http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (6): 1089-1097, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.1089.1097

Improving Whisper Intelligibility in Noise Environment Based on Joint Time Frequency Analysis

[1,2]Jian Zhou, [1]Huabin Wang, [1]Xianyong Fang, [1]Liang Tao and [2]Li Zhao
[1]Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Anhui University, Heifei, 230601, China
[2]Key Laboratory of Underwater Acoustic Signal Processing of Ministry of Education, Southeast University Nanjing, 210096, China

Abstract: A whisper intelligibility enhancement method was proposed based on the Real-valued Discrete Gabor Transform (RDGT) in the joint time frequency domain where the RDGT can change the density of the spectrum through the over sample rate parameter. An optimal logarithmic spectrum estimator was derived based on RDGT coefficients. Experimental results showed that the proposed method improved the intelligibility of enhanced whisper in an adverse noise environment when using a proper over sample rate value. The largest gains in intelligibility were got when the over sample rate value is set to 16 or 32. The proposed method outperforms other four speech enhancement algorithms in terms of intelligibility improvement. The experimental results also implied that the intelligibility performance of the conventional speech enhancement algorithms might be improved when increasing the density of the time-frequency spectrum properly.

Key words: Real-valued discrete gabor transform, joint time-frequency domain, speech intelligibility enhancement

INTRODUCTION

Recently, whispered speech has received much attention (Jou *et al*., 2005; Sharifzadeh *et al*., 2009, 2012; Ahmadi *et al*., 2008; Zhou *et al*., 2012a). In the library, for example, people often communicate using whispered speech for a loud voice is prohibited. Additionally, whispered speech is also used to avoid being overheard. Unlike normal phonated speech, whispered speech is a completely noise excited signal. The turbulent flow of the air exhaled from the lungs excites the open glottis and passes through the vocal tract, resulting in the whispered speech (Tartter, 1989). Whispered speech has much weaker energy than that of the phonated speech. So it is more susceptible to interference (Ito *et al*., 2005). In fact, improving the intelligibility of a noisy whisper rather than the quality (e.g., SNR improvements or comfort of the enhanced speech) is much more important for semantic information retrieve becomes the dominating purpose of whisper communication. Improving intelligibility of noisy whisper becomes a key preprocessing stage in various whispered based applications.

Nevertheless, there is little progress in designing algorithms of improving the speech intelligibility. Although the existed speech enhancement algorithms such as power subtraction and spectrum amplitude estimator based method are very powerful in terms of improving the listening comfort or the speech quality, they fail to improve the intelligibility of the enhanced speech (Hu and Loizou, 2007; Loizou and Kim, 2011; Lim, 1978).

Why the existed speech enhancement algorithms do not improve speech intelligibility is only partially understood (Hu and Loizou, 2007; Loizou and Kim, 2011). The difficulty of accurately estimating the noise spectrum is assumed as a key factor which contribute to little progress in the intelligibility improvement aspect. However, Zhou *et al*. (2012a) found that some spectrum components where the speech energy dominates are very important for improving the intelligibility of the enhanced whisper. So, the rough spectrogram used in the conventional speech enhancement algorithms may play a role in decreasing the intelligibility of the enhanced speech. The conventional speech enhancement algorithms such as the minimum mean square error spectrum estimator (Ephraim and Malah, 1984), the log based MMSE estimator (Ephraim and Malah, 1985) and the power subtraction method (Boll, 1979) use the Short Time discrete Fourier Transform (STFT) to compute the

Corresponding Author: Jian Zhou, Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Anhui University, Heifei, 230601, China

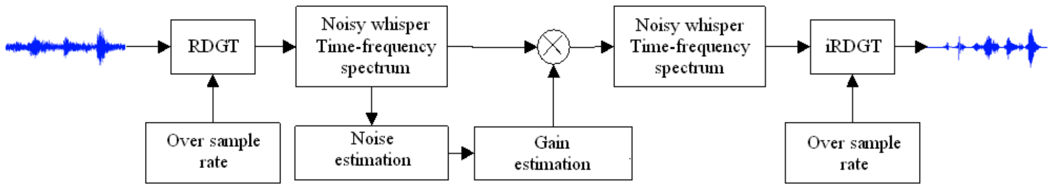

Fig. 1: Whisper intelligibility enhancement method

noisy speech spectrum. Each time frequency point of the STFT spectrum is attenuated with a gain function to get the enhanced speech spectrum, which is then used to synthesize the enhanced speech. When computing STFT spectrum in the discrete domain, the noisy signal is firstly framed with a fixed overlap (always 50% of frame length) which may make the spectrum too rough, resulting in that some key spectrum components are lost. However, these lost spectrum components are very important for whisper cognition in adverse noise environment.

In this study, a whisper intelligibility enhancement method based on Real-valued Discrete Gabor Transform (RDGT) is proposed. The motivations are three folds. Firstly, speech and noise are non-stationary signals and RDGT is a perfect technology to analysis signal of this type. Secondly, the RDGT can change the time-frequency spectrum density with a single parameter, so one can use RDGT get a spectrum of various spectral densities. Lastly, the RDGT is a Perfect Reconstruction (PR) method since the synthesis window and the analysis window satisfy the biorthogonality condition, while this is not true for the STFT.

Figure 1 shows the diagram of the proposed whispered speech intelligibility enhancement method. In the proposed method, the noisy whispered speech is decomposed into the joint time-frequency domain, where different densities of the noisy whisper spectrum are extracted using RDGT. Alogarithmic spectrum estimator is derived to get the enhanced whisper spectrum. The enhanced whisper wave is then synthesized using the attenuated noisy spectrum. Experimental results show that the proposed method can improve whisper intelligibility in various noise environments. Specifically, large gains of intelligibility are got using the spectrum extracted by RDGT with an over sample rate of 16.

REAL-VALUED DISCRETE GABOR TRANSFORM

Let x(k) denote a real finite and periodic discrete time signal with a period L, the real-valued discrete Gabor expansion is defined by Tao and Kwan (2012):

$$x(k) = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} a(m,n)\tilde{h}(k - m\bar{N})\,\mathrm{cas}\left(\frac{2\pi nk}{N}\right) \tag{1}$$

The coefficients a(m, n) can be obtained by:

$$a(m,n) = \sum_{k=0}^{L-1} x(k)\tilde{\gamma}(k - m\bar{N})\,\mathrm{cas}\left(\frac{2\pi nk}{N}\right) \tag{2}$$

In Eq. 1 and 2, $L = \bar{N}M = N\bar{M}$, M and N are the numbers of sampling points in time and frequency domains, respectively. $\bar{M}$ and $\bar{N}$ are the frequency and time sampling intervals, respectively. cas(.) = cos(.)+sin(.). $\tilde{h}(k)$ and $\tilde{\gamma}(k)$ are periodic signals with a period L and satisfy (Tao and Kwan, 2008):

$$\sum_{k=0}^{L-1} \tilde{h}(k + mN)\,\mathrm{cas}\left(\frac{2\pi nk}{\bar{N}}\right)\tilde{\gamma}(k) = \frac{L}{MN}\delta(m)\delta(n) \tag{3}$$

where, $0 \le m \le \bar{M}-1$, $0 \le n \le \bar{N}-1$, $\delta(m)$ and $\delta(n)$ are the Kronecker delta function.

In RDGT, $\tilde{h}(k)$ and $\tilde{\gamma}(k)$ are called synthesis window and analysis window respectively. Equation 3 is also called biorthogonality condition. The analysis window for a given synthesis window is usually computed via minimize l_2 norm which involved complicated inversion of large matrices. In our previous study, a new general and efficient approach was proposed to compute the analysis window for a given synthesis window in DGT (Zhou *et al*., 2012b). This method is based on a parallel lattice structure of block time-recursive structure and is also useful in RDGT. In (Zhou *et al*., 2012b), used fast discrete Fourier transform was used to compute the inverse of the block-circulant matrix. Parallel lattice structures of block time-recursive were thereafter used to compute the analysis window by exploiting the block structure of the matrices.

The condition $MN \ge L$ must be satisfied for a stable reconstruction. The critical sampling occurs when MN = L (The number of coefficients a(m, n) is equal to the number of time samples x(k)) and the over sampling occurs when $MN \ge L$. There may be a loss of information in an under sampling condition ($MN<L$).

The coefficients a(m, n) are periodic number series, that is, a(m+iM, n+jN) = a(m, n), i, j = 0, ±1, ±2, ±3,..., the relation between complex discrete Gabor transform coefficients c(m, n) and RDGT coefficients a(m, n) is described as:

$$\mathrm{Re}[c(m,n)] = \frac{a(m,n)+a(m,N-n)}{2}$$
$$\mathrm{Im}[c(m,n)] = -\frac{a(m,n)-a(m,N-n)}{2} \tag{4}$$

Tao and Kwan (2008), a unified parallel lattice structures of block time-recursive RDGT was developed not only for the computation of the DGT coefficients but also for the reconstruction of the original signal from the coefficients. They provide a more efficient and faster approach as compared to the existing RDGT algorithms for the computation of RDGT.

NOISY WHISPER SPECTRUM ATTENUATION USING RDGT SPECTRUM

Let the noisy speech y(n) = x(n)+d(n), where, x(n) and d(n) are clean speech and noise, respectively x(n) and d(n) are assumed uncorrelated. Let $Y_r(k, l)$, $X_r(k, l)$ and $D_r(k, l)$ be the RDGT coefficients of y(n), x(n) and d(n), respectively. The joint time-frequency spectrum of time k and frequency bin l is defined as:

$$Y(k,l) = \sqrt{\frac{1}{2}[Y_r(k,l)^2 + Y_r(k,N-l)^2]} \tag{5}$$

According to the uncorrelated relation between x(n) and d(n), the following equation is obtained:

$$Y(k,l) = X(k,l) + D(k,l) \tag{6}$$

where, X(k, l) and D(k, l) are spectrum of clean speech and noise at time k and frequency bin l, respectively.

The optimal logarithmic spectrum estimation is derived from:

$$\widehat{X}(k,l) = \underset{\widehat{X}}{\arg\min}\{E[\log(X(k,l)) - \log(\widehat{X}(k,l))]^2 \mid Y(k,l)\} \tag{7}$$

X(k, l) is then obtained as:

$$\widehat{X}(k,l) = \exp(E[\log X(k,l) \mid Y(k,l)]) \tag{8}$$

We assume that the spectrum of clean speech and noise are complex Gaussian variables, respectively. Given two hypotheses, H_0 and H_1, which indicate, respectively speech absence and presence at the time frequency point (k, l) in the joint time-frequency plane, and assuming a complex Gaussian distribution of the spectrum for both speech and noise, the conditional probability density functions (PDFs) of the observed signal are given by:

$$P(Y(k,\ell) \mid H_0(k,\ell)) = \frac{1}{\pi\lambda_d(k,l)}\exp\{-\frac{|Y(k,\ell)|^2}{\lambda_d(k,l)}\} \tag{9}$$

$$P(Y(k,l) \mid H_1(k,l)) = \frac{1}{\pi(\lambda_d(k,l)+\lambda_x(k,l))} \times \exp\{-\frac{|Y(k,l)|^2}{\lambda_d(k,l)+\lambda_x(k,l)}\} \tag{10}$$

where, $\lambda_x(k, l) = E\{X(k, l)^2|H_1(k, l)\}$ and $\lambda_d(k, l) = E\{D(k, l)^2\}$ are variances of the kth spectrum component of clean speech and noise signal, respectively.

Let the a posteriori SNR $\gamma(k, l)$ and a priori SNR $\xi(k, l)$ be defined by:

$$\gamma(k,\ell) = \frac{|Y(k,\ell)|^2}{\lambda_d(k,l)} \tag{11}$$

$$\xi(k,\ell) = \frac{\lambda_x(k,l)}{\lambda_d(k,l)} \tag{12}$$

Applying Bayes' rule for the conditional speech presence probability $p(k, l) = PH_1|\gamma$ to get:

$$p(k,\ell) = \left\{1 + \frac{q(k,\ell)}{1-q(k,\ell)}(1+\xi(k,\ell)) \times \exp(-\nu(k,\ell))\right\}^{-1} \tag{13}$$

where, $q(k, l) = P(H_0(k, l))$ is the a priori probability for speech absence and:

$$\nu = \frac{\gamma(k,\ell)\xi(k,\ell)}{1+\xi(k,\ell)}$$

A common noise estimation technique is to recursively average past spectral power values of the noisy measurement during periods of speech absence, and holds the estimate during speech presence. In this study, the recursive averaging is obtained by $\hat{\lambda}_x(k,\ell) = \beta\bar{\lambda}_x(k,\ell)$ such that the factor β compensates the bias when speech is absent.

$$\bar{\lambda}_x(k,\ell+1) = \bar{\alpha}_d(k,\ell)\bar{\lambda}_x(k,\ell) + [1-\bar{\alpha}_d(k,\ell)]|Y(k,\ell)|^2$$

where, $\alpha_d(k, l) = \alpha_d+(1-\alpha_d)p(k, l)$, α_d denotes the smoothing parameter ($0<\alpha_d<1$).

The improved minimum controlled recursive average (IMCRA) method was used to estimate the a priori speech absence probability (Cohen and Berdugo, 2001). In contrast to the minimum statistics based method, the smoothing of the noisy power spectrum is carried out in both time and frequency in IMCRA. This takes into account the strong correlation of the speech presence in neighboring frequency bins of consecutive frames. Furthermore, the proposed procedure comprises two iterations of smoothing and minimum tracking. The first iteration provides a rough voice activity detection in each frequency band. Then, the smoothing in the second iteration excludes relatively strong speech components, which makes the minimum tracking during speech activity robust, even when using a relatively large smoothing window.

Based on the binary hypothesis model, the spectral gain for the optimal modified logarithmic spectrum amplitude estimator is given by Cohen and Berdugo (2001):

$$G(k,\ell) = \{G_{H1}(k,\ell)\}^{p(k,\ell)} G_{min}^{1-p(k,\ell)} \tag{14}$$

where, G_{min} is a threshold which is determined by a subjective criteria for the noise naturalness when speech is absent. The conditional gain function when speech is present, is defined by Ephraim and Malah (1985):

$$G_{H1}(k,\ell) = \frac{\xi(k,\ell)}{1+\xi(k,\ell)} \exp\left(\frac{1}{2}\int_{v(k,\ell)}^{\infty} \frac{e^{-t}}{t} dt\right) \tag{15}$$

EXPERIMENTAL RESULTS

In order to evaluate the proposed algorithm, whispered speech data were collected. Fifty phonetically balanced sentences were used to produce whisper corpus. Each sentence was uttered by 3 male speakers and 3 female speakers in a quiet room, respectively. The format of the recordings is 16 kHz sampling rate, 2 bytes per sample and linear PCM.

Each clean whispered speech was artificially contaminated by noise at SNRs of -9 dB, -6 dB, -3 dB, 0 dB and 3 dB, respectively. Four types of noise recordings including Gaussian white noise, volvo car noise, babble noise and F16 fighter jet noise taken from the NOISEX-92 database were used as noise maskers (Varga and Steeneken, 1993). A noise segment of the same length as the speech signal is randomly cut out of the noise recordings and appropriately scaled to reach the desired SNR level. The Short Time Objective Intelligibility (STOI) (Taal *et al*., 2011) is computed to score the intelligibility performance for a given speech.

For comparison, the minimum mean-square error of short time spectrum amplitude based speech enhancement algorithm (denoted as MMSE-STSA) (Ephraim and Malah, 1984), the Minimum mean-square error log-spectral amplitude estimator (denoted as logMMSE) (Ephraim and Malah, 1985), the optimal gain modification based logarithmic spectrum estimator algorithm (denoted as OMLSA) (Cohen and Berdugo, 2001) and the Wiener algorithm (denoted by Wiener) (Scalart, 1996) were also used to do whisper intelligibility enhancement, respectively. In addition, the mean STOIs of the unprocessed noisy whispers (denoted as UN) was also computed.

Figure 2 ploted the whisper spectrum computed using STFT and RDGT with over sample rate of 2, 8, 16, 32 and 128, respectively. Figure 2a is whisper spectrum based on STFT. Figure 2b-f are whisper spectrum based on RDGT with over sample rate of 2, 8, 16, 32 and 128, respectively. Figure 2 indicates that the spectrum computed by RDGT with over sample rate of 32 retains more speech components than that of STFT. When the oversample rate is 128, more time-frequency point is labeled as speech component, which will not be attenuated in the enhancement stage, resulting more noise in the enhanced whisper.

Figure 3a ploted a clean whisper. Figure 3b plots the noisy whisper contaminated by Gaussian noise at SNR of -6 dB. Figure 3c-f plot the enhanced whisper using Gabor, MMSE-STSA, OMLSA and Wiener algorithm, respectively. The oversample rate of RDGT is set to 4.

Figure 4 ploted spectrum corresponding to each subplot in Fig. 3, respectively. From Fig. 3, one can see that the whisper enhanced using the proposed method have less residual noise than that obtained using other three algorithms. From Fig. 4, one can find that the enhanced whisper using RDGT retains more speech components than that enhanced using other three algorithms.

Figure 5 shows the mean STOIs of the whisper enhanced by the proposed algorithm with different over sample rates of RDGT (denoted by Gabor in Fig. 5 and the corresponded unprocessed noisy whisper (denoted by UN in Fig. 5, respectively. The noisy whisper is with SNR of -6 dB in the Gaussian noise environment.

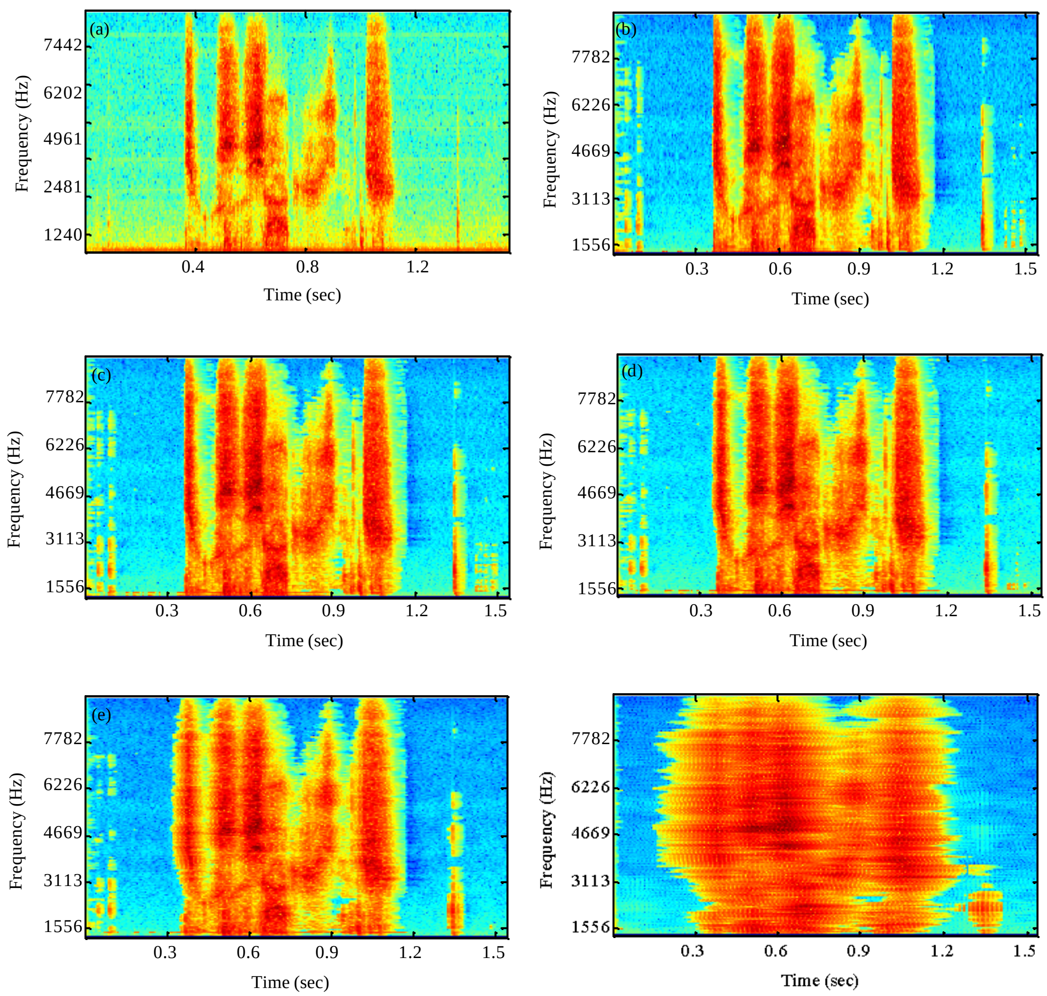

Fig. 2(a-f): Clean whisper spectrum computed via STFT and RDGT with different over sample rate β, (a) Whisper spectrum based on STFT and (b-f) Whisper spectrum based on RDGT with over sample rate of 2, 8, 16, 32 and 128, respectively

Larger over sample rate implies more time-frequency spectrum in a unit area in the time-frequency plane. From Fig. 5, one can find that, the mean STOI of Gabor is larger than that of the UN when the over sample rate is in [4, 64]. The largest difference is got when the sample rate is 16 and 32. However, the mean STOI of Gabor is smaller than that of the UN when over sample rate is 0.5 or 128. This is because that, when the over sample rate is 0.5, less noisy spectrum is used to do speech enhancement, some very important spectrum may be lost and this deduced decrease of the intelligibility of the enhanced whisper. However, when the over sample rate is too large, more unimportant noise spectrum is introduced to the enhancement stage and this also decrease the intelligibility of the enhanced whisper since the gain function does not attenuate these noise spectrum components to zero.

Since larger over sample rate also implies more computation load. In the following experiment, let over sample rate be 16 and compare the proposed algorithm to other speech enhancement algorithms.

Figure 6 plots mean STOIs of enhanced whispers using different algorithms in the context of Gaussian noise, Volvo Car noise, F16 jet noise and Babble noise, respectively. Figure 5 indicate that the proposed method gains large intelligibility improvement in all four types of noise environment with all different SNR conditions. In fact, the proposed method has the similar spectrum estimator to the logMMSE except that the noisy spectrum

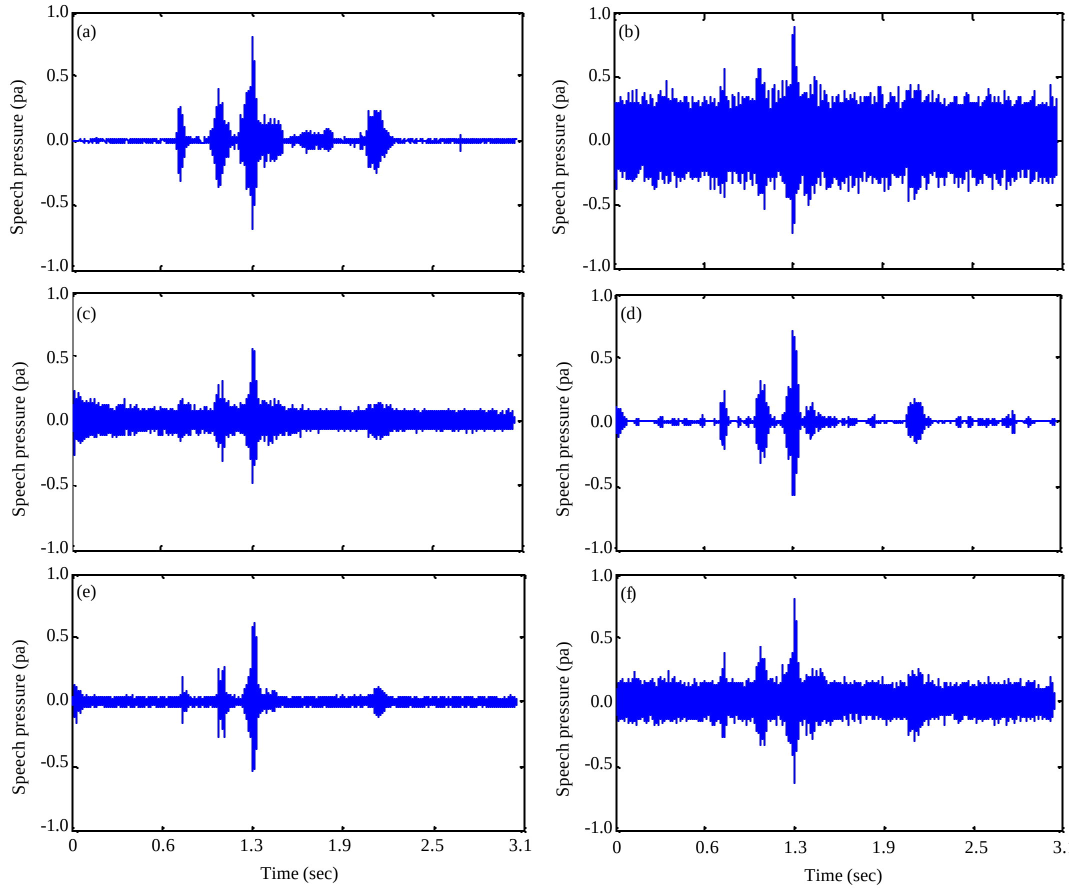

Fig. 3: Enhanced whisper wave using different algorithms in the context of Gaussian noise at SNR of -6 dB, (a) Clean whisper, (b) Noise whispers with SNR of -6 dB and (c-f) whisper enhanced using Gabor, MMSE-STSA, OMLSA and Wiener algorithm, respectively

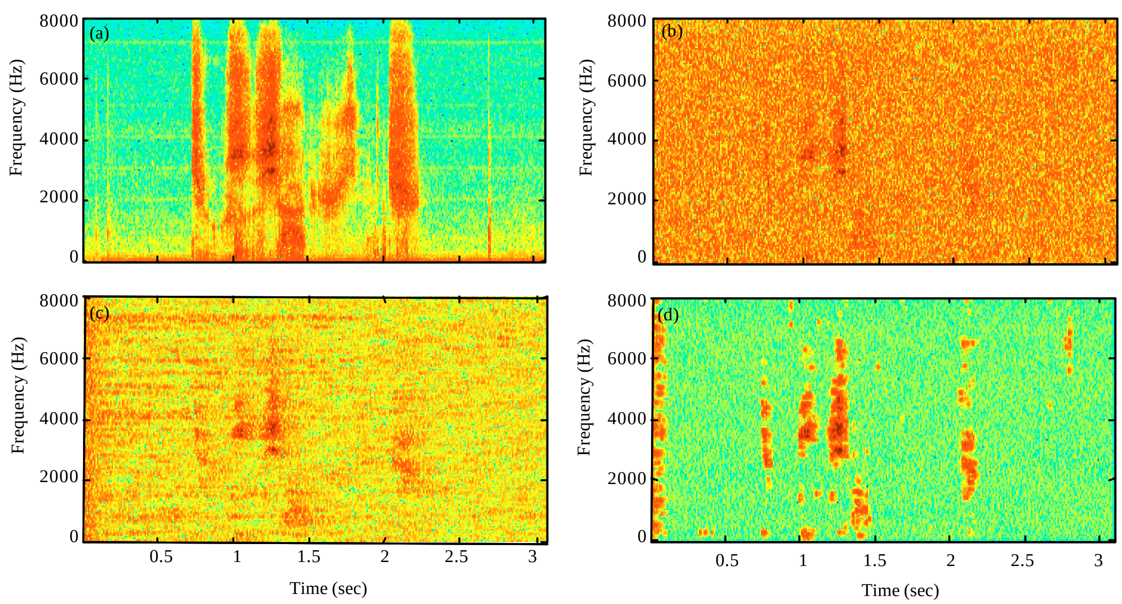

Fig. 4 (a-f): Continue

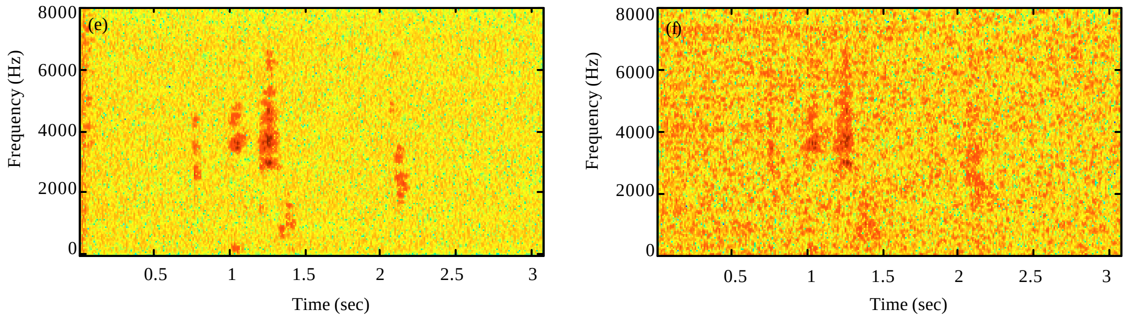

Fig. 4(a-f): Enhanced whisper spectrum corresponding to Fig. 3 using different algorithms in the context of Gaussian noise at SNR of -6 dB, (a) Spectrum of the clean whisper shown in Fig. 3a, (b) Spectrum of the noisy whisper shown in Fig. 3b (c-f) Spectrum of the enhanced whisper shown in Fig. 3c-f, respectively

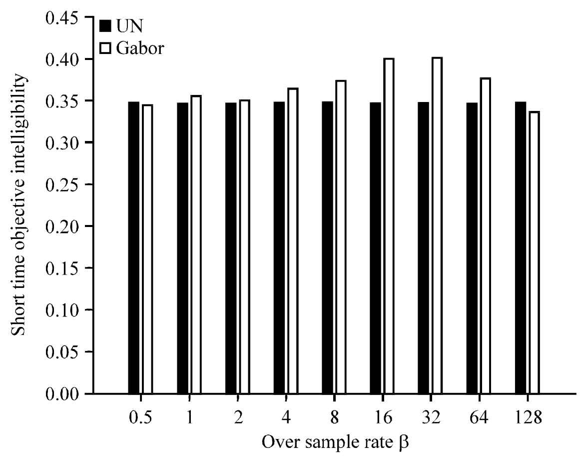

Fig. 5: Mean Short Time Objective Intelligibility (STOI) of enhanced whispers using RDGT (denoted as Gabor in the figure) under different over sample rates. The mean STOIs of the unprocessed noisy whisper (denoted as UN) is also plotted for comparison

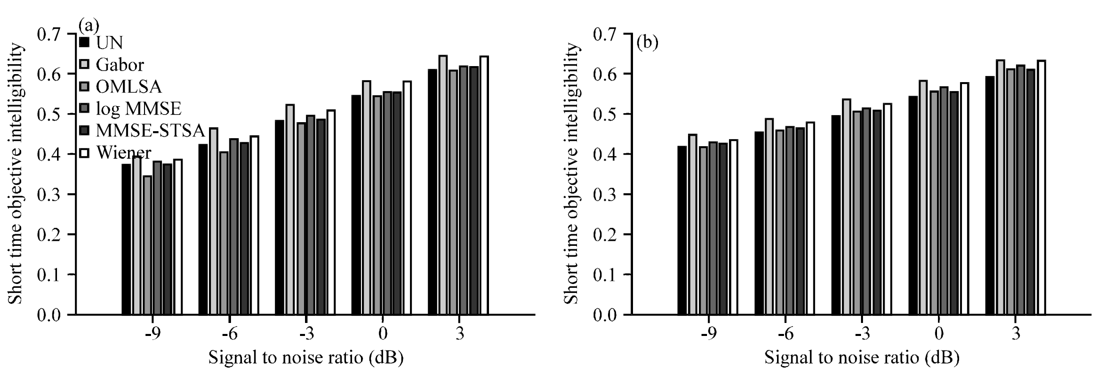

Fig. 6: Continue

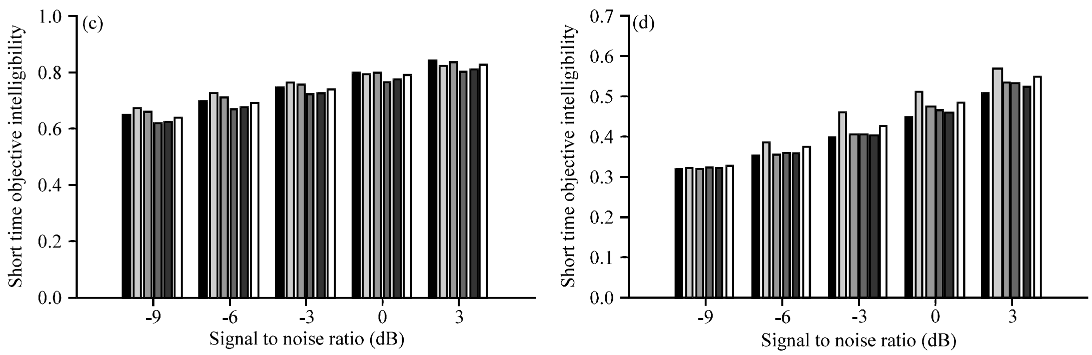

Fig. 6(a-d): Mean STOIs of enhanced whispers with different algorithms in different noise environments, (a) Gaussian white noise environment, (b) Babble noise environment, (c) Volvo car noise environment and (d) F16 fighter jet noise environment. The mean STOIs of unprocessed noisy whisper are also plotted for comparison

used in the two algorithms. The logMMSE uses STFT spectrum which is a special case of DGT and the proposed method uses spectrum computed by RDGT with over sample rate of 16. So this improvement is attributed to fine spectrum extraction by RDGT using over sample rate. This implies that when using more frequency point or increasing frame overlap in STFT may increase the intelligibility performance of traditional speech enhancement algorithms.

CONCLUSION

Whispered speech has received much attention recently. Whispered speech has much lower energy than the voiced speech. So improving the intelligibility rather than the quality of whisper in adverse noise environment is more important for its various applications. However, conventional speech enhancement algorithms failed to improve the intelligibility of the enhanced speech. One can hypothesize that this may attribute to the roughly spectrum extracted using Short Time discrete Fourier Transform (STFT).

In this study, noisy whisper spectrum was extracted based on real-valued discrete Gabor transform. There are some advantages when using RDGT to extract spectrum. Firstly, the synthesis and analysis windows are different and satisfy biorthogonal condition. Secondly there is a convenient parameter to control the unit size of spectrum in the time-frequency domain. The logarithmic spectrum estimator is derived based on spectrum with various densities using RDGT with different over sample rates.

Some conclusions can be deduced from the experimental results:

- The proposed whisper intelligibility enhancement method can improve the intelligibility of the enhanced whisper in an adverse noise environment when using a proper over sample rate value. Specially, larger gains of intelligibility are gotten when the over sample rate is set to 16 or 32
- The proposed method outperforms other three speech enhancement algorithms in terms of intelligibility improvement
- The experimental results imply that the intelligibility performance of the conventional speech enhancement may be improved when increasing the density of the time-frequency spectrum properly

ACKNOWLEDGMENTS

The Natural Science Foundation of China (No.60872073,No.51075068,No.60975017,No.61071169, No.61003131), the Doctoral Fund of Ministry of Education of China (No.20110092130004), the Autonomous Fund of Science and Technology on Acoustic Antagonizing Laboratory in 2009 (No.09ZD.2), the Science Foundation for Young Talents in the Educational Committee of Anhui province (No. 2010SQRL018) and the 211 project of Anhui university (No.2009QN027B), collegiate Natural Science Fund of Jiangsu Province (No. 10KJB510005).

REFERENCES

Ahmadi, F., I. McLoughlin and H. Sharifzadeh, 2008. Analysis-by-synthesis method for whisper-speech reconstruction. Proceedings of the IEEE Asia Pacific Conference on Circuits and Systems, November 30-December 3, 2008, Macao, China, pp: 1280-1283.

Boll, S.F., 1979. Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoust. Speech Signal Process., 27: 113-120.

Cohen, I. and B. Berdugo, 2001. Speech enhancement for non-stationary noise environments. Signal Process., 81: 2403-2418.

Ephraim, Y. and D. Malah, 1984. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. IEEE Trans. Acoust. Speech Signal Proc., 32: 1109-1121.

Ephraim, Y. and D. Malah, 1985. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. J. Acoust. Soc. Am., 33: 443-445.

Hu, Y. and P.C. Loizou, 2007. A comparative intelligibility study of single-microphone noise reduction algorithms. J. Acoust. Soc. Am., 122: 1777-1786.

Ito, T., K. Takeda and F. Itakura, 2005. Analysis and recognition of whispered speech. Speech Commun., 45: 139-152.

Jou, S.C., T. Schultz and A. Waibel, 2005. Whispery speech recognition using adapted articulatory features. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, March 18-23, 2005, IEEE, Philadelphia, PA, USA., pp: 1009-1012.

Lim, J., 1978. Evaluation of a correlation subtraction method for enhancing speech degraded by additive white noise. IEEE Trans. Speech Audio Proc., 26: 471-472.

Loizou, P.C. and G. Kim, 2011. Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions. IEEE Trans. Speech Audio Proc., 19: 47-56.

Scalart, P., 1996. Speech enhancement based on a priori signal to noise estimation. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, May 7-10, 1996, IEEE, Atlanta, Georgia, pp: 629-632.

Sharifzadeh, H.R., F. Ahmadi and I.V. McLoughlin, 2009. Speech reconstruction in post-laryngectomised patients by formant manipulation and pitch profile generation. Proceedings of the World Congress on Engineering, July 1-3, 2009, Newswood Limited, London, U.K., pp: 1838-1843.

Sharifzadeh, H.R., I.V. McLoughlin and M.J. Russell, 2012. A comprehensive vowel space for whispered speech. J. Voice, 26: 49-56.

Taal, C.H., R.C. Hendriks, R. Heusdens and J. Jensen, 2011. An algorithm for intelligibility prediction of time-frequency weighted noisy speech. IEEE Trans. Speech Audio Proc., 19: 2125-2136.

Tao, L. and H. Kwan, 2012. Multirate-based fast parallel algorithms for 2-D DHT-based real-valued discrete Gabor transform. IEEE Trans. Image Process., 21: 3306-3311.

Tao, L. and H.K. Kwan, 2008. Parallel lattice structures of block time-recursive discrete Gabor transform and its inverse transform. Signal Process., 88: 407-414.

Tartter, V., 1989. Whats in a whisper? J. Acoust. Soc. Am., 86: 1678-1683.

Varga, A. and H.J.M. Steeneken, 1993. Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems. Speech Commun., 12: 247-251.

Zhou, J., H. Wang, L. Tao and L. Zhao, 2012a. Dual window computation based on discrete Fourier transform and parallel lattice structures of block time-recursive in discrete Gabor transform. Acta Electronica Sinica, 40: 1839-1843.

Zhou, J., R. Liang, L. Zhao and C. Zou, 2012b. Whisper intelligibility enhancement using a supervised learning approach. Circuits Syst. Signal Process., 31: 2061-2074.