J. Biomedical Science and Engineering, 2010, 3, 187-192 **JBiSE**
doi:10.4236/jbise.2010.32024 Published Online February 2010 (http://www.SciRP.org/journal/jbise/).

Iterative spectral subtraction method for millimeter-wave conducted speech enhancement

Sheng Li[1,2], Jian-Qi Wang[1*], Ming Niu[1], Xi-Jing Jing[1], Tian Liu[1]

[1]Department of Biomedical Engineering, the Fourth Military Medical University, Xi'an, China;
[2]The Key Laboratory of Biomedical Information Engineering of Ministry of Education, and Department of Biomedical Engineering, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, China.
Email: [*]sheng@mail.xjtu.edu.cn

Received 4 November 2008; revised 4 December 2009; accepted 7 December 2009.

ABSTRACT

A non-air conducted speech detecting method has been developed in our laboratory by using millimeter wave radar technology. Because of the special attributes of the millimeter wave, this method may considerably extend the capabilities of traditional speech detecting methods. However, radar speech is substantially degraded by additive combined noises that include radar harmonic noise, electrocircuit noise, and ambient noise. This study, therefore, proposed an iterative spectral subtraction method which can be adaptively estimate noise spectrum at every iteration, and reduce the musical noise remained in the previous spectral subtraction process. Results from simulations as well as evaluations confirm that the proposed method satisfactorily reduces whole-frequency and musical noises and produces good speech quality.

Keywords: Millimeter Wave; Speech Detecting; Speech Enhancement

1. INTRODUCTION

The speech, which is produced by speech organ of human beings [1,2], is well known that it can be spread and perception by means of air, and can be detected and recorded by acoustic sensors. 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 using the bone-conduction sensors at special location [3]. Other medium, such as infrared ray, ultrasound wave, laser light also can be used to detecting the non-air spread speech or noise for some special applications [4].

Li Zong Wen (1996) [4] reported another medium, millimeter wave (MMW) radar, as well as light radar and laser radar, can detect and identify out exactly the existential speech signals in free space from a person speaking through the electromagnetic wave fields by principle and experiment. Since the microwave radar has low range attenuation, 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 the clinic to assist diagnosis or to measure speech articulator motions [6].

However, there has been little previous research work concentrated on the research of speech which is produced by MMW radar, some studies with respect to the MMW radar speech concentrated on the MMW non acoustic sensors [5,7], in order to measure speech articulator motions, such as vocal tract measurements and glottal excitation [5].

Although MMW radar provides another method to detect speech, 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 EMW 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.

The spectral subtraction method is the most widely used, and has been shown to be an effective approach for noise canceling, in order to improve the intelligibility and quality of digitally compressed speech. Due to the

simplicity of implementation, and low computational load, the spectral subtraction method is the primary choice for real time applications [8]. 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 [9,10,11,12,13], results performed with these algorithms show that there is a need for further improvement. Furthermore, in order to prevent destructive subtraction of the speech while removing most of the residual noise, it is necessary to propose a new approach to improve the subtraction procedure.

Therefore, the purpose of this investigation is motivated by the need of improving EMW radar speech, especially in the electronic environments. An iterative spectral subtraction algorithm is proposed to adaptively estimate noise spectrum at every iteration, and reduce the musical noise remained in the previous spectral subtraction process. The results suggest that for the proper iteration number, the proposed iteration number can significantly remove the musical noise, and improve the speech quality.

2. METHODS

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 6 dB directional coupler, a variable attenuator, a circulator, and then to a flat antenna. The 6 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 of the antenna is usually kept at a level of about 10–20 mW. The flat antenna radiates a microwave beam of about 9º 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 34G Hz 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 [14] and [15].

2.2. Iterative Spectral Subtraction Method

The iterative spectral subtraction algorithm is based on the assumption that the additive noise will 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.* (1979) [16] proposed the generalized spectral subtraction scheme is described as follows:

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

where $\alpha(\alpha > 1)$ is the over-subtraction factor [16], which is a function of the segmental SNR. $\beta(0 \le \beta \le 1)$ is the spectral floor, and γ is the exponent determining the transition sharpness. Here we set $\gamma = 2$ and $\beta =$ 0.002.

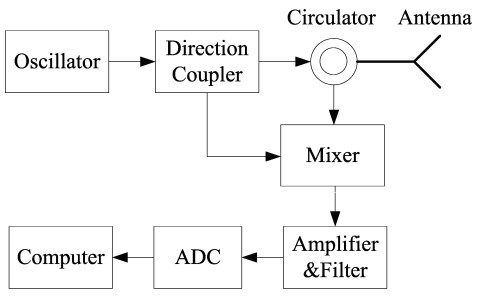

Figure 1. Schematic diagram of the speech-detection system.

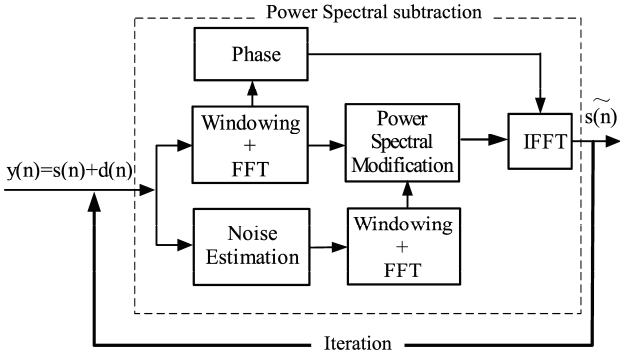

Figure 2. The proposed speech enhancement scheme.

In order to decrease the musical noise, which is produced by the speech enhancement procedure, an iterative spectral subtraction algorithm is proposed in this study. The iterative method is motivated from Wiener filtering which is one of the speech enhancement techniques [17,18]. In this study, the output of the enhanced speech using traditional spectral subtraction method is used as the input signal of the next iteration process.

Figure 2 shows the scheme of the proposed MMW speech enhancement algorithm. It can be seen from the figure that after the first spectral subtraction process, the type of the additive noise is changed to the musical noise. As the output signal is used as the input signal of the next iteration process, the musical noise is reestimated, this new estimated noise, furthermore, is been used to process the next spectral subtraction (that is, subtracted by the new noisy speech), therefore, an enhanced output speech signal can be obtained, and the iteration process goes on. If we regard the process of noise estimate and the spectral subtraction as a filter, then the output signal of the filter is used not only for designing the filter but also as the input signal of the next iteration process. More important, this filter can be refreshed adaptively by reestimate the musical noise so that to improve the speech quality effectively.

2.3. Noise Estimation

The noise in the radar speech, which included of each order of the EMW 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) [19] 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

$$\left|\hat{D}_{(k,l)}(\omega)\right|^2 = \lambda_D \left|\hat{D}_{(k-1,l)}(\omega)\right|^2 + (1-\lambda_D)\left|\hat{Y}_{(k,l)}(\omega)\right|^2$$
$$0 < \lambda_D < 1, \quad (4)$$

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) [19], then Eq. (4) implies

$$\left|\hat{D}_{(k,l)}(\omega)\right|^2 = \hat{\lambda}_D(k,l)\left|\hat{D}_{(k-1,l)}(\omega)\right|^2 + (1-\hat{\lambda}_D(k,l))\left|\hat{Y}_{(k,l)}(\omega)\right|^2 \quad (5)$$

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 [19,20].

3. EXPERIMENTS

Ten healthy volunteer speakers, 6 males and 4 females, participated in the radar speech experiment. All the subjects were native speakers of Mandarin Chinese. Their ages varied from 20 to 35, with a mean age of 28.1 (SD =12.05). All the experiments were conducted in accordance with the terms of the Declaration of Helsinki (BMJ 1991; 302, 1194), and appropriate consent forms were signed by the volunteers. Ten sentences of Mandarin Chinese were used as the speech material for acoustic analysis and acceptability evaluation. The lengths of the sentences varied from 6 words (5.6 s) to 30 words (15 s). The sentences were spoken by each participant in a quiet experimental environment. The speakers were instructed to read the speech material at normal loudness and speaking rates.

For the perceptual experiment, eight listeners were selected to evaluate the acceptability of each sentence based on the criteria of mean opinion score (MOS), which is a five-point scale (1, bad; 2, poor; 3, common; 4, good; 5, excellent). All the listeners were native speakers of Mandarin Chinese, had no reported history of hearing problems, and were unfamiliar with MMW radar speech. Their ages varied from 22 to 36, with a mean age of 26.37 (SD = 4.63).

In order to test the effectiveness of the proposed method, two different types of background noise, namely, white Gaussian noise and speech babble noise, were added to the enhanced MMW radar speech; both noises were taken from the Noisex-92 database. These two representative noises have a greater similarity to actual talking conditions than the other noises. Noises with SNRs of 0 dB were added to the original MMW radar speech signal.

4. RESULTS AND DISCUSSIONS

In order to evaluate and compare the performance of the proposed enhancement algorithm, two other algorithms are performed in this study, they are: traditional spectral subtraction method and noise-estimation algorithm [21]. For the purpose of analyzing the time-frequency distribution of the original/enhanced speech, speech spectrograms were provided since they have been identified as a well-suited tool for observing both the residual noise and speech distortion. In addition, results are also measured subjectively by Mean Opinion Score (MOS) in conditions of additive white Gaussian noise as well as Bobble noise (for MOS) for the algorithm evaluation.

Figure 3 shows the spectrograms of the original radar speech (a), the enhanced speech using traditional spectral subtraction algorithm (b), the enhanced speech using noise-estimation algorithm (c), and the proposed iterative spectral subtraction algorithm (d) (the iteration number is set to 5). The speech material is a Chinese sentence "Di si jun yi da xue" (in English, the Fourth Military Medical University).

Because of its different speech detecting theory and working conditions, non-air conducted speech has some special attributes. As stated earlier, the most important is that combined noises are introduced into the original MMW radar speech. These noises can be clearly seen in **Figure 3(a)**, especially in the speech-pause region. It can also be seen from the figure that the combined noise is mainly concentrated in the low-frequency components, roughly below 3 KHz. **Figure 3(b,c)** shows that the spectral subtraction algorithm and the noise estimate algorithm are effective in reducing the combined radar noises. However, there are still too much remnant noise in the enhanced speech, especially in the frequency section in which the noise are concentrated, suggesting that the noise reduction are not satisfactory. **Figure 3(d)** shows that the proposed algorithm not only greatly reduces the low-frequency noise, in which the combined radar noise is concentrated, but it also completely eliminates the high-frequency noise. It can be seen from the figure that in the speech-pause regions the residual noise

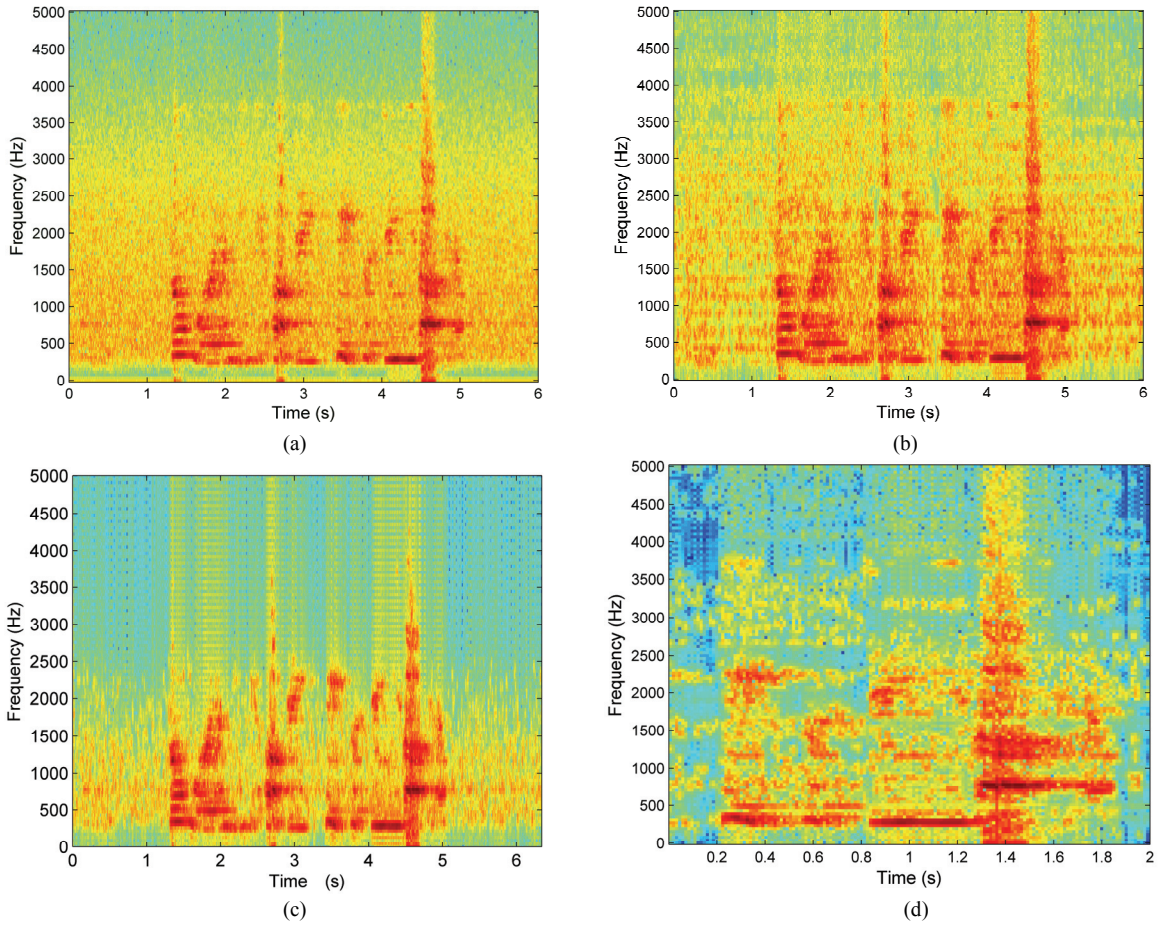

Figure 3. The Spectrogram of the sentence "Di Si Jun Yi Da Xue". (a) The original spectrogram of the millimeter wave conducted speech. (b) Enhanced radar speech obtained by the traditional spectral subtraction method. (c) Enhanced radar speech obtained by the noise estimated algorithm. (d) Enhanced radar speech obtained by the proposed algorithm.

is almost eliminated. Moreover, it is clear that the residual noise is greatly reduced and has lost its structure. These results suggest that the proposed algorithm achieves a better reduction of the whole-frequency noise than traditional spectral subtraction methods.

The perceptual score of the noisy speech and the enhanced noisy speech are shown in **Figure 4**. Mean Opinion Scores (MOS) were used for 100 sentences produced by ten volunteer speakers. The noisy speech in the cases of the additive white and babble noises had SNR inputs of 0 dB. It can be seen from the figure that the original noisy speech has "bad" perceptual effects, but the score of the enhanced speech obtained by using the proposed algorithm is much better. Comparing the two noises, it can be seen from the figure that the MOS for white noise is a little higher than for babble noise. This suggests that the proposed algorithm is more "sensitive" to white noise, however, the difference is small.

The iteration times is another important factor which has effects on the performance of speech enhancement. In order to explore the relationship between the performance of speech enhancement and the iteration times, the variation of the mean segmental SNR of the radar speech with iteration times are shown in **Figure 5**. It can be seen from **Figure 5** that the SNR increased as the iteration number increased, which suggest the larger iteration number will corresponding to the better speech enhancement performance and the less musical noise. However, both performed waveform and the corresponding spectrogram suggest that the larger iteration number would eliminate part of the normal speech component to some extent while it works overmuch effectively on reducing the musical noise, therefore, the proposed iteration number for the radar speech is 3 to 5.

As a single channel subtractive-type speech enhancement method, the algorithm proposed in this paper can be applied for the enhancement of non-air conducted speech using available electronics. For example, a millimeter wave conducted speech enhancing system, into

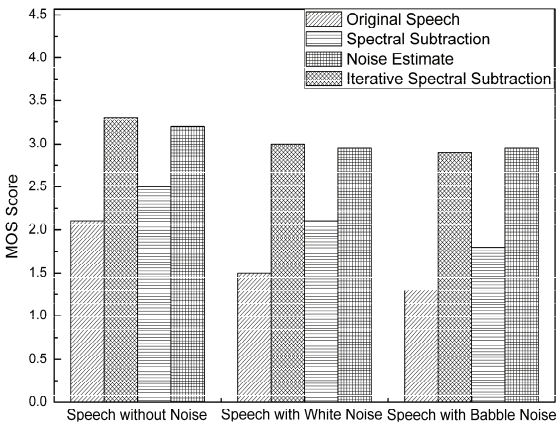

Figure 4. Perceptual results of the noisy and enhanced noisy speech based on the MOS criteria.

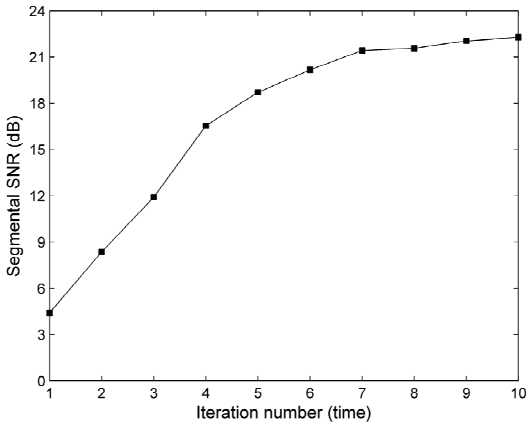

Figure 5. Variations of mean segmental SNR of the MMW radar speech with iteration times.

which this algorithm is embedded, can be developed. With the help of digital signal processing (DSP) technology, the speech enhancement function can be realized with a microprocessor and implanted into a radar-telephone, radar-microphone, or other electronic equipment. Different enhancement algorithms, suitable for different noise conditions, can be selected by a switch. With the development of efficient enhancement methods, the quality of non-air conducted speech will be vastly improved and will provide better perception.

5. CONCLUSIONS

As a non-air conducted speech, MMW radar speech may have greater advantage and wider applications than air conduct speech. However, the complex noises added in the radar speech decreased the speech quality to a large extent. Therefore, an improved spectral subtraction method, iterative spectral subtraction algorithm are used in this study in order to decrease the complex noise and the musical noise. The results from both simulation and evaluation suggest that for the proper iteration number, this method achieves a better reduction of the whole-frequency noise, the musical noise, and yields good speech quality.

6. ACKNOWLEDGEMENTS

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., Scherer, R.C., Wan, M., Wang, S. and Wu, H. (2006) The effect of glottal angle on intraglottal pressure. *Journal of the Acoustical Society of America*, **119(1)**, 539–548.

[2] Li, S., Scherer, R.C., Wan, M., Wang, S. and Wu, H. (2006) Numerical study of the effects of inferior and su-

perior vocal fold surface angles on vocal fold pressure distributions. *Journal of the Acoustical Society of America*, **119(5)**, 3003–3010.

[3] Yanagisawa, T. and Furihata, K. (1975) Pickup of speech signal utilization of vibration transducer under high ambient noise. *J. Acoust. Soc. Jpn*, **31(3)**, 213–220.

[4] Li, Z.-W., (1996) Millimeter wave radar for detecting the speech signal applications. *International journal of Infrared and Millimeter Waves*, **17(12)**, 2175–2183.

[5] Holzrichter, J.F., Burnett, G.C. and Ng, L.C. (1998) Speech articulator measurements using low power EM-wave sensors. *J. Acoust. Soc. Am*, **103(1)**, 622–625.

[6] Hu, R. and Raj, B. (2005) "A robust voice activity detector using an acoustic Doppler radar," in *IEEE Workshop on Automatic Speech Recognition and Understanding*, 319–324.

[7] Quatieri, T.F., Brady, K., Messing, D. and Campbell, J.P., (2006) Exploiting nonacoustic sensors for speech encoding. *IEEE Transactions on Audio, Speech and Language Processing*, **14(2)**, 533–544.

[8] Boll, S.F. (1979) Suppression of acoustic noise in speech using spectral subtraction, *IEEE Transactions on Acoustics, Speech and Signal Processing*, **27**, 113–120.

[9] Lockwood, P. and Boudy, J. (1992) Experiments with a nonlinear spectral subtractor (NSS), hidden Markov models and projection, for robust recognition in cars. *Speech Commun*, **11**, 215–228.

[10] Hansen, J.H.L., (1994) Morphological constrained feature enhancement with adaptive cepstral compensation (MCE-ACC) for speech recognition in noise and Lombard effect. *IEEE Trans. Speech Audio Process*, **2**, 598–614.

[11] Liu, H., Zhao, Q., Wan, M. and Wang, S. (2006) Application of spectral subtraction method on enhancement of electrolarynx speech. *J. Acoust. Soc. Am*, **120(1)**, 398–406.

[12] Kamath, S. and Loizou, P. (2002) A multi-band spectral subtraction method for enhancing speech corrupted by colored noise. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, **4**, 4160–4164.

[13] Udrea, R.M., Ciochina, S. and Vizireanu, D.N. (2005) Multi-band bark scale spectral over-subtraction for colored noise reduction. *International Symposium on Signals, Circuits and Systems*, **1**, 311–314.

[14] Wang, J.Q., Zheng, C.X., Jin, X.J. and Lu, G.H, (2004) Study on a non-contact life parameter detection system using millimeter wave. *Space Medicine & Medical Engineering*, **17(3)**, 157–161.

[15] Wang, J., Zheng, C., Lu, G.. and Jing, X. (2007) A new method for identifying the life parameters via radar. *EURASIP Journal on Advances in Signal Processing*, **2007(1)**, 8–16.

[16] Berouti, M., Schwartz, R. and Makhoul, J. (1979) Enhancement of speech corrupted by acoustic noise, Proc. *IEEE Int. Conf. Acoust., Speech, Signal Process*, 208–211.

[17] Lim, J.S. and Oppenheim, A.V. (1978) All-Pole Modeling of Degraded Speech. *IEEE Trans. Acoust. Speech, Signal Processing*, ASSP, **26**, 197–210.

[18] Ogata, S. and Shimamura, T. (2001) Reinforced spectral subtraction method to enhance speech signal, Electrical and Electronic Technology. *TENCON. Proceedings of IEEE Region 10 International Conference*, **1**, 242–245.

[19] Cohen, I. and Berdugo, B. (2002) Noise estimation by minima controlled recursive averaging for robust speech enhancement. *IEEE SIGNAL PROCESSING LETTERS*, **9(1)**, 12–15.

[20] Cohen, I. and Berdugo, B. (2001) Speech enhancement for non-stationarynoise environments. *Signal Processing*, **81**, 2403–2418.

[21] Rangachari, S. and Loizou, P. C. (2006) A noise-estimation algorithm for highly non-stationary environments. *Speech Communication*, **48**, 220–231.

