TELKOMNIKA, Vol.11, No.1, January 2013, pp. 130~135
ISSN: 2302-4046

Enhancement of Non-Air Conducted Speech Based on Wavelet-Packet Adaptive Threshold

Sheng Li*, Huijun Xue, Guohua Lu, Yang Zhang, Teng Jiao, Jianqi Wang, Xijing Jing
(School of Biomedical Engineering, the Fourth Military Medical University, Xi'an 710032, China)
*corresponding author, e-mail: sheng@mail.xjtu.edu.cn
The first two authors contributed equally to this work and should be regarded as co-first authors.

Abstract

This study developed a new kind of speech detecting method by using millimeter wave. Because of the advantage of the millimeter wave, this speech detecting method has great potential application and may provide some exciting possibility for wide applications. However, the MMW conduct speech is in less intelligible and poor audibility since it is corrupted by additive combined noise. This paper, therefore, also developed an algorithm of wavelet packet threshold by using hard threshold and soft threshold for removing noise based on the good capability of wavelet packet for analyzing time-frequency signal. Comparing to traditional speech enhancement algorithm, the results from both simulation and listening evaluation suggest that the proposed algorithm takes on a better performance on noise removing while the distortion of MMW radar speech remains acceptable, the enhanced speech also sounds more pleasant to human listeners, resulting in improved results over classical speech enhancement algorithms.

Keywords: *non-contact speech, wavelet packet, threshold, spectrogram.*

1. Introduction

It is well known that speech, which is produced by the larynx of human beings [1, 2], has significant effects for human beings for communication. Obtaining the accurate and reliable speech signal is necessary for human exchanging information, especially in various noising environments. The most common way for human beings to get the speech is hear by ears or detect by acoustic sensors, which is based on the basic principle, that is, speech can spread and be detected 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 provided using this method [3, 4]. Other medium, such as infrared ray, light wave, and laser also can be used to detect the non-air spread speech or acoustical vibrations, however, their application are limited since the materials in detail are usually difficult to obtain [5].

Our laboratory developed a new kind of method for detecting speech signal by using millimeter wave. Millimeter wave (MMW, as well as light and laser), was reported by previous study that they can 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 [5, 6]. Since the microwave radar has low range attenuation, better sense of direction, and has attribute of noninvasive, safe, fast, portable, low cost fashion [7, 8], 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. 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, 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

Received September 29, 2012; Revised November 22, 2012; Accepted November 30, 2012

speech enhancement method) to extent the traditional speech detecting method, so that to overcome the defects of traditional speech detecting method: (1) having good direction, and could check speech signal in the noisy background, such as busy market or in the tank; (2) permitting farther detection distance; (3) it has high sensitivity because of the high frequency we used (34GHz).

With the new method for testing 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 MMW and electro circuit 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.

During the last decades, a lot of speech enhancement methods have been proposed, such as spectral subtraction [9-12], hidden Markov modeling [13], signal subspace methods [14], wavelet-based methods [15] et al. But there were few literature reports about how to enhance the radar speech signal. Our research group has used nonlinear multiband spectral subtraction to reduce the colored electronic noise, which conducted by millimeter wave radar [16]. We also developed a novel non-air conducted speech detecting method based on millimeter wave radar technology and proposed an iterative spectral subtraction method to estimate noise and reduce the musical noise, which exist in the previous process [17]. Although the methods have promoted the quality of radar speech efficiently, their results suggest that there is a need to further improve the quality of radar speech.

Wavelet can be seen as an extension of function space level orthogonal triangulation. From matriculation analysis, we know that wavelet transform just decompose the part of low frequency rather than high frequency, so the resolution of the high frequency is low. It not only can process orthogonal decomposition in the low frequency component but also in the high frequency component. The wavelet packet transforms, which can be easily obtained by filtering a signal with multi-resolution filter banks [18, 19], has been applied to various research areas, including signal and image denoising, compression, detection, and pattern recognition [20].

In this paper, we propose a method by using wavelet packet to enhance the radar speech signal. Wavelet packet has some advantages: (1) proposing a more perfect analysis method for detecting signal, and separating frequency band to multi levels, (2) focusing on any signal detail between time domain and frequency domain, (3) selecting relevant band automatically according to feather of speech signal, that match the frequency of the original signal, therefore, strengthen the time and frequency's resolution. Moreover, we used the adaptive threshold method to enhance radar speech. The results show that the proposed method can remove noise effectively, and achieve the aim of enhancing radar speech.

2. Methods

2.1. Speech detecting equipment and experiment

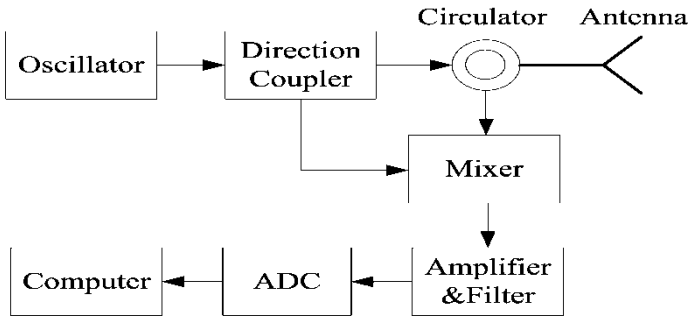

Figure 1. Schematic diagram of the non-air conducted speech detection system.

Millimeter wave radar (MMW) was used to detect the speech signal. Figure 1 shows the schematic diagram of the speech-detection system. 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.

Ten adults (age between 20 and 35) were selected as speakers (5 men and 5 women), ten sentences were spoken by each speaker, All of the subjects were native speakers of mandarin Chinese, the distance between the speaker and the radar is ten meters. All of the testers have signed the consent form according to the Declaration of Helsinki (BMJ 1991; 302: 1194).

$$\begin{cases} \phi(t) = \sum_{k \in Z} h_k \phi(2t-k) \\ \psi(t) = \sum_{k \in Z} g_k \psi(2t-k) \end{cases} \quad (1)$$

Where h_k and g_k are coefficients of the filter.

Wavelet packet decompose:

$$\begin{cases} d_l^{j,2n} = \sum_k a_{k-2l} d_k^{j+1,n} \\ d_l^{j,2n+1} = \sum_k b_{k-2l} d_k^{j+1,n} \end{cases} \quad (2)$$

Wavelet packet reconstitution:

$$d_l^{j+1,n} = \sum_k \left(h_{l-2k} d_k^{j,2n} + g_{l-2k} d_k^{j,2n+1} \right) \quad (3)$$

Block diagram of our proposed method for enhancing radar speech is illustrated in Figure 2.

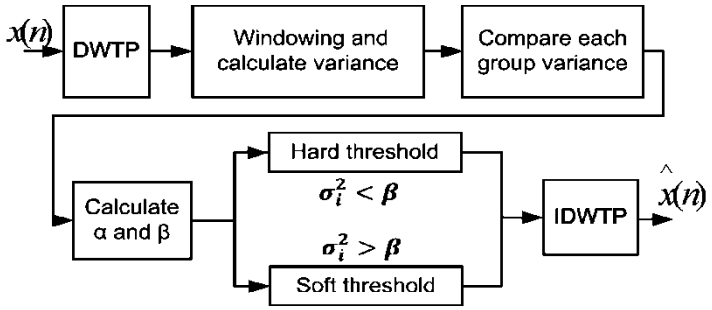

Figure 2. Block diagram of algorithm for enhancing radar speech

1) After decomposing the radar signal base on wavelet packet theory, we get twenty-five frequency bands and the coefficient of wavelet packet at each level. Generally, the signal that corrupted by noise was defined as:

$$x_m(n) = s_m(n) + d_m(n) \;\; n = 0,1,2,....,L-1 \tag{4}$$

Where m is the sequence of the frame, $x_m(n)$, $s_m(n)$ and $d_m(n)$ express corresponding speech signal with noise, pure speech and noise signal. L is the number of the samples in the scale.

2) This paper uses hard threshold and soft threshold to remove noise respectively. If the *N*th frame's variance σ of wavelet packet coefficient is very small, we consider this frame has few resonant peak component and it just contain noise component. However, according the resonant peak's intensity of speech signal, the value of the weighting coefficient α (α is between 0 and 1) could be decided.

3) Defining threshold is the ratio of current frame variance and maximum variance:

$$\beta_i(n) = \sigma^2_{median}(i) * \left(1 - \frac{\sigma_i^2(n)}{\sigma^2_{\max}(i)}\right) \quad i = 1,2,...,25; \;\; n = 1,2,...,N \tag{5}$$

When $\sigma_i^2 < \beta_i(n)$, pointing that this frame has few signal components and much noise component, it should be compressed by hard threshold. Otherwise, we should use soft threshold to avoid the loss of resonant peak component.

Hard threshold denoising method was used in the region, which has few resonant peak component, the speech signal is weak in this region:

$$w_{j,k} = w_{j,k} * \alpha * \min\left(\sqrt{\frac{\sigma^2_{\min}}{\sigma^2_{\max}}}, c\right) \tag{6}$$

where $\sigma^2_{\max}(i)$ and $\sigma^2_{\min}(i)$ are maximum and minimum of the *i*th level, the value c is 0.05.

Soft threshold was also used in the region, which have much resonant peak component, the speech signal is also strong in this region.

$$w_{j,k} = w_{j,k} * \alpha * \left(1 - \frac{\beta}{\sigma_i^2}\right) \tag{7}$$

4) In the end, the algorithm of wavelet packet reconstitution was performed to get the enhanced speech.

3. Results and Discussion

Generally, speech enhancement algorithm produces 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 perform the speech spectrogram of the other traditional speech enhancement algorithms, they are traditional spectral subtraction [21] and wiener filter [22]. On the other hand, in order to validate the objective performance evaluations,

the original and enhanced speech which were performed with different speech enhancement algorithms were presented to human listeners to obtain a subjective evaluation of the speech quality.

The spectrogram can express time-frequency character of a signal, thus, in this paper, spectrogram was used to evaluate the effect of speech enhancement. Figure 3 (a) is the spectrogram of original radar speech. Figure 3 (b) and (c) are the spectrograms of the signal processed by spectral subtraction [21] and wiener filter [22]. At last, figure 3 (d) is the spectrogram of the speech signal that was processed by wavelet packet transform.

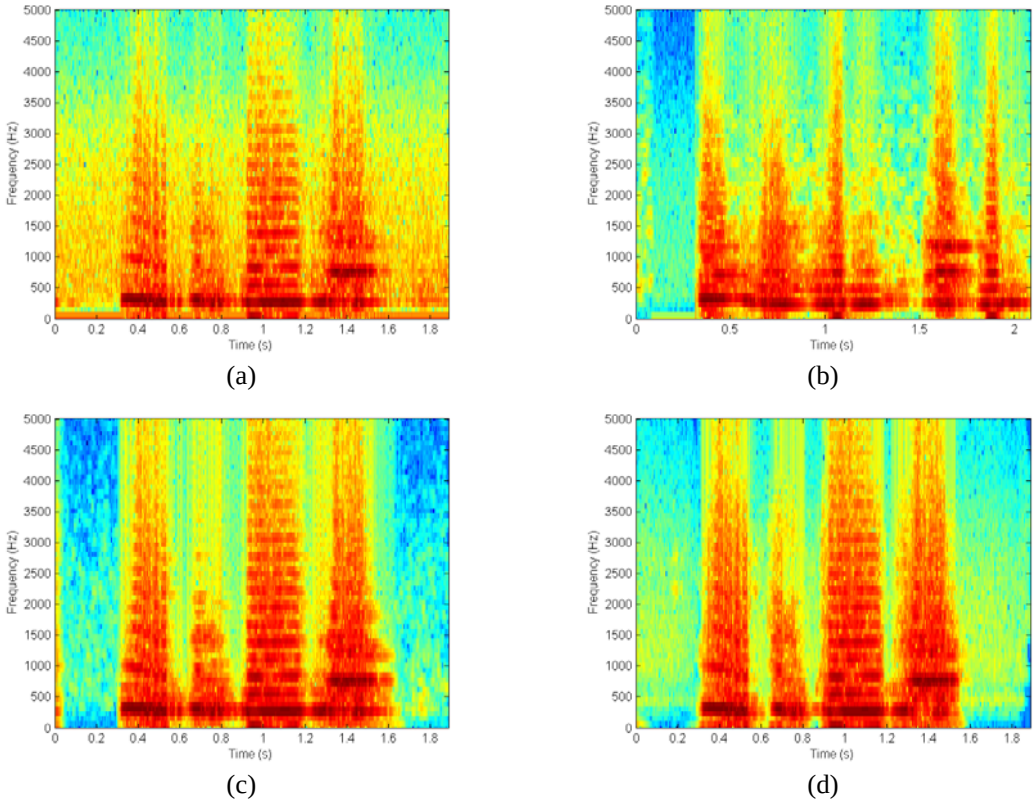

Figure 3. (a). The original speech corrupted by noise. (b). Enhanced speech obtained by spectral subtraction. (c). Enhanced speech obtained by wiener filter. (d). Enhanced speech obtained by wavelet packet transform.

Figure 3 (a) is the spectrogram of the original radar speech. It can be seen from the figure that the original radar speech signal was permeated with noise, which are electromagnetic noise and circuit noise produced by radar system. Two other speech enhancement algorithms were also performed for comparing purpose, they are: spectral subtraction and Wiener filter algorithm. Figure 3 (b) shows the results of radar speech enhancement by using spectral subtraction. It can be seen from the figure that the component of noise was reduced efficiently, but there was a lot of noise energy exists in speech signal, which shows that the noise was not removed completely. Furthermore, the component of high frequency energy less than that in original speech signal, which shows that the spectral subtraction method has damaged the high frequency component of radar speech. Also, due to spectral subtraction system defects, we can hear music noise after using this method. Figure 3 (c) shows the results of radar speech enhancement by using wiener filter. Compare to original speech signal, this method can remove component of noise efficiently, however, it created some new noise in the region of high frequency. Figure 3 (d) shows the results of radar speech enhancement by using wavelet packet denoising algorithm. Comparing to two methods stated

before, noise component was almost removed, and the speech signal was reserved. It can be seen from the figure that there was no new noise component produced, especially in non-speech section.

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

4. Conclusion

In this paper, a new method for detecting speech signal by radar sensor is developed. Comparing to traditional speech detecting method, this method may provide some exciting possibility of wide applications. However, it also introduces more annoying noise in speech than traditional speech detecting method. This study, therefore, also developed a new algorithm based on the wavelet packet threshold to remove radar noise. The results from both simulation and evaluation suggest that this algorithm is able to reduce the background noise efficiently and the residual noise is less structured while the distortion of MMW radar speech remains acceptable.

References

[1] Li S, Scherer RC, Minxi W, Wang s, Wu h. Numerical study of the effects of inferior and superior vocal fold surface angles on vocal fold pressure distributions. *J Acoust Soc Am*. 2006; 119(5): 3003-3010.
[2] Li S, Scherer RC, Wan M, Wang S, Wu H. The effect of glottal angle on intraglottal pressure. *J Acoust Soc Am.* 2006; 119(1): 539-548.
[3] Yanagisawa T, Furihata K. Pickup of speech signal utilization of vibration transducer under high ambient noise. *J Acoust Soc Jpn.* 1975; 31(3): 213-220.
[4] Fletcher H. Auditory patterns. *Rev Modern Phys.* 1940; 12: 47-65.
[5] Li Z-W. Millimeter wave radar for detecting the speech signal applications. *International journal of Infrared and Millimeter Waves.* 1996;17(12):2175-2183.
[6] Wang J, Zheng C, Lu G, Jing X. A New Method for Identifying the Life Parameters via Radar. *EURASIP Journal on Advances in Signal Processing.* 2007;2007(1):8-16.
[7] Holzrichter JF, Burnett GC, Ng LC. Speech articulator measurements using low power EM-wave sensors. *J Acoust Soc Am.* 1998; 103(1): 622-625.
[8] Wang Jq, Zheng Cx, Jin Xj, Lu Gh. Study on a Non-contact Life Parameter Detection System Using Millimeter Wave. *Space Medicine & Medical Engineering.* 2004; 17(3): 157-161.
[9] Boll SF. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans Acoust Speech Signal Process.* 1979; 27: 113-120.
[10] Liu H, Zhao Q, Wan M, Wang S. Enhancement of electrolarynx speech based on auditory masking. *IEEE Transactions on Biomedical Engineering.* 2006; 53(5): 865-874.
[11] Virag N. Single channel speech enhancement based on masking properties of the human auditory system. *IEEE Transon Speech and Audio Processing.* 1999; 7(2): 126-137.
[12] Liu H, Zhao Q, Wan M, Wang S. Application of spectral subtraction method on enhancement of electrolarynx speech. *J Acoust Soc Am.* 2006; 120(1): 398-406.
[13] Sameti H, Sheikhzadeh H, Deng L, Brennan RL. HMM-based strategies for enhancement of speech signals embedded in nonstationary noise. *IEEE Trans Speech Audio Process.* 1998; 6(5): 445-455.
[14] Klein M, Kabal P. Signal subspace speech enhancement with perceptual post-filtering. *Proc IEEE Internat Conf Acoust Speech Signal Process (ICASSP).* 2002; 1: 537-540.
[15] Donoho DL. De-noising by soft-thresholding. *IEEE Trans Inf.* 1995; 41(3): 613-627.
[16] Li S, Wang JQ, Jing XJ. The Application of Nonlinear Spectral Subtraction Method on Millimeter Wave Conducted Speech Enhancement. *Mathematical Problems in Engineering.* 2010; 2010: 1-12.
[17] Li S, Wang JQ, Niu M, Liu T, Jing XJ. Millimeter wave conduct speech enhancement based on auditory masking properties. *Microwave and Optical Technology Letters.* 2008; 50(8): 2109–2114.
[18] Guo B, Wen G. Periodic Time-Varying Noise in Current-Commutating Cmos Mixers. *Progress In Electromagnetics Research.* 2011; 117: 283-298.
[19] Polivka J, Fiala P, Machac J. Microwave noise field behaves like white light. *Progress In Electromagnetics Research.* 2011; 111: 311-330.
[20] Chong NR, Burnett IS, Chicharo JF. A new waveform interpolation coding scheme based on pitch synchronous wavelet transform decomposition. *IEEE Trans Speech Audio Process.* 2000; 8(3): 345-348.
[21] DL D. Denoising by soft thresholding. *IEEE Trans Inform Theory.* 1995; 41(3): 613-627.
[22] Berouti M, Schwartz R, Makhoul J. Enhancement of speech corrupted by acoustic noise. *In: Proc IEEE Internat Conf on Acoust Speech Signal Process (ICASSP).* 1979: 208-211.