Progress In Electromagnetics Research B, Vol. 9, 199–214, 2008

THE ENHANCEMENT OF MILLIMETER WAVE CONDUCT SPEECH BASED ON PERCEPTUAL WEIGHTING

S. Li, J. Q. Wang, M. Niu, T. Liu, and X. J. Jing †

Department of Biomedical Engineering
The Fourth Military Medical University
Xi'an 710032, China

Abstract—A new non-air conduct speech detecting method is introduced in this paper by means of millimeter wave (MMW) radar. Due to its special attribute, this method may provide some exciting possibility of wide applications. However, the resulting speech is of less intelligible and poor audibility since the present of the combined and colored additive noise. This paper, therefore, investigates the problem of the MMW radar speech enhancement by taking into account the frequency-domain masking properties of the human auditory system and reduces the perceptual effect of the residual noise. Considering the particular characteristics of MMW speech, the perceptual weighting technique is developed and incorporated into the traditional spectral subtraction algorithm to shape the residual noise and make it inaudible. The results from both acoustic and listening evaluation suggest that the background noise can be reduced efficiently while the distortion of MMW radar speech remains acceptable, the enhanced speech also sounds more pleasant to human listeners, suggesting that the proposed algorithm achieved a better performances of noise reduction over other subtractive-type algorithms.

1. INTRODUCTION

In order to detecting the speech signals which is produced by the vibration of the vocal folds [1, 2], traditional acoustic sensors, such as microphone, are well know to be used to detect this vibration of the vocal folds. The prerequisite of this method is that the speech can be

† The first author is also with The Key Laboratory of Biomedical Information Engineering of Ministry of Education, Department of Biomedical Engineering, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, P. R. China

spread by means of air. However, air is not the only medium which can spread and be used to detect speech. For example, voice content can be transmitted by way of bone vibrations. This vibration, therefore, can be picked up at the top of the skull using the bone-conduction sensors, strong voicing can be provide using this method [3]. Other medium, such as infrared ray, light wave, and laser also can be used to detect the non air conducted speech or acoustical vibrations, however, their application are limited since the materials in detail are usually difficult to obtain [4].

Another medium, millimeter wave (MMW, as well as light and laser), was reported by previous study that this medium can be used to detect and identify out exactly the existential speech or acoustical signals in free space from a person speaking through the electromagnetic wave fields by principle and experiment [4]. Since the microwave radar has low range attenuation, better sense of direction, and has attribute of noninvasive, safe, fast, portable, low cost fashion [5], it may extend traditional speech detecting method to a large extent, and provide some exciting possibility of wide applications: the speech and acoustic signal directional detection in complex and rumbustious acoustic environment, due to its better sense of direction; the tiny acoustic or vibrant signal detection which cannot be detected by traditional microphone; the microwave radar also can be used in clinic assistant diagnosis or measure speech articulator motions [6].

Nevertheless, there has been little previous research work concentrated on the MMW radar speech. Previous studies with respect to the MMW radar speech concentrated on the MMW non acoustic sensors [5, 7], and focused on the measurement of speech articulator motions, such as vocal tract measurements and glottal excitation [6], but not on the MMW speech itself. Therefore, there is a need to explore this new speech detecting way, as well as corresponding speech enhancement method to extent the traditional speech detecting method.

Although MMW radar provides another important method to detect speech or other acoustic signal, the MMW radar speech itself has several serious shortcomings including artificial quality, reduced intelligibility, and poor audibility. This is not only because some harmonic of the MMW and electrocircuit noise are combined in the detected speech due to the different detecting methods from traditional air conduct speech, but also the channel noise, as well as ambient noise combined in the MMW radar speech. These combined noise components are quite larger and more complex than traditional air conduct speech, and are the biggest problem which must be resolved for the application of the MMW radar speech. Therefore, speech enhancement is a challenging topic of MMW radar speech research.

Considering the limit effects on the speech enhancement by improving the hardware system, the researchers are impelled to consider the use of signal enhancement techniques to improve the resulting speech. Two main methods have been put forward to improve intelligibility and naturalness of MMW speech. One method is the subtractive-type algorithm [8], which is the most widely used, and has been shown to be an effective approach for noise canceling. Due to the simplicity of implementation, and low computational load, the spectral subtraction method is the primary choice for real time applications [9]. In general, this method enhanced the speech spectrum by subtracting an average noise spectrum from the noisy speech spectrum, here the noise is assumed to be uncorrelated and additive to the speech signal. The phase of the noisy speech is kept unchanged, since it is assumed that the phase distortion is not perceived by human ear. However, the serious draw back of this method is that the enhanced speech is accompanied by unpleasant musical noise artifact which is characterized by tones with random frequencies. Although many solutions have been proposed to reduce the musical noise in the subtractive-type algorithms [10–14], results performed with these algorithms show that there is a need for further improvement, especially at very low SNR condition. Furthermore, spectral subtraction algorithm is in general effective in reducing the noise but not in improving intelligibility, the subtraction parameters are also fixed and cannot be adapted frame to frame. The other method has been done by introducing knowledge on human perception in the enhancement process [15]. Some methods have been developed in this direction, by modeling several aspects of the enhancement function present in the auditory system [16–18], and is already widely used in perceptual wideband audio coding [19]. The auditory masking model is based on the masking phenomenon which is related to the notion of the critical band analysis, which takes into account the frequency domain masking properties of the human auditory system. The noise masking properties are modeled by calculating a noise masking threshold, which represents the maximum level of injected noise that will be inaudible when added to the input signal. However, this method relies on pre-enhancing the signal to estimate the thresholds, which will increase the computational load and decrease the computationally efficient.

Therefore, in this study, we propose a perceptually motivated approach for speech enhancement which does not need the pre-enhancing of the signal, furthermore, the subtraction parameters also can be adapted frame by frame. The proposed approach is based on the perceptual weighting technique used in low-rate analysis-by-synthesis speech coders [20, 21], which takes into account the frequency domain masking properties of the human auditory system. A minimized perceptually weighted error criterion was used by analysis-by-synthesis speech coders to obtain the optimal excitation for LPC (Linear Predictive Coding) synthesis in a closed-loop manner [21], this criterion, furthermore, can be calculated by frequency masking experiments [22] which showed that the noise near formant peaks is inaudible to the human ear, where the speech signal has high energy. Thus the noisy speech can be enhancement by distributed the quantizing noise unevenly across the frequency bands in a way that it is masked by the speech signal, this can be done by using a perceptual filter derived from the LPC analysis of the speech signal [20]. This method may be more effective than traditional spectral subtraction methods and the auditory masking algorithm not only because it allows one to find the best tradeoff between noise reduction, the speech distortion and the level of residual noise *in a perceptual sense*, but also increase the computationally efficient by decreasing the computational load.

2. METHOD

2.1. The Description of the System

The schematic diagram of the speech-detection system is shown in Figure 1. A phase-licked oscillator generates a very stable MMW at 34 GHz with an output power of 50 mW. The output of the amplifier is fed through a 10 dB directional coupler, a variable attenuator, a circulator, and then to a flat antenna. The 10 dB directional coupler branches out 1/4 of the amplifier output to provide a reference signal for the mixer. The variable attenuator controls the power level of the microwave signal to be radiated by the antenna. The radiated power

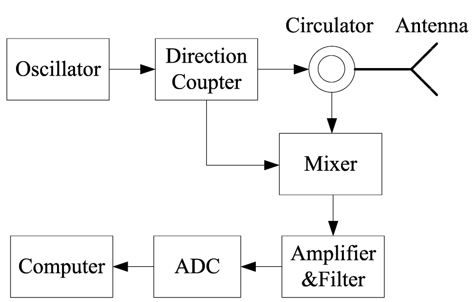


Figure 1. Schematic diagram of the speech-detection system.

of the antenna is usually kept at a level of about 10–20 mW. The flat antenna radiates a microwave beam of about 9° beam width aimed at the opposing human subjects standing or sitting directly in front of the antenna. The echo signal is received by the same antenna, which is a 34 GHz MMW signal modulated by the speech which is produced by the larynx of the opposing human subjects. This signal is then mixed with reference signal in a double-balanced mixer. The mixing of the amplified speech signal and a reference signal in the double-balanced mixer produces low-frequency signals and is amplified by a signal processor and then passed through a A/D converter before reaching computer to get further processor. For More details of description of the system, the reader is referred to [23] and [24].

2.2. Spectral Subtraction Algorithm

The spectral subtraction algorithm is based on the assumption that the additive noise to be stationary and uncorrelated with the clean speech signal. If $y(n)$, the noisy speech, is composed of the clean speech signal $s(n)$ and the uncorrelated additive noise signal $d(n)$, then:

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

The power spectrum of the corrupted speech can be approximately estimated as:

$$|Y(\omega)|^2 \approx |S(\omega)|^2 + |D(\omega)|^2 \tag{2}$$

where $|Y(\omega)|^2$, $|S(\omega)|^2$ and $|D(\omega)|^2$ represent the noisy speech short-time spectrum, the clean speech short-time spectrum, and the noise power spectrum estimate, respectively.

Most of the subtractive-type algorithms have different variations allowing for flexibility in the variation of the spectral subtraction. Berouti et al. [25] proposed the generalized spectral subtraction scheme is described as follows:

$$|\hat{S}(\omega)|^\gamma = \begin{cases} |Y(\omega)|^\gamma - \alpha|\hat{D}(\omega)|^\gamma, & \text{if } \dfrac{|\hat{D}(\omega)|^\gamma}{|Y(\omega)|^\gamma} < \dfrac{1}{\alpha+\beta} \\ \beta|\hat{D}(\omega)|^\gamma, & \text{otherwise,} \end{cases} \tag{3}$$

where $\alpha(\alpha > 1)$ is the over-subtraction factor [25], which is a function of the segmental SNR. $\beta(0 \leq \beta \leq 1)$ is the spectral floor, and γ is the exponent determining the transition sharpness. Here we set $\gamma = 2$, and $\beta = 0.002$.

The over-subtraction factor α is a function of the segmental noisy signal to noise ratio SNR which is calculated as:

$$SNR(\omega) = \frac{|Y(\omega)|^2}{|\hat{D}(\omega)|^2} \tag{4}$$

Both factors, α and β can be adjusted for different speech conditions to get better speech quality.

2.3. Speech Enhancement Based on Perceptual Weighting Technique

The main draw back of traditional spectral subtraction method is that the algorithms with fixed subtraction parameters are unable to adapt to varying noise levels and noise characteristics, however, the optimization of the parameters is a difficult task, because the spectrum of most of the noise which is added in speech is not "flat", but "colored" [13, 14]. An example of adaptation is multi-band spectral subtraction method, as an nonlinear spectral subtraction algorithm, this method adapts the over-subtraction factor α and β in time and frequency based on the SNR, leading to improved results [8]. The present algorithm also performed an adaptation of the parameters in frequency domain, but based on the perceptual weighting.

2.3.1. Perceptual Weighting Technique

The theory of psychoacoustics and auditory masking suggests that the auditory system has a limited ability to detect the quantization noise near the high energy regions of the speech spectrum, especially in the formant regions [20]. That is, quantization noise near the formant peaks is masked by the formant peaks, and is therefore not audible. Therefore, perceptual weighting technique can be exploited by shaping the frequency spectrum of the error criterion, so that less emphasis is placed near the formant peaks and more emphasis is placed on the spectral valleys, where any amount of noise present will be audible [26]. The error criterion, as stated before, has already been used in most low-rate speech in the excitation used for LPC synthesis, and can be shaped using the following filter [26]:

$$P(z) = \frac{A\left(\dfrac{z}{\sigma_1}\right)}{A\left(\dfrac{z}{\sigma_2}\right)} == \frac{1 - \displaystyle\sum_{k=1}^{p} a_k \sigma_1^k z^{-k}}{1 - \displaystyle\sum_{k=1}^{p} a_k \sigma_2^k z^{-k}} \tag{5}$$

where $A(z)$ is the LPC polynomial, a_k are the short-term linear prediction coefficients, σ_1 and σ_2 $(0 \leq \sigma_2 \leq \sigma_1 \leq 1)$ are parameters which control the energy of the error in the formant regions and p is the prediction order. Therefore, the perceptual weighting filter and its frequency response can be calculated by Eq. (5), where a_k is the linear prediction coefficients, which can be calculate from the noisy speech.

An example of the spectrum of the perceptual weighting filter for the vowel /a/ extracted from Chinese word "daxue" ("University" in English) is shown in Figure 2. It can be seen from the figure that more emphasis is placed in the spectral valleys and less emphasis is placed near the formant peaks. According to our experiments and analysis with different noise types and levels for selecting the appropriate values for parameters, the following values is choose to obtain a good tradeoff between residual noise and speech distortion: $p = 24$; $\sigma_1 = 1$ and $\sigma_2 = 0.8$.

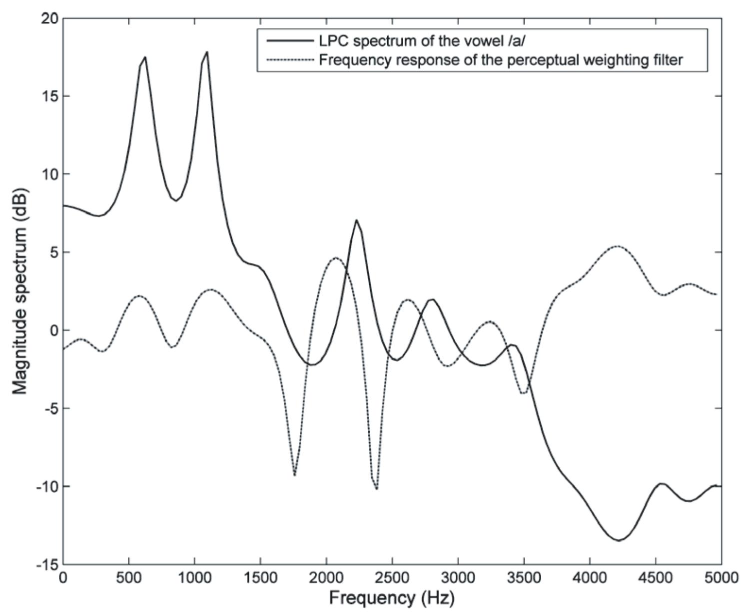


Figure 2. LPC spectrums of the vowel /a/, the frequency responses of the corresponding perceptual weighting filters are also shown, with p, σ_1 and σ_2 set to 24, 1 and 0.8, respectively.

The signal which is used to calculate $P(z)$ is not the pre-enhanced speech by traditional spectral subtraction method, but the noisy speech. This is because the procedure of speech enhancement will introduce a distortion in enhanced speech and lead to the inaccurate estimation of $P(z)$ [26]. The perceptual weighting technique is incorporated into the traditional spectral subtraction algorithm to shape the residual noise and make it inaudible. Therefore, the whole proposed MMW speech enhancement scheme is shown in Figure 3.

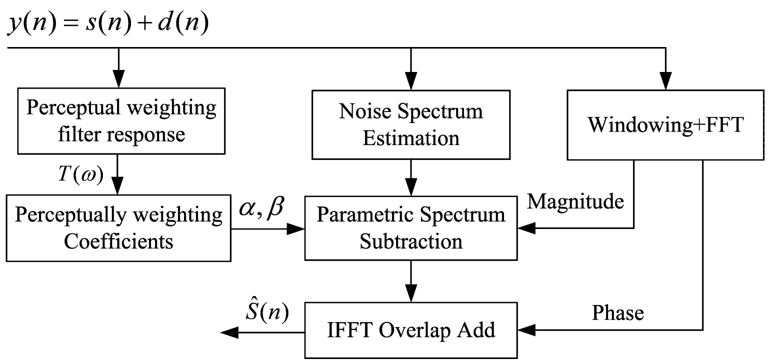


Figure 3. The proposed MMW speech enhancement scheme.

The perceptually weighting filter response, which is denoted as $T(\omega)$, can be calculated *directly* from noisy speech (compare to [8]). Therefore, the subtraction parameters α and β can be adapted in time and frequency based on $T(\omega)$, thus the enhanced speech can be obtained by the inverse transform with the combination of the calculated speech spectral magnitude and the noisy phrase.

2.3.2. Spectral Subtraction Parameters Adaptation

The subtraction parameters α and β can be adapted by the masking threshold $T(\omega)$, through which way we can find the best solution to choose the better enhancement parameters to put the residual noise stays below the masking threshold of the auditory system. That is, if the masking threshold is low, the subtraction parameters will be increased to reduce the noise, and if the masking threshold is high, the subtraction parameters will be decreased to keep distortion as low as possible [27]. The adaptation of the subtraction parameters is performed with the following relations [28]:

$$\alpha = \alpha_{\max}\left(\frac{T(\omega)_{\max} - T(\omega)}{T(\omega)_{\max} - T(\omega)_{\min}}\right) + \alpha_{\min}\left(\frac{T(\omega) - T(\omega)_{\min}}{T(\omega)_{\max} - T(\omega)_{\min}}\right) \tag{6}$$

$$\beta = \beta_{\max}\left(\frac{T(\omega)_{\max} - T(\omega)}{T(\omega)_{\max} - T(\omega)_{\min}}\right) + \beta_{\min}\left(\frac{T(\omega) - T(\omega)_{\min}}{T(\omega)_{\max} - T(\omega)_{\min}}\right) \tag{7}$$

where $T(\omega)_{\min} \leq T(\omega) \leq T(\omega)_{\max}$, where $\alpha_{\max}$, $\alpha_{\min}$, $\beta_{\max}$, $\beta_{\min}$, and $T(\omega)_{\min}$, $T(\omega)_{\max}$ are the minimal and maximal values of α, β and $T(\omega)$. It can be seen from equations that α, β achieves the maximal and the minimal values when $T(\omega)$ equals its minimal and maximal values.

According to our experiments and analysis with different noise types and levels for selecting the appropriate values for these parameters, the following values are chosen: $\alpha_{\min} = 1$; $\alpha_{\max} = 6$; $\beta_{\min} = 0$ and $\beta_{\max} = 0.02$.

2.4. Noise Estimation

The noise in the radar speech, which included of each order of the MMW harmonic, the channel noise, the ambient noise combined in the MMW radar speech, and so on, is highly nonstationary noise, it is imperative to update the estimate of the noise spectrum frequently. This study adopted the minimum-statistics method proposed by Cohen and Berdugo (2002) [29] for noise estimation, since this method is computationally efficient, robust with respect to the input signal-noise ratio (SNR), and have an ability to quick follow the abrupt changes in the noise spectrum. The minimum tracing is based on a recursively smoothed spectrum which is estimated using first-order recursive averaging

$$\left|\hat{D}_{(k,l)}(\omega)\right|^2 = \lambda_D \left|\hat{D}_{(k-1,l)}(\omega)\right|^2 + (1-\lambda_D)\left|\hat{Y}_{(k,l)}(\omega)\right|^2 \quad 0 < \lambda_D < 1, \tag{8}$$

where $\left|\hat{D}_{(k,l)}(\omega)\right|^2$ and $\left|\hat{Y}_{(k,l)}(\omega)\right|^2$ are the kth components of noise spectrum and noisy speech spectrum at the frame l, and λ_D is a smooth parameter. Let $p'(k,l)$ denote the conditional signal presence probability in Cohen and Berdugo (2002) [29], then Eq. (8) implies

$$\left|\hat{D}_{(k,l)}(\omega)\right|^2 = \hat{\lambda}_D(k,l) \left|\hat{D}_{(k-1,l)}(\omega)\right|^2 + (1-\hat{\lambda}_D(k,l))\left|\hat{Y}_{(k,l)}(\omega)\right|^2 \tag{9}$$

where $\hat{\lambda}_D(k,l) \triangleq \lambda_D + (1-\lambda_D)p'(k,l)$ is a time-varying smoothing parameter. Therefore, the noise spectrum can be estimated by averaging past spectral power values. For More details of description of this algorithm, the reader is referred to [29, 30].

3. EXPERIMENTS

3.1. Subjects

Ten healthy volunteer speakers participated in the radar speech experiment including 6 males and 4 females. All of the subjects were native speakers of mandarin Chinese, there ages varied from 20 to 35, with a mean age of 28.1 (SD = 12.05). All of the experiments are in terms of the consent form which was signed by volunteers according to the Declaration of Helsinki (BMJ 1991; 302: 1194).

The distance between the radar antenna and the human subject ranges from 2 m to 10 m, and one sentences of mandarin Chinese "Di Si Jun Yi Da Xue" (other sentences were also used, but they were not representative) uttered by the volunteer speakers were used to evaluate the proposed auditory masking approach.

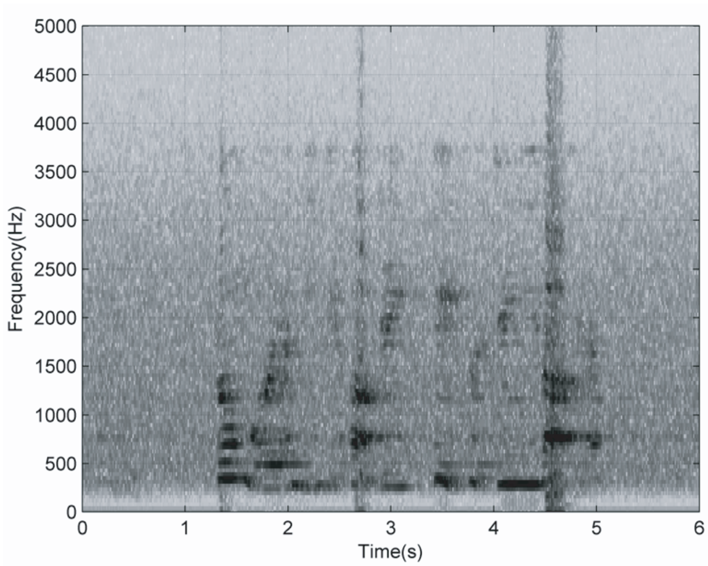
3.2. Evaluation

Generally, subtractive-type enhancement systems produce two main undesirable effects: residual noise and speech distortion. These two effects can be annoying to a human listener, and causes listeners fatigue. However, they are difficult to quantify. Therefore, it is important to analyze the time frequency distribution of the enhanced speech, in particular the structure of its residual noise. The speech spectrogram is a good tool to do this work, because it can give more accurate information about residual noise and speech distortion than the corresponding time waveforms. For comparative purposes, we also plot the performance of the traditional power spectral subtraction method as implemented by Berrouti et al. [25]. On the other hand, in order to validate the objective performance evaluations, the original and enhanced speech which were performed with the different subtractive-type algorithm were presented to human listeners to obtain a subjective evaluation of the speech quality.

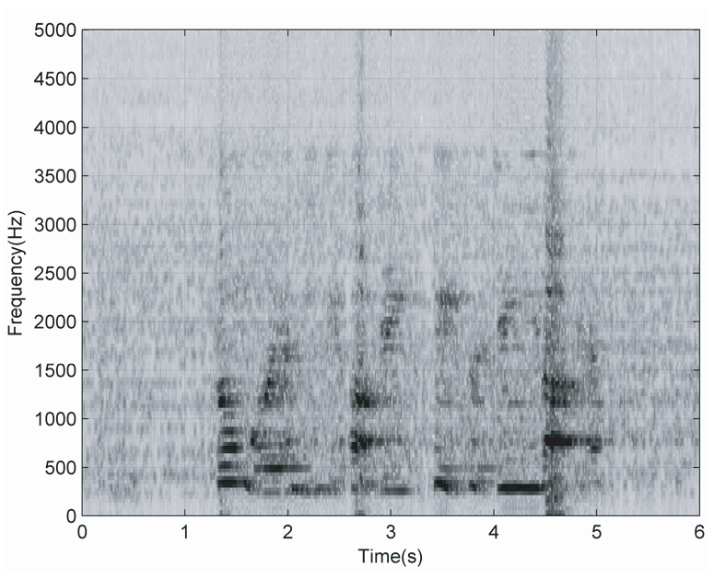
4. RESULTS AND DISCUSSIONS

As stated before, speech spectrograms constitute a well-suited tool for observing the residual noise and speech distortion. Figure 4 shows the spectrograms of the original radar speech (a), the enhanced speech using traditional spectral subtraction algorithm (b) and the proposed perceptual weighting algorithm in this study (c). The speech material is a Chinese sentence "Di si jun yi da xue" (the Fourth Military Medical University in English).

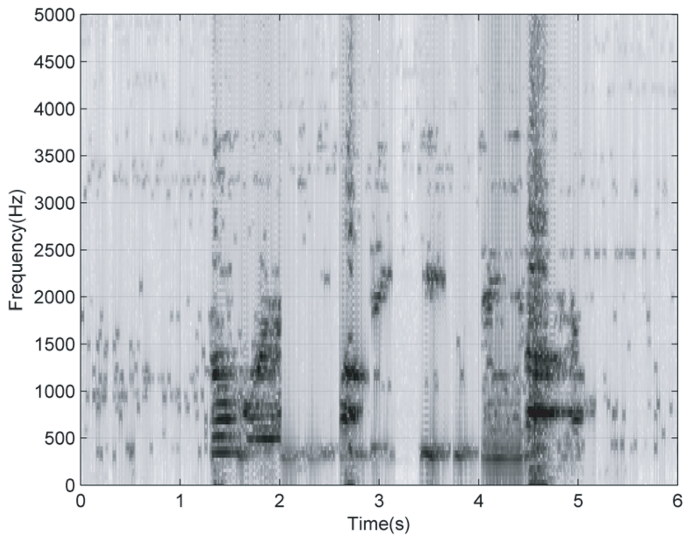
It can be seen from Figure 4(a) that a certain amount of the combined noises exist in the origin MMW radar speech because of the harmonic of the MMW, electrocircuit noise, as well as ambient noise combined in the MMW radar speech, and the combined noise is



(a)



(b)



(c)

Figure 4. The Spectrogram of the sentence "Di Si Jun Yi Da Xue". (a) The original MMW radar speech. (b) Enhanced speech obtained by the traditional spectral subtraction method. (c) The enhanced speech obtained by the proposed algorithm.

mainly concentrated on the low-frequency components roughly below 3 KHz. These noises can be obviously seen especially during speech pause. The spectrogram of traditional spectral subtraction algorithm is shown in Figure 4(b), the figure shows that this method is effective in reducing the combined radar noises, both in the speech and non-speech section. However, the algorithm decreases the noise uniformly for each frequency section, furthermore, it can be seen from the figure that much noise still remains in the enhanced speech, especially in the frequency section in which the noise is concentrated on, suggesting that the noise reduction is not satisfactory. Figure 4(c) shows that the proposed perceptual weighting algorithm not only reduces the low-frequency noise in which the combined radar noise is concentrated on, but also eliminates the high-frequency noise completely, it is clearly visible that the residual noise is reduced to a large extent and lost its structure. These results suggest that the proposed algorithm achieves a better reduction of the whole-frequency noise.

Informal listening tests also indicated that the enhanced speech with proposed perceptual weighting algorithm is more pleasant, the residual noise is better reduced, and with minimal, if any, speech distortion.

Moreover, the proposed perceptual weighting model is simple enough to avoid a great computational load. More important, the model does not rely on pre-enhancing MMW noisy speech. Considering its better effects on the MMW speech enhancement, the proposed algorithm is much more efficient than the traditional spectral-type algorithm [8] and the auditory masking algorithm [8].

Furthermore, the proposed perceptual weighting algorithm has strong flexibility to adapt complicated speech environment, this is because the two parameters of the spectral subtraction algorithm α and β can be self-adapted to fit other different or complex speech environment. This makes it possible to get better speech quality via speech enhancement under some rigorous speech environment. Additionally, when the two parameters in the proposed algorithm are constant at some appointed values, then the proposed perceptual weighting algorithm reduces to the traditional power spectral subtraction approach.

5. CONCLUSION

Besides air conducted speech, non-air conducted speech may extend the traditional speech detecting method, and provides some exciting possibility for wide applications. Based on the millimeter wave radar, a non-air conducted speech detecting method is introduced in this paper. However, the MMW conduct speech is in less intelligible and poor audibility since it is corrupted by additive combined and colored noise. This paper, therefore, proposed a simple and efficient way to take into account the frequency-domain masking properties of the human auditory system and reduces the perceptual effect of the residual noise. The perceptual weighting technique is developed and incorporated into the traditional spectral subtraction algorithm in the enhancement processes to shape the residual noise and make it inaudible. Therefore, the subtracted parameters of the spectral subtraction algorithm can be adaptive adjust in a perceptual sense. The results from both acoustic and listening evaluation suggest that the background noise can be reduced efficiently while the distortion of MMW radar speech remains acceptable, the enhanced speech also sounds more pleasant to human listeners, suggesting significant improvements over other speech enhancement methods.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (NSFC, No. 60571046), and the National postdoctoral Science Foundation of China (No. 20070411131). We also want to thank the participants from the E.N.T. Department, the Xi Jing Hospital, the Fourth Military Medical University, for helping with data acquisition and analysis.

REFERENCES

1. Li, S., R. C. Scherer, M. Wan, S. Wang, and H. Wu, "The effect of glottal angle on intraglottal pressure," *Journal of the Acoustical Society of America*, Vol. 119, No. 1, 539–548, 2006.
2. Li, S., R. C. Scherer, M. Wan, S. Wang, and H. Wu, "Numerical study of the effects of inferior and superior vocal fold surface angles on vocal fold pressure distributions," *Journal of the Acoustical Society of America*, Vol. 119, No. 5, 3003–3010, 2006.
3. Yanagisawa, T. and K. Furihata, "Pickup of speech signal utilization of vibration transducer under high ambient noise," *J. Acoust. Soc. Jpn.*, Vol. 31, No. 3, 213–220, 1975.
4. Li, Z.-W., "Millimeter wave radar for detecting the speech signal applications," *International journal of Infrared and Millimeter Waves*, Vol. 17, No. 12, 2175–2183, 1996.
5. Holzrichter, J. F., G. C. Burnett, and L. C. Ng, "Speech articulator measurements using low power EM-wave sensors," *J. Acoust. Soc. Am.*, Vol. 103, No. 1, 622–625, 1998.
6. Hu, R. and B. Raj, "A robust voice activity detector using an acoustic Doppler radar," *IEEE Workshop on Automatic Speech Recognition and Understanding*, 319–324, 2005.
7. Quatieri, T. F., K. Brady, D. Messing, and J. P. Campbell, "Exploiting nonacoustic sensors for speech encoding," *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 14, No. 2, 533–544, 2006.
8. Li, S., J. Wang, M. Niu, T. Liu, and X. Jing, "The enhancement of millimeter wave conduct speech using multi-band spectral subtraction method," *Computers & Electrical Engineering*, 2007 (submitted).
9. Boll, S. F., "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process*, Vol. 27, No. 2, 113–120, 1979.
10. Lockwood, P. and J. Boudy, "Experiments with a nonlinear

spectral subtractor (NSS), hidden Markov models and projection, for robust recognition in cars," *Speech Commun.*, Vol. 11, 215–228, 1992.

11. Hansen, J. H. L., "Morphological constrained feature enhancement with adaptive cepstral compensation (MCE-ACC) for speech recognition in noise and Lombard effect," *IEEE Trans. Speech Audio Process*, Vol. 2, 598–614, 1994.
12. Liu, H., Q. Zhao, M. Wan, and S. Wang, "Application of spectral subtraction method on enhancement of electrolarynx speech," *J. Acoust. Soc. Am.*, Vol. 120, No. 1, 398–406, 2006.
13. Kamath, S. and P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 4160–4164, 2002.
14. Udrea, R. M., S. Ciochina, and D. N. Vizireanu, "Milti-band bark scale spectral over-subtraction for colored noise reduction," *International Symposium on Signals, Circuits and Systems*, 311–314, 2005.
15. Li, S., J. Wang, M. Niu, T. Liu, and J. Xijing, "Millimeter wave conduct speech enhancement based on auditory masking properties," *Microwave and Optical Technology Letters*, Vol. 50, No. 8, 2109–2114, 2008.
16. Tsoukalas, D., M. Paraskevas, and J. Mourjopoulos, "Speech enhancement using psycho-acoustic criteria," *Proc. IEEE ICASSP*, 359–361, Minneapolis, MN, Apr. 1993.
17. Usagawa, T., M. Iwata, and M. Ebata, "Speech parameter extraction in noisy environment using a masking model," *Proc. IEEE ICASSP*, 81–84, Adelaide, Australia, Apr. 1994.
18. Nandkumar, S. and J. H. L. Hansen, "Dual-channel iterative speech enhancement with constraints on an auditory-based spectrum," *IEEE Trans. Speech Audio Processing*, Vol. 3, 22–34, Jan. 1995.
19. Johnston, J. D., "Transform coding of audio signal using perceptual noise criteria," *IEEE J. Select. Areas Commun.*, Vol. 6, No. 2, 314–323, 1988.
20. Atal, B. S. and M. R. Schroeder, "Predictive coding of speech and subjective error criteria," *IEEE Trans. Acoust., Speech, Signal Process*, Vol. 27, No. 3, 247–254, 1979.
21. Kroon, P. and B. S. Atal, "Predictive coding of speech using analysis-by-synthesis techniques," *Advances in Speech Signal Processing*, 141–164, Marcel Dekker, 1992.

22. Schroeder, M. R., B. S. Atal, and J. L. Hall, "Optimizing digital speech coders by exploiting masking properties of the human ear," *J. Acoust. Soc. Am.*, Vol. 66, No. 6, 1647–1651, Dec. 1979.
23. Wang, J. Q., C. X. Zheng, X. J. Jin, and G. H. Lu, "Study on a non-contact life parameter detection system using millimeter wave," *Space Medicine & Medical Engineering*, Vol. 17, No. 3, 157–161, 2004.
24. Wang, J., C. Zheng, G. Lu, and X. Jing, "A new method for identifying the life parameters via radar," *EURASIP Journal on Advances in Signal Processing*, Vol. 2007, No. 1, 8–16, 2007.
25. Berouti, M., R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process*, 208–211, 1979.
26. Hu, Y. and P. C. Loizou, "A perceptually motivated approach for speech enhancement," *IEEE Trans. Speech Audio Process*, Vol. 11, No. 5, 457–464, 2003.
27. Virag, N., "Single channel speech enhancement based on masking properties of the human auditory system," *IEEE Trans. on Speech and Audio Processing*, Vol. 7, No. 2, 126–137, 1999.
28. Liu, H., Q. Zhao, M. Wan, and S. Wang, "Enhancement of electrolarynx speech based on auditory masking," *IEEE Transactions on Biomedical Engineering*, Vol. 53, No. 5, 865–874, 2006.
29. Cohen, I. and B. Berdugo, "Noise estimation by minima controlled recursive averaging for robust speech enhancement," *IEEE Signal Processing Letters*, Vol. 9, No. 1, 12–15, 2002.
30. Cohen, I. and B. Berdugo, "Speech enhancement for non-stationarynoise environments," *Signal Processing*, Vol. 81, 2403–2418, 2001.