Journal of Southeast University (English Edition) Vol. 30, No. 3, pp. 261 – 266 Sept. 2014 ISSN 1003—7985

Intelligibility evaluation of enhanced whisper in joint time-frequency domain

Zhou Jian[1,2] Wei Xin[3] Liang Ruiyu[4] Zhao Li[2]

([1]Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Anhui University, Hefei 230601, China)
([2]Key Laboratory of Underwater Acoustic Signal Processing of Ministry of Education, Southeast University, Nanjing 210096, China)
([3]College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)
([4]College of Computer and Information, Hohai University, Nanjing 210098, China)

Abstract: Some factors influencing the intelligibility of the enhanced whisper in the joint time-frequency domain are evaluated. Specifically, both the spectrum density and different regions of the enhanced spectrum are analyzed. Experimental results show that for a spectrum of some density, the joint time-frequency gain-modification based speech enhancement algorithm achieves significant improvement in intelligibility. Additionally, the spectrum region where the estimated spectrum is smaller than the clean spectrum, is the most important region contributing to intelligibility improvement for the enhanced whisper. The spectrum region where the estimated spectrum is larger than twice the size of the clean spectrum is detrimental to speech intelligibility perception within the whisper context.

Key words: whispered speech enhancement; intelligibility evaluation; real-valued discrete Gabor transform; joint time-frequency analysis

doi: 10.3969/j. issn. 1003 – 7985. 2014. 03. 001

Recently, processing of whispered speech has received much attention[1–3]. Whisper is often used in public places where normal speech is not allowed or the speaker wants to avoid being overheard. Particularly, whisper is the only path to communication for aphonic individuals who cannot produce normal speech. An earlier study on whispers focused on phonetics and medical needs. With the rapid development of mobile communication technology, more attention has been paid to whisper applications such as whispers to normal speech transformation, whisper recognition, and whisper emotion analysis, etc.

A whisper is produced by a turbulent like excitation airflow from the lungs with no vocal cord vibration. The energy of whispered speech is much lower than that of normal speech. As a consequence, whispered speech is more susceptible to interference and canceling noise from whisper is a considerable challenge for whisper based applications in the noisy environment.

The aim of speech enhancement is to improve quality and/or intelligibility. Much progress has been made in improving speech quality in the past decade. However, there has been little progress in improving speech intelligibility[4]. Powerful speech enhancement algorithms such as the power subtraction method, the minimum mean-square error spectrum amplitude estimator method, and the Wiener method cannot improve speech intelligibility but even reduce it slightly[5–6].

The reasons why existing speech enhancement algorithms do not improve speech intelligibility is partially known. Loizou et al. [7] suggested that an over-estimation of speech in enhancement stage is a key factor in that the enhanced speech has no improvement in the aspect of intelligibility. Wang et al. [8] also found that the enhanced speech obtains greater intelligibility when the spectrum component where speech energy is larger than that of the noise spectrum is used to synthesize the enhanced speech[8]. However, these studies focused on voiced speech, and it is not clear whether these results follow in the context of whispers. Additionally, the effect of the density of spectrum used in speech algorithms on speech intelligibility is not considered in previous studies.

In this paper, we evaluate some factors affecting intelligibility of the enhanced whisper in the joint time-frequency domain. We first propose a joint time-frequency gain modification based speech enhancement algorithm where the real-valued discrete Gabor transform (RDGT) is used to obtain the time-frequency spectrum of different densities. The inspiration for using the RDGT rather than the short time Fourier transform is that the former has the ability to extract different levels of the speech spectrum density in the joint time-frequency domain with a simple parameter. Different levels of spectrum and speech over-

Received 2014-03-11.

Biographies: Zhou Jian (1981—), male, doctor, lecturer; Zhao Li (corresponding author), male, doctor, professor, zhaoli@ seu. edu. cn.

Foundation items: The National Natural Science Foundation of China (No. 61301295, 61273266, 61301219, 61201326, 61003131), the Natural Science Foundation of Anhui Province (No. 1308085QF100, 1408085MF113), the Natural Science Foundation of Jiangsu Province (No. BK20130241), the Natural Science Foundation of Higher Education Institutions of Jiangsu Province (No. 12KJB510021), the Doctoral Fund of Anhui University.

Citation: Zhou Jian, Wei Xin, Liang Ruiyu, et al. Intelligibility evaluation of enhanced whisper in joint time-frequency domain[J]. Journal of Southeast University (English Edition), 2014, 30(3): 261 – 266. [doi: 10.3969/j. issn. 1003 – 7985. 2014. 03. 001]

estimation are evaluated and analyzed for their effect on speech intelligibility, respectively.

The evaluation system is described in Fig. 1. The noisy whisper is first transformed into a joint time-frequency domain via the RDGT, where the noise spectrum is estimated. The sample rate parameter is used to control the spectrum density. The larger the over-sample rate, the more dense is the spectrum. The region control parameter is used to extract different parts of the enhanced spectrum, which reflects the over-estimation and under-estimation of speech spectrum after processing by the proposed speech enhancement algorithm. Both the spectrum density and speech over-estimation/under-estimation are evaluated for their effect on whispered speech intelligibility.

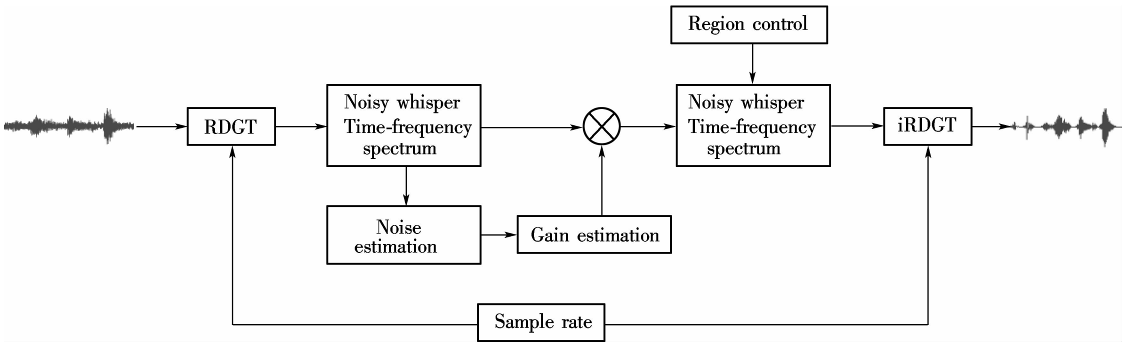

Region control
RDGT
Noisy whisper Time-frequency spectrum
Noisy whisper Time-frequency spectrum
iRDGT
Noise estimation
Gain estimation
Sample rate

Fig. 1 Diagram of the whisper intelligibility enhancement evaluation system

1 Deriving Logarithmic based Whisper Spectrum by RDGT

The RDGT[9] is defined as

$$a(m, n) = \sum_{k=0}^{L-1} z(k) \tilde{g}(k - m\overline{N}) \operatorname{cas}(2\pi nk/N) \tag{1}$$

where $a(m, n)$ are the RDGT coefficients of the signal $z(k)$. The analysis window $\tilde{g}(k)$ is a real finite and periodic discrete time signal with a period of L.

Let $x(n)$ and $d(n)$ be the sampled uncorrelated clean speech and the noise signal, respectively. The noisy whisper $y(n) = x(n) + d(n)$. The RDGT coefficients of $y(n)$, $x(n)$ and $d(n)$ are denoted as $Y_{\mathrm{r}}(k, l)$, $X_{\mathrm{r}}(k, l)$, $D_{\mathrm{r}}(k, l)$, respectively. The joint time-frequency spectrum of $y(n)$ is defined as

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

$X(k, l)$ is always estimated from $Y(k, l)$ by minimizing a non-negative error function $d(\epsilon) = d(X(k, l) - \hat{X}(k, l))$ at frequency bin k and time index l when only one microphone source is provided. The Bayesian risk of $d(\epsilon)$ is given by

$$R_{\mathrm{B}} = E\{d(X(k, l), \hat{X}(k, l))\} =$$
$$\iint d(X(k, l), \hat{X}(k, l)) p(X(k, l), Y(k, l)) \mathrm{d}X(k, l) \mathrm{d}Y(k, l) =$$
$$\int \Big[\int d(X(k, l), \hat{X}(k, l)) p(X(k, l) \mid Y(k, l)) \mathrm{d}X(k, l) \Big] \cdot$$
$$p(Y(k, l)) \mathrm{d}Y(k, l) \tag{3}$$

In general, the speech enhancement algorithms are distinct from each other in terms of different cost functions. In this paper, the logarithmic cost function $d_{\mathrm{LOG}}(X(k, l), \hat{X}(k, l)) = (\log(X(k, l)) - \log(\hat{X}(k, l)))^2$ is used as it is more suitable for speech enhancement.

By minimizing R_{B}, the logarithmic spectrum estimation $\hat{X}(k, \ell)$ is derived as

$$\hat{X}(k, \ell) = \exp(E[\log X(k, \ell) \mid Y(k, \ell)]) \tag{4}$$

where $X(k, l) = \sqrt{\frac{1}{2}[X_{\mathrm{r}}(k, l)^2 + X_{\mathrm{r}}(k, N-l)^2]}$.

Assume that the spectrum of the clean whisper and noise are complex Gaussian variables respectively. Given two hypotheses, H_0 and H_1, which indicate speech absence and presence at the time-frequency point (k, l) in the joint time-frequency plane, respectively, and assuming a complex Gaussian distribution of the spectrum for both speech and noise; the spectral gain for the logarithmic spectrum amplitude estimator is derived similarly to Ref. [10] as follows:

$$G(k, \ell) = G_{\mathrm{H1}}(k, \ell)^{p(k, \ell)} G_{\min}^{1-p(k, \ell)} \tag{5}$$

where $G_{\min}$ is a threshold which is determined by a subjective criteria for the noise naturalness when speech is absent. Also, $p(k, \ell)$ is computed by the Bayesian rule,

$$p(k, \ell) =$$
$$\left\{1 + \frac{q(k, \ell)}{1 - q(k, \ell)}(1 + \xi(k, \ell)) \exp(-\nu(k, \ell))\right\}^{-1} \tag{6}$$

In Eq. (6), $q(k, \ell) = P(\mathrm{H}_0(k, \ell))$ is the *a priori* probability for speech absence, $\nu = \frac{\gamma(k, \ell)\xi(k, \ell)}{1 + \xi(k, \ell)}$, $\gamma(k, \ell) = \frac{|Y(k, \ell)|^2}{\lambda_{\mathrm{d}}(k, l)}$ is the *a posteriori* SNR, $\xi(k, \ell) = \frac{\lambda_x(k, l)}{\lambda_{\mathrm{d}}(k, l)}$ is the *a priori* SNR. The gain function $G_{\mathrm{H1}}(k, \ell)$ is derived as

$$G_{\mathrm{H1}}(k, \ell) = \frac{\xi(k, \ell)}{1 + \xi(k, \ell)} \exp\left(\frac{1}{2}\int_{\nu(k, \ell)}^{\infty} \frac{\mathrm{e}^{-t}}{t} \mathrm{d}t\right) \tag{7}$$

Once the estimated speech spectrum is obtained, the enhanced whisper is synthesized using the inverse RDGT as

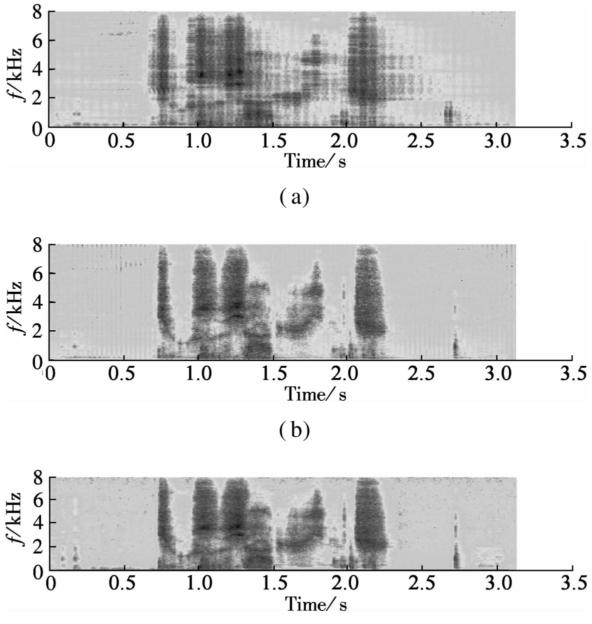
$$x(k) = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} a(m, n)\tilde{h}(k - m\bar{N})\mathrm{cas}(2\pi nk/N) \quad (8)$$

In Eq. (1) and Eq. (8), $L = \bar{N}M = N\bar{M}$, where 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 $(\cdot)$ = cos $(\cdot)$ + sin $(\cdot)$; $\tilde{h}(k)$ and $\tilde{g}(k)$ are the periodic synthesis window and analysis window with a period L, respectively and satisfy

$$\sum_{k=0}^{L-1}\tilde{h}(k + mN)\{\mathrm{cas}(2\pi nk/\bar{N})\tilde{g}(k) = \frac{L}{MN}\delta(m)\delta(n) \quad (9)$$

where $0 \leqslant m \leqslant \bar{M} - 1$; $0 \leqslant n \leqslant \bar{N} - 1$; $\delta(m)$ and $\delta(n)$ are Kronecker delta functions, respectively.

The advantage of using the RDGT to conduct a spectrum analysis is that it can compute spectrums of different densities with the sample rate defined as $\beta = MN/L$. $\beta = 1$ denotes critical sampling. $\beta < 1$ denotes under sampling and $\beta > 1$ denotes over sampling. An example is illustrated in Fig. 2, where spectrograms of three levels of densities for a whisper are plotted. As can be seen from Fig. 2, the more dense the spectrum, the more speech components which are retained in the spectrum domain. In this paper, the spectra of different densities are derived using β of different values.



(a)

(b)

(c)

Fig. 2 Whisper spectrum computed by the RDGT with different sampling rate β. (a) $\beta = 0.5$; (b) $\beta = 1$; (c) $\beta = 8$

2 Intelligibility Evaluation for Enhanced Whisper

2.1 Corpus and evaluation measurement

20 sentences were used to produce the whisper corpus. Three male and three female speakers uttered each sentence once in a soundproof environment. Four types of noise, i. e., Gaussian white noise, F16 cockpit noise, Babble noise and M109 tank noise, were used to synthesize the noisy whispers with prescribed SNRs. Noise-free speech signal and noise signal were both down-sampled to 16 kHz. Clean whispers were contaminated by noise signals at SNRs of −9, −6, −3, 0 and 3 dB, respectively. A listening test can lead to an evaluation as observed by the intended group of users. However, such tests are costly and time consuming. Other objective intelligibility measures such as the articulation index (AI) and the speech transmission index (STI) are also less appropriate for methods where noisy speech is processed by the time-frequency gain function. Recently, Taal et al.[11] proposed a short-time objective intelligibility measure (STOI) which shows a high correlation with the intelligibility of noisy and time-frequency weighted noisy speech. The STOI is a function of a time-frequency dependent intermediate intelligibility measure, which compares the temporal envelopes of clean and degraded speech in short-time regions by means of a correlation coefficient. The average of the intermediate intelligibility measure over all bands and frames is calculated as

$$d = \frac{1}{JM}\sum_{j,m}\frac{(\boldsymbol{x}_{j,m} - \mu_{\boldsymbol{x}_{j,m}})^{\mathrm{T}}(\bar{\boldsymbol{y}}_{j,m} - \mu_{\bar{\boldsymbol{y}}_{j,m}})}{\|(\boldsymbol{x}_{j,m} - \mu_{\boldsymbol{x}_{j,m}})\| \ \|\bar{\boldsymbol{y}}_{j,m} - \mu_{\bar{\boldsymbol{y}}_{j,m}}\|} \quad (10)$$

where $\boldsymbol{x}_{j,m}$ and $\boldsymbol{y}_{j,m}$ are the frame based envelope spectrum of the clean speech and the enhanced speech, respectively. $\mu(\cdot)$ refers to the sample average of the corresponding vector. M represents the total number of frames and J the number of one-third octave bands. In this paper, the STOI is used to evaluate the performance of enhanced whispers in the aspect of intelligibility.

2.2 Effect of spectrum density on speech intelligibility

Fig. 3 plots time domain waves of enhanced whispers with different algorithms in the context of Gaussian noise at SNR of −6 dB. Fig. 3(a) plots a clean whisper. Fig. 3(b) plots the noisy whisper contaminated by Gaussian noise at SNR of −6 dB. Figs. 3(c) to (f) plot the enhanced whisper using the Gabor based spectrum, MMSE-STSA[12], OMLSA[10], and the Wiener algorithm[13], respectively. The sampling rate of the RDGT is set to be 4. In addition, the spectrograms of the enhanced whispers in Fig. 3 are plotted in Fig. 4. As can be seen from Fig. 3 and Fig. 4, the enhanced whisper using RDGT retains more speech components than that with MMSE-STSA and more noise is cancelled than that with OMLSA and Wiener.

Fig. 5(a) plots the mean STOI value of enhanced whispers as a function of sample rate β. The mean STOI value of the unprocessed noisy whispers is also plotted for comparison. As can be seen from Fig. 5(a), large gains in intelligibility are achieved with the spectrum derived by

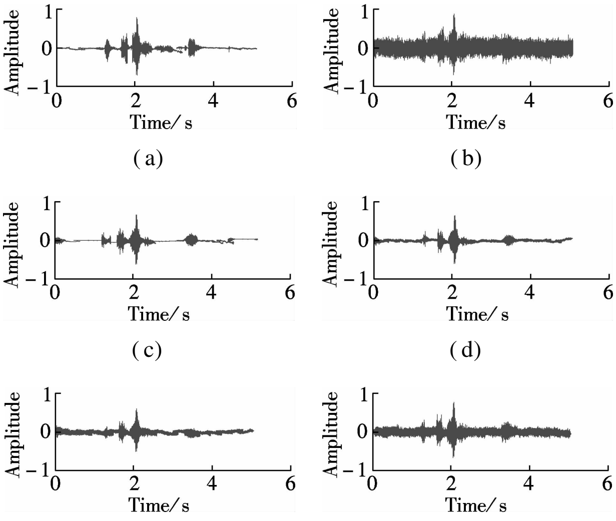


Fig. 3 Time domain waves of enhanced whispers using different algorithms in context of Gaussian noise at SNR of −6 dB. (a) Clean whisper; (b) Noisy whisper contaminated by Gaussian noise at SNR of −6 dB; (c) Enhanced whisper using Gabor based spectrum; (d) MMSE-STSA[12]; (e) OMLSA[10]; (f) Wiener algorithm[13]

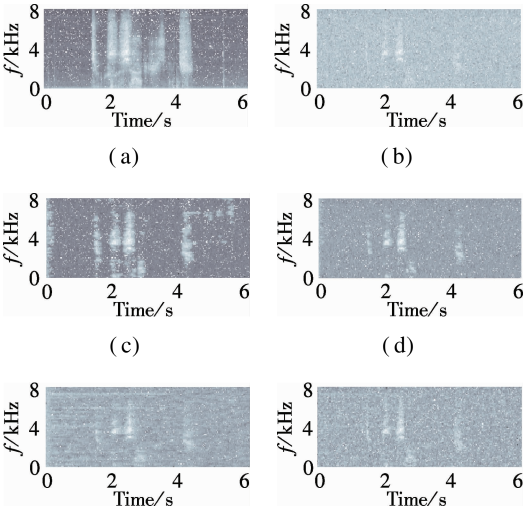


Fig. 4 Spectrograms of enhanced whispers using different algorithms in context of Gaussian noise at SNR of −6 dB. (a) Clean whisper; (b) Noisy whisper contaminated by Gaussian noise at SNR of −6 dB; (c) Enhanced whisper using Gabor based spectrum; (d) MMSE-STSA[12]; (e) OMLSA[10]; (f) Wiener algorithm[13]

the RDGT with $\beta = 32$. This implies that the conventional speech enhancement algorithms maybe improve both the quality and intelligibility at the same time when using a more dense spectrum. Fig. 5(b) plots mean STOI value of the enhanced whispers as the function of SNR. The noisy whispers are contaminated by Gaussian white noise. The estimated whispers are obtained using the RDGT with $\beta = 32$. The mean STOI value of the unprocessed whispers is also plotted for comparison. As can be seen from Fig. 5(b), the conventional speech enhancement algorithms can improve speech intelligibility when the spectrum of some appropriate density is used.

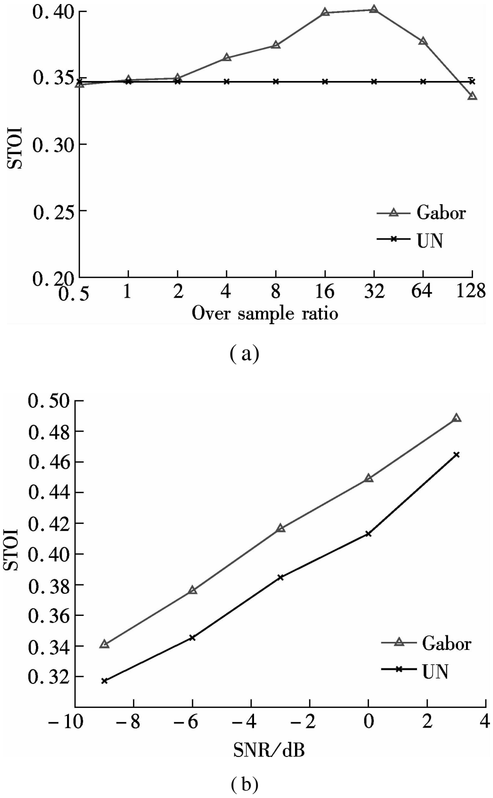


Fig. 5 The effect of sampling rate and SNR on STOI value of enhanced whisper (denoted as Gabor) and unprocessed whisper (denoted as UN). (a) STOI value as a function of sampling rate; (b) STOI value as a function of SNR

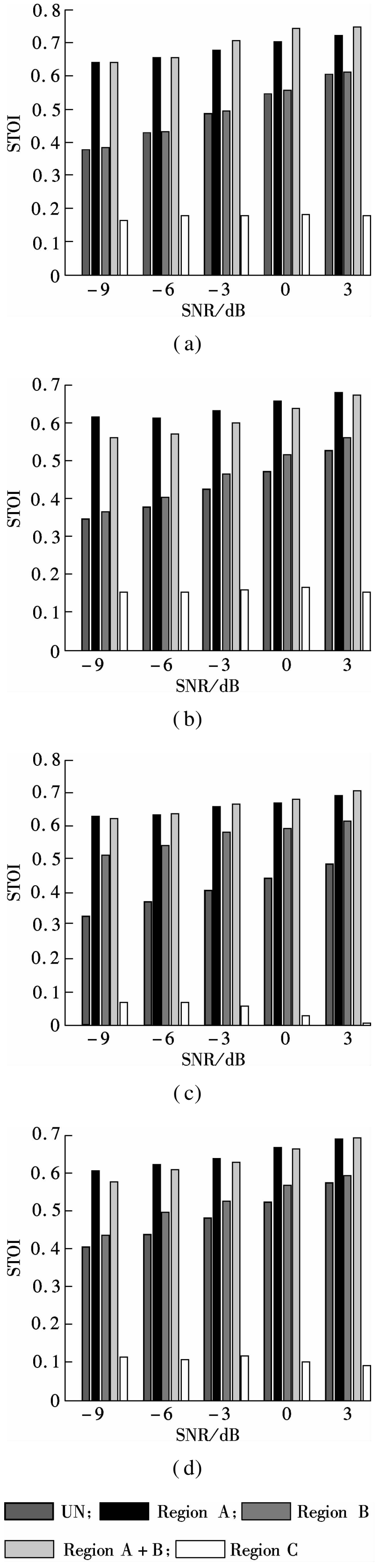
2.3 Effect of over-estimation/underestimation on speech intelligibility

In order to evaluate the effect of spectrum components of different regions on speech intelligibility, we divide the enhanced spectrum into three disjoint regions:

$$
\begin{aligned}
&\text{A}: \tilde{X}(k,l) = \begin{cases} \hat{X}(k,l) & \hat{X}(k,l) < X(k,l) \\ 0 & \text{otherwise} \end{cases} \\
&\text{B}: \tilde{X}(k,l) = \begin{cases} \hat{X}(k,l) & X(k,l) \leqslant \hat{X}(k,l) \leqslant 2X(k,l) \\ 0 & \text{otherwise} \end{cases} \\
&\text{C}: \tilde{X}(k,l) = \begin{cases} \hat{X}(k,l) & \hat{X}(k,l) > 2X(k,l) \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
\tag{11}
$$

The regions of A, B, C and A + B are then used to synthesize the enhanced whisper, respectively, and the intelligibility of which is then evaluated using the STOI.

Fig. 6 plots the mean STOI value of enhanced whispers using different regions of the enhanced spectrum which are derived using the proposed method in different noise environments. For comparison, the mean STOI of unprocessed whispers (denoted as UN) is also plotted in Fig. 6. As can be seen from Fig. 6, the enhanced whisper reconstructed by region A gains large intelligibility improvement under different test conditions. The region A + B has similar intelligibility performance to region A. Region C, however, has a detrimental effect on intelligibility improvement.



(a)

(b)

(c)

(d)

(e)

Fig. 6 Mean STOI value of enhanced whispers using different regions of enhanced spectrum in different noise environments. (a) Gaussian; (b) F16; (c) Babble; (d) M109

Fig. 7 plots parametric gain curves under different *a posteriori* SNRs. As can be seen from Fig. 7, the *a priori* SNR becomes lower with the decrease in G_{H1}, no matter what the *a posteriori* SNR γ is. It implies that speech distortion (i. e., under-estimation) occurs easily in low SNR and noise distortion (i. e., over-estimation) occurs easily in high SNR. As a consequence, in a low SNR environment, the time-frequency unit with high local SNR (>0 dB) will be underestimated. This is confirmed in Tab. 1. As can be seen from Tab. 1, 74. 03% of the speech-dominated time-frequency units fall into region A after processing. Most noise dominated time-frequency units also fall into this area after processing. This may be another factor which improves speech intelligibility in the whisper context.

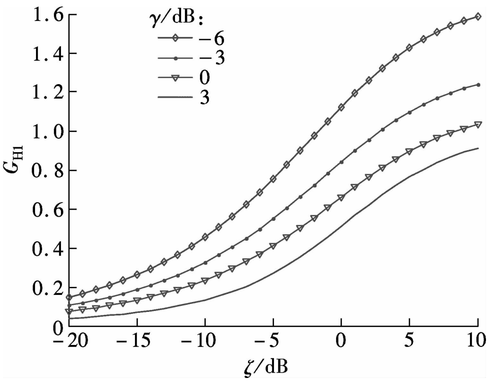


Fig. 7 Parametric gain curves of Eq. (7) as a function of the *a prior*i SNR

Tab. 1 Spectrum components falling into three regions after processing %

Dominated area type	Region A	Region B	Region C
Speech dominated area ($\xi>1$)	74.03	25.86	0.11
Noise dominated area ($\xi\leqslant 1$)	76.52	6.36	17.72

Region B represents the estimated clean spectrum $\hat{X}(k, l)$ satisfying $X(k, l) \leqslant \hat{X}(k, l) \leqslant 2X(k, l)$. In region B, the estimated speech spectrum has been over-estimated. There is no speech distortion in this region but much residual noise of some level is retained in the enhanced spectrum. It implies that in the whisper context, the speech amplification distortion of less than 6. 02 dB is insignificant for whispered speech intelligibility perception when compared with the unprocessed whisper.

The intelligibility of the enhanced whisper reconstructed by the spectrum of region A + B does not have distinct improvement when comparing with that of region A. This is because region A represents an under-estimation and region B represents an over-estimation. When the spectrum of region A + B is used to reconstruct the enhanced whisper, speech distortion and residual noise coexist in the enhanced whisper, resulting in no distinct intelligibility improvement.

3 Conclusion

We evaluate the intelligibility of the enhanced whispered speech in the joint time-frequency domain in different noisy environments. A more dense spectrum is beneficial for conventional single channel speech enhancement algorithms in terms of speech intelligibility improvement. The rough spectrum used by the conventional speech enhancement algorithms, however, is a detrimental factor resulting in intelligibility decrease. We find from the experiments that the region with $\hat{X} \leqslant X$ is the most important area for whisper information cognition. Therefore, low amplification distortion of less than 6.02 dB is harmless to speech intelligibility. However, the region with $\hat{X} \geqslant 2X$ is detrimental to intelligibility improvement.

References

[1] Remijn G, Kikuchi M, Yoshimura Y, et al. Cortical hemodynamic response patterns to normal and whispered speech [J]. *The Journal of the Acoustical Society of America*, 2013, **133**(5):3606－3606.

[2] Ruggles D, Riddell A, Freyman R L, et al. Intelligibility of voiced and whispered speech in noise in listeners with and without musical training [C]//*Proceedings of Meetings on Acoustic*. Montreal, Canada, 2013: 50－64.

[3] Sarria-Paja M, Falk T H. Whispered speech detection in noise using auditory-inspired modulation spectrum features [J]. *IEEE Signal Processing Letters*, 2013, **20**(8):783－786.

[4] Loizou P. *Speech enhancement: theory and practice* [M]. New York: CRC, 2007.

[5] Hu Y, Loizou P. A comparative intelligibility study of single-microphone noise reduction algorithms [J]. *The Journal of the Acoustical Society of America*, 2007, **122**(3):1777－1786.

[6] Li J, Yang L, Zhang J, et al. Comparative intelligibility investigation of single-channel noise-reduction algorithms for Chinese, Japanese, and English [J]. *The Journal of the Acoustical Society of America*, 2011, **129**(5):3291－3301.

[7] Loizou P, Kim G. Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions [J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2011, **19**(1):47－56.

[8] Wang D, Kjems U, Pedersen M, et al. Speech intelligibility in background noise with ideal binary time-frequency masking [J]. *The Journal of the Acoustical Society of America*, 2009, **125**(4): 2336－2347.

[9] Tao L, Kwan H. Multirate-based fast parallel algorithms for 2-D DHT-based real-valued discrete Gabor transform [J]. *IEEE Transactions on Image Processing*, 2012, **21**(7):3306－3311.

[10] Cohen I, Berdugo B. Speech enhancement for non-stationary noise environments [J]. *Signal Processing*, 2001, **81**(11):2403－2418.

[11] Taal C, Hendriks R, Heusdens R, et al. An algorithm for intelligibility prediction of time-frequency weighted noisy speech [J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2011, **19**(7):2125－2136.

[12] Ephraim Y, Malah D. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator [J]. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1984, **32**(6):1109－1121.

[13] Scalart P. Speech enhancement based on a priori signal to noise estimation [C]//*Proceedings of Acoustics, Speech, and Signal Processing*. Atlanta, USA, 1996: 629－632.

联合时频域中增强后耳语音的可懂度评估

周 健[1,2] 魏 昕[3] 梁瑞宇[4] 赵 力[2]

([1] 安徽大学智能计算与信号处理教育部重点实验室,合肥 230601)
([2] 东南大学水声信号处理教育部重点实验室,南京 210096)
([3] 南京邮电大学通信与信息工程学院,南京 210003)
([4] 河海大学计算机与信息学院,南京 210098)

摘要:对在联合时频域影响增强后耳语音可懂度的因素进行了评估. 分析了耳语音时频谱密度和增强后耳语音时频谱中不同区域对耳语音可懂度的影响. 实验结果表明,在基于增益修正的时频域语音增强算法中,采用密度较高的耳语音谱可提高增强后耳语音可懂度. 此外,在增强后的耳语音的时频谱中,频谱幅度小于干净耳语音时频谱的频谱区域对增强后的耳语音的可懂度提高最为重要,而那些频谱幅度大于2倍干净耳语音频谱的频谱区域对增强后的耳语音的可懂度具有消极作用.

关键词:耳语音增强;可懂度评价;实值离散 Gabor 变换;联合时频分析

中图分类号:TN912.35