An Improved Speex Speech Compression Algorithm for Wireless Sensor Network

Jiang Ming Zhang Haifeng Tang Jialu Hu Ruiying

MEMS Center, Harbin Institute of Technology, Harbin, Heilongjiang, China
E-MAIL: mjiang@hit.edu.cn

Abstract

An improved Speex speech compression algorithm is proposed in this paper, which can be applied to speech transmission through wireless sensor networks (WSN). To overcome disadvantage of traditional Speex speech compression algorithm, the improved algorithm used the wavelet filter and Minimum Mean Square Error (MMSE) to replace the perceptual weighting filter and minimum perceptual weighted error module respectively, which are used in traditional algorithm. The simulation results show that the speech quality has been significantly enhanced after using this improved algorithm.

Keywords: speech compression; wireless sensor network; wavelet; MMSE

1. Introduction

The appearance of WSN can well meet the requirement of people for information transmission equipments. Because the WSN has limited band width, highly efficient speech compression algorithm which meet the band width requirement is necessary in speech transmission. The speech compression coding algorithm determines quality of speech transmission in WSN. In the 1970s, the linear prediction coding technique was applied in grouped speech communication and the coding speed used was 3.5kbps[1,2]. In 1985, Code exciting linear prediction (CELP) coding was proposed by B.S.Atal and M.R.Schmeder, which is a closed loop analysis speech compression algorithm, then the closed analysis algorithm became the main research aspect[3,4]. In the 1990s, variant speed speech coding was developed. Speex speech compression algorithm is a speech coding algorithm with multiple modes and multiple speeds based on CELP algorithm[5]. In this paper, the proposed algorithm enhances the performace of traditional Speex speech compression algorithm and it is more suitable for the application in speech transmission through WSN.

2. Improvement of Speex Speech Compression Algorithm

Speex speech compression algorithm is designed for three kinds of sample rate which are 8kHz, 16kHz and 32kHz. These three kinds of frequencies are on behalf of narrow band, wide band and ultra wide band, respectively. CELP speech coding technique mainly includes linear prediction model analysis, adaptive code book searching, constant code book searching and sensor weighting filter closed searching. In traditional Speex speech compression algorithm, the filter module uses sensor weighting filter and its sensor weighting factor is invariant. For the design of sensor weighting filter, the details of different frequencies of signal including the sensitivity of ear for different frequencies are not considered.

At the same time, the sensor weighting coefficient of sensor weighting filter can not be changed according to actual speech cases, so the final received speech quality is affected. In this section, in order to improve speech quality, the sensor weighting filter in the Speex speech compression algorithm was substituted for wavelet filter and minimum sensor weighting error module was substituted for MMSE module.

The speech signal is non-steady signal and contains noise. The wavelet transformation is usually used to eliminate the noise in the speech signal. The wavelet transformation can process the signal in the time and frequency domain, so it is widely used to eliminate the noise of non-steady signal. Generally, a speech signal contained noise can be expressed as:

$$x_n = f_n + e_n \tag{1}$$

Where f_n is the primarily true speech signal, e_n is noise signal. The schematic diagram of wavelet transformation is shown in Fig.1. The aim of wavelet transformation is to separate f_n and e_n in x_n, and obtain a speech signal F, in which noise is eliminated. So F is a speech signal which similar to f_n.

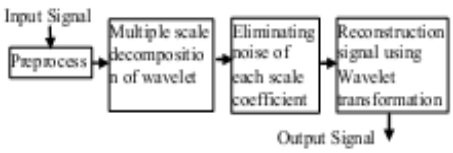


Fig.1: Schematic diagram of wavelet transformation.

The schematic diagram of MMSE algorithm is shown in Fig.2. Assume that the short time spectrum of speech signal x_n is subject to Gauss distribution and contains noise e_n ,and the true signal is f_n, then their kth spectrum components are $Y_k = R_k \exp(j\theta_k)$, N_k and $S_k = A_k \exp(ja_k)$. The estimation formula of amplitude spectrum A_k can be obtained according to MMSE:

$$\begin{aligned}\hat{A}_k &= E\{A_k | y(n), 0 \le n \le N-1\} \\ &= E\{A_k | Y_0, Y_1, \cdots\}\end{aligned} \tag{2}$$

Assume that the spectrum components are mutually independent, we can have:

$$\hat{A}_k = E(A_k | Y_k) = \frac{\int_0^\infty \int_0^{2\pi} a_k p(a_k, a_k) p(Y_k | a_k, a_k) da_k da_k}{\int_0^\infty \int_0^{2\pi} p(a_k, a_k) p(Y_k | a_k, a_k) da_k da_k} \tag{3}$$

For the speech application, the MMSE estimation based on logarithm spectrum is usually used. So the amplitude spectrum of speech in frequency domain can be expressed as:

$$\hat{A}_k = \exp\{E[InA_k | Y_k], 0 \le t \le T\} \tag{4}$$

The estimation formula of amplitude spectrum is:

$$\hat{A}_k = \frac{\xi_k}{1+\xi_k} \exp[\frac{1}{2}\int_k^\infty \frac{e^{-t}}{t} dt] R(k) \tag{5}$$

The gain function G_{MMSE} can be expressed as:

$$G_{MMSE}(\xi_k, \gamma_k) = \frac{\hat{A}_k}{R_k} = \frac{\xi_k}{1+\xi_k} \exp[\frac{1}{2}\int_{\nu_k}^\infty \frac{e^{-t}}{t} dt] \tag{6}$$

The simplified gain function G_{MMSE} can be expressed as:

$$G_{MMSE}(\xi_k, \gamma_k) = \frac{\xi_k}{1+\xi_k} \exp[\frac{1}{2} E\,\mathrm{int}(\nu_k)] \tag{7}$$

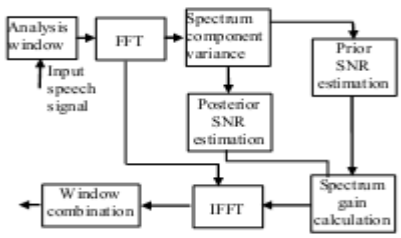


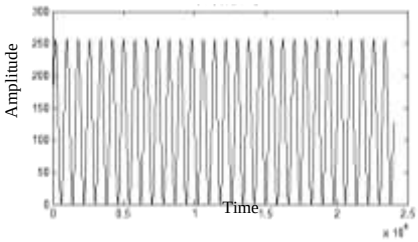
Fig.2: Schematic diagram of MMSE algorithm.

3. Software Simulation of the Improved Algorithm

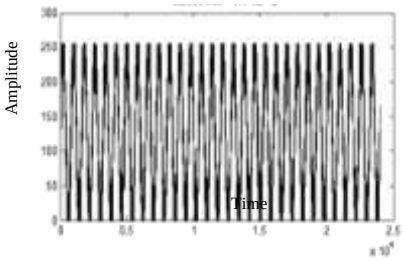
In order to testify the performance of improved algorithm, this paper used Matlab software to carry out simulation. The Gauss white noise was added to exciting signal. The exciting signals were sinuous signal and Blocks signal. Here the SNR is defined as:

$$SNR = 10\log\frac{\sum_{i=1}^{N} X^2(i)}{\sum_{i=1}^{N}[X(i)-\tilde{X}(i)]^2} \tag{8}$$

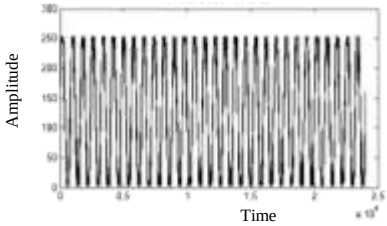
For the sinuous signal, the frequency is 10Hz, the sampling rate is 8kHz. The primary signal contained the sinuous signal and noise shown in Fig.3(b). The processed results of traditional Speex algorithm and improved algorithm are shown in Fig.3(c) and Fig.3(d).The SNR of primary signal is 27.52dB, the SNR of signal processed by traditional Speex algorithm is 48.07dB and the SNR of signal processed by improved Speex algorithm is 71.41dB. It can be seen that the improved algorithm can enhance SNR of the sinuous signal.



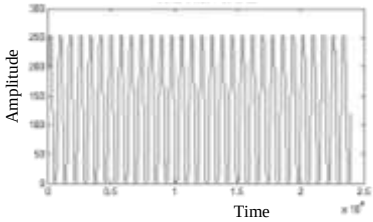
(a) Sinuous signal



(b) Primary signal



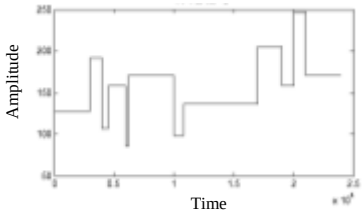
(c) Results of traditional Speex algorithm



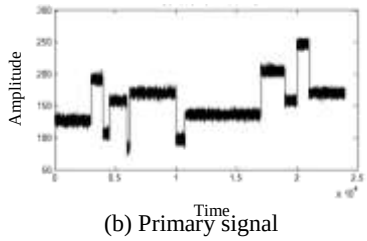
(d) Results of improved Speex algorithm

Fig.3: Simulation results of sinuous signal.

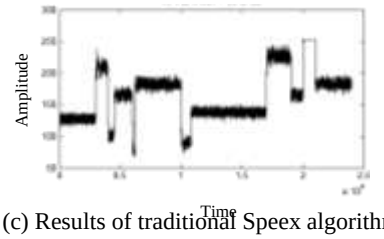
In order to testify the reliability of speech signal transmission, Blocks signal was also selected as the tested signal. Blocks signal is not continuous signal with break. The waveform of a Blocks signal is shown in Fig. 4(a) , the primary signal is Blocks signal with noise shown in Fig.4(b). The processed result of traditional Speex algorithm and improved Speex algorithm are shown in Fig. 4(c) and Fig. 4(d). The SNR of primary signal is 69.07dB, the SNR of signal processed by traditional Speex algorithm is 46.21dB, and the SNR of signal processed by improved Speex algorithm is 78.06dB. From the above simulation results, it can be seen that traditional algorithm will add noise, while the improved algorithm can decrease it.



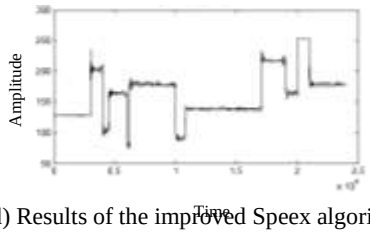
(a) Blocks signal



(b) Primary signal



(c) Results of traditional Speex algorithm



(d) Results of the improved Speex algorithm

Fig.4: Simulation results of Blocks signal.

4. Conclusion

In this paper, an improved Speex speech compression algorithm is proposed. On the one hand, this algorithm inherits advantages of traditional Speex speech compression algorithm: multiple modes and multiple speeds. On the other hand, it enhances the performance of Speex speech compression algorithm by using the wavelet filter and MMSE to replace the perceptual weighting filter and minimum perceptual weighted error module respectively. The simulation results show that the proposed algorithm is more suitable for the application in fields of speech transmission through WSN.

5. Acknowledgement

The authors would like to acknowledge the financial support from the national high technology research and development program of china (Grant No. SS 2012AA062902), Major Science and Technology Program for Water Pollution Control and Treatment (Grant No. 2012ZX07201003-01), Key Science and Technology Project of Harbin (Grant No. 2011AA1CG003).

6. References

[1] Vijayasenan D, Valente F, “An Information Theoretic Approach to Speaker Diarization of Meeting Data”, *Audio, Speech, and Language Processing*, pp. 1382-1393, 2009.

[2] Irino T, Patterson R D, “A Dynamic Compressive Gammachirp Auditory Filterbank”, *Audio, Speech, and Language Processing*, pp. 2222-2232, 2006.

[3] F.C.A.Brooks, Lajos Hanzo, “A Multiband Excited Waveform-Interpolated 2.35kb/s Speech Coder for Bandlimited Channels”, *IEEE Transaction on Vehicular Technology*, pp. 766-772, 2000.

[4] P C Meuse, “A 2400bps Multi-band Excitation Vocoder”, *IEEE Int. Conf. Acoustic, Speech, Signal Processing*, pp. 9-12, 2004.

[5] Srivastava Pooja, Babu Kishore, Osv Tejaswi, “Performance evaluation of Speex audio codec for wireless communication networks”, *IEEE Int. Conf. Wireless and Optical Communications Networks*, pp. 1-5, 2011.