Enhancement of non-air conduct speech based on multi-band spectral subtraction method*

Sheng Li, Ming Niu, JianQi Wang*, Tian Liu, XiJing Jing
Department of Biomedical Engineering, the Fourth Military Medical University, Xi'an 710032, China

E-mail: sheng@mail.xjtu.edu.cn

Abstract

The spread of speech is not only depending upon the air, but by means of other medium. This study reports a new speech detecting method by using EMW Radar. However, the combined noise which is added in the detected speech decreased the speech quality to a large extent, since noise is mostly colored and does not affect the speech signal uniformly over the entire spectrum. This study, therefore, propose a multi-band spectral subtraction approach which takes into account the fact that colored noise affects the speech spectrum differently at various frequencies. The results suggest that this method achieves a better reduction of the whole-frequency noise, as well as musical noise, and yields good speech quality.

1. Introduction

It is well known that speech, which is produced by speech organ of human beings [1, 2], 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].

Li Zong Wen [4] reported another medium, ElectroMagnetic Wave (EMW), which is produced by light radar, laser radar and Millimeter Wave radar, can detect and identify out exactly the existential speech signals in free space from a speaking person. Since the microwave radar has low range attenuation, and has attributes of safe, fast, and portable, EMW radar speech may has more advantage and special applications than air-conduct speech.

Although EMW radar provides another method to detect speech, however, EMW radar speech has several serious shortcomings including artificial quality, reduced intelligibility, and poor audibility. This is not only because of 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 EMW radar speech. These combined noise components are quite larger and more complex than air conduct speech, and are the biggest problem which must be resolved for the application of the EMW radar speech. Therefore, speech enhancement is a challenging topic of EMW radar speech research.

The spectral subtraction method 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 [5].

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 [6-9], results performed with these algorithms show that there is a need for further improvement. Furthermore, unlike white Gaussian noise, which has a flat spectrum, the spectrum of EMW radar noise is not flat. Thus, the noise signal does not affect the speech signal uniformly over the whole spectrum. Some frequencies are affected more adversely than others. This means that this kind of noise is "colored". Therefore, it is necessary to propose a non-linear approach to improve the subtraction procedure.

So, this study proposed a multi-band spectral subtraction algorithm which takes into account the variation of signal-to-noise ratio across the speech spectrum using a different over-subtraction factor in each frequency band to reduce colored noise. The algorithm significantly removed the colored noise and improved the speech quality to a large extent.

* This work was supported by the National Natural Science Foundation of China (NSFC) 60571046.

978-0-7695-3119-9/08 $25.00 © 2008 IEEE
DOI 10.1109/CISP.2008.274

2. Method

2.1 Description of the system

A phase-licked oscillator was used to generate a very stable EM wave. The output of the amplifier is fed through a directional coupler, a variable attenuator, a circulator, and then to a flat antenna. The flat antenna radiates a microwave beam aimed at the opposing human subjects standing or sitting directly in front of the antenna. The echo signal was received by the same antenna, which is modulated by the speech and is produced by the larynx of the opposing human subjects. This signal is then mixed with reference signal in a double-balanced mixer. This mixing, therefore, produced low-frequency signals and was amplified by a signal processor and then passed through an A/D converter before reaching computer to get further processor. For More details of description of the system, the reader is referred to [10].

2.2. Multi-band spectral subtraction method

The multi-band is based on the assume 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 [11] proposed the generalized spectral subtraction scheme is described as follows:

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

where $\alpha(\alpha > 1)$ is the over-subtraction factor [11], 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 β =0.002.

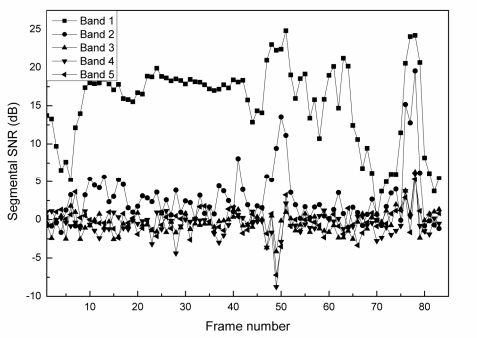

Figure 1. The segmental SNR for five frequency bands of EMW radar speech.

This implementation assumes that noise affects the speech spectrum uniformly, the over-subtraction factor α, furthermore, subtracts an over-estimate of the noise over the whole spectrum. However, the noise in the non-air conduct speech, which is produced by Millimeter Wave radar maybe colored and does not affect the speech signal uniformly over the entire spectrum. Figure 1 shows the estimated segmental SNR for five frequency bands (60～300Hz, 300～1KHz, 1K～2K, 2K～3K, 3K～5K) of radar speech corrupted by radar noise. It can be seen from Figure 1 that the SNR of the low frequency band (Band 1, 2) was significantly higher than the SNR of the high frequency band (Band 3-5). The largest SNR difference among the SNR was more than 30 dB, a large difference. This phenomenon suggests that the noise signal does not affect the speech signal uniformly over the whole spectrum, therefore, subtracting a constant factor of noise spectrum over the whole frequency spectrum may remove speech also.

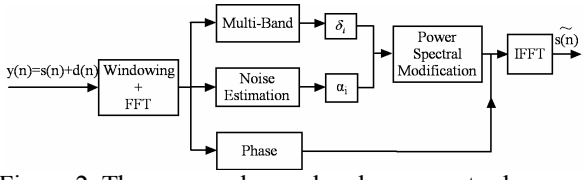

Figure 2. The proposed speech enhancement scheme.

In order to take into account the fact that colored noise affects the speech spectrum differently at various frequencies, it becomes imperative to estimate a suitable factor that will subtract just the necessary amount of the noise spectrum from each frequency sub-band. In this study, the speech spectrum was divided into N (N=5) non-overlapping bands, and spectral subtraction is performed independently in each band. Hence the estimate of the clean speech spectrum in the *i*th band is obtained by:

$$|\hat{S}_i(k)|^2 = |Y_i(k)|^2 - \alpha_i \delta_i |\hat{D}_i(k)|^2, \quad b_i \le k \le e_i \tag{4}$$

Where α_i is the over-subtraction factor of the ith frequency band, and δ_i is a tweaking factor that can be individually set for each frequency band to customize the noise removal properties. b_i and e_i are the beginning and ending frequency of the ith frequency band. Therefore, the whole algorithm can be shown in Figure 2.

The band specific over-subtraction factor α_i is a function of the segmental noisy signal to noise ratio SNR_i of the ith frequency band which is calculated as:

$$SNR_i(dB) = 10\log_{10}\frac{\sum_{k=b_i}^{e_i}|Y_i(k)|^2}{\sum_{k=b_i}^{e_i}|\hat{D}_i(k)|^2} \quad (5)$$

According to the SNR_i value calculated in Eq. (5), also consistent with Kamath et al.[8] and Udrea et al.[9], the over-subtraction factor α_i is calculated as:

$$\alpha_i = \begin{cases} 5 & SNR_i < 5 \\ 4 - \frac{3}{20}(SNR_i) & -5 \le SNR_i \le 20 \\ 1 & SNR_i > 20 \end{cases} \quad (6)$$

The use of this over-subtraction factor α_i can provides a degree of control over the noise subtraction level in each band. Another factor δ_i, which is shown in Eq. (4) can be used to provide an additional degree of control within each band, since most of the speech energy is present in the lower frequencies, smaller δ_i values were used for the low frequency bands in order to minimize speech distortion. The values of δ_i were empirically determined and set to:

$$\delta_i = \begin{cases} 1 & 60Hz \le f_i \le 300Hz \\ 1.2 & 0.3KHz < f_i \le 1KHz \\ 1.5 & 1KHz < f_i \le 2KHz \\ 2.5 & 2kHz < f_i \le 3kHz \\ 1.5 & 3kHz < f_i \le 5kHz \end{cases} \quad (7)$$

Both factors, α_i and δ_i can be adjusted for each band for different speech conditions to get better speech quality.

3. Experiments

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 8 m, and one sentences of mandarin Chinese "Di Si Jun Yi Da Xue" (other sentences were also used, but they are not representative) uttered by the volunteer speakers were used to evaluate the proposed multi-band spectral subtraction approach.

4. Results

In order to analyze the time-frequency distribution of the origin radar speech and the enhanced speech, speech spectrograms was presented to give accurate information about residual noise and speech distortion. For comparative purposes, the performance of the traditional power spectral subtraction method was also plot as implemented by Berrouti et al [11].

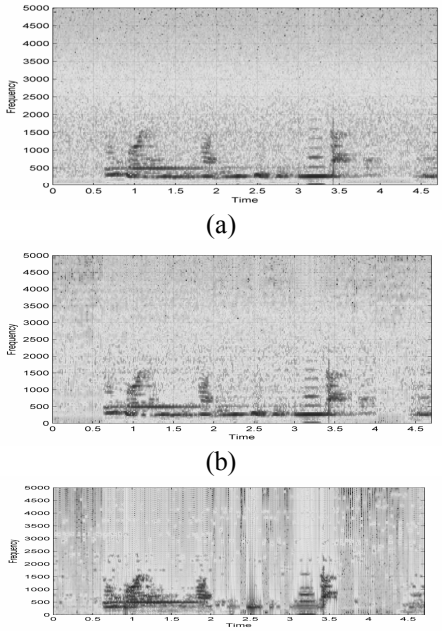

Figure 3. The Spectrogram of the sentence "Di Si Jun Yi Da Xue". (a) The original EMW radar speech; (b) enhanced signal obtained by the traditional spectral subtraction method. (c) The enhanced signal obtained by the multi-band spectral subtraction algorithm.

Figure 3 shows the spectrograms of the original radar speech (a), the enhanced speech using spectral subtraction algorithm (b) and the multi-band spectral subtraction algorithm in this study (c). Figure 3(a) shows that a certain amount of the combined noise exists in the origin radar speech because of the harmonic of the EMW, electrocircuit noise, as well as

ambient noise combined in the EMW radar speech. This noise can be obviously seen during speech pause. Figure 3(b) shows that the spectral subtraction algorithm is effective in reducing the combined noise below 2 kHz as well as an amount of noise during speech pauses, but not to reduce the high-frequency noise. Fig. 3(c) shows that the multi-band spectral subtraction algorithm not only reduces the combined noise but also eliminates both the low- and the high-frequency noise completely. That is, the multi-band spectral subtraction algorithm achieves a better reduction of the whole-frequency noise as compared to the spectral subtraction algorithm.

Informal listening tests also indicated that the multi-band approach yielded very good speech quality with very little trace of musical noise and with minimal, if any, speech distortion.

Moreover, multi-band spectral subtraction method has strong flexibility to adapt complicated speech environment by adjusting the two parameters of α_i and δ_i easily. In addition, when the total number of bands is one, then the approach of multi-band spectral subtraction algorithm reduces to the traditional power spectral subtraction approach.

5. Conclusion

As a non-air conduct speech, EMW radar speech has greater advantage and may have wider applications than air conduct speech. However, the complex noise added in the radar speech decreased the speech quality to a large extent. Therefore, an improved spectral subtraction method, multi-band spectral subtraction algorithm are used in this study in order to takes into account the non-uniform effect of colored noise on the spectrum of radar speech. The results from both simulation and evaluation suggest that this method achieves a better reduction of the whole-frequency noise, the musical noise, and yields good speech quality.

6. References

[1] S. Li, 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*. 2006, 119 (1): 539-548.

[2] S. Li, 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*. 2006, 119 (5): 3003-3010.

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

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

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

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

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

[8] S. Kamath and P. Loizou A multi-band spectral subtraction method for enhancing speech corrupted by colored noise Conference. 2002. 4160-4164.

[9] R. M. udrea, S. Ciochina, and D. N. Vizireanu Milti-band Bark Scale Spectral over-subtraction for Colored Noise Reduction Conference. 2005. 311-314.

[10] J. q. WANG, 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*. 2004, 17 (3): 157-161.

[11] M. Berouti, R. Schwartz, and J. Makhoul. "Enhancement of speech corrupted by acoustic noise". *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process*. 1979: 208-211.