Efficient β-order Perceptually Motivated Spectral Amplitude Bayesian Estimator Based On Chi-distribution for Speech Enhancement

Huan Zhao, Yong Yang, Zhiqiang Lu
School of Information Science and Engineering, Hunan University, Changsha, China
Email: { hzhao, yyang, zqlu }@hnu.edu.cn

Fei Yu
Jiangsu Provincial Key Laboratory of Computer Information Processing Technology, Suzhou, China
Email: hunanyufei @126.com

***Abstract*—The traditional Bayesian estimator of short-time spectral amplitude is based on the minimization of the squared-error cost function under the common Gaussian probability density function (pdf). The Gaussian distribution, however, is not the optimal probability distribution. To overcome this phenomenon, we considered to replace the traditional distribution hypothesis of spectral amplitude of speech in this paper. More precisely, we proposed a β -order perceptive Bayesian spectral amplitude estimator which incorporated the assumption of Super-Gaussian chi-distributed spectral amplitude. The new weighting function incorporated the perceptive property as well as the different importance of the spectral valley and peak. Experiments showed that the proposed estimator can achieve a more significant noise reduction and yield a better spectral estimation over the most of latest enhancement algorithms.**

***Index Terms*—Bayesian Estimation, β -order, Perception, Speech Enhancement**

I. INTRODUCTION

The speech signals are frequently contaminated by noise in practical environments, resulting in reduction of speech quality. Especially when the signal to noise ratio (SNR) is low or the disturbing noise is nonstationary, the signals would be plagued by speech distortions and unnatural sounding or fluctuating residual background noises. To reduce listener fatigue and improve the recognition rate, improving speech quality is essential. In order to eliminate the noise influence more effective, many approaches in the time/transform domain have been investigated to date. Firstly, speech enhancement technique, which was one of most workable and effective methods in dealing with noise, was commonly used in improving speech quality and intelligibility. Secondly, robust speech features were developed for some applications: speech recognition, hands-free systems, voice command systems. Thirdly, noise compensation method based on adaptive model parameter was used to dealing with noise influence.

In terms of the methodology adopted, the most popular methods for speech enhancement can be broadly categorized as single-channel and multi-channel enhancement techniques. Single channel enhancement techniques, which have received significant interest due to their low complexity and relatively good performance, applied to situations in which one acquisition channel is available and can especially used in mobile communication applications. Instead, due to limitations such as cost and size considerations, multi-channel enhancement techniques was not widely used in comparing with single channel enhancement techniques.

Among most of the single channel speech enhancement algorithms, the statistical model-based enhancement techniques have recently received much attention. These algorithms, which relayed on the optimal pdf of speech to deduce the spectral weighting function, made use of the prior knowledge of speech and noise as much as possible. Compared with the Spectral Subtraction [1] based on Maximum Likelihood, these methods allowed better and more suppression of the musical noise. Ephraim et al. proposed the minimum mean square estimation (MMSE) short-time spectral amplitude estimation method for speech enhancement by minimizing the Bayesian risk in [2-3] based on Gaussian distribution of speech Discrete Fourier Transform (DFT) amplitude coefficients and statistical independence assumption. The MMSE estimators have been shown to be successful in eliminating musical noise. Moreover, MMSE-based algorithm has the advantage of mathematical calculability and can be deduced effectively. However, the estimator originated from Bayesian risk based on square error function may not be subjectively meaningful, in that, small and large squared estimation errors might not necessarily correspond to good and poor

National Science Foundation of China (Grant NO. 61173106), the Key Program of Hunan Provincial Natural Science Foundation of China (Grant No.10JJ2046), the Planned Science and Technology Key Project of Hunan Province, China (Grant No.2010GK2002).

Corresponding author: Huan Zhao(Email: hzhao@hnu.edu.cn)

doi:10.4304/jcp.7.11.2829-2835

speech quality respectively. Martin et al. [4-5] extended the MMSE idea to super-Gaussian distribution for modeling the spectral coefficients and proposed some new Bayesian estimators. Loizou was perhaps the first to propose the basic perceptive estimator by weighting the traditional squared-error in [6], where Gaussian model was used. Recently, You proposed β-order MMSE and β-order perceptive MMSE estimator in [7] and [8] respectively, which were both based on pdf of Gaussian distribution. More recently, Optimal estimators for other probability distribution-based spectral amplitude MMSE estimator were given in [9]. Chen et al [10] proposed a laplacian-based MMSE algorithm under speech presence uncertainty. It is showed that the Laplacian-based MMSE estimator yielded less residual noise than the traditional Gaussian-based estimator. In Ref. [11], a log and super-Gaussian estimator was proposed, which origined from the basic log-MMSE algorithm. In literatures [12-13], a β-order perceptive estimator based on Gaussian probability distribution was proposed for the first time. Recently, a Bayesian estimator [14] based on Gaussian mixture model (GMM) was proposed which modeled speech more precisely. In Ref. [15], a power spectral based bayesian estimator was proposed, the innovation of the algorithm lied in the hypothesis of the exponential distribution for spectral amplitude.

Generally speaking, most of the traditional estimators were derived under constant one of β. However, this wasn't optimal. To overcome the shortcomings and make use of the prior knowledge effectively, we proposed a subjectively meaningful β-order perceptive weighting function incorporating the chi-distribution and β-order theory for speech enhancement. The new β-order perceptive Bayesian estimator substituted Chi-distribution for Gaussian distribution, incorporated the perceptive characteristic and difference by emphasizing the importance of spectral valley, improved the performance effectively.

II. BAYESIAN ESTIMATION

In a general way, Bayesian estimator was deduced by assuming two usual and commonly used error criterions, which were MMSE and maximum a posteriori (MAP). Both MMSE and MAP were expressed in mathematical sense. In fact, the MMSE-based method can not be always deduced, and MAP-based method was proposed for approximating.

The goal of speech enhancement is to compute an estimate of the clean speech signal $\hat{x}(n)$, given a noisy signal $y(n)$, which was composed of a clean signal $x(n)$ degraded by noise $n(n)$, with n being the sample index. The basic assumption of the algorithm is treating the noise as uncorrelated additive noise. The noisy speech signal can be expressed as (1). Short time spectrum of the noisy speech can be expressed as (2). $Y(\omega)$, $X(\omega)$, $N(\omega)$ are the frequency description of noisy signal, clean signal and noise accordingly.

$$y(n) = x(n) + n(n) \tag{1}$$

$$Y(\omega) = X(w) + N(w) \tag{2}$$

By assuming the error with $\delta = A - \hat{A}$ at every coefficient, then $\cos t(\delta) = d(A, \hat{A})$ means cost function which is non-negative. Mean cost function was given by (3).

$$\begin{aligned} Risc &= E[d(A,\hat{A})] \\ &= \iint d(A,\hat{A}) p(A,Y) dA dY \\ &= \int [\int d(A,\hat{A}) p(A \mid Y) dA] p(Y) dY \end{aligned} \tag{3}$$

By minimizing expression (3) about $\hat{A}$, we can get the estimator of spectral amplitude. And we can also know that by solving the internal integral in (3), we can get the weighting function. For example, MMSE-based estimator was proposed by Ephraim et al. under cost function $d(A,\hat{A}) = (A - \hat{A})^2$.

III. BAYESIAN ESTIMATOR BASED ON CHI-DISTRIBUTION

A. *Probability Distribution of Speech DFT Coefficients*

There were three types of statistic estimator derived from different speech distributions, which were Gamma, Laplace and Gaussian Probability Distribution. Experiments showed that gamma-based algorithm can model the speech more accurate, and gain better enhancement performance. In this paper, we proposed a new estimator which incorporated the chi-distribution into the β-order perceptive Bayesian estimator.

First, the speech DFT coefficients have been assumed to be Gamma distributed. By letting the parameter of γ to be 2 in (4), we got the super-Gaussian distribution, which named as chi-distribution. $\sigma_X^2 = E[X^2]$ stand for the variance of speech. Because $\nu < 1$, so the speech is fit for the super-Gaussian property. Because the positive value of spectral kurtosis implicates super-distribution property of spectral amplitude.

$$f_A(a) = \frac{\gamma \beta^{\nu}}{\Gamma(\nu)} a^{\gamma\nu - 1} \exp(-\beta a^{\gamma}) \tag{4}$$

$$f_A(a) = \frac{2\beta^{\nu}}{\Gamma(\nu)} a^{2\nu - 1} \exp(-\beta a^2) \tag{5}$$

where β is defined as $\beta = \nu / \sigma_X^2$ in (4) and (5).

B. Probability Distribution of Noise DFT Coefficients

In this paper, noise DFT coefficients have been assumed Gaussian distributed. According to the experiments and the central-limit theorem, we can know that the distribution of noise DFT coefficients is often close to Gaussian-distribution with the variance of zero. The pdf of Gaussian model is expressed as (6) while expression (7) is the variant version. $\sigma_N^2 = E[N^2]$ stands for the variance of noise. Therefore, the pdf can be expressed with (8). r means noisy speech amplitude, ϕ , a means the phase and amplitude of clean speech respectively, θ , n stands for the phase and amplitude of noise respectively.

$$f_N(n) = \frac{1}{\pi\sigma_N^2}\exp(-\frac{|n|^2}{\sigma_N^2}) \tag{6}$$

$$f_{Y|N}(y \mid n) = \frac{1}{\pi\sigma_N^2}\exp(-\frac{|y-x|^2}{\sigma_N^2}) \tag{7}$$

$$f_{Y|A,\Phi}(y \mid a,\phi) = \frac{1}{\pi\sigma_N^2}\exp(\frac{2ar\cos(\phi-\theta)-r^2-a^2}{\sigma_N^2}) \tag{8}$$

C. Derivation of the β -order Bayesian Estimator

Compared with the estimator which emphasized the spectral peak, the estimator emphasizing spectral valley yielded better performance: less musical noise and better speech quality [6]. In this paper, we considered the similar problem. To make the traditional distortion measure more meaningful and effective, we introduced chi-distribution probability distribution into a β -order estimator, which was used in Gaussian distribution [13] before. The cost function was subjectively meaningful by incorporating the parameter in denominator and numerator.

First, we defined the perceptually error with expression (9), which was used in reference [10] aiming at Gaussian distributed speech. We can know that the variant distortion measure penalizes the estimation error differently when A is spectral valley and peak. In this paper, however, in order to deducing the new weighting gain estimator, the remaining progresses was carried out under chi-distributed pdf. A stands for the amplitude of the speech, and Y stands for noisy speech.

$$d(A,\hat{A}) = \left(\frac{A^\beta - \hat{A}^\beta}{A^p}\right)^2 \tag{9}$$

Then, we got the Bayesian error in expression (10).

$$Risc = \int_0^\infty \left(\frac{A^\beta - \hat{A}^\beta}{A^p}\right)^2 p(A \mid Y)dA \tag{10}$$

Expression (10) is a conditional expectation, $p(A \mid Y)$ stands for conditional pdf. The perceptive estimator under chi-distribution was obtained by desolving the partial derivative about $\hat{A}^\beta$. Partial derivative was given by (11) by letting the partial derivative to be 0.

$$\frac{\partial(Risc)}{\partial\hat{A}^\beta} = -2\int_0^\infty \left(\frac{A^\beta - \hat{A}^\beta}{A^p}\right)p(A \mid Y)dA = 0 \tag{11}$$

The perceptive estimator under chi-distribution was expressed with (12).

$$\hat{A}^\beta = \frac{E\left\{A^{\beta-2p}\right\}}{E\left\{A^{-2p}\right\}} \tag{12}$$

Then, the amplitude estimator was formulated as (13).

$$\hat{A} = \left(\frac{E\left\{A^{\beta-2p}\right\}}{E\left\{A^{-2p}\right\}}\right)^{1/\beta} \tag{13}$$

The weighting function or estimated spectrum was obtained by dividing the numerator with the denominator, which were expressed as (14) and (15) respectively.

$$E\{A^{\beta-2p} \mid Y\} = \frac{\int_0^{+\infty}\int_0^{2\pi} a^{\beta-2p} f_{Y|A,\Phi}(y \mid a,\phi) f_A(a)d\phi da}{\int_0^{+\infty}\int_0^{2\pi} f_{Y|A,\Phi}(y \mid a,\phi) f_A(a)d\phi da} \tag{14}$$

$$E\{A^{-2p} \mid Y\} = \frac{\int_0^{+\infty}\int_0^{2\pi} a^{-2p} f_{Y|A,\Phi}(y \mid a,\phi) f_A(a)d\phi da}{\int_0^{+\infty}\int_0^{2\pi} f_{Y|A,\Phi}(y \mid a,\phi) f_A(a)d\phi da} \tag{15}$$

By inserting (5) and (8) into (14) and (15), and utilizing the series table, the initial β -order amplitude can be derived and formulated as (16). $\Gamma(x)$ signify gamma function, and $M(x,y;Z)$ stands for confluent hyper-geometric distribution function. In function $M(x,y;Z)$, Z had been derived into vector. In the expression, r stand for the amplitude of noisy speech. Then the Bayesian estimator can be formulated as follows:

$$\frac{E\{A^{\beta-2p}\}}{E\{A^{-2p}\}} = \frac{\Gamma(\nu+\beta/2-p)\left(\frac{\zeta\xi}{(\nu+\xi)}\right)^{\beta/2} \mathrm{M}(\nu+\beta/2-p;1;\frac{\zeta\xi}{\nu+\xi})}{\Gamma(\nu-p)\mathrm{M}(\nu-p;1;\frac{\zeta\xi}{\nu+\xi})}\frac{r^\beta}{\zeta^\beta}$$

(16)

And the β -order estimator is (17).

$$\hat{A}=\frac{1}{\zeta}\left(\frac{\zeta\xi}{(\nu+\xi)}\right)^{1/2}\left(\frac{\Gamma(\nu+\beta/2-p)\mathrm{M}(\nu+\beta/2-p;1;\frac{\zeta\xi}{\nu+\xi})}{\Gamma(\nu-p)\mathrm{M}(\nu-p;1;\frac{\zeta\xi}{\nu+\xi})}\right)^{1/\beta} r \tag{17}$$

In order to getting the uniform expression, (17) was expressed as (18) by using $\varepsilon=(\xi\zeta)/(\xi+\nu)$. The weighting function can be formulated as (18)

$$\hat{A}=\left(\frac{E\{A^{\beta-2p}\}}{E\{A^{-2p}\}}\right)^{1/\beta}=\frac{\sqrt{\varepsilon}}{\zeta}\left(\frac{\Gamma(\nu+\beta/2-p)\mathrm{M}(\nu+\beta/2-p;1;\varepsilon)}{\Gamma(\nu-p)\mathrm{M}(\nu-p;1;\varepsilon)}\right)^{1/\beta} r \tag{18}$$

$$G=\left(\frac{E\{A^{\beta-2p}\}}{E\{A^{-2p}\}}\right)^{1/\beta}=\frac{\sqrt{\varepsilon}}{\zeta}\left(\frac{\Gamma(\nu+\beta/2-p)\mathrm{M}(\nu+\beta/2-p;1;\varepsilon)}{\Gamma(\nu-p)\mathrm{M}(\nu-p;1;\varepsilon)}\right)^{1/\beta} \tag{19}$$

ξ , ζ stands for a prior SNR and a posterior SNR respectively. $\xi=\sigma_X^2/\sigma_N^2$, $\zeta=Y^2/\sigma_N^2$ are the definition expression accordingly. According to the spectral kurtosis theory, In order to insuring the super-Gaussian property of probability distribution of speech, the parameter should be controled. In this paper, we assumed that the value of parameter $\nu=0.6$, and got an estimator like (20).

$$\hat{A}=\frac{\sqrt{\varepsilon}}{\zeta}\left(\frac{\Gamma(0.6+(1/3)/2-0.5)\mathrm{M}(0.6+(1/3)/2-0.5;1;\varepsilon)}{\Gamma(0.6-0.5)\mathrm{M}(0.6-0.5;1;\varepsilon)}\right)^{3} r \tag{20}$$

In the expression, $\varepsilon=(\xi\zeta)/(\xi+0.6)$ 。

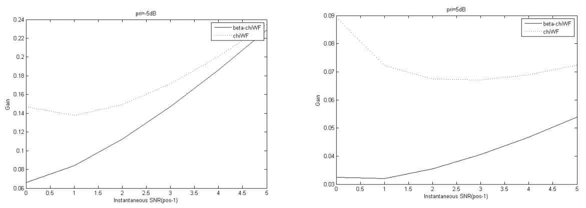

Figure 1. Gain comparison for new estimator

In Fig.1, the left and right sections were on behalf of the gain coefficients versus Instantaneous SNR under a priori SNR of -5dB and 5dB SNR respectively. The range from 0 to 5 in horizontal ordinate stood for Instantaneous SNR. We also gave the commonly used amplitude estimator under chi-distribution for comparing. In Fig. 1, beta-chiWF stood for the new weighting function proposed in this paper, and the chi-WF was the estimator for comparing. It was apparent that the gain value of the new weighting function was below the common chi-distribution-based estimator and can suppress the noise more effectively. In Fig. 2, we showed the block diagram of the new proposed algorithm. Dotted line marked out the core of the estimator.

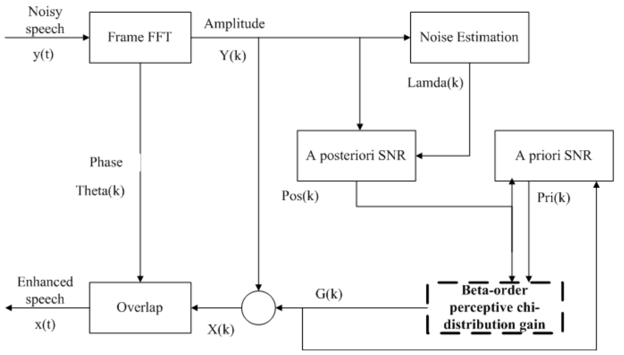

Figure 2. Block diagram of New Estimator of Speech Enhancement

IV. BAYESIAN ESTIMATION

A. The Value of Parameter β

Compared to literature [13], the new proposed filter brought parameter β into chi- distribution-based spectral weighting rule instead of Gaussian-distribution-based filter. Then the speech can be modeled more precisely. Furthermore, from the analysis of literature [13], we know that the value of β wasn't connected with the model, for this reason, the value of β =1/3 [13] can also be used in our paper. And this is sufficient for the purpose of experiments.

B. Experiment Evaluation

In order to evaluating the improved method and comparing it with existing method, we should choose proper speech corpus. In this paper, we chose NOISEUS [16] as the speech evaluation material. NOISEUS, which was developed by UTDALLAS, is used to performance comparing. This speech corpus contains 30 sentences which were chose from IEEE Speech corpus. These sentences contain the entire phoneme in American English, and have the low foresee-ability about speech content.

The background noises were taken from the NOISEx-92 database, designed for speech recognition in noisy environments. In this database, there are as many as 15 kinds of noise, such as babble, F16, white and so on. It is suitable for speech evaluating. For our enhancement experiments, the noise signal were down-sampled to 8kHz, then noise was added to the clean speech files at -5, 0, 5, 10dB SNR).

The proposed estimator was applied to 32-ms duration frames of speech. In order to keeping the transition of speech signal, we chose 16-ms (50% overlap between frames) duration as the frame shift. Subjective evaluation of speech enhancement for evaluating the performance of the algorithms is the most accurate and reliable provided it is performed under professional methodology and excellent equipment. However, it is costly and time consuming. Therefore, researchers tried their best to find appropriate objective measures to substitute for this. In this paper, we chose segSNR and perceptual evaluation of speech quality (PESQ) as the objective measures for evaluating. segSNR has a clear meaning, and simple in computation, can signify noise reduction clearly. PESQ has the subjective meaning.

C. Effective SNR Improvement

segSNR considered the level of remaining noise and decrease of speech quality, the definition is given as follows:

$$SegSNR = \frac{1}{N}\sum_{n=0}^{N-1} 10\log_{10}\frac{\sum_{t=Tn}^{Tn+T-1} x^2(t)}{\sum_{t=Tn}^{Tn+T-1}(\hat{x}(t)-x(t))^2} \tag{21}$$

In (21), N stood for the number of speech frame, t mean the index of time frame, and segSNR was the mean at every kind of noise and SNR.

In our experiments, all the SNR of the corpus was global SNR form, and the relation between the global SNR and the segmental SNR was given in Table I.

TABLE I.
COMPARISONS BETWEEN GLOSNR AND SEGSNR FOR NOISY MATERIAL

segSNR/dB	-15	-10	-5	0
gloSNR/dB	-5	0	5	10

The experiments were repeated under many kinds of noises. We chose four representative noises lastly. Chi-WF is the common estimator under chi-distribution, GWF is the weighting function under Gaussian-distribution, and chi-beta-order stands for the new proposed filter. Fig. 3 showed the output segSNR for different input segSNRs with four kinds of noises. Horizontal and vertical coordinates stood for segSNR of noisy speech and processed speech respectively. It was obvious that Fig. 3 showed that segSNR improved obviously and the new estimator gain better effect under nearly all kinds of noises. Compared with the previous methods, the average value improved is above 0.7dB, and 2dB at some conditions.

D. Assuring and Acceptable PESQ

Fig. 4 showed the PESQ distortion measure. From the curve we can see that the quality was guaranteed in the same level in comparing with the reference algorithm. It was showed that the new estimator can eliminate noise influence effectively while maintaining a high level of speech quality. Informal listening test proved that the new method outperformed the reference algorithm.

E. Waveform and Spectrogram

Fig. 5 and Fig. 6 demonstrated examples of the waveform and spectrogram for comparison under white and pink noise. From the spectrogram, we can see the proposed method suppressed more noise obviously.

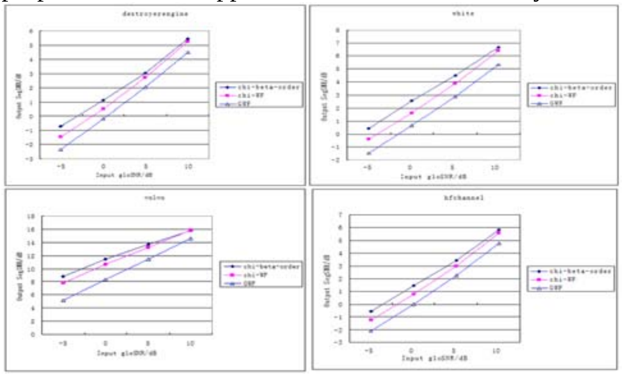

Figure 3. segSNR comparisons under four kinds of noises

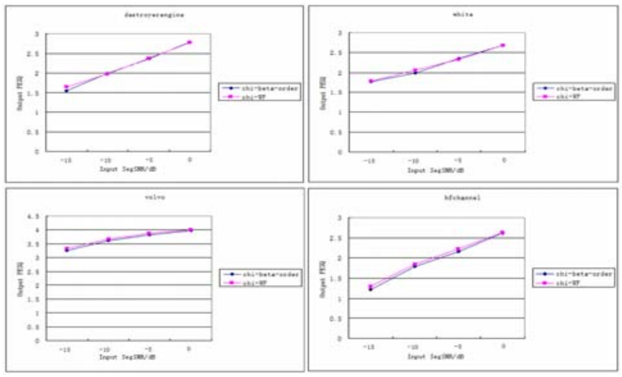

Figure 4. PESQ value comparisons under four kinds of noises

V. CONCLUSION

In this paper, we derived a new β-order perception filter based on chi-distribution and β-order idea for speech enhancement. The new weighting function gives up the traditional Gaussian probability distribution about amplitude by substituting the chi-distribution probability density function for it. Meanwhile, features of perception were incorporated into the cost function. In a word, the present study focused on the derivation of perceptually-motivated β-order Bayesian estimator of spectral amplitude. The prior information of speech, as well as the different perceptive properties of spectral valley and peak were considered carefully in the new β-order estimator. It is showed that the new filter has excellent effect in segSNR improvement and also can ensure the quality of speech.

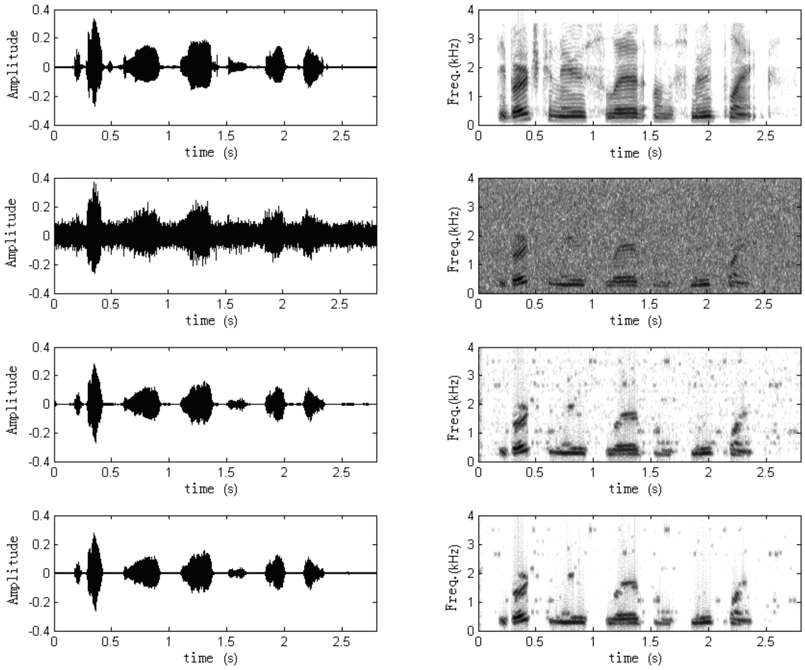

Figure 5. Waveform and spectrogram under white noise(From Top to Bottom: clean speech、0dB noisy speech under white noise, chi-WF processed speech, chi-beta-order processed speech)

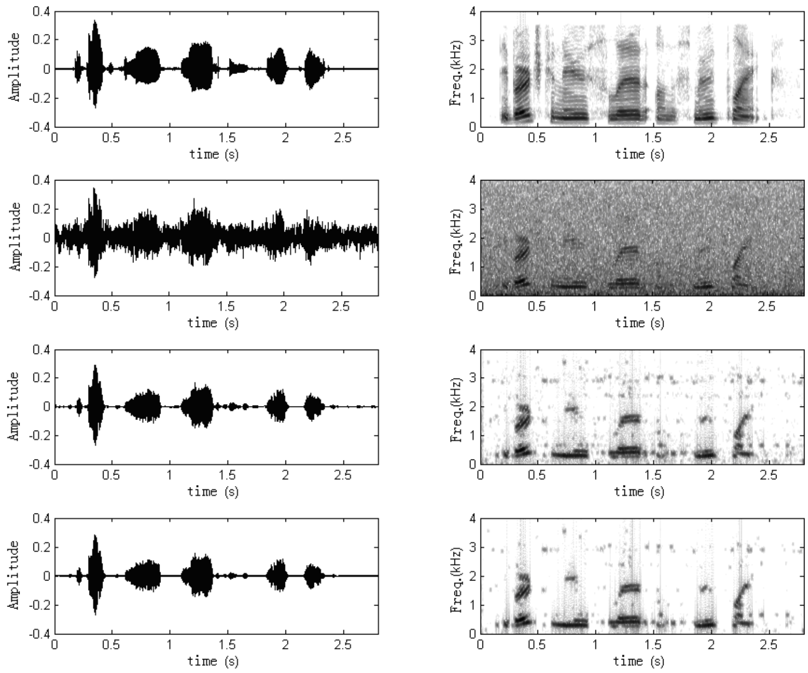

Figure 6. Waveform and spectrogram under pink noise(From Top to Bottom: clean speech、0dB noisy speech under pink noise, chi-WF processed speech, chi-beta-order processed speech)

ACKNOWLEDGMENT

This work is supported by the National Science Foundation of China (Grant NO. 61173106) and the Key Program of Hunan Provincial Natural Science Foundation of China (Grant No.10JJ2046), and the Planned Science and Technology Key Project of Hunan Province, China (Grant No.2010GK2002).

REFERENCES

[1] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoustic., Speech, Signal Process.*, vol. ASSP-27, no. 2, pp. 113-120, Apr. 1979.

[2] Y. Ephraim, D. Malah, "Speech Enhancement Using a Minimum Mean Square Error Short-Time Spectral Amplitude Estimator," *IEEE Trans. Acoustic., Speech,*

Signal Process., vol. ASSP-32, no. 6, pp. 1109-1121, Dec. 1984.
[3] Y. Ephraim, D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoustic., Speech, Signal Process.*, vol. ASSP-33, no. 2, pp. 443-445, Apr. 1985.
[4] R. Martin, "Speech enhancement using MMSE short time spectral estimation with gamma distributed speech priors," in *proceedings IEEE Int. Conf. Acoustic., Speech, Signal Process.*, May 2002, Orlando, FL, USA, pp. I253-I256.
[5] R. Martin, C. Breithaupt, "Speech enhancement in the DFT domain using Laplacian speech priors," in *proceedings of the International Workshop on Acoustic Echo and Noise Control (IWAENC2003)*, Sept.2003, Kyoto,Japan, pp. 87-90.
[6] P. Loizou, "Speech enhancement based on perceptually motivated Bayesian estimators of the magnitude spectrum," *IEEE Trans. Speech and Audio Process.*, vol. 13, no. 5, pp. 857-869, Sep. 2005.
[7] C.H. You, S.N. Koh, S. Rahardja, "β-order MMSE spectral amplitude estimation for speech enhancement," *IEEE Transactions Speech and Audio Processing*, vol. 13, no. 4, pp. 475-486, 2005.
[8] C.H. You, S.N. Koh, S. Rahardja, "Masking-based [beta]-order MMSE speech enhancement," *Speech communication*, vol. 48, no. 1, pp. 57-70,2006.
[9] C. Breithaupt, M. Krawczyk, R. Martin, "Parameterized MMSE spectral magnitude estimation for the enhancement of noisy speech," in *proceedings IEEE Int. Conf. Acoust., Speech, Signal Process.*, March 31 2008-April 4 2008, Las Vegas, NV, pp. 4037-4040.
[10] B. Chen, P. Loizou, "A Laplacian-based MMSE estimator for speech enhancement," *Speech communication*, vol. 49, no. 2, pp. 134-143, 2007.
[11] R.C. Hendriks, R. Heusdens, J. Jensen, "Log-Spectral Magnitude MMSE Estimators under Super-Gaussian Densities," in *proceedings of the 10th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Sept. 2009, Brighton, United Kingdom.
[12] E. Plourde, Bayesian short-time spectral amplitude estimators for single-channel speech enhancement(2009), Ph.D. dissertation, McGill University, Montreal, Canada.
[13] E. Plourde, B. Champagne, "Auditory-Based Spectral Amplitude Estimators for Speech Enhancement," *IEEE Transactions Audio, Speech, and Language Processing*, vol. 16, no. 8, pp. 1614-1623, 2008.
[14] J. Hao, H. Attias, S. Nagarajan, et al. "Speech enhancement, gain, and noise spectrum adaptation using approximate Bayesian estimation," *IEEE Transactions Audio, Speech, and Language Processing*,, vol. 17, no.1, pp. 24-37, 2009.
[15] Y. Lu, P. Loizou, "Estimators of The Magnitude-Squared Spectrum and Methods for Incorporating SNR Uncertainty ," *IEEE Transactions Audio, Speech, and Language Processing*, vol. 19, no. 5, pp. 1123-1137, 2011.
[16] P. Loizou, NOIZEUS: a noisy speech corpus for evaluation of speech enhancement algorithms. http:// www.utdallas.edu/~loizou/speech/noizeus/, 2011-7-25.

Huan Zhao is a professor at the School of Information Science and Engineering, Hunan University. She obtained her B.Sc. degree, M.S. degree and Ph.D. in Computer Science and Technology at Hunan University in 1989, 2004 and 2010, respectively. Her current research interests include speech information processing, embedded system design and embedded speech recognition. She served as visiting scholar at the University of California, San Diego (UCSD), USA during the period of March 2008 to September 2008. The visiting scholarship was appointed and sponsored by the China Scholarship Council (CSC). Prof. Zhao is a Senior Member of China Computer Federation, Governing of Hunan Computer Society, China and China Education Ministry Steering Committee Member of Computer Education on Arts. She has published more than 40 papers and 9 books.

Yong Yang received his B.Sc. degree in Computer Science and Technology at the College of Information Science & Technology, Qingdao University of Science & Technology, P. R. China in 2009. Currently, he is a M.S. candidate of Hunan University, P. R. China. His current research interests include speech information processing and speech enhancement.

Zhiqiang Lu received his B.Sc. degree in communications engineering at the School of Xiangtan University, P. R. China in 2010. Currently, he is a M.S. candidate of Hunan University, P. R. China. His current research interests include speech information processing and speech enhancement.

Fei Yu was born in Ningxiang, China, on February 06, 1973. Before Studying in Peoples' Friendship University of Russia, Russia, He joined and worked in Hunan University, Zhejiang University, Hunan Agricultural University, China. He have taken as a guest researcher in State Key Laboratory of Information Security, Graduate School of Chinese Academy of Sciences, Guangdong Province Key Lab of Electronic Commerce Market Application Technology, Jiangsu Provincial Key Lab of Image Processing and Jiangsu Provincial Key Laboratory of Computer Information Processing Technology.
He has wide research interests, mainly information technology. In these areas he has published above 90 papers in journals or conference proceedings and a book has published by Science Press, China (Fei Yu, Miaoliang Zhu, Cheng Xu, et al. Computer Network Security, 2004). Above 70 papers are indexed by SCI, EI. He has won various awards in the past.
He served as many workshop chair, advisory committee or program committee member of various international ACM/IEEE conferences, and chaired a number of international conferences such as IITA'07, IITA'08;ISIP'08, ISIP'09, ISIP'10, ISIP'11;ISECS'08, ISECS'09, ISECS'10, ISECS'11; WCSE'08, WCSE'09, WCSE'10, WCSE'11, WCSE'12 and ISISE'08, ISISE'09, ISISE'10.