ATLANTIS PRESS

Advances in Engineering Research, volume 114

International Conference on Advances in Materials, Machinery, Electrical Engineering (AMMEE 2017)

A Speech Enhance Method of Spectral Subtraction Based on Objective Evaluation

Zitong Yu [a], Longxiang Guo and Huifeng Li

College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001, China;

[a] yuzitong@hrbeu.edu.cn

Keywords: Objective Evaluation, Spectrum Subtraction, Speech Coding, Acoustic Communication.

Abstract. Speech coding is a technique that transforms speech signal of big data to digital signal of small data. This characteristic is suitable for the underwater acoustic transmission of narrowband communication. It's one of the support to underwater acoustic high speed voice communication on the source code level. This paper expounds MELP, and applies PESQ, an objective speech quality assessment, to evaluate and mark the quality of speech coding. According to the evaluation score to judge the quality of speech communication. It is suppressed that the stabile noise components in decode speech by using spectrum subtraction method. Experimental results show that it can improve the PESQ evaluation score and obtain better auditory perception.

1. Introduction

Language is the most intuitive way of human communication, voice is natural, convenient, rich information communication carrier. In the field of underwater acoustic communication, underwater platform information communication and connection between members of underwater operation, are rely on the speech coding and underwater acoustic communication technique. Owing to the complex characteristics of the underwater acoustic channel, how to make the speech coding better adapt to the environment of underwater acoustic communication, has become the key to the underwater voice communication.

The traditional subjective evaluation standard MOS (Mean Opinion Score) method does not suit because of the limitation of underwater voice communication environment, the small number of samples and the confidentiality of communication contents. So, a kind of evaluation method which objective and does not rely on a lot of manual work is needed to assess decode speech signal. PESQ (Perceptual evaluation of speech quality) method, which the correlation with MOS is 0.98 and a good way to replace the MOS method is provided by ITU-T Recommendation P.862 standard.

In subjective listening experience, 2.4kbps MELP (Mixed-Excitation Linear Predictive) speech coding is close to 4.8kbps CELP (Code Excited Linear Prediction) method, and additionally contains the speaker's individual information, therefore it is more suitable for underwater voice communication. In consideration of the obvious stability noise, the spectrum subtraction can be a good elimination of this noise, making the voice more clearly, higher score evaluation, and better subjective listening feel.

2. Main Method

2.1 PESQ.

In this paper, PESQ, the objective evaluation standard, combines PSQM (Perceptual Speed Quality Measure) and PAMS (Perceptual Analysis Measurement System) the advantages of two methods, considering the end-to-end delay, applicable to a variety of phonetic system of quality evaluation, and has good robust to the environmental noise and communication time delay. The algorithm flow is shown in Fig.1 [1].

PESQ algorithm works as follows: The model begins by level aligning both signals to a standard listening level. They are filtered (using an FFT) with an input filter to model a standard telephone handset. The signals are aligned in time then processed through an auditory transform similar to that

of PSQM [2]. Then calculate the spectral distortion measure of them as a disturbance value. Finally, frame disturbances D_n and DA_n are obtained from the cognitive model, and their time domain averages to obtain disturbance density d_{SYM} and asymmetrical disturbance density d_{ASYM}, and the PESQ score can be expressed as [3]

$$PESQ_{MOS} = 4.5 - 0.1 \bullet d_{SYM} - 0.0309 \bullet d_{ASYM} \quad (1)$$

It shows that PESQ score is from -0.5 to 4.5, but normally not less than 1.0.

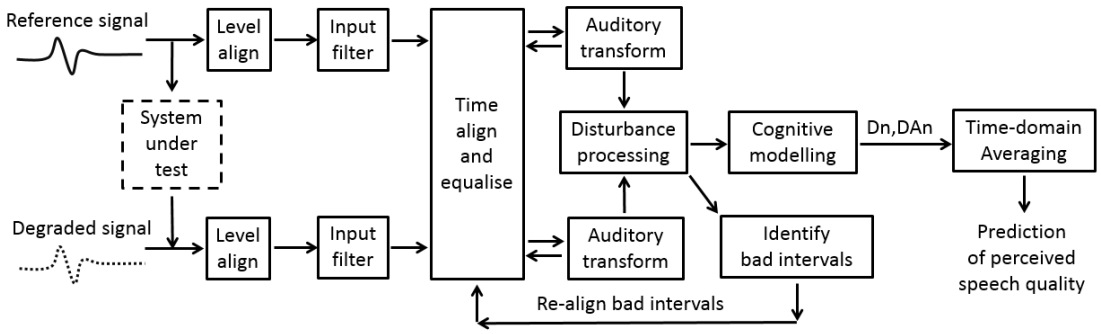

Fig. 1 PESQ flow

2.2 MELP.

Modern speech coding technology mainly to the direction of low-speed development, which representative of the algorithm include: STC (Sine Transform Coding), MBE (Multi-band excitation), MELP, WI (Waveform Interpolation), e.t.

Among them, MELP encoder has the advantages of low coding rate and good quality of speech synthesis. Compared with other encoder with coding rate up to 2.4kbps below, it has good speech quality, a high degree of natural, recognizable voices of male and female, distinguishable mood of voice of the speaker, and including individual information. And in noisy environment, it has good robustness. Compromised coding rate and restore the natural nature of voice, MELP can be used as a mainly way to implement underwater voice source coding [4].

2.3 Spectral Subtraction.

Figure (2) shows that amplifying wave-form of decoding speech signal. It obviously shows that there is a steady noise in the entire speech signal after restoration. For this kind of stationary noise, spectral subtraction is the optimization method which is the least operation and easy to implement, and can enhance the decoded speech obviously.

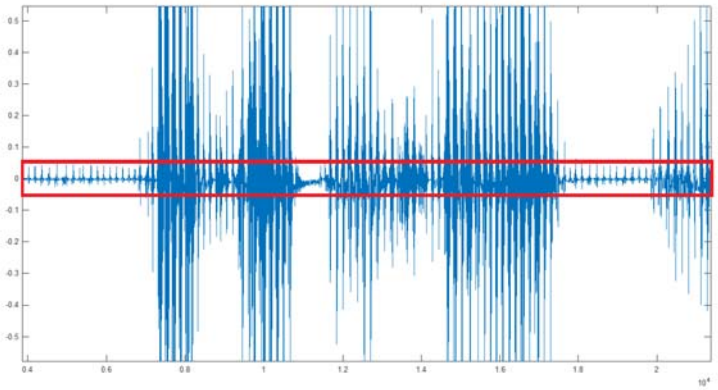

Fig. 2 Stationary noise

Usually, we consider speech signal $s(t)$ and additive noise signal $n(t)$ are irrelevant, defining

$$x(t) = s(t) + n(t) \quad (2)$$

To obtain a 'pure' speech signal, it is only necessary to subtract the power spectrum of the noise $n(\omega)$ from the power spectrum of the speech signal $x(\omega)$, as expressed by [5]

$$s(\omega) = x(\omega) - n(\omega) \quad (3)$$

In practical applications, $n(\omega)$ is usually gained through the quiet section window flaming, and find the average time average energy. The estimate of the 'pure' speech amplitude spectrum is

$$\left|\widehat{X}(\omega)\right| = \begin{cases} |Y(\omega)| - \left|\widehat{N}(\omega)\right| & |Y(\omega)| - \left|\widehat{N}(\omega)\right| \geq 0 \\ 0 & otherwise \end{cases} \quad (4)$$

3. Simulation Results

This section mainly discusses the performance of spectral subtraction.

3.1 Simulated Spectral Subtraction.

Simulation step:

First: encode, decode speech signals by using MELP, gain restored signals;

Second: Spectral Subtraction, get 'pure' signal;

Third: assess by PESQ, comparing score shown as:

Table 1 Spectral Subtraction performance

Signal	Restored Signals	'Pure' Signal	Improvement
Test1.wav	3.046	3.625	19.01%
Test2.wav	2.837	3.258	14.84%
Test3.wav	2.962	3.364	13.57%

Fourth: Take 'Test1.wav' as an example, compare the wave-form of speech signal, restored signal, and 'pure' signal. The result shows as

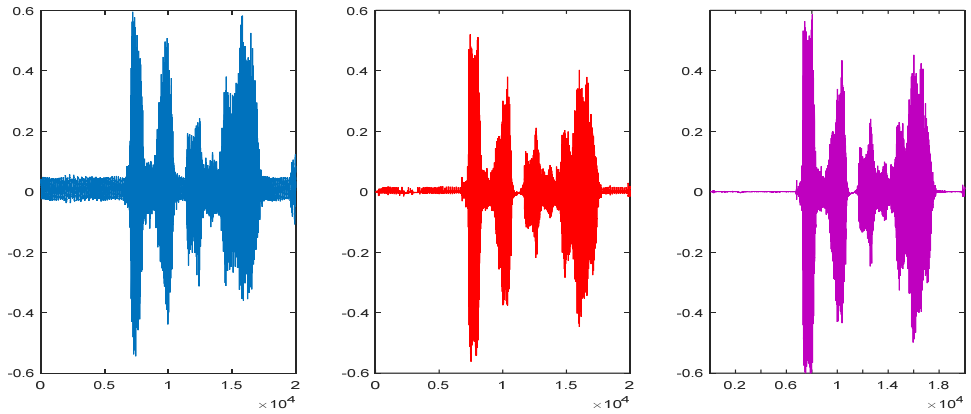

Fig. 3 Compare the wave form

Results analysis:

1. From table(1), different speech signals voice quality are improved significantly after make spectral subtraction.
2. From figure(3), in the same ordinate scale, noise of 'pure' signal is suppressed obviously, so has better subjective listening feel.

3.2 Spectral Subtraction in Error Code.

Table 2 Spectral Subtraction performance in different error rate

Signal	Error Rate	Restored Signal	'Pure' Signal	Improvement
Test1.wav	0%	3.046	3.625	19.01%
	1%	2.709	3.115	14.99%
	2%	2.338	2.632	12.57%
	3%	2.138	2.378	11.23%

In fact, due to the error code by the channel transmission, the quality of voice communication will reduce. This section simulates performance of spectral subtraction at different communication error rates. The results are showed as table (2). Spectral subtraction processing can also significantly improve the quality of speech with communication error.

4. Conclusion

MELP is an effective method for the application of speech coding in underwater acoustic communication, but, there is obvious stable noise in the decoded speech. It is verified that the subtraction can effectively reduce the noise which leads to better hearing feeling, and improve the quality of the speech signal. And in the case of errors existing in the experiment of communication, spectral subtraction is also good performance.

Acknowledgements

Public science and technology research funds projects of ocean (201405036). National Natural Science Foundation of China (Grant No. 50509059). Funding in part of Science and Technology on Underwater Acoustic Antagonizing Laboratory

References

[1]. Information on: http://www.opticom.de/

[2]. Rix A W, Beerends J G, Hollier M P, et al. Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs[C]. IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001. Proceedings. IEEE Xplore, Vol.2 (2001), p .749-752.

[3]. Zhang Weiwei: Research on Voice Quality Evaluation in Communication System [D] (Dr., Beijing University of Posts and Telecommunications, China, 2014). P. 12-28.

[4]. Wang Hong, Tang Kai. Low-rate Speech Coding [M]. Defense Industry Press, 2006, p. 211-236.

[5]. Song Zhiyong. The Application of MATLAB in Speech Signal Analysis and Synthesis [M]. Beijing University of Aeronautics and Astronautics Press, 2013, p.183-191