scale mixture (GSM) representation of a Laplace random vector $\mathbf{x}$ is usually used in practice [12], i.e.,

$$(10) \qquad \mathbf{x} = \sqrt{z}\mathbf{u}\,, \qquad \mathbf{x}, \mathbf{u} \in \mathbb{R}^d, \quad z \in \mathbb{R}\ with\ z \geq 0$$

where $\mathbf{u}$ is a d-dimensional zero-mean Gaussian random vector with covariance matrix $\sigma^2 \boldsymbol{I_d}$,

$$(11) \qquad p_{\mathbf{u}}(\mathbf{u}) = \frac{1}{(2\pi\sigma^2)^{d/2}} \exp\left(-\frac{\|\mathbf{u}\|^2}{2\sigma^2}\right)$$

and z is a unit mean exponential random variable,

$$(12) \qquad p_z(z) = \exp(-z)\,, \quad z \geq 0$$

Therefore, the PDF of $\mathbf{x}$ can be expressed as

$$(13) \qquad p_{\mathbf{x}}(\mathbf{x}) = \int_0^\infty p_a(a) \frac{1}{a^d} p_{\mathbf{u}}\left(\frac{\mathbf{x}}{a}\right) da$$

where $a=\sqrt{z}$, and the PDF of a is

$$(14) \qquad p_a(a) = 2ap_z(a^2) = 2a\exp(-a^2), \qquad a \geq 0\,.$$

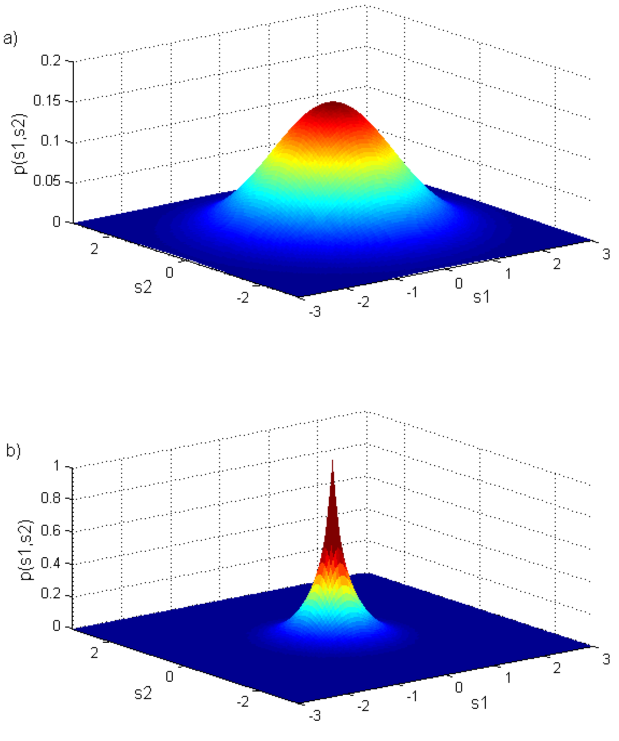

Fig.1. (a) Multivariate Laplace distribution versus (b) multivariate Gassuian distribution with d = 2.

The comparison of multivariate Laplace distribution and multivariate Gaussian distribution for d=2 is illustrated in Fig. 1. It is clear that the multivariate Laplace distribution is more peaked and heavy-tailed compared to the multivariate Gaussian distribution, and thus better fitted to the histogram of speech spectral coefficients. In addition, as a random vector composed of a set of random variables, it is able to characterize the dependencies between the spectral components implicitly.

4 The Proposed algorithm

In this section, we derive the MMSE estimator and the speech prensence uncertainty under the multivarite Laplace speech model. Taking into account the assumption that the real and imaginary parts of speech DFT coefficients, S_R and S_I, are statistically independent, we will only consider the estimation of S_R; a similar procedure can be followed for S_I. To enhance the brevity of the following results, we will drop the subscripts R and I.

4.1 MMSE estimator with multivariate Laplace speech prior

we assume that a d-component vector $\mathbf{S}$, composed of the real parts of the adjacent speech DFT coefficients, is modeled as multivariate Laplace distribution,

$$p_{\mathbf{S}}(\mathbf{S}) = \frac{1}{\pi\sigma_S^2}\left(\frac{1}{\sqrt{2}\pi\sigma_S\|\mathbf{S}\|}\right)^{d/2-1} K_{d/2-1}\left(\frac{\sqrt{2}}{\sigma_S}\|\mathbf{S}\|\right) \tag{15}$$

where σ_S^2 denotes the variance of each speech spectral component in $\mathbf{S}$. The noise is assumed to be additive zero-mean white Gaussian noise with covariance matrix $\sigma_D^2\boldsymbol{I_d}$,

$$p_{\mathbf{D}}(\mathbf{D}) = \frac{1}{(2\pi\sigma_D^2)^{d/2}}\exp\left(-\frac{\|\mathbf{D}\|^2}{2\sigma_D^2}\right) \tag{16}$$

Then the PDF of $\mathbf{Y}$, the real parts of DFT coefficients of noisy speech, also a random vector with d components, is obtained by the multivariate convolution,

$$p_{\mathbf{Y}}(\mathbf{Y}) = \int_{\mathbb{R}^d} p_{\mathbf{S}}(\mathbf{S})p_{\mathbf{D}}(\mathbf{Y}\text{-}\mathbf{S})\,d\mathbf{S} \tag{17}$$

Using the GSM representation of $\mathbf{S}$ as derived in (13), we get

$$p_{\mathbf{Y}}(\mathbf{Y}) = \frac{1}{(2\pi\sigma_S^2)^{d/2}}\exp\left(\frac{1}{\xi}\right)\Gamma\left(1-\frac{d}{2},\frac{1}{\xi};\frac{\|\mathbf{Y}\|^2}{2\sigma_S^2}\right) \tag{18}$$

where $\xi = \sigma_S^2/\sigma_D^2$ denotes the *a prior* SNR, and $\Gamma(\alpha, x; b)$ is the generalized incomplete gamma function which is defined as

$$\Gamma(\alpha, x; b) = \int_x^\infty t^{\alpha-1}\exp\left(-t-\frac{b}{t}\right)dt \tag{19}$$

The MMSE estimator of speech spectral components given $\mathbf{Y}$ is

$$\begin{aligned}\hat{S}_i &= E\{S_i \mid \mathbf{Y}\} = \int_{\mathbb{R}^d} S_i p_{\mathbf{S}|\mathbf{Y}}(\mathbf{S}\mid\mathbf{Y})d\mathbf{S}\\ &= \frac{1}{p_{\mathbf{Y}}(\mathbf{Y})}\int_{\mathbb{R}^d} S_i p_{\mathbf{D}}(\mathbf{Y}-\mathbf{S})p_{\mathbf{S}}(\mathbf{S})d\mathbf{S}\end{aligned} \tag{20}$$

where S_i is the ith scalar component of $\mathbf{S}$.

Using (15), (16) and (18) in (20), we get,

$$\hat{S}_i = E\{S_i \mid \mathbf{Y}\} = Y_i\left[1-\frac{1}{\xi}\frac{\Gamma\left(-\frac{d}{2},\,\frac{1}{\xi};\,\frac{\|\mathbf{Y}\|^2}{2\sigma_S^2}\right)}{\Gamma\left(1-\frac{d}{2},\,\frac{1}{\xi};\,\frac{\|\mathbf{Y}\|^2}{2\sigma_S^2}\right)}\right] \tag{21}$$

where Y_i is the ith scalar component of $\mathbf{Y}$.

The equation (21) is the analytical solution for the MMSE estimator under the assumption of multivariate Laplace speech prior. Note that if d is set to 1 in (21), the estimator will degenerate to the estimator proposed in [3]. Therefore, the new estimator can be regarded as a generalization of MMSE estimator with univariate Laplace prior, and the well-known estimator derived in [3] is a special case under the condition that the dimension d=1.

4.2 Speech presence uncertainty

Motivated by the fact that speech is not surely present at all times and at all frequencies, the speech presence uncertainty is usually incorporated into MMSE estimator to improve the performance[1]. Therefore, we derive the speech presence possibility under the new multivariate Laplace prior model in this subseciton.

A two-state model is considered for speech events, i.e., that either speech is present at a particular frequency bin (hypothesis H^1) or that is not (hypothesis H^0). According to [1], the modified MMSE estimator under speech presence uncertainty is given by

$$\widehat{S}_{i\,\mathrm{mod}} = \frac{\Lambda(\mathbf{Y},q)}{1+\Lambda(\mathbf{Y},q)}E\{S_i \mid \mathbf{Y}, H^1\} \tag{22}$$

where $E\{S_i|\mathbf{Y},H^1\}$ equals to $E\{S_i|\mathbf{Y}\}$ which has been derived in equation (21), and $\Lambda(\mathbf{Y}, q)$ is a generalized likelihood ratio which is defined as

$$\Lambda(\mathbf{Y},q) = \frac{1-q}{q}\frac{p(\mathbf{Y}\mid H^1)}{p(\mathbf{Y}\mid H^0)} \tag{23}$$

where q denotes the *a priori* probability of speech absence. Under the assumption in the last subsection that the speech spectral coefficients is modeled as a multivariate Laplace distribution and that the noise is additive multivariate Gaussian, we obtain,

$$p(\mathbf{Y}\mid H^1) = p_{\mathbf{Y}}(\mathbf{Y}) \tag{24}$$

$$p(\mathbf{Y}\mid H^0) = p_{\mathbf{D}}(\mathbf{Y}) \tag{25}$$

Using (16) and (18) in (23) gives,

$$\begin{aligned}\Lambda(\mathbf{Y},q) &= \frac{1-q}{q}\frac{p_{\mathbf{Y}}(\mathbf{Y})}{p_{\mathbf{D}}(\mathbf{Y})}\\ &= \frac{1-q}{q}\left(\frac{1}{\xi}\right)^{d/2}\exp\left(\frac{\|\mathbf{Y}\|^2}{2\sigma_D^2}+\frac{1}{\xi}\right)\Gamma\left(1-\frac{d}{2},\frac{1}{\xi};\frac{\|\mathbf{Y}\|^2}{2\sigma_S^2}\right)\end{aligned} \tag{26}$$

5 Experimental results

The proposed algorithm was implemented in MATLAB, with the following experimental setup. The sampling frequency was 8000Hz. A frame length of 256 samples with 50% overlap was used. The frames were windowed using a Hanning window. The test set consisted of 16 Chinese speech utterances, spoken by two male and two female speakers. Three types of background noise in the NOISEX-92 database were used in the experiments: white Gaussian noise, M109 tank noise and F16 aircraft noise. The noises were resampled at 8000Hz and then used to corrupt the speech utterances at a SNR level varying from -5 dB to 10 dB.

We compare the performance of the proposed algorithm (PA) to the algorithm based on univariate Laplacian speech model (ULAP) described in [3] and the multidimensional Bayesian estimator (MDB) proposed in [10]. To determine the variance of the noise, a minimum statistics [13] noise estimator is employed in all the three algorithms. The *a priori* SNR is estimated using the “decision-directed” approach of [1] with a fixed smoothing parameter of α = 0.95. A fixed value of q = 0.2 is adopted in this paper, with reference to [1]. In addition, the dimension of the random vector is set as d = 12 empirically for the proposed algorithm. The performances of speech enhancement algorithms are evaluated in terms of segmental SNR (SSNR), Log spectral distance (LSD), perceptual evaluation of speech quality (PESQ) and speech spectrogram.

5.1 Segmental SNR improvement

The amount of noise attenuation is generally measured by segmental SNR (SSNR), which is defined as

$$SSNR = \frac{1}{M}\sum_{m=0}^{M-1} \mathcal{T}\left(10\log_{10}\frac{\|\mathbf{s}_m\|^2}{\|\mathbf{s}_m-\hat{\mathbf{s}}_m\|^2}\right) \tag{27}$$

where $\mathbf{s}_m$ and $\hat{\mathbf{s}}_m$ denote the original clean and enhanced speech signal frame respectively, M is the total number of frames, and $\mathcal{T}(\cdot)$=max(min($\cdot$, 35), -10), confining the local SNR to a perceptual meaningful range [-10dB, 35dB].

Fig. 2 gives the average SSNR improvements for the three algorithms under white Gaussian, M109 tank, F16 aircraft noises at various noise levels. The results show that the proposed algorithm provides a noticeable improvement in noise suppression, across noise types and levels, relative to ULAP and MDB.

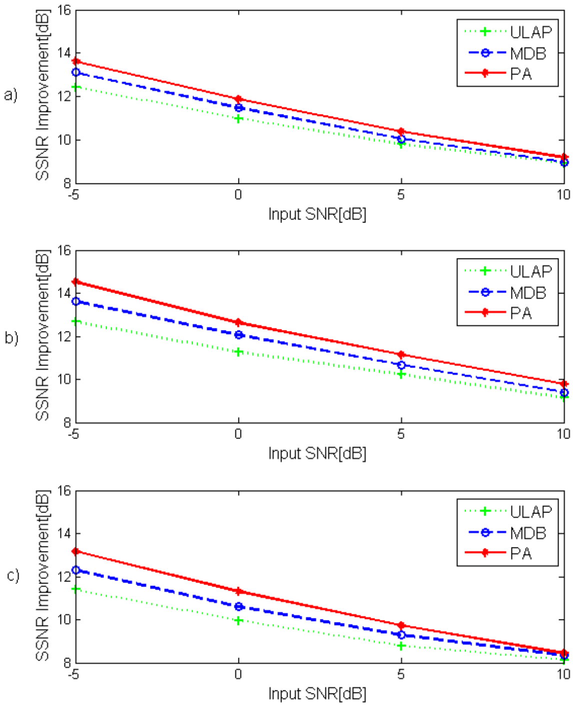

Fig.2. SSNR improvements of the three algorithms in (a) white Gaussian noise, (b) M109 tank noise, and (c) F16 aircraft noise.

5.2 Log spectral distance (LSD) performance

The LSD measures the dissimilarity between the spectra of clean speech and enhanced speech, which is expressed by the following equation,

$$LSD=\frac{1}{M}\sum_{l=0}^{M-1}\sqrt{\frac{1}{N/2+1}\sum_{k=0}^{N/2}\left[10\log_{10}\frac{|S(k,l)|}{|\hat{S}(k,l)|}\right]^2} \tag{28}$$

where $|S(k,l)|$ and $|\hat{S}(k,l)|$ are the magnitude spectra of the original clean and the enhanced speech signals of the lth frame, respectively. Note that large value of LSD implies bad performance.

The average LSD performances for the three algorithms are presented in Fig. 3, which shows that the proposed algorithm outperformed the other two algorithms in terms of speech distortion. Specificially, much more improvements are obtained at low SNR levels such as -5dB and 0dB, especially under white Gaussian noise and M109 tank noise conditions.

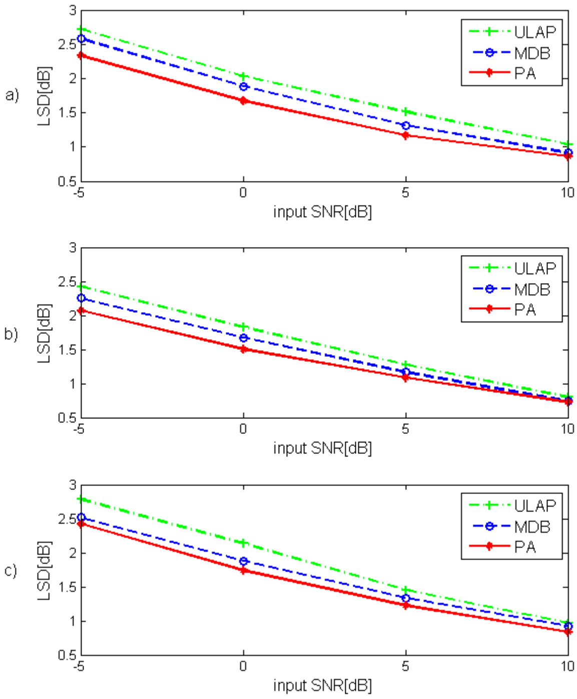

Fig.3. LSD comparison of the three algorithms in (a) white Gaussian noise, (b) M109 tank noise, and (c) F16 aircraft noise.

Table1. PESQ scores of the three algorithms

Noise type	Input SNR/dB	ULAP	MDB	PA
white Gaussian noise	-5	2.16	2.23	2.33
	0	2.68	2.73	2.84
	5	3.04	3.09	3.16
	10	3.24	3.29	3.34
M109 tank noise	-5	2.31	2.46	2.55
	0	2.83	2.94	3.02
	5	3.23	3.32	3.38
	10	3.49	3.58	3.57
F16 aircraft noise	-5	2.21	2.32	2.44
	0	2.66	2.78	2.85
	5	3.01	3.12	3.20
	10	3.28	3.36	3.41

5.3 Overall speech quality

The PESQ is a measure designed to predict the subjective opinion score of a degraded speech utterance and it is recommended by ITU-T for speech quality assessment [14]. It has been proven to be more reliable than some traditional objective measures. Therefore, PESQ measure is adopted as an excellent objective measure tool for predicting the overall quality of enhanced speech.

The PESQ results of the proposed algorithm are given in Table 1. It is obvious that the quality of enhanced speech by the proposed algorithm is better than that by ULAP. Even though compared with MDB, the PESQ improvements seem marginal, the perceptual quality of enhanced speech for the proposed algorithm is much better, which has been illustrated by informal subjective evaluations.

5.4 Spectrograms

The spectrograms of the clean, noisy, and enhanced speech by the three algorithms are shown in Fig. 4, in which the speech is corrupted by white Gaussian noise at

SNR=5dB. It demonstrates that the proposed algorithm suppresses a more significant amount of background noise and preserves most parts of the speech in comparison with the other two algorithms. Specifically, The ULAP leads to more visible residual noise, while the MDB introduces slightly more noticeable speech distortion than the proposed algorithm.

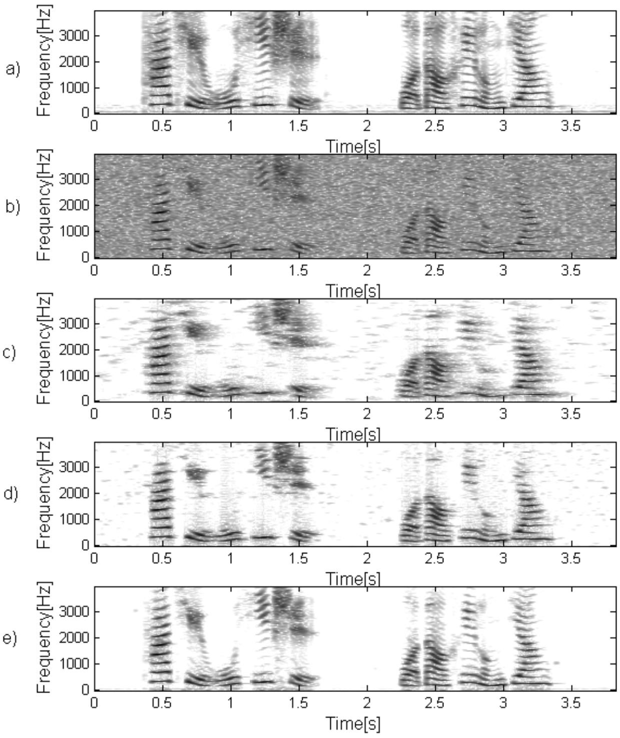

Fig.4. The spectrograms of the clean, noisy, and enhanced speech by the three algorithms. (a) Clean speech; (b) noisy speech corrupted by 5dB white Gaussian noise; (c) enhanced speech by ULAP; (d) enhanced speech by MDB; (e) enhanced speech by PA.

6 Conclusion

This paper proposed a new short-time spectrum estimation algorithm for single channel speech enhancement. Instead of assuming that the spectral components of speech are independent with each other, a new speech prior model, multivariate Laplace distribution model, was utilized to characterize the dependencies between frequency bins. The MMSE estimator based on the new speech model was derived using the GSM representation of random vectors and, furthermore, the speech presence probability was incorprated to modify the estimator. Note that if the dimension in the proposed estimator was set to 1, the new estimator degraded to the well-known Laplacian model-based MMSE estimator proposed in [3]. Therefore，the proposed estimator was regarded as a generalization of the MMSE estimator with univariate Laplace speech model. The Experimental results in terms of SNR, LSD and PESQ measures have demonstrated the effectiveness of the proposed algorithm under various noise conditions.

Acknowledgement

The authors wish to acknowledge the financial support of Natural Science Foundation of Jiangsu Province in 2009(BK2009059)

REFERENCES

[1] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 6, pp.1109–1121, Dec. 1984.

[2] S. Gazor and W. Zhang, "Speech probability distribution," *IEEE Signal Process. Lett.*, vol. 10, no.7, pp. 204–207, Jul. 2003.

[3] R. Martin, "Speech enhancement based on minimum mean-square error estimation and supergaussian priors," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 845–856, Sep. 2005.

[4] R. C. Hendriks, R. Heusdens and J. Jensen, "Log-spectral magnitude MMSE estimators under super-Gaussian densities," in *Proc. INTERSPEECH*, 2009, pp. 1319-1322.

[5] K. Paliwal, B. Schwerin, and K. Wojcicki, "Single channel speech enhancement using MMSE estimation of short-time modulation magnitude spectrum," in *Proc. of INTERSPEECH*, Florence, Italy, 2011, pp. 1209-1212.

[6] J. S. Erkelens, R. C. Hendriks, R. Heusdens, and J. Jensen, "Minimum mean-square error estimation of discrete Fourier coefficients with generalized Gamma priors," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 6, pp. 1741–1752, Aug. 2007.

[7] T. Esch and P. Vary, "Model-based speech enhancement using SNR dependent MMSE estimation," in *Proc. IEEE Int. Conf. Acoust.,Speech, Signal Process.*, Prague, Czech, May 2011, pp. 4652-4655.

[8] B. J. Borgstrom and A. Alwan, "Log-spectral amplitude estimation with generalized Gamma distributions for speech enhancement," in *Proc. IEEE Int. Conf. Acoust.,Speech, Signal Process.*, Prague, Czech, May 2011, pp. 4756-4759.

[9] C. Li and S. V. Andersen, "A block-based linear MMSE noise reduction with a high temporal resolution modeling of the speech excitation," *EURASIP J. Appl. Signal Process.*, vol. 18, pp. 2965–2978, 2005.

[10] E. Plourde and B. Champagne, "Multi-dimensional Bayesian STSA estimators for the enhancement of speech with correlated frequency components," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3013-3024, Jul. 2011.

[11] I. W. Selesnick, "The estimation of Laplace random vectors in additive white Gaussian noise," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3482-3496, Aug. 2008.

[12] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.

[13] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 504–512, Jul. 2001.

[14] 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 Rec. P. 862, 2001.

***Authors**:*
Bin ZHOU, Postgraduate Team 2, Institute of Command Automation, Haifu Xiang 1, Baixia District, Nanjing, China, 210007, E-mail: binzhou86@yahoo.com.cn;
Xiong-wei ZHANG, PLA University of Science and Technology.
Xia ZOU, PLA University of Science and Technology.
Gaihua ZHAO, PLA University of Science and Technology.